

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

Patrícia Lopes

**METODOLOGIA PARA APRENDIZAGEM DE ONTOLOGIAS:
uma proposta no contexto do governo eletrônico**

Belo Horizonte

2025

PATRÍCIA LOPES

**METODOLOGIA PARA APRENDIZAGEM DE ONTOLOGIAS:
uma proposta no contexto do governo eletrônico**

Tese apresentada ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento, Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito para a obtenção do título de Doutora em Gestão e Organização do Conhecimento.

Área de Concentração: Ciência da Informação.

Linha de Pesquisa: Arquitetura & Organização do Conhecimento (AOC).

Orientadora: Profa. Dra. Gercina Ângela de Lima.

Belo Horizonte

2025

L864m

Lopes, Patrícia.

Metodologia para aprendizagem de ontologias [recurso eletrônico] : uma proposta no contexto do governo eletrônico / Patrícia Lopes. - 2025.

1 recurso online (227 f. : il., color.) : pdf.

Orientadora: Gercina Ângela de Lima..

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

Referências: f. 214-223.

Apêndice: f. 224-227.

Exigência do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Ontologias – Teses. 3. Governo eletrônico – Teses. 4. Organização do conhecimento – Teses. 5. Interoperabilidade semântica – Teses. I. Lima, Gercina Ângela de. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.

CDU 004.4



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

ATA DA DEFESA DE TESE DA ALUNA

PATRÍCIA LOPES FERREIRA FRANÇA

Realizou-se, no dia 27 de fevereiro de 2025, às 15:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *Metodologia para aprendizagem de ontologias: uma proposta no contexto do Governo Eletrônico*, apresentada por PATRÍCIA LOPES FERREIRA FRANÇA, número de registro 2020660428, graduada no curso de PEDAGOGIA, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Gercina Angela de Lima - ECI/UFMG (Orientadora), Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG, Prof(a). Cristiane Mendes Netto - UNIVALE, Prof(a). Frederico Cesar Mafra Pereira - ECI/UFMG, Prof(a). Webert Junio Araujo - CEFET-MG, Prof(a). Sérgio Mariano Dias - SERPRO.

A Comissão considerou a tese:

Aprovada

Reprovada

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

Belo Horizonte, 27 de fevereiro de 2025.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Gercina Angela de Lima, Coordenador(a) de curso de pós-graduação**, em 07/03/2025, às 16:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Mendes Netto, Usuário Externo**, em 10/03/2025, às 14:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 10/03/2025, às 14:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Frederico Cesar Mafra Pereira, Professor do Magistério Superior**, em 10/03/2025, às 15:23, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sérgio Mariano Dias, Usuário Externo**, em 10/03/2025, às 17:52, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Webert Junio Araujo, Usuário Externo**, em 19/03/2025, às 14:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3999932** e o código CRC **3F7D7D50**.



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

FOLHA DE APROVAÇÃO

Metodologia para aprendizagem de ontologias: uma proposta no contexto do Governo Eletrônico

PATRÍCIA LOPES FERREIRA FRANÇA

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

Aprovada em 27 de fevereiro de 2025, por videoconferência, pela banca constituída pelos membros:

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

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

Prof(a). Cristiane Mendes Netto
UNIVALE

Prof(a). Frederico Cesar Mafra Pereira
ECI/UFMG

Prof(a). Webert Junio Araujo
CEFET-MG

Prof(a). Sérgio Mariano Dias
SERPRO

Belo Horizonte, 27 de fevereiro de 2025.



Documento assinado eletronicamente por **Gercina Angela de Lima, Coordenador(a) de curso de pós-graduação**, em 07/03/2025, às 16:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Mendes Netto, Usuário Externo**, em 10/03/2025, às 14:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 10/03/2025, às 14:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Frederico Cesar Mafra Pereira, Professor do Magistério Superior**, em 10/03/2025, às 15:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sérgio Mariano Dias, Usuário Externo**, em 10/03/2025, às 17:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Webert Junio Araujo, Usuário Externo**, em 19/03/2025, às 14:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3999941** e o código CRC **A3F6908E**.

AGRADECIMENTOS

Concluir esta pesquisa foi uma jornada intensa, permeada por desafios, aprendizados e, sobretudo, pela colaboração de muitas pessoas que, direta ou indiretamente, contribuíram para que este trabalho se tornasse realidade. A todas elas, expresso minha sincera gratidão.

Agradeço à minha orientadora, Prof.^a Dr.^a Gercina Ângela de Lima, pelo suporte contínuo, pela paciência e pelas contribuições, que enriqueceram significativamente este estudo. Sua orientação criteriosa foi fundamental para que esta pesquisa alcançasse a maturidade necessária.

Aos professores e colegas do Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Universidade Federal de Minas Gerais, pela troca de conhecimentos e pelas discussões instigantes que ampliaram minha compreensão sobre o campo da Ciência da Informação.

À minha família, pelo apoio incondicional ao longo desta trajetória; pela compreensão nos momentos de ausência, pelo incentivo nos períodos de desânimo e pela celebração a cada pequena conquista. Sem vocês, este percurso teria sido muito mais árduo.

Aos amigos que, de diferentes formas, acompanharam-me durante essa caminhada, oferecendo suporte, incentivo e palavras de motivação nos momentos mais desafiadores.

Aos profissionais e instituições que colaboraram com este estudo, fornecendo dados e suporte técnico essencial para a condução da pesquisa.

A todos e todas que, de alguma maneira, contribuíram para que este trabalho se concretizasse, meu profundo agradecimento.

RESUMO

As ontologias têm ganhado crescente reconhecimento em diferentes domínios, sendo consideradas fundamentais para a organização e representação do conhecimento. Contudo o aumento exponencial de dados não estruturados na Web tornou o processo de construção ontológica altamente complexo, demorado e dispendioso. No contexto do governo eletrônico, essa complexidade se intensifica diante da necessidade de estruturar e integrar grandes volumes de dados de forma padronizada, transparente e semanticamente interoperável. Esta tese teve como objetivo elaborar uma metodologia de aprendizagem de ontologias voltada à organização e representação do conhecimento e da informação no âmbito do governo eletrônico brasileiro. Para isso, foram realizados: a análise de estudos sobre geração automática e semiautomática de ontologias; a identificação de Sistemas de Organização do Conhecimento adotados por países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da Organização das Nações Unidas; o inventário de ontologias desenvolvidas no contexto governamental brasileiro; e a análise comparativa de quatro metodologias de construção de ontologias elaboradas no âmbito do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais. Como principal resultado, propôs-se a Onto Merge Methodology (OMM), metodologia que possibilita a automatização parcial de etapas relacionadas à aquisição e extração de conhecimento, utilizando técnicas como mineração de texto e processamento de linguagem natural. A OMM foi validada por meio de uma prova de conceito aplicada ao domínio das compras públicas, demonstrando ganhos de eficiência e viabilidade de replicação em outros contextos governamentais. Concluiu-se que, embora a automação completa da construção ontológica ainda enfrente desafios técnicos e institucionais, metodologias como a OMM representam um avanço importante, ao oferecer uma abordagem mais ágil, replicável e alinhada às exigências de escalabilidade, padronização e interoperabilidade da gestão pública digital.

Palavras-chave: Organização e Representação do Conhecimento e da Informação; Sistemas de Organização do Conhecimento (SOC); ontologia; aprendizagem de ontologias; governo eletrônico.

ABSTRACT

Ontologies have gained increasing recognition across various domains, being considered fundamental for the organization and representation of knowledge. However, the exponential growth of unstructured data on the Web has made the ontology construction process highly complex, time-consuming, and costly. In the context of e-government, this complexity is heightened by the need to structure and integrate large volumes of data in a standardized, transparent, and semantically interoperable manner. This dissertation aimed to develop a methodology for ontology learning focused on the organization and representation of knowledge and information within the Brazilian e-government environment. To this end, the following steps were carried out: analysis of studies on automatic and semi-automatic ontology generation; identification of Knowledge Organization Systems adopted by countries ranked highest in the United Nations E-Government Development Index; inventory of ontologies developed within the Brazilian government context; and comparative analysis of four ontology construction methodologies developed within the Graduate Program in Knowledge Management and Organization at the Federal University of Minas Gerais. As a main result, the Onto Merge Methodology (OMM) proposed a systematic and modular approach that enables partial automation of tasks related to knowledge acquisition and extraction, using techniques such as text mining and natural language processing. The OMM was validated through a proof of concept applied to the domain of public procurement, demonstrating efficiency gains and the potential for replication in other governmental contexts. It is concluded that, although full automation of ontology construction still faces technical and institutional challenges, methodologies such as the OMM represent a significant advancement by offering a more agile, replicable approach aligned with the scalability, standardization, and interoperability demands of digital public management.

Keywords: Knowledge and Information Organization and Representation; Knowledge Organization Systems (KOS); ontology; learning; e-government.

RESUMEN

Las ontologías han ganado un reconocimiento creciente en diferentes dominios, siendo consideradas fundamentales para la organización y representación del conocimiento. Sin embargo, el crecimiento exponencial de datos no estructurados en la Web ha convertido el proceso de construcción ontológica en una tarea altamente compleja, lenta y costosa. En el contexto del gobierno electrónico, esta complejidad se intensifica debido a la necesidad de estructurar e integrar grandes volúmenes de datos de forma estandarizada, transparente e interoperable desde el punto de vista semántico. Esta tesis tuvo como objetivo desarrollar una metodología de aprendizaje de ontologías orientada a la organización y representación del conocimiento y de la información en el ámbito del gobierno electrónico brasileño. Para ello, se llevaron a cabo: el análisis de estudios sobre la generación automática y semiautomática de ontologías; la identificación de Sistemas de Organización del Conocimiento adoptados por los países mejor posicionados en el Índice de Desarrollo del Gobierno Electrónico de las Naciones Unidas; el inventario de ontologías desarrolladas en el contexto gubernamental brasileño; y el análisis comparativo de cuatro metodologías de construcción de ontologías desarrolladas en el marco del Programa de Posgrado en Gestión y Organización del Conocimiento de la Universidad Federal de Minas Gerais. Como resultado principal, se propuso la Onto Merge Methodology (OMM), una metodología sistemática y modular que permite la automatización parcial de etapas relacionadas con la adquisición y extracción de conocimiento, utilizando técnicas como la minería de textos y el procesamiento del lenguaje natural. La OMM fue validada mediante una prueba de concepto aplicada al dominio de las compras públicas, demostrando ganancias de eficiencia y viabilidad de replicación en otros contextos gubernamentales. Se concluye que, aunque la automatización completa del proceso de construcción ontológica enfrenta desafíos técnicos e institucionales, metodologías como la OMM representan un avance importante al ofrecer un enfoque más ágil, replicable y alineado con las exigencias de escalabilidad, estandarización e interoperabilidad de la gestión pública digital.

Palabras clave: Organización y Representación del Conocimiento y la Información; Sistemas de Organización del Conocimiento (SOC); ontología; aprendizaje de ontologías; gobierno electrónico.

LISTA DE FIGURAS

Figura 1 – Passos da Etapa 1	97
Figura 2 – Página inicial do documento Orçamento Federal – Orçamento Fiscal e da Seguridade Social	149
Figura 3 – Página de detalhamento do documento Orçamento Federal – Orçamento Fiscal e da Seguridade Social	150
Figura 4 – Tela de busca da plataforma GovBR	151
Figura 5 – Tela de pesquisa complementar na plataforma GovBR	151
Figura 6 – Página de pesquisa do Portal da Transparência.....	153
Figura 7 – Detalhamento de um documento de empenho do Portal da Transparência	154
Figura 8 – Resultado de busca na Biblioteca Digital do Senado	155
Figura 9 – Detalhamento dos metadados no padrão <i>Dublin Core</i>	156
Figura 10 – Resultado de busca na página da biblioteca do IBGE	157
Figura 11 – Detalhamento de documento recuperado na Biblioteca do IBGE	157
Figura 12 – Resultado de busca no Repositório Institucional da ENAP	158
Figura 13 – Detalhamento de documento recuperado no Repositório Institucional da ENAP.....	159
Figura 14 – Recorte de documento recuperado no Repositório Institucional da ENAP	159
Figura 15 – Primeiro resultado de busca na Biblioteca de Teses e Dissertações ...	160
Figura 16 – Documento recuperado na Biblioteca de Teses e Dissertações	161
Figura 17 – Segundo resultado de busca na Biblioteca de Teses e Dissertações ..	162
Figura 18 – Terceiro resultado de busca na Biblioteca de Teses e Dissertações ...	163
Figura 19 – Resultado de busca no SciELO.....	164
Figura 20 – Página inicial da ontologia OntoAccount	168
Figura 21 – Visão consolidada das fases e etapas das metodologias analisadas ..	171
Figura 22 – Visão Geral da Metodologia Onto Merge	183

LISTA DE QUADROS

Quadro 1 – Resumo Protocolo da Revisão Sistemática.....	29
Quadro 2 – Termos de busca Etapa 1	29
Quadro 3 – Critérios de Inclusão.....	30
Quadro 4 – Critérios de Exclusão.....	30
Quadro 5 – Resumo dos trabalhos relacionados a metodologias de aprendizagem de ontologias	41
Quadro 6 – Resumo das metodologias identificadas	47
Quadro 7 – Etapas e fases identificadas	52
Quadro 8 – Diferenciais, lacunas e resultados	57
Quadro 9 – Termos de busca Etapa 2	99
Quadro 10 – Termos de busca Etapa 3	101
Quadro 11 – Fontes de pesquisa Etapa 3.....	102
Quadro 12 – Comparação entre as quatro metodologias.....	179
Quadro 13 – Fases da Onto Merge Methodology (OMM)	183
Quadro 14 – Integração das Etapas para construção de ontologias com base nas Metodologias de Referência.....	186
Quadro 15 – Etapas detalhadas da Fase Avaliação da Necessidade.....	189
Quadro 16 – Etapas detalhadas da Fase Planejamento e Especificação	190
Quadro 17 – Etapas detalhadas da Fase Aquisição de Conhecimento	192
Quadro 18 – Etapas detalhadas da Fase Modelagem Conceitual	194
Quadro 19 – Etapas detalhadas da Fase Implementação e formalização	196
Quadro 20 – Etapas detalhadas da Fase Avaliação	198
Quadro 21 – Etapas detalhadas da Fase Documentação.....	199
Quadro 22 – Etapas detalhadas da Fase Publicação	201
Quadro 23 – <i>Prompt 1</i>	205
Quadro 24 – <i>Prompt 2</i>	206
Quadro 25 – <i>Prompt 3</i>	207

LISTA DE TABELAS

Tabela 1 – Relação dos países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU	98
Tabela 2 – Documentos recuperados nos <i>sites</i> de dados abertos.....	105
Tabela 3 – Menções aos SOC – <i>site</i> de dados abertos da Nova Zelândia	106
Tabela 4 – Menções aos SOC – <i>site</i> de dados abertos da Austrália	111
Tabela 5 – Menções aos SOC – <i>site</i> de dados abertos dos Estados Unidos.....	115
Tabela 6 – Menções aos SOC – <i>site</i> de dados abertos do Reino Unido.....	121
Tabela 7 – Menções aos SOC – <i>site</i> de dados abertos de Singapura	125
Tabela 8 – Menções aos SOC – <i>site</i> de dados abertos de Malta.....	126
Tabela 9 – Menções aos SOC – <i>site</i> de dados abertos da Espanha	127
Tabela 10 – Menções aos SOC – <i>site</i> de dados abertos da Irlanda	131
Tabela 11 – Menções aos SOC – <i>site</i> de dados abertos do Canadá.....	134
Tabela 12 – Menções aos SOC – <i>site</i> de dados abertos do Uruguai.....	139
Tabela 13 – Menções aos SOC – <i>site</i> de dados abertos do Brasil	144

LISTA DE SIGLAS E ABREVIATURAS

ACM	Association for Computing Machinery
AOC	Arquitetura e Organização do Conhecimento
API	Application Programming Interface
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BFO	Basic Formal Ontology
BIM	Building Information Modeling
CSO	Computer Science Ontology
DPCS	Density-Peak Clustering
DSRM	Design Science Research Methodology
EGDI	E-Government Development Index
ENAP	Escola Nacional de Administração Pública
GDPR	Regulamento Geral sobre a Proteção de Dados
HCEN	História Clínica Electrónica Nacional
HCI	Índice de Capital Humano
HTML	HyperText Markup Language
IBGE	Instituto Brasileiro de Geografia e Estatística
IBTS	International Bottom Trawl Survey
IoT	Internet of Things
IWs	Mundos Institucionais Interconectados
KE4WoT	Knowledge Extraction for Web of Things
LAM	Learning Algorithm Modulator
LDA	Latent Dirichlet Allocation
LGPD	Lei Geral de Proteção de Dados
LLM	Large Language Model
LOM	Learned Ontology Model
LSA	Análise Semântica Latente
LSTM	Long Short-Term Memory
MeSH	Medical Subject Headings
MMR	Maximal Marginal Relevance
NPRI	National Pollutant Release Inventory
OBM	Ontologia Brasileira de Medicamentos

OGP	Parceria para Governo Aberto
OMM	Onto Merge Methodology
ONU	Organização das Nações Unidas
ORSD	Ontology Requirements Specification Document
OSI	Índice de Serviços Online
OWL	Web Ontology Language
PLN	Processamento de Linguagem Natural
PRONTO	Projeto de Recursos Ontológicos
PRSL	Protocolo de Revisão Sistemática da Literatura
RAG	Retrieval-augmented Generation
RDF	Resource Description Framework
RSL	Revisão Sistemática de Literatura
SciELO	Scientific Electronic Library Online
SIM	Sistema Integrado de Metadatos
SKOS	Simple Knowledge Organization System
SOC	Sistemas de Organização do Conhecimento
SPARQL	Protocol and RDF Query Language
SUS	Sistema Único de Saúde
TCU	Tribunal de Contas da União
TIC	Tecnologias de Informação e Comunicação
TII	Índice de Infraestrutura de Telecomunicações
UN DESA	Departamento de Assuntos Econômicos e Sociais
VRAM	Video Random Access Memory
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Map Service

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Problema	20
1.2 Justificativa.....	22
1.3 Pressupostos.....	23
1.4 Objetivos	24
1.4.1 Objetivo geral	24
1.4.2 Objetivos específicos.....	24
1.5 Estrutura da tese	25
2 REVISÃO DE LITERATURA	27
2.1 Introdução	27
2.2 Procedimentos metodológicos adotados na Revisão de Literatura.....	28
2.3 Trabalhos relacionados a metodologias de aprendizagem de ontologias	31
2.4 Resultados e discussão.....	43
2.4.1 Metodologias de aprendizagem de ontologias	44
2.4.2 Etapas/fases das metodologias de aprendizagem de ontologias.....	49
2.4.3 Diferenciais, lacunas e principais resultados das metodologias de aprendizagem de ontologias	55
3 FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA	62
3.1 Organização e Representação do Conhecimento e da Informação	62
3.1.1 Sistemas de Organização do Conhecimento (SOC)	70
3.2 Ontologias	75
3.2.1 Metodologias para construção de ontologias	77
3.2.2 Aprendizagem de Ontologias	84
3.3 Governo Eletrônico.....	88
3.4 Plataforma de serviços públicos Gov.br	90
3.5 Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI)	92

4 METODOLOGIA	95
4.1 Caracterização da pesquisa	95
4.2 Delimitação do universo da pesquisa	95
4.3 Método	96
4.4 Visão geral dos procedimentos metodológicos	97
4.4.1 Procedimentos metodológicos adotados na Etapa 2	98
4.4.2 Procedimentos metodológicos adotados na Etapa 3	100
4.4.3 Procedimentos metodológicos adotados na Etapa 4	103
5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	104
5.1 Documentos recuperados em países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da Organização das Nações Unidas (EGDI)	104
5.1.1 Documentos recuperados no <i>site</i> de dados abertos do governo da Nova Zelândia	106
5.1.2 Documentos recuperados no <i>site</i> de dados abertos do governo da Austrália	110
5.1.3 Documentos recuperados no <i>site</i> de dados abertos do governo dos Estados Unidos da América	114
5.1.4 Documentos recuperados no <i>site</i> de dados abertos do governo do Reino Unido	120
5.1.5 Documentos recuperados no <i>site</i> de dados abertos do governo de Singapura	124
5.1.6 Documentos recuperados no <i>site</i> de dados abertos do governo de Malta.....	125
5.1.7 Documentos recuperados no <i>site</i> de dados abertos do governo da Espanha	127
5.1.8 Documentos recuperados no <i>site</i> de dados abertos do governo da Irlanda...	130
5.1.9 Documentos recuperados no <i>site</i> de dados abertos do governo do Canadá .	133
5.1.10 Documentos recuperados no <i>site</i> de dados abertos do governo do Uruguai	138
5.1.11 Documentos recuperados no <i>site</i> de dados abertos do governo do Brasil...	143
5.2 Discussão dos resultados.....	145

5.3 Ontologias desenvolvidas no contexto do governo brasileiro.....	148
5.3.1 Apresentação dos resultados	148
5.3.1.1 Portal de dados abertos do governo brasileiro	148
5.3.1.2 Portal de serviços do governo brasileiro – GovBR	150
5.3.1.3 Portal da transparência	152
5.3.1.4 Biblioteca Digital do Senado.....	154
5.3.1.5 Biblioteca do Instituto Brasileiro de Geografia e Estatística (IBGE)	156
5.3.1.6 Repositório Institucional da ENAP	158
5.3.1.7 Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).....	160
5.3.1.8 SciELO – <i>Scientific Electronic Library Online</i>	163
5.3.1.9 Anais Seminário de Pesquisa em Ontologia no Brasil – Ontobras.....	164
5.3.1.10 Google Acadêmico	164
5.3.2 Discussão dos resultados.....	165
5.4 Metodologias de construção de ontologias: uma análise comparativa.....	169
5.4.1 Descrição das metodologias	170
5.4.1.2 Fases e Etapas da Metodologia OntONeo	173
5.4.1.3 Fases e Etapas da Metodologia OntoAccount	174
5.4.1.4 Fases e Etapas da Metodologia PRONTO	175
5.4.2 Análise comparativa das quatro metodologias selecionadas	177
6 ONTO MERGE METHODOLOGY (OMM)	181
6.1 Resumo do Percurso Metodológico	181
6.2 Fases propostas da OMM	182
6.3 Apresentação e descrição das etapas de construção de Ontologias não automatizadas	186
6.3.1 Etapas detalhadas da Fase Avaliação da Necessidade.....	188
6.3.2 Etapas detalhadas da Fase Planejamento e Especificação	189
6.3.3 Etapas detalhadas da Fase Aquisição de Conhecimento	192

6.3.4 Etapas detalhadas da Fase Modelagem Conceitual	194
6.3.5 Etapas detalhadas da Fase Implementação e Formalização	195
6.3.6 Etapas detalhadas da Fase Avaliação	197
6.3.7 Etapas detalhadas da Fase Documentação	199
6.3.8 Etapas detalhadas da Fase Publicação	201
6.4 Prova de conceito para Aquisição de Conhecimento no Domínio de Compras Públicas.....	202
7 CONSIDERAÇÕES FINAIS	210
REFERÊNCIAS.....	214
APÊNDICE A – PROTOCOLO DE REVISÃO SISTEMÁTICA DE LITERATURA .	224

1 INTRODUÇÃO

A organização e a representação do conhecimento constituem fundamentos essenciais da Ciência da Informação, sobretudo em uma era marcada pelo aumento exponencial de dados digitais e pela complexidade das demandas informacionais nos setores público e privado (Brascher; Café, 2008; Hjørland, 2008; Glushko, 2013; Lima; Maculan, 2024).

No campo da gestão pública, essas questões tornam-se ainda mais críticas. Com a consolidação do governo eletrônico como estratégia de modernização administrativa, os órgãos estatais passaram a produzir e compartilhar volumes massivos de dados que, para serem úteis, precisam estar organizados, acessíveis e semanticamente interoperáveis (Cruz, 2001; Castro *et al.*, 2017; Silva *et al.*, 2020; Mcguinness; O'neill; Cross, 2020).

Nesse contexto, as ontologias — entendidas como representações formais e compartilhadas de um domínio do conhecimento (Gruber, 1993) — emergem como ferramentas promissoras para estruturar, integrar e representar dados de maneira padronizada e inteligível por humanos e sistemas. Sua aplicação no governo eletrônico tem potencial para aprimorar a recuperação da informação, diminuir redundâncias, qualificar a interoperabilidade entre sistemas e ampliar a transparência institucional (Wang; Wang; Wang, 2019). No entanto, apesar de seu potencial estratégico, a adoção de ontologias no contexto governamental brasileiro ainda é tímida, dispersa e desprovida de uma diretriz metodológica estruturada.

Embora reconhecidas por seu valor teórico e prático, ontologias apresentam desafios expressivos quando construídas manualmente. O processo é tecnicamente exigente, intensivo em tempo e altamente dependente de especialistas (Deb *et al.*, 2018; Nguyen; Tettamanzi, 2019; Mizoguchi; Sunagawa; Kozaki, 2019; Drissi *et al.*, 2021; Sattar *et al.*, 2021; Rajaonarivo; Arakawa, 2021). Tais características tornam sua adoção restrita, especialmente em países com infraestrutura tecnológica limitada ou carência de profissionais capacitados. A literatura especializada já aponta que o avanço da aprendizagem de ontologias — sobretudo por meio de abordagens automatizadas e semiautomatizadas — pode ser determinante para ampliar sua aplicabilidade e viabilidade (Yadav *et al.*, 2018; Enea *et al.*, 2018). Além disso, observa-se que, embora muitas metodologias de construção de ontologias abranjam etapas como análise de domínio, conceituação, implementação e avaliação, há uma

carência de detalhamento quanto às técnicas e atividades envolvidas. Persistem lacunas significativas em aspectos fundamentais, como manutenção, documentação, colaboração, reutilização, interoperabilidade e estimativa de recursos humanos, o que limita sua aplicabilidade em contextos mais exigentes (Sattar *et al.*, 2021).

Diante desse cenário, esta pesquisa parte da seguinte pergunta: como uma metodologia de aprendizagem de ontologias pode contribuir para a organização e a representação do conhecimento no governo eletrônico brasileiro? Para respondê-la, propõe-se a construção da Onto Merge Methodology (OMM), fundamentada na análise crítica de metodologias consolidadas, visando identificar fases e etapas passíveis de automatização. A OMM busca integrar boas práticas metodológicas com recursos computacionais para tornar o processo de construção de ontologias mais ágil, replicável e adaptável.

O governo eletrônico demanda soluções organizacionais robustas, capazes de lidar com altos volumes de dados e de integrar serviços entre instituições de forma transparente e eficiente. Nesse sentido, a OMM foi desenhada com base em princípios de modularidade, clareza metodológica e potencial de automação. Para validar sua aplicabilidade, uma de suas fases foi implementada no domínio das compras públicas, setor que envolve regulamentações complexas, requisitos de transparência e intensa produção de dados estruturados e não estruturados.

Assim, esta tese pretende oferecer uma contribuição teórica e prática ao campo da Ciência da Informação, com especial atenção ao domínio da aprendizagem de ontologias e à qualificação da gestão do conhecimento no setor público digital brasileiro.

1.1 Problema

A presente pesquisa parte de uma problemática concreta e recorrente no contexto da administração pública brasileira: a carência de métodos estruturados e adaptáveis para a construção de ontologias voltadas para a organização e para a representação do conhecimento em ambientes governamentais. Em um cenário marcado pela crescente digitalização dos serviços públicos, produção massiva de dados e busca por maior transparência e eficiência administrativa, torna-se indispensável desenvolver soluções que permitam integrar e reutilizar informações de

maneira consistente, padronizada e semântica (Cruz, 2001; Castro *et al.*, 2017; Silva *et al.*, 2020; Mcguinness; O'neill; Cross, 2020).

A adoção de ontologias nesse contexto representa uma oportunidade estratégica. Esses instrumentos viabilizam a construção de um vocabulário comum entre sistemas, facilitando a interoperabilidade e o reuso de dados por diferentes órgãos e plataformas. No entanto a realidade observada nas instituições públicas brasileiras ainda revela um cenário de fragmentação tecnológica e informacional. A ausência de uma política nacional orientada à modelagem do conhecimento, aliada à baixa maturidade institucional no uso de recursos semânticos, compromete o potencial das ontologias como Sistemas de Organização do Conhecimento (SOC) eficazes e reutilizáveis.

Mesmo nos casos em que ontologias são desenvolvidas, predominam iniciativas isoladas, desarticuladas entre si e, frequentemente, desenvolvidas de forma *ad hoc*. Tal situação torna o processo de modelagem do conhecimento dependente de profissionais altamente especializados, o que dificulta a replicabilidade, eleva custos e compromete a escalabilidade dos resultados.

A literatura consultada aponta para avanços na área da aprendizagem de ontologias, especialmente por meio de abordagens automatizadas e semiautomatizadas, baseadas em técnicas de extração de conhecimento, mineração de texto e processamento de linguagem natural (Deb *et al.*, 2018; Rodriguez; Aguilar, 2018; Yadav *et al.*, 2018; Noura *et al.*, 2019; Salatino *et al.*, 2020; Gómez-Suta; Echeverry-Correa; Soto-Mejía, 2020; Santosa; Miyazaki; Han, 2021; Sadeghianasl *et al.*, 2021; Drissi *et al.*, 2021). Entretanto esses avanços ainda não se traduziram, de forma significativa, em soluções metodológicas adaptadas à realidade do governo eletrônico brasileiro.

Nesse sentido, o problema desta pesquisa reside na ausência de uma metodologia que, ao mesmo tempo, seja sistemática, flexível e parcialmente automatizável, capaz de orientar o processo de aprendizagem de ontologias no setor público. A resposta a essa lacuna exigiu a formulação de uma proposta metodológica — a Onto Merge Methodology (OMM) — desenvolvida com base na análise de quatro metodologias existentes, buscando integrar boas práticas e identificar oportunidades de automatização. A aplicação da OMM no domínio das compras públicas visa validar sua aplicabilidade e evidenciar seu potencial de generalização para outros setores da administração pública.

1.2 Justificativa

A presente pesquisa justifica-se pela necessidade urgente de superar as limitações metodológicas que comprometem a organização e a representação do conhecimento em ambientes governamentais no Brasil. Em um contexto global no qual a informação se consolida como um ativo estratégico, a capacidade de estruturar dados de forma semântica, acessível e interoperável é indispensável para a construção de governos mais transparentes, responsivos e centrados no cidadão.

Experiências internacionais demonstram que a adoção de Sistemas de Organização do Conhecimento (SOC), em especial as ontologias, tem desempenhado papel crucial na modernização de governos digitais. A análise dos países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI) da ONU evidencia investimentos consistentes em estruturas semânticas interoperáveis e ontologias reutilizáveis. Entretanto, no Brasil, apesar de avanços na disponibilização de dados abertos, a adoção de ontologias ainda é tímida, desarticulada e carente de diretrizes metodológicas consolidadas.

Essa lacuna decorre, em parte, da dificuldade em replicar processos manuais de construção de ontologias, que demandam alta especialização técnica, tempo e recursos. A ausência de metodologias adaptadas à realidade institucional brasileira e com potencial de automatização contribui para a fragmentação de iniciativas e para a limitação de sua escalabilidade e efetividade (Asim *et al.*, 2018; Liao; Wu; King, 2019).

Diante disso, a Onto Merge Methodology (OMM) foi concebida para responder à necessidade de uma abordagem sistemática, adaptável e escalável. A OMM resulta da integração de boas práticas metodológicas extraídas de quatro propostas desenvolvidas no PPGGOC/UFMG — OntoForInfoScience, OntONeo, OntoAccount e PRONTO — com foco na automatização parcial da fase de aquisição de conhecimento, otimizando o processo e reduzindo a dependência de especialistas.

A originalidade desta pesquisa está na articulação entre teoria, prática e tecnologia: contribui com o campo emergente de *ontology learning*; propõe soluções viáveis de automação; e valida a proposta em um domínio real e relevante, como o das compras públicas. Ao fazer isso, fortalece o campo da Arquitetura e Organização do Conhecimento e oferece contribuições significativas para a inovação na gestão pública, promovendo uma administração mais eficiente e orientada por dados.

1.3 Pressupostos

Tendo em vista o problema de pesquisa apresentado, são levantados os pressupostos a seguir.

O primeiro pressuposto que orienta esta pesquisa é que os estudos no Brasil sobre geração automática ou semiautomática de ontologias são poucos. Embora as ontologias sejam reconhecidas como ferramentas fundamentais para a organização e representação do conhecimento, a produção científica nacional ainda apresenta uma lacuna no que se refere ao desenvolvimento de abordagens que utilizem técnicas automatizadas para a sua construção. Essa escassez limita o avanço da área e demonstra a necessidade de propostas metodológicas que incorporem tecnologias emergentes no processo de aprendizagem de ontologias.

O segundo pressuposto considera que as iniciativas de construção de ontologias no contexto do governo brasileiro podem ser consideradas ações isoladas e sem a adoção de uma metodologia unificada. Apesar da existência de experiências pontuais em diferentes instituições públicas, essas ações não seguem um modelo metodológico comum que favoreça a padronização, a interoperabilidade e a reutilização dos artefatos ontológicos. Essa fragmentação resulta em soluções desintegradas, o que compromete a consistência semântica dos dados governamentais e limita a criação de uma infraestrutura de conhecimento compartilhada entre os diversos órgãos e esferas do governo.

O terceiro pressuposto baseia-se na constatação de que os países que estão mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU adotam algum tipo de ontologia para organização e representação do conhecimento e da informação. A utilização de ontologias nesses contextos tem contribuído para a promoção de serviços digitais mais integrados, interoperáveis e orientados ao cidadão. Esses países têm investido em soluções ontológicas como parte de estratégias estruturadas de transformação digital, o que sugere que a presença de tais instrumentos está associada a níveis mais elevados de maturidade em governo eletrônico.

Por fim, o quarto pressuposto assume que a inexistência de uma metodologia que automatize etapas do processo de construção de ontologia pode ser um dos fatores que induz ao baixo uso desse tipo de Sistema de Organização do Conhecimento (SOC) no contexto do Governo Eletrônico no Brasil. A construção de

ontologias é uma tarefa complexa e, quando realizada de forma inteiramente manual, torna-se demorada, custosa e de difícil replicação. A ausência de metodologias que ofereçam mecanismos automatizados ou semiautomatizados pode desestimular sua adoção, sobretudo em ambientes institucionais que demandam escalabilidade e eficiência.

1.4 Objetivos

Buscando-se esclarecer o problema de pesquisa apresentado, a seguir são descritos o objetivo geral e os objetivos específicos desta pesquisa.

1.4.1 Objetivo geral

Elaborar uma metodologia de aprendizagem de ontologias para organização e representação do conhecimento e da informação no contexto do governo eletrônico brasileiro.

1.4.2 Objetivos específicos

- Evidenciar o estágio atual da pesquisa com relação à aprendizagem de ontologias.
- Identificar os Sistemas de Organização do Conhecimento que têm sido adotados nos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU.
- Inventariar ontologias desenvolvidas no contexto do governo brasileiro.
- Identificar possibilidades de automatização de fases e etapas de quatro metodologias específicas de construção de ontologias.

A fim de ampliar o conhecimento sobre o campo de pesquisa em relação à geração automática e semiautomática de ontologias, conduziu-se uma Revisão Sistemática de Literatura, conforme detalhado nos procedimentos descritos no Apêndice A.

Para identificar os SOC adotados nos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU e no Brasil, foi

realizada uma análise documental conduzida mediante a consulta de dados abertos disponíveis na internet.

No intuito de compilar um inventário das ontologias desenvolvidas no contexto do governo brasileiro, foi também realizada uma análise documental utilizando dados abertos disponíveis na internet.

Com o propósito de identificar oportunidades de otimização nas fases e etapas da geração manual de ontologias desenvolvidas e catalogadas no contexto do governo brasileiro, foi realizada uma análise documental visando inventariar e documentar ontologias desenvolvidas nesse contexto.

Por fim, foi também realizada análise documental de quatro metodologias para construção de ontologias no intuito de identificar fases, etapas e demais elementos previstos em cada proposta que possam compor a nova metodologia proposta nesta pesquisa.

1.5 Estrutura da tese

Esta pesquisa está estruturada em sete capítulos, organizados da seguinte forma:

No Capítulo 1, são apresentados a introdução, a justificativa, o problema de pesquisa, os objetivos e a estrutura da tese.

No Capítulo 2, é apresentada a revisão de literatura, abrangendo os procedimentos adotados, a apresentação dos estudos selecionados, as metodologias de aprendizagem de ontologias, etapas/fases das metodologias identificadas, vantagens e diferenciais dessas metodologias, bem como as lacunas existentes e os principais resultados evidenciados pelos autores.

No Capítulo 3, é apresentado o referencial teórico e metodológico da pesquisa, que engloba temáticas como organização e representação do conhecimento e da informação, SOC, ontologias e governo eletrônico.

No Capítulo 4, é apresentada a caracterização e a descrição da metodologia, contemplando a caracterização da pesquisa, a delimitação do seu universo, os métodos e os procedimentos metodológicos adotados.

No Capítulo 5, são apresentados os resultados provenientes dos procedimentos metodológicos, contemplando a análise documental das práticas de organização e representação do conhecimento adotadas nos dez países mais bem

posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da Organização das Nações Unidas; a análise documental das ontologias desenvolvidas no contexto do governo brasileiro e o resultado da análise das quatro metodologias de construção de ontologias selecionadas.

No Capítulo 6, buscou-se apresentar uma proposta inicial de automatização de parte das etapas de construção de ontologias, incluindo possíveis fases e etapas, bem como a apresentação de uma prova de conceito.

Por fim, no Capítulo 7, são apresentadas as considerações finais, as limitações da pesquisa e as sugestões de trabalhos futuros.

2 REVISÃO DE LITERATURA

Este capítulo tem como objetivo identificar metodologias sobre aprendizagem de ontologias, incluindo etapas e métodos, vantagens e lacunas apresentadas pelos autores, bem como os principais resultados. O conceito de metodologia adotado nesta pesquisa corresponde a um conjunto de procedimentos, técnicas, ferramentas e documentos que podem auxiliar os autores no processo de implementação de um novo sistema de informação, podendo ser organizado em fases, que se subdividem em subfases ou etapas.

2.1 Introdução

Nesta seção, é apresentada uma Revisão Sistemática da Literatura (RSL) sobre a geração automática e semiautomática de ontologias, área conhecida como aprendizagem de ontologias ou *ontology learning*, buscando agregar conhecimento sobre as abordagens, metodologias e ferramentas mais recentes. A revisão de artigos científicos visa identificar metodologias, etapas, vantagens e lacunas apontadas pelos autores, bem como os principais resultados. Compreender a dinâmica e o estado atual da pesquisa nesse campo do conhecimento contribuirá para o desenvolvimento e o aprimoramento dessa área na representação do conhecimento.

Para a organização e gestão das referências, utilizou-se o *software* Mendeley¹, enquanto o Atlas TI² foi empregado para a análise qualitativa dos estudos selecionados. A pesquisa visou identificar trabalhos que aplicaram métodos de aprendizagem de ontologias entre 2018 e 2023, abordando questões como a identificação de metodologias específicas, etapas envolvidas, vantagens, diferenciais e lacunas existentes. O período selecionado justifica-se em razão de possível obsolescência de métodos e técnicas computacionais, pressupondo-se, assim, que os trabalhos publicados nos últimos cinco anos poderiam ser mais relevantes. As bases de dados consultadas incluem *ACM Digital Library*, *IEEE Xplore Digital Library*,

¹ Mendeley é um *software* gratuito de gestão de referências e PDFs, que permite organizar, citar e compartilhar fontes acadêmicas. Integra-se com editores de texto e facilita a colaboração entre pesquisadores. Disponível em: www.mendeley.com. Acesso em: 27 abr. 2025.

² O Atlas.ti é um *software* utilizado para análise qualitativa de dados, permitindo organizar, codificar e interpretar informações provenientes de textos, entrevistas, documentos, áudios e vídeos, com o objetivo de identificar padrões, categorias e significados no material analisado. O *software* pode ser acessado por meio do seguinte endereço: <https://atlasti.com>. Acesso em: 27 abr. 2025.

Scopus Preview e *Web of Science*, utilizando termos de busca em inglês e português para garantir a abrangência e a relevância dos resultados.

O protocolo da RSL definiu critérios de inclusão e exclusão para assegurar a qualidade e a pertinência dos estudos analisados. Foram considerados apenas trabalhos com relação direta à temática de aprendizagem de ontologias, disponíveis integralmente e publicados em inglês, português ou espanhol. A exclusão de duplicatas e a filtragem de trabalhos irrelevantes resultaram em 183 textos, reduzidos a 170 após a verificação de disponibilidade do texto completo. A análise final concentrou-se em 22 estudos, selecionados pela sua relevância e aderência aos critérios estabelecidos. O detalhamento dos procedimentos metodológicos adotados na revisão está descrito na Seção 4 – Metodologia, e o Protocolo completo utilizado está disponível no Apêndice A.

Esta revisão contextualiza a importância da geração automática de ontologias, destacando sua aplicação prática e potencial de inovação no campo do governo eletrônico. A revisão detalhada das metodologias existentes permite identificar avanços e desafios, contribuindo para o desenvolvimento de abordagens mais eficazes e robustas. A clareza metodológica e a replicabilidade são aspectos fundamentais para a evolução dessa área, proporcionando uma base sólida para futuras pesquisas e implementações práticas.

2.2 Procedimentos metodológicos adotados na Revisão de Literatura

Os procedimentos metodológicos para a Revisão Sistemática da Literatura foram respaldados no método proposto por Kitchenham (2004). Para organização e gestão das referências, foi utilizado o *software* Mendeley. O *software* Atlas TI foi utilizado para análise qualitativa do material selecionado. As pesquisas referenciadas foram selecionadas considerando maior aproximação dos resultados apresentados com a temática desta pesquisa e estão organizadas nas categorias de análise previamente identificadas no protocolo da revisão, descrito no Quadro 1, a seguir.

Quadro 1 – Resumo Protocolo da Revisão Sistemática

TEMÁTICA PRINCIPAL	GERAÇÃO AUTOMÁTICA DE ONTOLOGIAS
Áreas de Conhecimento	Ciência da Informação, Ciência da Computação.
Protocolo de Revisão Sistemática da Literatura (PRSL)	Apresenta a estrutura metodológica para a execução da Revisão da Literatura sobre métodos para o processo de aprendizagem de ontologias.
Pergunta Principal do Estudo	Quais trabalhos aplicaram métodos de aprendizagem de ontologias entre os anos de 2018 e 2023?
Perguntas Secundárias	(1) É possível identificar a utilização de alguma metodologia de aprendizagem de ontologias?
	(2) Quais as etapas/fases da metodologia ora adotada?
	(3) Quais as vantagens e os diferenciais da metodologia adotada?
	(4) Quais as lacunas existentes nessas metodologias?
	(5) Quais os principais resultados evidenciados?

Fonte: elaborado pela autora (2024).

Partindo-se das perguntas de pesquisa, seguiu-se para o levantamento dos termos de busca, considerando inicialmente as variações nos idiomas inglês e português, conforme disposto no Quadro 2, a seguir.

Quadro 2 – Termos de busca Etapa 1

TERMOS EM INGLÊS	TERMOS EM PORTUGUÊS
[Ontology Learning] OR [Ontologies Learning]	[Aprendizagem de Ontologia] OR [Aprendizagem de Ontologias]
[Ontology Engineering] OR [Ontologies Engineering]	[Engenharia de ontologia] OR [Engenharia de ontologias]
[Ontology Extraction] OR [Ontologies Extraction]	[Extração de Ontologia] OR [Extração de Ontologias]
[Automatic ontology Generation] OR [Automatic ontologies Generation]	[Geração automática de ontologia] OR [Geração automática de ontologias]

Fonte: elaborado pela autora (2024).

Após a definição dos termos de busca, considerando a aderência à temática da pesquisa, foi realizado o levantamento das bases de dados, chegando-se às seguintes bases: *ACM Digital Library*; *IEEE Xplore Digital Library*; *Scopus Preview* e *Web of Science*. Inicialmente, executaram-se as expressões de busca indicadas no Quadro 3

em cada uma das bases de dados. A partir da lista de estudos retornados, os resultados foram exportados em formato RIS para importação na ferramenta Mendeley.

Na primeira etapa de seleção dos textos, foram excluídas as duplicatas, resultando-se em 183 textos. Considerando a exclusão dos trabalhos que não disponibilizavam o texto completo, chegou-se a um total de 170 textos. Essas etapas foram realizadas entre os meses de setembro e outubro de 2023.

Em seguida, foram identificados os trabalhos aderentes aos critérios de inclusão e exclusão, indicados nos Quadros 4 e 5, refletindo o *corpus* final para a etapa de análise, chegando-se a 22 textos. Essa etapa foi realizada entre novembro de 2023 e janeiro de 2024.

Quadro 3 – Critérios de Inclusão

CRITÉRIO	DESCRIÇÃO DO CRITÉRIO DE INCLUSÃO
CI1	Título possui relação com a temática aprendizagem de ontologias.
CI2	Resumo possui relação com a temática aprendizagem de ontologias.
CI3	Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas.
CI4	Serão incluídos trabalhos que tenham aplicado um método de geração automática ou semiautomática de ontologias.
CI5	Trabalhos nos idiomas inglês, português e espanhol.
CI6	Período de publicação compreendido entre 2018 e 2023.
CI7	Serão considerados trabalhos dos seguintes tipos: artigo de periódico, capítulo de livro, artigo de evento.

Fonte: elaborado pela autora (2024).

Quadro 4 – Critérios de Exclusão

CRITÉRIO	DESCRIÇÃO DO CRITÉRIO DE EXCLUSÃO
CE1	Título não possui relação com a temática aprendizagem de ontologias.
CE2	Resumo não possui relação com a temática aprendizagem de ontologias.
CE3	Serão excluídos trabalhos que não apresentem resumo/ <i>abstract</i> .
CE4	Serão excluídos <i>short papers</i> e trabalhos disponibilizados em bases preprint.
CE5	Serão excluídos documentos que não estejam em português, inglês ou espanhol.
CE6	Serão excluídos trabalhos que não estejam disponíveis em texto completo.
CE7	Serão excluídos os trabalhos que não apliquem o processo de geração automática ou semiautomática de ontologias.

Fonte: elaborado pela autora (2024).

Finalizada a etapa de extração dos dados, foram realizadas a análise, a interpretação e a documentação dos resultados encontrados nos estudos, tendo em vista a pergunta principal e as perguntas secundárias previstas nesta RSL. Essa etapa foi realizada entre os meses de janeiro e fevereiro de 2024 com o apoio do *software* Atlas TI, conforme disposto a seguir.

2.3 Trabalhos relacionados a metodologias de aprendizagem de ontologias

Com o avanço da tecnologia da informação, a ontologia tornou-se uma ferramenta amplamente aplicada em diversos campos para a representação do conhecimento. Diante desse cenário, a construção e a extensão de ontologias emergem como uma área de pesquisa de grande relevância. Yadav *et al.* (2018) abordam a importância desse campo, destacando a necessidade de ir além do pensamento de alguns engenheiros de ontologia, promovendo a criação de ontologias comuns que atendam às exigências de diferentes usuários em domínios específicos. A colaboração é uma peça-chave nesse processo, visto que a engenharia de ontologia deve ser um esforço conjunto para possibilitar um desenvolvimento mais ágil. Com o crescimento da Web Social, a simplicidade tem se mostrado eficaz na atração de uma participação em massa. Nesse contexto, os autores propõem a plataforma "EasyOnto", que oferece uma interface gráfica simples e acessível. Essa plataforma permite que os usuários contribuam de maneira colaborativa para o desenvolvimento de uma ontologia semiformal, facilitando assim a representação do conhecimento de forma eficiente e intuitiva. O estudo explora as nuances desse projeto e suas implicações na evolução da ontologia e compartilhamento de informações na era digital.

Gyrard, Zimmermann e Sheth (2018) focam na utilização de catálogos de ontologias no processo de *design* de aplicações voltadas para a Internet das Coisas (IoT) em contextos de cidades inteligentes. Apresentam uma análise de quatro catálogos de ontologias relevantes, enriquecendo-os com ontologias correlatas, avaliando sua utilidade e qualidade e delineando sugestões para possíveis melhorias. As contribuições do artigo englobam o desenvolvimento de catálogos de ontologias, o aprimoramento da qualidade das ontologias, o enriquecimento dos catálogos e a avaliação específica de um catálogo em particular. O catálogo proposto, LOV4IoT, emerge como uma coleção abrangente, composta por mais de 440 ontologias

relacionadas à IoT e cidades inteligentes. Esse catálogo facilita a reutilização e mantém-se em atualização contínua.

Deb *et al.* (2018) apresentam um estudo que delinea um modelo voltado ao aprendizado de ontologia a partir de dados taxonômicos no contexto do sistema agrícola. Em consonância com a proposta, são apresentadas diversas referências sobre construção automática e semiautomática de ontologias, abrangendo um espectro temporal e geográfico variado. Os temas englobam questões pertinentes ao aprendizado de ontologia e tópicos correlatos, incluindo, mas não se limitando a editores de ontologia, sistemas de suporte, *crawlers* e sistemas de classificação de texto. Com base no conjunto de pesquisas já realizadas, os autores destacam as limitações inerentes aos sistemas preexistentes de aprendizado de ontologia, culminando na proposição de uma abordagem que se concentra especificamente em textos de natureza taxonômica. Os autores detalham a arquitetura do *framework* proposto, que incorpora algoritmos dedicados à extração de termos, conceitos, relações hierárquicas e não hierárquicas, bem como ao mapeamento da ontologia.

Em Rodriguez e Aguilar (2018), é apresentado um sistema destinado à extração de conhecimento proveniente de documentos no formato *HyperText Markup Language* (HTML) não estruturados. Para efetuar essa tarefa, empregam-se técnicas avançadas de Processamento de Linguagem Natural (PLN) com o objetivo de converter o texto contido nos documentos em conhecimento estruturado, o qual é semanticamente descrito em uma ontologia. A organização do conhecimento extraído é viabilizada por meio da aplicação de um Grafo de Aprendizagem. O artigo fornece exemplos concretos de aplicações práticas desse sistema e oferece uma descrição detalhada da arquitetura que o sustenta.

Malchiodi e Tettamanzi (2018) abordam a temática concernente à previsão de pontuações atribuídas a axiomas candidatos no âmbito do aprendizado de ontologias. A abordagem proposta fundamenta-se em um algoritmo modificado de agrupamento de vetores de suporte, com uma medida de similaridade semântica aplicada a axiomas. O método desenvolvido revela eficácia ao aprender pontuações em uma base de conhecimento composta por axiomas candidatos da Linguagem de Ontologia Web (OWL). O texto oferece uma exposição sobre o problema do aprendizado de ontologias, as heurísticas de pontuação aplicáveis a axiomas e a medida de similaridade adotada. Os resultados experimentais corroboram a viabilidade e a eficácia do método proposto. A abordagem apresentada focaliza a computação de

similaridades entre axiomas de subordinação, considerando tanto as afirmativas quanto as negativas, no contexto da linguagem OWL. A computação dessas similaridades é realizada por meio de consultas SPARQL de contagem. Os experimentos conduzidos revelam que a inferência de pontuações probabilísticas para axiomas candidatos OWL, com base na metodologia proposta, produz resultados altamente satisfatórios. O artigo propõe um método para inferir um preditor destinado à pontuação probabilística de axiomas de subordinação candidatos na linguagem OWL. O algoritmo de aprendizado é concebido para mapear a tarefa de inferência à estimativa da função de associação a um conjunto difuso. A aplicação desse procedimento em uma base de conhecimento demonstra um desempenho notável, evidenciado pela precisão dos preditores induzidos. Ademais, os resultados destacam um subconjunto de axiomas que se revelam mais desafiadores de aprender quando comparados ao restante da base de conhecimento.

Já Malchiodi, da Costa Pereira e Tettamanzi (2018) propuseram uma metodologia para antecipar a pontuação probabilística de axiomas na linguagem OWL, utilizando a técnica de regressão de vetor de suporte. Os autores concentram-se especificamente em axiomas de subordinação, empregando uma medida de similaridade inspirada para atribuir pontuações a esses axiomas. A comparação dos resultados obtidos com um método de inferência fundamentado na indução de associação difusa revela que a abordagem da regressão de vetor de suporte supera tanto em precisão quanto em eficiência computacional. Entretanto destacam a possível presença de fórmulas que variam em dificuldade de aprendizado, indicando a necessidade de investigações adicionais para compreender esse fenômeno. Os pesquisadores expandiram a análise para incluir axiomas adicionais, empregando a técnica de regressão, identificando assim 30 axiomas desafiadores, alguns dos quais se sobrepõem aos encontrados no estudo original. A conclusão é que a técnica de regressão adotada oferece resultados superiores em comparação com outras técnicas, além de proporcionar um tempo de treinamento mais eficiente. A observação sobre a existência de um subconjunto de axiomas mais complexos de pontuar é reiterada, sugerindo que investigações futuras poderiam explorar diferentes métodos de previsão e reformular o problema de pontuação como uma tarefa de classificação binária.

Aggoune (2018), por sua vez, apresenta uma abordagem para a aprendizagem automática de ontologias a partir de bancos de dados relacionais heterogêneos,

focalizando especificamente no contexto dos riscos alimentares. A estratégia proposta se baseia no uso do WordNet e na medida de similaridade semântica denominada Wup para avaliar a afinidade semântica entre elementos presentes em bancos de dados. O artigo oferece uma descrição do processo para a geração e evolução da ontologia, destacando sua aplicabilidade em ambientes que envolvem dados relacionados a riscos alimentares. A avaliação da abordagem comparativamente a ferramentas similares revela sua capacidade de gerar ontologias de elevada qualidade a partir de múltiplos bancos de dados relacionais. A abordagem proposta se sustenta em quatro processos sequenciais, a saber: geração de classes, geração de propriedades de tipo de dado, geração de propriedades de objeto e integração de registros. A implementação prática da abordagem, concretizada em um protótipo, é comparativamente analisada com outras ferramentas similares, evidenciando sua eficácia e eficiência.

Enea *et al.* (2018) propõem uma abordagem para aprimorar o processo de criação e enriquecimento de ontologias, envolvendo operadores humanos. A estratégia delineada compreende a extração automática de triplas de fontes heterogêneas, as quais são subsequentemente submetidas à validação por um operador humano. O sistema incorpora representações gráficas do nível de incerteza associado à ontologia, proporcionando ao usuário a execução de consultas hipotéticas realizadas por meio da linguagem *SPARQL Protocol and RDF Query Language*. O operador humano desempenha um papel estratégico ao aceitar ou rejeitar triplas com base em sua análise pessoal, sendo essencial na conclusão eficiente da tarefa de criação da ontologia. O artigo também apresenta o LAM (Modulador de Algoritmo de Aprendizado), uma ferramenta que ajuda a transformar os resultados de um modelo automático de ontologia (LOM) em versões mais consistentes e utilizáveis. Esse processo é feito em conjunto com o usuário, que pode interagir com o sistema por meio de uma interface simples. O LAM permite que a pessoa defina filtros, faça perguntas ao sistema (“e se...?”) e avalie as respostas obtidas. Além disso, é possível visualizar, alterar e confirmar informações da ontologia. Os dados são armazenados em um banco gráfico temporário, facilitando o acesso e a manipulação das informações durante o uso da ferramenta. A interface do usuário apresenta gráficos para visualizar a distribuição de incertezas na ontologia, bem como a proporção de declarações certas e incertas em uma consulta SPARQL. A abordagem proposta pelo LAM confere flexibilidade ao processo de aprendizado e validação de ontologias,

possibilitando que os usuários refinem e validem iterativamente a ontologia aprendida com base em seu conhecimento e suposições.

Em Nguyen e Tettamanzi (2019), é proposta uma abordagem evolutiva fundamentada na utilização da técnica computacional de Evolução Gramatical e da Teoria Matemática da Possibilidade, visando à descoberta automática de axiomas de classes disjuntos em ontologias a partir de dados no formato RDF. A metodologia proposta foi submetida a uma avaliação comparativa em relação a um Padrão Ouro construído manualmente por especialistas, evidenciando elevada precisão e abrangência satisfatória. No decorrer do artigo, são abordados trabalhos correlatos no âmbito do aprendizado de ontologia e nos fundamentos da Web Semântica. Os autores, mediante a análise e experimentação, concluem que a abordagem proposta se revela eficaz na descoberta de axiomas de classes disjuntos em conjuntos de dados RDF, embora sugiram a possibilidade de aprimoramentos para enfrentar axiomas de maior complexidade.

Vedula, Maneriker e Parthasarathy (2019), por sua vez, apresentam um *framework* intitulado BOLT-K, que adota uma abordagem baseada em *Long Short-Term Memory* (LSTM) para o aprendizado de ontologias em um domínio específico. Esse *framework* realiza uma adaptação e transferência de conhecimento proveniente de um domínio relacionado, empregando-o para inicializar o processo de aprendizado ontológico. O artigo provê detalhes elucidativos acerca dos componentes do *framework*, procedimentos de treinamento adotados e técnicas de aumento de dados implementadas. Adicionalmente, apresenta os resultados de avaliação, promovendo uma comparação entre o desempenho do BOLT-K e outras abordagens de referência. O trabalho inclui representações gráficas e tabelas que ilustram o desempenho e o impacto de diferentes estratégias adotadas pelo *framework* BOLT-K.

Noura *et al.* (2019) propõem uma metodologia denominada KE4WoT, destinada à extração automatizada de conhecimento com o propósito de edificar aplicações na esfera da Web Semântica das Coisas (WoT). A metodologia abarca etapas que compreendem a seleção de ontologias, seu pré-processamento, a extração do vocabulário, o cálculo de frequências de termos e a aplicação de um algoritmo para a identificação de associações entre palavras. Os tópicos resultantes dessa extração são suscetíveis de serem empregados para descrever ontologias preexistentes e aprimorar a interoperabilidade. O artigo ainda incorpora três cenários que visam

ilustrar a aplicação do conhecimento extraído. A metodologia é submetida à avaliação e evidencia sua eficácia na identificação de conceitos pertinentes.

Gómez-Suta, Echeverry-Correa e Soto-Mejía (2020) propõem uma metodologia destinada à extração e validação de estruturas ontológicas a partir de textos em língua espanhola, com um enfoque específico no contexto do conflito armado colombiano. A metodologia compreende etapas como o pré-processamento textual, a seleção de termos índice e a extração de conceitos e relações por meio de técnicas de agrupamento semântico. Diferentes propostas de ponderação estatística são comparadas para a construção do vocabulário, e as estruturas extraídas são avaliadas por meio de diversas métricas. Os experimentos são conduzidos em uma plataforma específica, e os resultados e as abordagens implementadas estão disponíveis para análises adicionais. O artigo também oferece uma extensa lista de referências relacionadas à criação de ontologias, agrupamento semântico de documentos de texto, extração de conceitos e aprendizado ontológico. Essas referências abordam diversos aspectos desses temas e incluem discussões sobre a aplicação dessas técnicas em domínios específicos, tais como violações de direitos humanos, análise de texto em mídia noticiosa e desenvolvimento de cidades inteligentes. Essa fundamentação bibliográfica robusta enriquece a compreensão das abordagens adotadas no presente estudo, situando-o dentro do contexto mais amplo da pesquisa científica.

Salatino *et al.* (2020) apresentam a *Computer Science Ontology (CSO)*, uma ontologia elaborada de maneira automática para as áreas de pesquisa em ciência da computação. A construção da CSO envolve a aplicação de técnicas de aprendizado de máquina e processamento de linguagem natural em uma extensa base de dados composta por artigos científicos. Esse sistema abarca uma vasta gama de tópicos que não são contemplados em outras classificações existentes, destacando-se pela capacidade de atualização automática. Segundo os autores, a CSO tem sido empregada em diversas aplicações, incluindo a classificação automática de artigos de pesquisa, a identificação de comunidades de pesquisa e a projeção de tendências em investigação científica. Adicionalmente, o artigo apresenta duas ferramentas associadas à CSO: o *CSO Classifier*, destinado à classificação automática de artigos, e o *CSO Portal*, uma aplicação Web que possibilita a exploração e o fornecimento de *feedback* sobre a ontologia. A avaliação abrangente da CSO e de suas ferramentas destaca sua eficácia e utilidade em diversas tarefas específicas no contexto científico.

Alharbi, Tamma e Grasso (2021) investigam a disparidade entre a teoria e a prática no âmbito da reutilização de ontologias. Para tal, um questionário foi aplicado a desenvolvedores e profissionais especializados em ontologia, com o intuito de coletar percepções acerca de abordagens preferenciais para a reutilização, tipos de ontologias frequentemente empregadas nesse processo, critérios fundamentais na decisão de qual ontologia reutilizar e os principais obstáculos enfrentados nesse contexto. Os resultados revelam que há uma demanda premente por diretrizes abrangentes. Observou-se que os desenvolvedores com vasta experiência têm preferência por reutilizar ontologias de nível superior, enquanto aqueles com conhecimento moderado priorizam a reputação e a expertise dos desenvolvedores responsáveis pela ontologia. Conflitos com outras ontologias importadas e a complexidade inerente ao processo são identificados como os principais obstáculos à reutilização. Os resultados desse estudo possuem relevância significativa ao fornecer orientações para o desenvolvimento de melhores práticas e diretrizes voltadas para a reutilização de ontologias na comunidade da Web Semântica.

Santosa, Miyazaki e Han (2021) expõem uma abordagem integralmente automatizada destinada à expansão de ontologias no domínio da ciência da computação por meio da aplicação de técnicas de classificação. Os autores conduzem uma análise comparativa entre duas estratégias distintas, denominadas "Direta" e "Indireta", e avaliam sua eficácia empregando modelos convencionais de aprendizado de máquina e *deep learning*. Adicionalmente, introduzem novos procedimentos de avaliação, pautados na referência da *Wikipedia*, para a análise da tarefa de extensão ontológica. Os resultados obtidos revelam que a abordagem Direta, aliada a um modelo denominado *Naive Bayes Gaussiano*, é capaz de gerar os *links* mais válidos e confiáveis. A conclusão do artigo é fundamentada na expansão da ontologia de ciência da computação, incorporando novos tópicos e *links* de relevância.

Na pesquisa de Sattar *et al.* (2021), é apresentada uma metodologia destinada ao desenvolvimento de ontologias em contextos dos chamados Mundos Institucionais Interconectados (IWs), concentrando-se em fases cruciais, como estimativa de recursos, construção colaborativa, localização, integração, interoperabilidade, versionamento e população. A aplicação dessa metodologia é exemplificada por meio da concepção de uma ontologia voltada para o domínio do Gerenciamento de Resíduos, sendo sua eficácia avaliada mediante a implementação prática em uma aplicação Web denominada SmartBinAnalytics. Uma análise comparativa é conduzida

entre a metodologia proposta e abordagens existentes, destacando-se suas respectivas vantagens e desvantagens. O artigo contextualiza a metodologia proposta em relação a outras pesquisas relevantes no âmbito do desenvolvimento e da engenharia de ontologias, fazendo referência a ferramentas consagradas, como o Protégé, e enfatiza a importância da internacionalização e localização no contexto de ontologias. Em termos gerais, o artigo proporciona uma abordagem abrangente para o desenvolvimento de ontologias, evidenciando sua aplicabilidade em um domínio específico, notabilizando-se pela consideração minuciosa das diversas fases inerentes ao processo de desenvolvimento de ontologias.

Visando à construção semiautomática de uma ontologia de dados financeiros, Drissi *et al.* (2021) propõem uma metodologia denominada Norms2Onto. A metodologia proposta incorpora técnicas de aprendizagem de máquina, visando otimizar a compreensão e a atualização de informações de natureza financeira. Os autores comparam a abordagem com uma classificação manual realizada por um contador, especialista no domínio, evidenciando a superioridade do método proposto. Adicionalmente, os pesquisadores disponibilizam um módulo de visualização destinado a representar as relações existentes entre os conceitos contidos na ontologia. Experiências e avaliações da metodologia são conduzidas mediante a utilização de um conjunto de padrões contábeis. Os autores empregaram três critérios para avaliação do desempenho na metodologia proposta: a pertinência dos conceitos extraídos, a subcategoria de cada conceito e as hierarquias conceituais. Os achados revelam que a metodologia Norms2Onto supera a classificação manual em termos de número de conceitos extraídos, pertinência dos conceitos e consideração de subcategorias. Em síntese, os autores sustentam que sua abordagem para a construção e o enriquecimento de ontologias financeiras, utilizando aprendizado de máquina, é superior à classificação manual.

Em Khemiri *et al.* (2021), é delineado o método denominado Learn2Construct, desenvolvido para a construção automatizada de ontologias fundamentadas na modelagem de tópicos, empregando o algoritmo de *Latent Dirichlet Allocation* (LDA). Essa abordagem visa superar as limitações identificadas nas metodologias preexistentes que se baseiam em LDA, especialmente no que concerne à extração de conceitos e relações, bem como na capacidade de adaptabilidade a alterações no *corpus*. A revisão sistemática da literatura incorpora uma análise comparativa entre o Learn2Construct e outras abordagens existentes, empregando critérios como

recursos de conhecimento, tipos de termos, nível de automação, tarefas, domínio de aplicação e medidas de avaliação. O estudo relata os resultados de experimentos realizados para validar o método Learn2Construct, comparando-os com um método preexistente em um conjunto de dados específico. A avaliação é conduzida através de medidas de coerência, recall, precisão e F-score, revelando resultados promissores tanto em termos de construção de ontologia quanto de desempenho em relação a metodologias existentes.

Em Rajaonarivo e Arakawa (2021), é descrita uma metodologia para a geração automatizada de uma ontologia de eventos a partir de dados provenientes de redes sociais. O escopo principal consiste na detecção de eventos locais por meio da análise conjunta dos movimentos de indivíduos e das postagens em plataformas de redes sociais. A abordagem delineada compreende etapas específicas, a saber: a descoberta de entidades, a aplicação de filtros, o enriquecimento de informações e a correspondência entre entidades e elementos ontológicos. A ontologia gerada é submetida a uma avaliação a fim de verificar sua confiabilidade e utilidade. A aplicação da abordagem proposta emprega dados oriundos da dinâmica de turistas e informações obtidas da rede social Twitter para gerar a ontologia. O processo de geração da ontologia é composto por cinco fases distintas: coleta de *tweets*, descoberta de entidades, filtragem de entidades e *tweets*, enriquecimento de entidades e, por fim, geração/atualização da ontologia. A avaliação da ontologia é conduzida utilizando métricas do OntoQA, aliada a um estudo de avaliação humana. Os resultados da avaliação indicam que a ontologia em questão se destaca por sua capacidade de classificação, abrangendo uma extensa gama de conhecimento geral e apresentando conexões significativas entre indivíduos e classes.

Já Hananto, Serdült e Kryssanov (2021) propõem um sistema para a construção de um modelo de conhecimento em turismo por meio da análise de avaliações online. O sistema adota a modelagem de tópicos, empregando o algoritmo de processamento de linguagem natural *Latent Dirichlet Allocation* (LDA), a fim de realizar a rotulação automática de documentos e treinar classificadores. Constatou-se que a abordagem de rotulação automática utilizando o LDA apresentou uma precisão de 70% na classificação de avaliações online relacionadas ao turismo. O sistema revelou-se capaz de construir um modelo de conhecimento robusto no âmbito do turismo, proporcionando conjuntos de dados de treinamento de qualidade aceitável para sistemas de recomendação turística. A pesquisa empregou, de maneira

específica, a técnica de LDA para a construção do referido modelo de conhecimento, utilizando os tópicos gerados como rótulos de classes. Posteriormente, um procedimento de rotulação automática fundamentado nos resultados obtidos foi conduzido nas avaliações analisadas. Esse processo de rotulação automática demonstrou uma precisão de 70%, superando as pontuações de desempenho alcançadas mediante a anotação manual em tarefas de classificação. O estudo ressalta que o referido processo de rotulação automática, ancorado no modelo de conhecimento proposto, pode ser eficazmente empregado para o desenvolvimento de sistemas de recomendação turística.

No estudo de Sadeghianasl *et al.* (2021), é introduzida uma abordagem para a criação de ontologias de atividades, visando aprimorar a qualidade dos rótulos associados às atividades em conjuntos de dados de processos. Os autores empregam estratégias de *crowdsourcing* e *gamificação* para envolver especialistas do domínio na tarefa de aprendizado ontológico. A ontologia resultante é formalizada em OWL e incorpora relações semânticas entre as distintas atividades. O sistema *gamificado* motiva os participantes a se engajarem na tarefa de aprendizado ontológico por meio da implementação de diferentes níveis, pontos e distintivos. A avaliação do sistema evidencia sua eficácia na mobilização de especialistas do domínio, resultando na produção de ontologias de alta qualidade. O artigo introduz a abordagem proposta e realiza uma revisão de técnicas e aplicações em aprendizado ontológico, além de discutir pesquisas correlatas sobre motivação, qualidade de dados e ontologia. A recomendação dos autores para investigações futuras nesta área reflete a complexidade e relevância do tema abordado. As referências citadas abrangem uma variedade de tópicos pertinentes, incluindo motivação, qualidade de dados, cidadã, ontologia, *gamificação* e tecnologias Web, proporcionando uma base sólida para contextualizar o estudo no panorama científico atual.

Por fim, em um artigo de revisão centrado na sumarização extrativa de múltiplos documentos, Jalil, Nasir e Nasir (2021) apresentam uma revisão de literatura que abarca diversas técnicas, abordando, entre outras, a *Sentence Density-Peak Clustering* (DPCS), agrupamento baseado em *Density-Peak*, Análise de *Big Data*, sumarização hierárquica, Análise Semântica Latente (LSA), *Maximal Marginal Relevance* (MMR), sumarização interativa e várias outras metodologias pertinentes. O artigo realiza uma análise comparativa, destacando os pontos fortes e as limitações inerentes a cada técnica, além de proporcionar referências a conjuntos de dados e

métodos de avaliação. A seguir, no Quadro 5, um resumo das metodologias identificadas nos estudos referenciados.

Quadro 5 – Resumo dos trabalhos relacionados a metodologias de aprendizagem de ontologias

(Continua...)

ABORDAGEM	REFERÊNCIA	DESCRIÇÃO
Catálogos de Ontologias para IoT e Cidades Inteligentes	Gyrard, Zimmermann e Sheth, 2018	Análise e aprimoramento de catálogos de ontologias, destacando a utilidade e atualizações contínuas do catálogo LOV4IoT.
Previsão de Pontuações de Axiomas	Malchiodi e Tettamanzi, 2018; Malchiodi, Da Costa Pereira e Tettamanzi, 2018	Utilização de algoritmos de suporte vetorial e técnicas de inferência para prever pontuações de axiomas OWL, mostrando eficácia em comparação a métodos tradicionais.
Aprendizado de Ontologia em Dados Taxonômicos	Deb <i>et al.</i> , 2018	Extração de termos, conceitos e relações de textos taxonômicos no sistema agrícola, abordando limitações dos sistemas existentes.
Extração de Conhecimento de Documentos HTML	Rodriguez e Aguilar, 2018	Aplicação de PLN para converter texto de documentos HTML em conhecimento estruturado, utilizando grafos de aprendizagem.
LAM (Modulador de Algoritmo de Aprendizado)	Enea <i>et al.</i> , 2018	Combinação de extração automática de triplas com validação humana, utilizando SPARQL para consultas hipotéticas e visualizações de incertezas na ontologia.
Aprendizagem de Ontologias de Bancos de Dados Relacionais Heterogêneos	Aggoune, 2018	Utilização de WordNet e medidas de similaridade semântica para gerar ontologias no contexto de riscos alimentares, apresentando processos sequenciais detalhados e comparações com outras ferramentas.
Plataforma EasyOnto	Yadav <i>et al.</i> , 2018	Promoção da criação colaborativa de ontologias semiformais por meio de uma interface gráfica simples, facilitando a participação em massa e o desenvolvimento ágil de ontologias.

(Continua...)

Evolução Gramatical e Teoria da Possibilidade	Nguyen e Tettamanzi, 2019	Descoberta de axiomas disjuntos em ontologias a partir de dados RDF, mostrando precisão e abrangência satisfatórias.
KE4WoT	Noura <i>et al.</i> , 2019	Foco na extração automatizada de conhecimento para aplicações na Web Semântica das Coisas (WoT), com avaliações mostrando eficácia na identificação de conceitos pertinentes.
Framework BOLT-K	Vedula, Maneriker e Parthasarathy, 2019	Base em LSTM, adaptação de conhecimento de domínios relacionados para o aprendizado de ontologias, com avaliações comparativas e representações gráficas detalhadas.
Computer Science Ontology (CSO)	Salatino <i>et al.</i> , 2020	Construção de ontologia automaticamente para áreas de pesquisa em Ciência da Computação, empregando aprendizado de máquina e PLN, com ferramentas associadas para classificação de artigos e exploração da ontologia.
Extração de Estruturas Ontológicas em Língua Espanhola	Gómez-Suta, Echeverry-Correa e Soto-Mejía, 2020	Extração de conceitos e relações em textos sobre o conflito armado colombiano, utilizando técnicas de agrupamento semântico e avaliação de estruturas extraídas.
Reutilização de Ontologias	Alharbi, Tamma e Grasso, 2021	Investigação da prática <i>versus</i> teoria na reutilização de ontologias, com avaliação de desenvolvedores quanto a obstáculos identificados.
Expansão Automática de Ontologias em Ciência da Computação	Santosa, Miyazaki e Han, 2021	Avaliação de estratégias de classificação direta e indireta, utilizando técnicas de aprendizado de máquina e <i>deep learning</i> para expandir ontologias existentes.
Sumarização Extrativa de Múltiplos Documentos	Jalil, Nasir e Nasir, 2021	Revisão de técnicas de sumarização extrativa, comparando metodologias e destacando futuras linhas de pesquisa.

(Conclusão.)

Geração de Ontologias de Eventos de Redes Sociais	Rajaonarivo, Mine e Arakawa, 2021	Detecção de eventos locais por meio da análise de movimentação de indivíduos e postagens em redes sociais, utilizando cinco fases distintas de processamento e avaliação baseada em métricas do OntoQA.
Metodologia para Mundos Institucionais Interconectados	Sattar <i>et al.</i> , 2021	Uso de fases de construção colaborativa, integração, versionamento e aplicação prática em gerenciamento de resíduos, destacando-se pela internacionalização e localização.
Learn2Construct	Khemiri <i>et al.</i> , 2021	Construção de ontologias baseada em LDA, validada por experimentos comparativos mostrando resultados promissores.
Sistema de Análise de Avaliações Online	Hananto, Serdült e Kryssanov, 2021	Utilização do algoritmo LDA para rotulação automática de documentos no turismo, com precisão de 70% na classificação de avaliações online.
Criação de Ontologias de Atividades com Crowdsourcing e Gamificação	Sadeghianasl <i>et al.</i> , 2021	Envolvimento de especialistas do domínio em uma abordagem <i>gamificada</i> para criar ontologias de alta qualidade, formalizadas em OWL.
Norms2Onto	Drissi <i>et al.</i> , 2021	Incorporação de técnicas de aprendizado de máquina para construir uma ontologia de dados financeiros, comparada à classificação manual e com módulo de visualização.

Fonte: elaborado pela autora (2024).

2.4 Resultados e discussão

Nesta seção, são apresentados os principais resultados obtidos a partir das perguntas secundárias previstas no protocolo da revisão, consideradas enquanto categorias de análise.

2.4.1 Metodologias de aprendizagem de ontologias

Nesta seção, são apresentadas, em ordem temporal, as metodologias de aprendizagem de ontologias identificadas na revisão de literatura.

Em 2018, Deb *et al.* propuseram o uso de taxonomias para criação automática de ontologias no sistema agrícola, diferenciando dos métodos convencionais que utilizam dados não estruturados. A validação experimental no domínio agrícola mostrou que essa abordagem pode aprimorar significativamente a precisão e a eficiência na gestão de dados agrícolas.

Outro estudo de 2018 foi desenvolvido por Yadav *et al.*, que criaram a aplicação EasyOnto, uma plataforma colaborativa que facilita a construção de ontologias por meio de uma interface gráfica intuitiva. A colaboração entre usuários permite a criação e o aprimoramento contínuo das ontologias, promovendo uma abordagem democrática e inclusiva.

Também em 2018, Enea *et al.* apresentaram uma metodologia baseada na cognição humana para gerenciar incertezas na aprendizagem de ontologias, utilizando extração automática de triplas e validação humana. Essa abordagem garante a precisão e a validade das ontologias, integrando o raciocínio humano no processo automatizado.

Ainda em 2018, Rodriguez e Aguilar propuseram um sistema para extração de conhecimento de textos HTML não estruturados, organizando-os semanticamente em um Grafo de Aprendizagem. Essa metodologia é relevante para transformar dados não estruturados em conhecimento organizado e acessível.

Malchiodi, Da Costa Pereira e Tettamanzi (2018) aplicaram modelos de regressão para prever pontuações de axiomas em ontologias, mostrando resultados satisfatórios. Essa abordagem é essencial para a validação de ontologias complexas, garantindo sua precisão e confiabilidade.

Gyrard, Zimmermann e Sheth, em 2018, propuseram uma metodologia para enriquecer catálogos de ontologias em IoT e cidades inteligentes. A análise e a melhoria de catálogos existentes demonstraram a importância de ontologias bem estruturadas para o desenvolvimento de cidades inteligentes baseadas em tecnologias IoT.

Ainda em 2018, Aggoune desenvolveu uma abordagem semântica para a aprendizagem automática de ontologias a partir de bancos de dados relacionais

heterogêneos no domínio de riscos alimentares. Utilizando uma sequência de algoritmos, essa metodologia assegura a extração e a integração precisas de classes e relações, essenciais para a segurança alimentar.

Em 2019, Nguyen e Tettamanzi apresentaram uma metodologia para descobrir axiomas a partir de dados RDF, utilizando um algoritmo computacional para identificar axiomas de classes disjuntos. Esse método é essencial para garantir a consistência lógica das ontologias, permitindo a detecção de inconsistências e a inferência de informações implícitas, essencial para a Web Semântica.

Também em 2019, Noura *et al.* introduziram a metodologia KE4WoT para extração de conhecimento, utilizando um *corpus* de 4.500 artigos científicos para identificar tópicos importantes. Essa abordagem não só automatiza a identificação de tópicos, mas também aprimora a interoperabilidade e a padronização, essenciais para o desenvolvimento de aplicações semânticas.

Vedula, Maneriker e Parthasarathy, ainda em 2019, apresentaram o *framework* BOLT-K, que utiliza redes neurais para aprender hierarquias ontológicas em novos domínios, transferindo conhecimento existente. Essa metodologia reduz a necessidade de especialistas, tornando o processo de construção de ontologias mais eficiente.

Em 2020, Gómez-Suta, Echeverry-Correa e Soto-Mejía propuseram um método para aprender ontologias a partir de textos sobre o conflito armado colombiano, utilizando agrupamento semântico para extrair e validar conceitos. Essa abordagem é vital para organizar e interpretar informações em contextos específicos e complexos.

Salatino *et al.* (2020) apresentaram uma metodologia para produzir ontologias de tópicos de pesquisa a partir de metadados acadêmicos, oferecendo um portal Web para explorar e fornecer *feedback* sobre a ontologia desenvolvida. A capacidade de focar em ramos específicos ou produzir uma ontologia completa destaca a flexibilidade e a utilidade dessa abordagem.

Avançando para 2021, Drissi *et al.* introduziram a metodologia Norms2Onto, que utiliza técnicas de *machine learning* para construção automática e semiautomática de ontologias a partir de um *corpus* textual. Comparada a métodos tradicionais, essa abordagem se destaca por sua eficiência e precisão, automatizando tarefas que tradicionalmente exigiriam intervenção humana intensiva.

Na área de turismo, Hananto, Serdült e Kryssanov, também em 2021, aplicaram a técnica de modelagem de tópicos para automatizar a rotulação de conteúdo, facilitando a construção de ontologias. Essa metodologia propõe melhorias para a organização da informação, e contribui para sistemas de recomendação mais eficazes, demonstrando a aplicabilidade prática das ontologias em setores específicos.

No domínio da Ciência da Computação, Santosa, Miyazaki e Han (2021) propuseram uma extensão automática de ontologias, avaliando diferentes algoritmos para identificar a sequência mais adequada. Essa pesquisa contribui para a expansão contínua das ontologias, garantindo sua relevância e atualidade.

Sattar *et al.* (2021) destacaram-se por sua flexibilidade para lidar com a complexidade e o tamanho das bases de conhecimento. Essa abordagem é particularmente útil em domínios nos quais a base de conhecimento está em constante expansão e diversificação.

Rajaonarivo, Mine e Arakawa (2021) propuseram uma abordagem dinâmica para a construção de ontologias a partir de publicações no *Twitter*, utilizando algoritmos que convertem *tweets* em instâncias na ontologia e atualizam dinamicamente com novos dados. Essa metodologia é especialmente relevante para domínios nos quais os dados são gerados continuamente e precisam ser incorporados rapidamente.

Khemiri *et al.* (2021) introduziram uma abordagem para a construção automática de ontologias usando LDA, criando conceitos e inter-relacionamentos a partir de textos. A estabilidade do modelo, independentemente das mudanças no *corpus*, destaca a robustez dessa metodologia.

Sadeghianasl *et al.* (2021) utilizaram técnicas de *crowdsourcing* e *gamificação* para aprender ontologias de atividades de processo a partir de *logs* de eventos, promovendo o engajamento dos usuários. Essa metodologia demonstra como a participação ativa dos usuários pode melhorar a qualidade das ontologias.

Ainda em 2021, Jalil, Nasir e Nasir forneceram uma revisão abrangente de técnicas de sumarização extrativa de múltiplos documentos, contribuindo para a gestão de grandes volumes de dados.

Assim, as metodologias revisadas demonstram evolução contínua e diversificadas técnicas para a construção de ontologias, abrangendo uma ampla gama de domínios e abordagens para melhorar a estruturação e a interpretação de

informações complexas. No Quadro 6, a seguir, é apresentado um resumo das metodologias identificadas.

Quadro 6 – Resumo das metodologias identificadas

(Continua...)

ABORDAGEM	REFERÊNCIA	DESCRIÇÃO
Modelos de Regressão para Axiomas	Malchiodi, Da Costa Pereira e Tettamanzi, 2018	Previsão de pontuações de axiomas usando técnicas de regressão.
Taxonomias no Sistema Agrícola	Deb <i>et al.</i> , 2018	Uso de taxonomias como entrada para a criação automática de ontologias, diferindo dos métodos convencionais que utilizam dados não estruturados.
Extração de Conhecimento de HTML	Rodriguez e Aguilar, 2018	Transformação de textos não estruturados em HTML em ontologias semânticas.
Cognição Humana e Simulação Conceitual	Enea <i>et al.</i> , 2018	Gestão de incertezas no processo de aprendizagem de ontologias baseada em raciocínio 'e se'.
Ontologias de Bancos de Dados Relacionais	Aggoune, 2018	Aprendizagem automática de ontologias a partir de bancos de dados relacionais heterogêneos.
Construção Colaborativa – EasyOnto	Yadav <i>et al.</i> , 2018	Aplicação Web EasyOnto para construção colaborativa de ontologias.
Enriquecimento de Catálogos de Ontologias	Gyrard, Zimmermann e Sheth, 2018	Enriquecimento de catálogos de ontologias para Internet das Coisas e cidades inteligentes.
Framework BOLT-K	Vedula, Maneriker e Parthasarathy, 2019	Adaptação e transferência de conhecimento de ontologias existentes para novos domínios usando redes neurais.
Algoritmo Computacional para RDF	Nguyen e Tettamanzi, 2019	Descoberta automática de axiomas a partir de dados RDF usando um algoritmo computacional.
KE4WoT	Noura <i>et al.</i> , 2019	Extração de conhecimento para a Web of Things usando técnicas de <i>machine learning</i> , analisando 4.500 artigos científicos.

(Conclusão.)

ABORDAGEM	REFERÊNCIA	DESCRIÇÃO
Ontologia de Tópicos de Pesquisa	Salatino <i>et al.</i> , 2020	Produção de ontologias de tópicos de pesquisa a partir de metadados acadêmicos e fontes externas.
Textos em Espanhol	Gómez-Suta, Echeverry-Correa e Soto-Mejía, 2020	Aprendizagem de ontologias a partir de textos sobre o conflito armado colombiano.
Extensão Automática de Ontologia	Santosa, Miyazaki e Han, 2021	Avaliação de diferentes algoritmos para extensão automática de ontologias na ciência da computação.
Extração de Publicações do Twitter	Rajaonarivo, Mine e Arakawa, 2021	Construção de uma ontologia específica a partir de publicações em redes sociais, convertendo <i>tweets</i> em instâncias na ontologia.
DSRM	Sattar <i>et al.</i> , 2021	Desenvolvimento de ontologias com base na Design Science Research Methodology (DSRM), adequada para diversos aspectos da construção de ontologias.
Algoritmo LDA	Khemiri <i>et al.</i> , 2021	Construção automática de ontologias a partir de um <i>corpus</i> textual usando o algoritmo LDA.
Modelagem de Tópicos	Hananto, Serdült e Kryssanov, 2021	Utilização da técnica de modelagem de tópicos para automatizar a rotulação de conteúdo no domínio do turismo.
Crowdsourcing e Gamificação	Sadeghianasl <i>et al.</i> , 2021	Aprendizagem de ontologias de atividades de processo usando <i>logs</i> de eventos com apoio de <i>crowdsourcing</i> e <i>gamificação</i> .
Norms2Onto	Drissi <i>et al.</i> , 2021	Construção automática e semiautomática de ontologias utilizando técnicas de <i>machine learning</i> a partir de um <i>corpus</i> textual.

Fonte: elaborado pela autora (2024).

2.4.2 Etapas/fases das metodologias de aprendizagem de ontologias

Nesta seção, são apresentadas as etapas e as fases das metodologias identificadas na revisão de literatura. Buscou-se captar tanto os estudos que detalham todas as etapas quanto etapas isoladas, para obter direcionamentos para esta pesquisa.

Em 2018, Deb *et al.* produziram uma metodologia para o domínio agrícola que utiliza taxonomias organizadas em uma base de conhecimento. As etapas envolvem categorização dos dados taxonômicos, pré-processamento de textos, construção de uma biblioteca de algoritmos para extração de relacionamentos e mapeamento ontológico, e definição de restrições do modelo com a ajuda de especialistas.

Yadav *et al.* (2018) desenvolveram a aplicação EasyOnto, uma plataforma colaborativa que facilita a construção de ontologias em sete etapas, desde a escolha do domínio até a criação de vínculos entre categorias. A interface gráfica intuitiva permite que usuários contribuam com novas instâncias, categorias e relações, promovendo uma abordagem inclusiva.

Outro estudo de 2018, de Enea *et al.*, introduziu um método baseado em jogos interrogativos para validar ontologias, utilizando raciocínio indutivo e dedutivo. A interação entre a aplicação e o operador humano melhora o processo de aprendizagem de ontologias.

Rodriguez e Aguilar (2018) desenvolveram uma aplicação para extração de conhecimento de textos não estruturados, utilizando técnicas de processamento de linguagem natural para separar textos em frases, analisar entidades e relações, e atualizar um Grafo de Aprendizagem.

Gyrard, Zimmermann e Sheth (2018) propuseram uma metodologia para enriquecer catálogos de ontologias em IoT e cidades inteligentes, dividida em cinco etapas: investigação de novos domínios, atualização do dicionário, atualização do catálogo RDF, atualização da página HTML e desenvolvimento de aplicações baseadas no catálogo.

Ainda em 2018, Aggoune desenvolveu uma abordagem semântica para a aprendizagem automática de ontologias a partir de bancos de dados relacionais heterogêneos no domínio de riscos alimentares. Utilizando uma sequência de algoritmos, essa metodologia assegura a extração e a integração precisas de classes e relações, essenciais para a segurança alimentar.

Nguyen e Tettamanzi (2019) apresentaram uma metodologia para descobrir axiomas a partir de dados RDF, utilizando um algoritmo computacional para identificar axiomas de classes disjuntos. Esse método é essencial para garantir a consistência lógica das ontologias, permitindo a detecção de inconsistências e a inferência de informações implícitas, essencial para a Web Semântica.

No domínio da *Web of Things*, Noura *et al.* (2019) detalharam a metodologia KE4WoT em dez etapas, desde a seleção de ontologias até a identificação de tópicos populares, demonstrando a aplicabilidade prática em diversos domínios.

Ainda em 2019, Vedula, Maneriker e Parthasarathy (2019) descreveram o *framework* BOLT-K, que identifica pares de conceitos e as relações entre eles, usando dados anotados para construir ontologias completas. A abordagem é promissora, mas necessita de um detalhamento mais exaustivo das etapas.

No ano de 2020, foram recuperados dois estudos. O primeiro, de Gómez-Suta, Echeverry-Correa e Soto-Mejía (2020), no qual os autores propuseram um método para aprendizagem de ontologias no contexto do conflito armado colombiano, utilizando técnicas de agrupamento semântico para extrair e validar conceitos. A proposta é acessível e aplicável a dados e domínios com recursos de conhecimento estruturados limitados. O segundo, de Salatino *et al.* (2020), que detalharam uma metodologia para gerar ontologias de tópicos de pesquisa, processando metadados acadêmicos e fontes externas. A criação de um portal para explorar e fornecer *feedback* sobre a ontologia desenvolvida demonstra um compromisso com a evolução contínua e a participação da comunidade científica.

Mas foi em 2021 que se obteve um maior número de estudos relacionados à metodologia de aprendizagem de ontologias, somando-se oito documentos no total. Drissi *et al.* (2021) introduziram a metodologia Norms2Onto, que utiliza técnicas de *machine learning* para construção automática e semiautomática de ontologias a partir de um *corpus* textual. Comparada a métodos tradicionais, essa abordagem se destaca por sua eficiência e precisão, automatizando tarefas que tradicionalmente exigiriam intervenção humana intensiva.

Os autores Hananto, Serdült e Kryssanov (2021) propuseram um padrão de modelagem de tópicos no setor de turismo, dividido em três partes: descoberta de tópicos, etiquetagem automatizada e treinamento de classificadores. Os tópicos são mapeados em um modelo de conhecimento construído a partir de uma revisão de

literatura e uma ontologia geral existente, utilizando uma taxonomia da DBPedia para identificar classes de destinos turísticos.

Utilizando a *Design Science Research*, Sattar *et al.* (2021) detalharam uma abordagem *Methodology* (DSRM), dividida em três estágios: pré-conceitualização, conceitualização e pós-conceitualização. Cada estágio abrange várias fases, como estruturação de um Gráfico de Gantt, definição de ferramentas, aquisição de conhecimento, validação, construção colaborativa e manutenção.

Rajaonarivo, Mine e Arakawa (2021) apresentaram uma metodologia de seis fases para extração de ontologias a partir de dados do *Twitter*, incluindo coleta de *tweets*, descoberta de entidades, filtragem de entidades, enriquecimento de informações, geração/atualização de ontologias e avaliação. A clareza das fases facilita a replicação e a validação da metodologia.

No domínio da Ciência da Computação, Santosa, Miyazaki e Han (2021) discutiram a extensão automática de ontologias, composta por cinco etapas: extração de tópicos, criação de um gráfico de tópicos, estruturação de um dicionário de tópicos, análise do nível dos tópicos e filtragem de *links* inadequados.

Khemiri *et al.* (2021) propuseram uma abordagem para construção automática de ontologias usando LDA, dividida em cinco módulos: pré-processamento, extração de termos, modelagem de tópicos, extração de conceitos e relações, e visualização de ontologias. A metodologia destaca-se pela robustez e capacidade de responder a questões básicas sobre a ontologia.

Já Sadeghianasl *et al.* (2021) utilizaram técnicas de *crowdsourcing* e *gamificação* para aprender ontologias de atividades de processo a partir de *logs* de eventos, promovendo o engajamento dos usuários. Essa metodologia demonstra como a participação ativa dos usuários pode melhorar a qualidade das ontologias.

Por fim, Jalil, Nasir e Nasir (2021) forneceram uma revisão abrangente de técnicas de sumarização extrativa de múltiplos documentos, contribuindo para a gestão de grandes volumes de dados. A seguir, no Quadro 7, é possível identificar um consolidado das etapas e fases identificadas nas metodologias referenciadas.

Quadro 7 – Etapas e fases identificadas

(Continua...)

ABORDAGEM	REFERÊNCIA	DETALHAMENTO DAS ETAPAS/FASES
Uso de Taxonomias no Sistema Agrícola	Deb <i>et al.</i> , 2018	<ol style="list-style-type: none"> 1. Categorização dos dados taxonômicos. 2. Pré-processamento de textos. 3. Extração de relacionamentos e mapeamento da ontologia. 4. Construção de classes e subclasses. 5. Identificação das propriedades. 6. Indicação das restrições do modelo.
Extração de Conhecimento de HTML	Rodriguez e Aguilar, 2018	<ol style="list-style-type: none"> 1. Separação do texto em frases. 2. Análise de frases (reconhecimento de nomes próprios, análise morfosintática, reconhecimento de entidades e relacionamentos). 3. Atualização do Grafo de Aprendizagem.
Extrações e Validação de Informações	Enea <i>et al.</i> , 2018	<ol style="list-style-type: none"> 1. Extração automática de triplas de fontes heterogêneas. 2. Validação por operador humano por meio de representações gráficas e consultas SPARQL.
Geração Automática de Ontologias de Riscos Alimentares	Aggoune, 2018	<ol style="list-style-type: none"> 1. Geração de classes. 2. Extração de propriedades de tipo de dados. 3. Geração de propriedades de objetos. 4. Integração dos registros. 5. Detecção de modificações nas bases de dados e atualização da ontologia.
Enriquecimento de Catálogos de Ontologias	Gyrard, Zimmermann e Sheth, 2018	<ol style="list-style-type: none"> 1. Investigação de um novo domínio de aplicativo IoT. 2. Atualização do dicionário. 3. Atualização do conjunto de dados do catálogo RDF. 4. Atualização do Catálogo de Ontologia HTML. 5. Desenvolvimento de aplicações baseadas nos conjuntos de dados do catálogo.

(Continua...)

ABORDAGEM	REFERÊNCIA	DETALHAMENTO DAS ETAPAS/FASES
Construção Colaborativa – EasyOnto	Yadav <i>et al.</i> , 2018	<ol style="list-style-type: none"> 1. Escolha de domínio. 2. Adição de instâncias, categorias ou relações. 3. Enriquecimento com anotações semânticas. 4. Vinculação de instâncias a categorias. 5. Escolha de relações. 6. Atribuição automática de classes pai. 7. Criação de vínculos entre categorias.
Extração de Axiomas de Classes RDF	Nguyen e Tettamanzi, 2019	<ol style="list-style-type: none"> 1. Projeto da gramática para gerar axiomas. 2. Processo evolutivo: geração aleatória de axiomas candidatos. 3. Identificação de axiomas na forma de expressões lógicas.
KE4WoT	Noura <i>et al.</i> , 2019	<ol style="list-style-type: none"> 1. Seleção de Ontologias. 2. Pré-processamento. 3. Extração de vocabulário. 4. Extração de Termos. 5. Cálculo da Frequência. 6. Algoritmo Word2vec. 7. Agrupamento K-Means. 8. Agregação de resultados. 9. Atribuição do nome do tópico. 10. Identificação do Tópico Popular.
Framework BOLT-K	Vedula, Maneriker e Parthasarathy, 2019	<ol style="list-style-type: none"> 1. Identificação de pares de conceitos relacionados. 2. Filtragem de conceitos. 3. Ampliação das instâncias de treinamento. 4. Aprendizado das relações entre os conceitos-alvo restantes.
Ontologia de Tópicos de Pesquisa	Salatino <i>et al.</i> , 2020	<ol style="list-style-type: none"> 1. Processamento de informações de metadados acadêmicos e fontes externas. 2. Inferência de relações semânticas entre palavras-chave. 3. Geração automática de ontologias para o domínio da Ciência da Computação.

(Continua...)

ABORDAGEM	REFERÊNCIA	DETALHAMENTO DAS ETAPAS/FASES
Aprendizagem de Ontologias no Conflito Armado Colombiano	Gómez-Suta, Echeverry-Correa e Soto-Mejía, 2020	<ol style="list-style-type: none"> 1. Identificação automática de <i>tags</i> de classes gramaticais. 2. Lematização e normalização. 3. Remoção de palavras irrelevantes, pontuação e caracteres numéricos. 4. Padronização de termos semânticos relacionados.
Construção Automática com LDA	Khemiri <i>et al.</i> , 2021	<ol style="list-style-type: none"> 1. Pré-processamento. 2. Extração de termos. 3. Modelagem de tópicos. 4. Extração de conceitos e relações. 5. Visualização de Ontologias.
Norms2Onto	Drissi <i>et al.</i> , 2021	<ol style="list-style-type: none"> 1. Módulo de pré-processamento: limpeza e normalização dos dados. 2. Módulo de aprendizagem: extração de conceitos relevantes. 3. Módulo de visualização: desenvolvimento de interfaces intuitivas.
DSRM para Desenvolvimento de Ontologias	Sattar <i>et al.</i> , 2021	<ol style="list-style-type: none"> 1. Pré-conceitualização: definição de ferramentas, técnicas, recursos, análise de domínio, etc. 2. Conceitualização: definição de endurantes e perdurantes. 3. Pós-conceitualização: construção colaborativa, implementação, verificação, controle de versão, etc.
Modelagem de Tópicos	Hananto, Serdült e Kryssanov, 2021	<ol style="list-style-type: none"> 1. Descoberta de tópicos por meio de análises online. 2. Processo automatizado de etiquetagem. 3. Treinamento de classificadores e avaliação do modelo.

(Conclusão.)

ABORDAGEM	REFERÊNCIA	DETALHAMENTO DAS ETAPAS/FASES
Extração de Publicações do Twitter	Rajaonarivo, Mine e Arakawa, 2021	<ol style="list-style-type: none"> 1. Coleta de <i>tweets</i>. 2. Descoberta de entidades. 3. Filtragem de entidades e <i>tweets</i>. 4. Enriquecimento de entidades. 5. Geração/atualização de ontologias. 6. Avaliação da ontologia.
Extensão Automática de Ontologia	Santosa, Miyazaki e Han, 2021	<ol style="list-style-type: none"> 1. Extração de novos tópicos de resumos. 2. Proposição de um gráfico de tópicos. 3. Estruturação de um dicionário de tópicos. 4. Análise do nível dos novos tópicos. 5. Filtragem de <i>links</i> inadequados e previsão de <i>links</i> ocultos.

Fonte: elaborado pela autora (2024).

2.4.3 Diferenciais, lacunas e principais resultados das metodologias de aprendizagem de ontologias

Nesta seção, são apresentados os diferenciais, as lacunas e os principais resultados das metodologias propostas pelos autores, identificados na revisão de literatura. Os diferenciais das metodologias de geração automática de ontologias são evidentes em diversos estudos.

Yadav *et al.* (2018) enfatizam que o sistema proposto pelos autores acelerou a fase de aquisição de ontologias, eliminando a necessidade de especialistas nas fases iniciais. Rodriguez e Aguilar (2018) afirmam que o sistema criado permitiu atualizar ontologias e léxicos simultaneamente. Gyrard, Zimmermann e Sheth (2018) destacam que a metodologia proposta promove a reutilização de ontologias e a interoperabilidade semântica em aplicações de cidades inteligentes. Deb *et al.* (2018) argumentam que sua aplicação pode ajudar na aprendizagem automática de ontologias a partir de textos taxonômicos e pode ser usada em outras áreas devido à sua natureza genérica. Aggoune (2018) destaca a eficiência na geração de ontologias completas, sem erros técnicos, com baixo custo e tempo de aprendizado, além de um protótipo que ocupa menos espaço em disco.

Noura *et al.* (2019) afirmam que o KE4WoT reduz o tempo de desenvolvimento e a carga de trabalho humano, promovendo a reutilização de tópicos e mantendo as ontologias atualizadas. Vedula, Maneriker e Parthasarathy (2019) ressaltam que sua aplicação é flexível e generalizável para aprender ontologias em subdomínios emergentes, reduzindo a necessidade de conhecimento manual e detectando novos tipos de relacionamentos. Nguyen e Tettamanzi (2019) confirmam a precisão e a cobertura de sua abordagem para extração de axiomas, destacando a geração de axiomas complexos.

Gómez-Suta, Echeverry-Correa e Soto-Mejía (2020) afirmam que sua abordagem permite validar ontologias abstratas sem avaliação manual, sendo acessível em termos de tempo e recursos, e aplicável a domínios com poucos recursos de conhecimento estruturado. Salatino *et al.* (2020) detalham uma metodologia para gerar ontologias de tópicos de pesquisa, processando metadados acadêmicos e fontes externas. A criação de um portal para explorar e fornecer *feedback* sobre a ontologia desenvolvida demonstra um compromisso com a evolução contínua e a participação da comunidade científica.

Drissi *et al.* (2021) destacam que a metodologia Norms2Onto extraiu automaticamente 2.000 conceitos de cinco normas contábeis, enquanto a classificação manual extraiu apenas 50. A aplicação também identificou 1.900 conceitos relevantes e 1.860 subcategorias validadas pelo dicionário Investopedia, em contraste com 45 conceitos relevantes identificados manualmente. Hananto, Serdült e Kryssanov (2021) mencionam que a abordagem baseada no modelo LDA alcançou uma precisão média de 70% na rotulagem automatizada, superando a anotação manual. Khemiri *et al.* (2021) destacam que a ontologia desenvolvida facilita a anotação automática de documentos acadêmicos, proporcionando um processo mais consistente. Santosa, Miyazaki e Han (2021) ressaltam que a abordagem por eles proposta foi avaliada empiricamente e por especialistas, demonstrando melhor desempenho global.

Em contrapartida, poucos autores indicaram lacunas em suas metodologias. Hananto, Serdült e Kryssanov (2021) observaram que o modelo LDA não conseguiu dividir revisões em tópicos suficientemente distintos, resultando em anotações incorretas em algumas classes. Vedula, Maneriker e Parthasarathy (2019) apontaram que a aplicação BOLT-K requer intervenção manual para definir domínios relacionados durante o aprendizado de ontologias.

Com relação aos resultados, Drissi *et al.* (2021) mostraram que o desempenho do modelo Norms2Onto superou a classificação manual, apresentando oportunidades para novas pesquisas. Deb *et al.* (2018) argumentam que sua aplicação pode ajudar na aprendizagem automática de ontologias a partir de textos taxonômicos e pode ser usada em outras áreas devido à sua natureza genérica. Nguyen e Tettamanzi (2019) confirmam a precisão e a cobertura de sua abordagem para extração de axiomas, destacando a geração de axiomas complexos. Rajaonarivo, Mine e Arakawa (2021) validaram a confiabilidade das informações na ontologia desenvolvida. Santosa, Miyazaki e Han (2021) demonstraram a viabilidade da extensão automática de ontologias no domínio da Ciência da Computação. Vedula, Maneriker e Parthasarathy (2019) ressaltam que sua aplicação é flexível e generalizável para aprender ontologias em subdomínios emergentes, reduzindo a necessidade de conhecimento manual e detectando novos tipos de relacionamentos. Gyrard, Zimmermann e Sheth (2018) destacam que a metodologia proposta promove a reutilização de ontologias e a interoperabilidade semântica em aplicações de cidades inteligentes. Gómez-Suta, Echeverry-Correa e Soto-Mejía (2020) afirmam que sua abordagem permite validar ontologias abstratas sem avaliação manual, sendo acessível em termos de tempo e recursos, e aplicável a domínios com poucos recursos de conhecimento estruturado. A seguir, no Quadro 8, um resumo dos diferenciais, das lacunas e dos resultados identificados.

Quadro 8 – Diferenciais, lacunas e resultados

(Continua...)

ABORDAGEM	REFERÊNCIA	DIFERENCIAIS	LACUNAS	RESULTADOS
Aprendizado Automático de Ontologias	Aggoune (2018)	Geração de ontologias completas, sem erros técnicos, com baixo custo e tempo de aprendizado.	Não mencionado.	Não mencionado.
EasyOnto	Yadav <i>et al.</i> (2018)	Aceleração da fase de aquisição de ontologias, eliminando a necessidade de especialistas nas fases iniciais.	Não mencionado.	Não mencionado.

(Continua...)

ABORDAGEM	REFERÊNCIA	DIFERENCIAIS	LACUNAS	RESULTADOS
Extração de Conhecimento	Rodriguez e Aguilar (2018)	Permissão para atualizar ontologias e léxicos simultaneamente.	Não mencionado.	Não mencionado.
Aprendizagem de Ontologias a partir de Textos Taxonômicos	Deb <i>et al.</i> (2018)	Não mencionado.	Não mencionado.	Aplicação pode ajudar na aprendizagem automática de ontologias a partir de textos taxonômicos.
Reutilização de Ontologias	Gyrard, Zimmermann e Sheth (2018)	Não mencionado.	Não mencionado.	Promoção da reutilização de ontologias e a interoperabilidade semântica em aplicações de cidades inteligentes.
KE4WoT	Noura <i>et al.</i> (2019)	Redução do tempo de desenvolvimento e da carga de trabalho humano.	Não mencionado.	Não mencionado.
Extração de Axiomas	Nguyen e Tettamanzi (2019)	Não mencionado.	Não mencionado.	Precisão e cobertura de sua abordagem para extração de axiomas, destacando a geração de axiomas complexos.
BOLT-K	Vedula, Maneriker e Parthasarathy (2019)	Não mencionado.	Requisição da intervenção manual para definir domínios relacionados durante o aprendizado de ontologias.	Aplicação flexível e generalizável para aprender ontologias em subdomínios emergentes, detectando novos tipos de relacionamentos.

(Conclusão.)

ABORDAGEM	REFERÊNCIA	DIFERENCIAIS	LACUNAS	RESULTADOS
Validação de Ontologias	Gómez-Suta, Echeverry-Correa e Soto-Mejía (2020)	Não mencionado.	Não mencionado.	Validação de ontologias abstratas sem avaliação manual, sendo acessível em termos de tempo e recursos.
Norms2Onto	Drissi <i>et al.</i> (2021)	Capacidade de extrair automaticamente 2.000 conceitos de cinco normas contábeis.	Não mencionado.	Desempenho do modelo superou a classificação manual, apresentando oportunidades para novas pesquisas.
Modelagem de Tópicos no Turismo	Hananto, Serdült e Kryssanov (2021)	Abordagem baseada no modelo LDA obteve uma precisão média de 70% na rotulagem automatizada.	Modelo LDA não conseguiu dividir revisões em tópicos suficientemente distintos.	Não mencionado.
Extensão Automática de Ontologias	Santosa, Miyazaki e Han (2021)	Avaliada empiricamente e por especialistas, demonstrando melhor desempenho global.	Não mencionado.	Viabilidade da extensão automática de ontologias no domínio da Ciência da Computação.
Anotação Automática de Documentos	Khemiri <i>et al.</i> (2021)	Facilitação da anotação automática de documentos acadêmicos, proporcionando um processo mais consistente.	Não mencionado.	Não mencionado.
Geração de Ontologias a partir de Tweets	Rajaonarivo, Mine e Arakawa (2021)	Não mencionado.	Não mencionado.	Confiabilidade das informações na ontologia desenvolvida.

Fonte: elaborado pela autora (2024).

Finalizando este capítulo, observa-se que os resultados obtidos oferecem um panorama abrangente das metodologias de aprendizagem de ontologias, permitindo identificar os principais diferenciais, lacunas e contribuições de cada estudo analisado. A sistematização dos dados, com base nos protocolos previamente definidos, possibilitou a categorização das abordagens adotadas nos 22 trabalhos selecionados, servindo como base sólida para as etapas subsequentes da pesquisa.

Entre essas metodologias, a técnica de *Latent Dirichlet Allocation* (LDA) destacou-se como a mais recorrente. Empregada em estudos como os de Khemiri *et al.* (2021) e Hananto *et al.* (2021), o LDA demonstrou eficácia na modelagem de tópicos, mostrando-se útil na organização de grandes volumes de textos e na geração de conceitos estruturados para compor ontologias. Esse destaque evidencia a valorização de abordagens que conciliam robustez matemática com aplicabilidade em contextos diversos.

A maioria dos estudos analisados apresentou uma abordagem prática, com o desenvolvimento de protótipos, ferramentas e validações aplicadas a domínios específicos — como agricultura, turismo, cidades inteligentes, análise de conflitos, dados financeiros, entre outros. Soluções como a plataforma EasyOnto (Yadav *et al.*, 2018), o sistema Norms2Onto (Drissi *et al.*, 2021) e o modelo baseado em publicações no Twitter (Rajaonarivo *et al.*, 2021) são exemplos concretos de como as pesquisas têm extrapolado o campo teórico para entregas aplicáveis e inovadoras. Em muitos casos, essas propostas foram avaliadas com base em dados reais, métricas reconhecidas ou *feedback* de especialistas, demonstrando um compromisso consistente com a aplicabilidade e a efetividade das ontologias produzidas.

Além do LDA, outras técnicas também receberam destaque, como algoritmos de regressão e de suporte vetorial (Malchiodi *et al.*, 2018), empregados na previsão de pontuações de axiomas, bem como modelos de redes neurais, como o BOLT-K (Vedula *et al.*, 2019), baseados em LSTM, e abordagens de *deep learning* voltadas para a expansão de ontologias (Santosa *et al.*, 2021). Essas aplicações evidenciam a crescente sofisticação e diversidade das soluções computacionais utilizadas para apoiar a aprendizagem ontológica.

Por fim, embora alguns estudos já tenham explorado redes neurais profundas, não foi identificado o uso direto de *Large Language Models* (LLM) nos trabalhos analisados. Considerando que a literatura da época já sinalizava o potencial dessas tecnologias, sua ausência pode estar relacionada ao momento inicial de sua

popularização e às limitações de acesso técnico e financeiro. Ainda assim, os avanços recentes no uso de LLM para extração e representação de conhecimento indicam um caminho promissor para pesquisas futuras. Diante desse panorama, o próximo capítulo apresenta os fundamentos teóricos e metodológicos que sustentam esta investigação, abordando os principais conceitos relacionados à organização e representação do conhecimento e da informação, às ontologias, ao governo eletrônico, à plataforma Gov.br e ao Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI).

3 FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA

Neste capítulo, são apresentados os fundamentos teóricos e metodológicos que orientaram esta pesquisa. Na seção 3.1 Organização e Representação do Conhecimento e da Informação, busca-se diferenciar os processos de organização e representação do conhecimento e da informação e apresentar uma breve contextualização acerca de alguns conceitos adotados na Biblioteconomia e Ciência da Informação (BCI), como categorização, classificação. Além disso, foram abordados também os chamados Sistemas de Organização do Conhecimento (SOC). Na seção 3.2 Ontologias, é apresentada uma contextualização da temática, incluindo duas sub-seções, sendo a primeira sobre metodologias para construção de ontologias e a segunda sobre aprendizagem de ontologias. Na seção 3.3 Governo Eletrônico, discute-se o histórico e os principais conceitos relacionados à temática, destacando-se como o governo eletrônico utiliza tecnologias digitais para melhorar a eficiência administrativa, a transparência e a participação cidadã, transformando a relação entre governo e sociedade. Na seção 3.4 Plataforma de serviços públicos Gov.br, apresenta-se a plataforma desenvolvida pelo governo brasileiro que tem como propósito unificar serviços digitais do governo federal, simplificando a interação dos cidadãos com o Estado e promovendo eficiência, transparência e participação. A centralização de serviços visa facilitar o acesso a informações governamentais e criar canais para *feedback* e colaboração, alinhando-se às tendências globais de governança digital. Por fim, na seção 3.5 Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI), é brevemente descrito o referido índice que avalia a prontidão e a capacidade dos países em usar TICs para serviços públicos, considerando a prestação de serviços online, infraestrutura de telecomunicações e capital humano.

3.1 Organização e Representação do Conhecimento e da Informação

O ato de organizar, presente nas atividades cotidianas, é exemplificado na organização de roupas ou na classificação de livros (White, 2011). As decisões subjacentes a essas práticas envolvem segregação e categorização, destacando a imposição de uma ordem específica. A distinção entre organizar coisas, conhecimento e informações é uma ação importante (Glushko, 2013). Ao organizar livros em uma

biblioteca, além da disposição física, é necessário criar modelos de organização do conhecimento, evidenciando a sua prévia estruturação.

A clareza na representação desses modelos é vital para a compreensão e a adaptação pelos usuários. Adaptações, como o uso de cores para representar gêneros literários, exemplificam essa necessidade. Catálogos, fontes de informações sobre os livros, devem seguir padrões para facilitar a utilização desses modelos pelos usuários, destacando a importância da representação eficaz na organização de informações. O processo de organização vai além das atividades diárias, alcançando a organização do conhecimento. A distinção entre organizar coisas e informações sobre elas é importante para entender como os modelos de organização são concebidos e representados. A representação desses modelos facilita a utilização e permite adaptações às necessidades dos usuários.

Com o aumento de uma coleção de elementos, a necessidade de identificação explícita torna-se ainda mais importante, que pode se manifestar por meio de ícones, códigos ou outras formas, representando informações sobre os elementos. Isso engloba registros bibliográficos, metadados diferenciando livros e mecanismos organizacionais, como diretórios de lojas, catálogos de bibliotecas e índices, que proporcionam compreensão da coleção e facilitam a localização dos elementos (Glushko, 2013).

A estrutura resultante do processo de organização do conhecimento vai além da expressão do autor, representando uma visão consensual do domínio a ser abordado. Esse modelo de abstração do mundo real é construído com um propósito específico (Brascher; Café, 2008).

Conforme salientado por Hjørland (2008), a organização do conhecimento abrange diversas perspectivas. Em uma visão mais abrangente, essa organização pode refletir a divisão social do trabalho, instituições sociais, linguagens e sistemas simbólicos, sistemas conceituais, teorias e gêneros literários, entre outros elementos. Em uma abordagem mais específica, a organização do conhecimento volta-se para atividades como descrição de documentos, indexação e classificação realizadas em bibliotecas, bases de dados bibliográficos, arquivos e outras instituições de memória. Portanto engloba tanto a organização social, associada a profissões, negócios e disciplinas, quanto a organização intelectual ou cognitiva, que se relaciona à estruturação do conhecimento em conceitos, sistemas e teorias conceituais. Ainda na mesma linha de visão de Hjørland (2008), White (2011) argumenta que a organização

do conhecimento tem uma tradição em olhar o conhecimento e a organização separadamente e que a combinação dessas duas áreas criou a disciplina de Organização do Conhecimento, que é estudada por diversos pesquisadores dentro e fora da BCI.

Segundo Brascher e Café (2008), a organização do conhecimento é delineada como um processo de modelagem que visa à construção de representações, tendo como base a análise do conceito e de suas características para o estabelecimento de sua posição desse conceito em um determinado domínio do conhecimento.

A modelagem, também conhecida como modelização, refere-se ao processo de elaboração de modelos que buscam descrever e explicar os fenômenos observados, fundamentando-se em princípios específicos a serem seguidos durante a atividade de modelagem. Esses princípios, delineados por Campos (2004), incluem o método de raciocínio, o objeto de representação, as relações entre os objetos e as formas gráficas adotadas por um modelo.

No que concerne aos métodos de raciocínio, destaca-se a dicotomia entre os métodos indutivo e dedutivo. O método indutivo viabiliza a construção de modelos ao representar inicialmente os elementos/objetos e suas relações em um contexto específico. Esse processo ocorre de forma ascendente, partindo das partes para o todo, seguindo uma abordagem *bottom-up*. Por outro lado, o método dedutivo propõe a elaboração de abstrações para conceber, em primeiro lugar, o domínio/contexto, independentemente dos elementos e suas inter-relações. Nesse caso, o movimento se dá de forma descendente, do todo para as partes, adotando uma abordagem *top-down*. É importante considerar uma abordagem integrada desses dois métodos de raciocínio, conforme preconizado por Campos (2004), que sugere uma perspectiva complexa e sistêmica, envolvendo processos distintos de análise e síntese.

O conceito de objeto de representação do conhecimento refere-se à menor unidade de manipulação e representação em um contexto específico. As relações entre os objetos em um determinado contexto apresentam uma natureza variada, dependendo dos objetivos da modelagem, podendo assumir formas categorial, hierárquica, partitiva, entre categorias e de equivalência. No que diz respeito às formas de representação gráfica, é notável a persistência de uma lacuna na disponibilidade de estratégias para representação visual de modelos estruturados na área de BCI (Campos, 2004).

Conforme destacado por Davis, Shrobe e Szolovits (1993), a definição da representação do conhecimento suscita a questão fundamental sobre o papel desempenhado por esse processo. Segundo esses autores, a representação do conhecimento pode assumir diversos papéis, como: servir como substituto para entidades do mundo real; estabelecer compromissos ontológicos, que refletem decisões sobre como interpretar o mundo e o que considerar nele; constituir uma teoria fragmentária do Raciocínio Inteligente, sendo motivada por uma visão específica sobre como as pessoas raciocinam de maneira inteligente ou com crenças sobre o significado razoável da razão; funcionar como meio para computação eficiente, partindo de uma perspectiva mecanicista que encara o raciocínio em máquinas (e, possivelmente, em seres humanos) como um processo computacional; e atuar como meio de expressão humana.

Observa-se que os papéis indicados por Davis, Shrobe e Szolovits (1993) focam em uma visão ampla e não restrita da representação do conhecimento, visão essa defendida por Hjørland (2008). Nessa mesma linha, entende-se também que está situada a abordagem de Campos (2004), que retoma e interpreta os papéis sugeridos por Davis, Shrobe e Szolovits (1993), para definir representação do conhecimento:

1. Uma representação de conhecimento é um mecanismo usado para se raciocinar sobre o mundo, em vez de agir diretamente sobre ele.
2. Uma representação de conhecimento é uma resposta à pergunta “Em que termos devo pensar sobre o mundo?”, isto é, um conjunto de compromissos ontológicos. Uma vez que toda representação é uma aproximação imperfeita da realidade, ao selecionarmos uma representação, estamos tomando um conjunto de decisões sobre como e o que ver no mundo.
3. Uma representação de conhecimento é uma teoria fragmentada de raciocínio que especifica que inferências são válidas e quais são recomendadas. Uma representação é motivada por alguma percepção de como as pessoas argumentam ou por alguma crença sobre o que significa raciocinar de forma inteligente.
4. Uma representação de conhecimento é um meio de computação pragmaticamente eficiente.
5. Uma representação de conhecimento é um meio de expressão, isto é, uma linguagem na qual se pode dizer coisas sobre o mundo (Campos, 2004, p. 24).

Ao transitar do domínio das ideias para o domínio dos objetos, ocorre uma mudança de perspectiva no processo de organização da informação. Esse processo implica na organização sistemática de objetos informacionais, considerando tanto a descrição física quanto a semântica do conteúdo contido nesses objetos. A

organização da informação envolve a disposição sistemática de um conjunto de objetos informacionais em coleções, visando facilitar o acesso ao conhecimento contido na informação. Para alcançar esse propósito, torna-se necessário realizar a descrição física e de conteúdo dos objetos informacionais. O resultado desse procedimento descritivo é a representação da informação, compreendida como um conjunto de elementos descritivos que representam os atributos do objeto (Brascher; Café, 2008).

A organização e representação do conhecimento estão intrinsecamente ligadas ao processo cognitivo. Quando um sujeito se depara com uma nova informação em determinado contexto, inicia um processo de comparação dessa informação com conhecimentos já internalizados. Em seguida, ele categoriza essa informação na estrutura cognitiva, gerando, assim, um novo conhecimento. Essas categorias podem ser criadas na mente, conforme a necessidade de classificar um novo conhecimento, ou podem ser estabelecidas como modelos a serem adotados na organização e representação do conhecimento. Partindo de uma abstração do conhecimento humano, Aristóteles propôs, de forma dedutiva, um conjunto de categorias que, em tese, poderiam abranger todo o conhecimento humano. O modelo aristotélico foi e continua sendo uma referência significativa para o processo de organização do conhecimento (Lima; Maculan, 2024).

Conforme Lima (2010), categorizar é uma ação intrínseca ao ser humano, sendo diretamente influenciada pela interação do sujeito com o meio em que está inserido. Esse processo cognitivo é dinâmico e interativo, uma vez que, à medida que se aprende, novas estruturas e conexões são assimiladas, e novas informações são incorporadas às estruturas de conhecimento preexistentes.

A categorização é compreendida como um processo de agrupar elementos do mundo em conjuntos, cujos membros compartilham alguma semelhança em um contexto específico. Como já mencionado, livros podem ser agrupados por autores, gênero, ano ou qualquer outra categoria relevante. A categorização representa um mecanismo cognitivo fundamental que simplifica a experiência individual com o ambiente (Jacob, 2004).

As semelhanças e diferenças entre as coisas podem determinar a necessidade de criar subcategorias ou facetas. Por exemplo, livros de um mesmo autor podem ser subdivididos por ano ou por gênero. De acordo com Lima e Raghavan (2014), a categorização implica um propósito e, dependendo do contexto, as categorias podem

mudar para refletir diferentes perspectivas. Além disso, é essencial que os membros de uma categoria possuam características necessárias e suficientes para fazer parte dela. A categorização, como processo cognitivo, relaciona-se ao significado semântico de uma determinada unidade do conhecimento, estabelecendo-se no mundo das ideias e não no mundo dos objetos físicos. Nesse contexto, a categorização é reconhecida como uma das principais atividades no processo de organização do conhecimento.

A organização da informação desempenha um papel estratégico tanto para os usuários, que buscam encontrar e compreender informações, quanto para os sistemas computacionais, que processam dados para decisões, inferências e recomendações. O grande desafio reside em assegurar a interoperabilidade entre os esquemas de diferentes sistemas de organização do conhecimento e os metadados adotados, permitindo que distintos sistemas interajam eficientemente com as informações. O processo de organização da informação requer a identificação de um conjunto de características, também denominadas propriedades, apresentadas por meio de declarações que conectam entidades (nomes) por meio de relações (verbos). Tanto os nomes quanto as relações precisam ser explicitados conforme padrões predefinidos, garantindo assim a interoperabilidade (Soergel, 2004).

A descrição dos elementos a serem organizados pode ocorrer a partir das características físicas ou do conteúdo semântico. No contexto das características físicas de um objeto informacional, há uma base nos métodos, técnicas e padrões estabelecidos pela Representação Descritiva. Essa disciplina, inserida na área de BCI, tem como propósito fornecer diretrizes para que os indexadores descrevam as informações sobre um objeto informacional de maneira padronizada, facilitando, assim, o processo de recuperação.

A representação descritiva emprega diversos padrões, variando de acordo com o tipo de objeto informacional a ser indexado. Em contextos como a catalogação de itens em acervos bibliográficos, o formato MARC é um padrão utilizado na Representação Descritiva. Esse formato facilita a recuperação da informação e viabiliza a importação e exportação de dados entre sistemas computacionais diversos, promovendo interoperabilidade entre diferentes *hardwares* e *softwares* disponíveis no mercado. A identificação de objetos informacionais é realizada por meio de metadados, descrições estruturadas que buscam representar as propriedades essenciais de dados específicos em um banco de dados e dos objetos de dados que

compõem a informação na Web. A inserção de metadados pode seguir diferentes padrões, sendo o padrão Dublin Core³ um exemplo destacado. Esse esquema de metadados é especialmente projetado para descrever objetos digitais, principalmente quando o objetivo é contextualizá-los na Web, permitindo que sistemas de busca os recuperem eficientemente (Lourenço, 2005).

Nem sempre a descrição física de um objeto é suficiente para atender às necessidades do usuário. A busca pode ocorrer mediante o conteúdo de natureza semântica, ou seja, o usuário pode necessitar realizar a busca em uma base de dados pelo assunto de um determinado objeto, não apenas pelo conteúdo de natureza sintática, como extensão do arquivo, tamanho, tipo etc. Para contemplar a representação do assunto contido em um objeto informacional, utiliza-se a representação temática. Essa abordagem envolve a análise manual ou automatizada do objeto informacional, buscando-se identificar os assuntos e as temáticas abordadas.

Há diferentes padrões para identificação e seleção dos assuntos/temas de um objeto informacional, e o processo de seleção pode variar conforme o tipo de objeto sendo indexado, o contexto da indexação, o perfil do público-alvo e os conhecimentos prévios do indexador. Os temas/assuntos selecionados no processo de indexação têm como propósito a composição de catálogos, índices e outros instrumentos de busca, que podem ser consultados diretamente pelos usuários do sistema informacional ou processados por sistemas computadorizados.

Uma vez que as categorias foram estabelecidas, o próximo passo é classificar os itens nesse modelo. Segundo Jacob (2004), o termo "classificação" na BCI engloba diferentes conceitos: um sistema de classes, ordenado de acordo com princípios predefinidos, utilizado para organizar entidades; um grupo ou classe dentro de um sistema de classificação; e o processo de atribuição de entidades a classes em um sistema de classificação. Concentrando-se no primeiro conceito, um sistema de classificação é uma ferramenta representacional empregada para organizar coleções de recursos informacionais.

³ O Padrão *Dublin Core* é um conjunto de elementos de metadados utilizado para descrever recursos informacionais, físicos ou digitais, de forma padronizada e interoperável. Desenvolvido originalmente em 1995, é amplamente adotado em bibliotecas, arquivos e repositórios digitais, e inclui elementos como título, autor, data, assunto e formato, facilitando a organização, a recuperação e o compartilhamento de informações entre diferentes sistemas.

Complementando essa definição, Tristão, Fachin e Alarcon (2004) referem-se à classificação como o ato e efeito de organizar, ordenar e dispor itens em classes. As classes agrupam itens que compartilham alguma característica comum, sendo a determinação e seleção das classes essencialmente relacionadas às necessidades de utilização do esquema estruturado.

A classificação, enquanto processo, compreende a atribuição ordenada e sistemática de cada entidade a uma única classe dentro de um sistema mutuamente exclusivo, seguindo princípios previamente estabelecidos de maneira organizada. O esquema elaborado para a classificação é considerado artificial e arbitrário, uma vez que tem o propósito de estabelecer uma organização significativa e adotar critérios para definir as classes, refletindo uma perspectiva única do domínio (Jacob, 2004). Segundo White (2011), alguns autores utilizam o termo "classificação" para se referir à organização, pressupondo que o processo de classificação significa simplesmente agrupar coisas que compartilham características comuns. Por outro lado, Soergel (2004) considera a classificação uma estrutura que organiza conceitos em uma hierarquia significativa. O autor exemplifica classificações com taxonomias, como a taxonomia dos seres vivos, e ontologias, que podem ser vistas como taxonomias com maior detalhamento nos tipos de relacionamentos entre os itens que compõem a classificação.

Uma classificação hierárquica geralmente apresenta uma estrutura de classes subordinantes, coordenadas e subordinadas, divididas com base em princípios específicos, resultando em classes simples ou compostas. Essas classificações podem ser especializadas ou gerais, dependendo do objetivo ou assunto em questão, ou facetadas, também conhecidas como classificações por facetas. As classificações facetadas, amplamente adotadas na BCI, originaram-se dos estudos do bibliotecário indiano Shiyali Ramamrita Ranganathan, que estabeleceu cinco categorias fundamentais para a classificação de qualquer tipo de assunto. Além disso, Ranganathan propôs uma série de princípios para orientar o processo de classificação facetada (Tristão; Fachin; Alarcon, 2004; Lima, 2015).

Segundo Soergel (2004), a classificação lida com conceitos, sendo necessário empregar termos para expressá-los. O termo, nesse contexto, é o símbolo linguístico utilizado para representar um conceito, que pode ser simples ou complexo. Por exemplo, o conceito de "Organização do conhecimento" envolve dois conceitos distintos (organização e conhecimento) e um terceiro conceito emergente dessa

combinação. De acordo com o autor, a existência de diferentes termos para se referir ao mesmo conceito, ou o mesmo termo sendo utilizado para diferentes conceitos destaca a importância de vocabulários controlados para mitigar ambiguidades.

Apesar das similaridades entre os processos de categorização e classificação, é importante reconhecer as diferenças significativas entre esses dois conceitos, pois elas têm implicações distintas nos ambientes informacionais. A confusão decorrente do uso indiscriminado dos termos categoria/categorização e classe/classificação como sinônimos amplia a complexidade da discussão. A categorização é empregada para segmentar o mundo da experiência em grupos ou categorias, nos quais os membros compartilham alguma semelhança imediata dentro de um contexto específico. Por outro lado, a classificação opera dividindo um conjunto de entidades em um sistema de exclusão mútua, organizando classes de acordo com princípios estabelecidos, sem acumulá-las. A estabilidade da referência fornecida por um sistema de classificação baseia-se na invariabilidade tanto do contexto quanto da composição dessas classes (Jacob, 2004). Sendo assim, entende-se que a categorização se relaciona ao processo de Organização e Representação do Conhecimento, ao passo que a classificação está relacionada ao processo de Organização e Representação da Informação.

3.1.1 Sistemas de Organização do Conhecimento (SOC)

Segundo Brascher e Café (2008), a representação do conhecimento pode ser realizada por meio de diversos tipos de SOC, que são sistemas conceituais estruturados para representar um domínio por meio da organização de conceitos e das relações semânticas entre eles.

Os mecanismos de representação do conhecimento proporcionam a formalização de objetos e suas relações em contextos predefinidos, permitindo a criação de linguagens documentárias verbais e notacionais. Essas linguagens visam à recuperação de informações e à organização de conteúdos informacionais de documentos no âmbito da BCI (Campos, 2004).

O termo SOC abrange uma variedade de esquemas destinados a organizar informações e facilitar a gestão do conhecimento. Diferentes categorias de SOC, como tesouros, esquemas de classificação, sistemas de cabeçalho e taxonomias, são

amplamente reconhecidas e aplicadas em sistemas de informação tanto modernos quanto tradicionais (Zeng; Hodge, 2011).

Carlan e Brascher (2011) explicam que os SOC são sistemas conceituais semanticamente estruturados, abrangendo termos, definições, relacionamentos e propriedades de conceitos. Sua principal finalidade é a padronização terminológica para facilitar a indexação e a recuperação de informações pelos usuários. São considerados ferramentas semânticas com vocabulários formalizados, utilizados no tratamento e recuperação de informações, tanto em ambientes web quanto tradicionais, inclusive na Web Semântica.

Os SOC podem ser interpretados como uma nova designação para as chamadas linguagens documentárias, integrando as tecnologias contemporâneas aos instrumentos previamente adotados na área de BCI. Por meio dos SOC, os indexadores têm a capacidade de criar esquemas e modelos de forma sistemática, utilizados nos processos de organização e representação do conhecimento e da informação (Carlan; Brascher, 2011).

Os SOC podem ser agrupados com base na complexidade de suas estruturas, variando desde estruturas mais simples, como dicionários, glossários e anéis de sinônimos, até modelos mais elaborados, exemplificados pelos modelos de relacionamento, que incluem tesouros, redes semânticas e ontologias. Esses sistemas de organização do conhecimento podem ser categorizados de acordo com funções específicas, como eliminação de ambiguidade, controle de sinônimos, estabelecimento de relações hierárquicas, estabelecimento de relações associativas e apresentação de propriedades (Lei Zeng, 2008).

Um Sistema de Organização do Conhecimento (SOC) fornece uma estrutura ou esquema para armazenar e organizar dados, informações e conhecimento sobre o mundo, bem como para a compreensão, recuperação ou descoberta, o raciocínio e outros propósitos. Esse tipo de sistema aborda conceitos, categorias, classes e os relacionamentos entre eles, além de termos ou outras designações para esses conceitos e relacionamentos. Em muitos casos, os SOCs incorporam uma quantidade significativa de asserções empíricas e teóricas, tornando-se desafiador estabelecer limites claros entre um SOC, uma base de conhecimento, um dicionário ou uma enciclopédia. A informação contida em muitos SOCs serve tanto como fonte de informação independente quanto como guia para localizar outras informações (Soergel, 2014).

Os SOCs podem ser vistos como uma iniciativa para aprimorar o acesso aos recursos digitais por meio do controle de vocabulário e organização do conhecimento. O controle de vocabulário desempenha um papel estratégico na redução de ambiguidades, especialmente porque as pessoas tendem a adotar termos diversos para expressar o mesmo conceito. Esses sistemas organizam e estruturam conceitos por meio de diferentes tipos de relacionamentos semânticos. A apresentação hierárquica dos conceitos, entre outras estruturas semânticas, auxilia tanto o indexador quanto o pesquisador na seleção do conceito mais adequado para seus propósitos (Tudhope; Nielsen, 2006).

Quanto aos padrões de construção de Sistemas de Organização do Conhecimento (SOC), o World Wide Web Consortium (W3C) desempenhou um papel fundamental no desenvolvimento de especificações e padrões que respaldam a utilização desses sistemas, tais como tesouros, esquemas de classificação, sistemas de cabeçalho de assunto e taxonomias no contexto da chamada Web Semântica. Esse conjunto de diretrizes e normas é conhecido como *Simple Knowledge Organization System* – SKOS (Carlan; Brascher, 2011).

Existem vários tipos de SOC, cada um designado para um propósito específico. Destacam-se as taxonomias, devido à sua ampla aplicação no processo de organização e representação do conhecimento e da informação. As ontologias também merecem destaque, considerando o seu potencial para representar estruturas complexas (Carlan; Brascher, 2011).

A taxonomia, um tipo de SOC, compreende termos preferidos que são conectados e organizados de forma hierárquica em categorias, dispostas numa hierarquia da mais geral para a mais específica (Lei Zeng, 2008). O termo foi inicialmente empregado por Carolus Linnaeus, que desenvolveu uma taxonomia para agrupar espécies biológicas conforme características físicas compartilhadas (Lei Zeng, 2008), resultando na conhecida organização taxonômica em gênero e espécie dos seres vivos.

Ao longo do tempo, o conceito de taxonomia evoluiu, especialmente em sua aplicação em diversos domínios. Na BCI, o termo taxonomia ganhou destaque para identificar classificações baseadas em hierarquias de assuntos, desempenhando um papel estratégico como ferramenta organizacional e de recuperação de informações, ao mesmo tempo que assegura o controle do vocabulário específico da área. Em BCI, as taxonomias têm a função de otimizar o tempo dedicado ao acesso à informação,

promovendo eficiência na recuperação de dados. Além disso, auxiliam no controle conceitual durante o processo de indexação, contribuem para a padronização terminológica e facilitam a identificação de termos relacionados, sinônimos e referências, agregando valor à linguagem utilizada em buscas dentro de uma base taxonômica (Aganette, 2010).

Campos e Gomes (2007) recomendam determinadas etapas no processo de desenvolvimento de uma taxonomia:

1. Captura do conhecimento – que pode se dar através de entrevistas com especialistas da área, de documentos existentes na instituição, de outros instrumentos classificatórios ou terminológicos;
2. Análise dos documentos/informações que serão agregados à taxonomia – adequação das informações existentes nos acervos a terminologia apresentada na taxonomia;
3. Elaboração da estrutura classificatória da taxonomia – que deve se pautar nos princípios teóricos de categorização, de formação e de ordenação das classes de conceitos;
4. Validação – se apoia como base para a certificação da proposta classificatória, visando atender às necessidades da comunidade para quem se destina. Esta ação permitirá que possamos atingir critérios de comunicabilidade, estimulação e compatibilidade (Campos; Gomes, 2007, p. 11).

O emprego de taxonomias viabiliza a criação de padrões abrangentes para ordenação e classificação de informações e contribui para que organizações identifiquem e relacionem atividades de valor agregado, otimizando esforços tanto na produção quanto na utilização do conhecimento. Ao fornecer uma estrutura de conceitos, a taxonomia permite que os usuários aprendam, possibilitando a compreensão dos termos adotados em um vocabulário específico de um domínio. A hierarquia apresentada nas taxonomias facilita a identificação de termos relacionados em classes e subclasses, sendo especialmente útil para usuários não familiarizados com o vocabulário do domínio (Campos; Gomes, 2007).

Embora as taxonomias sejam valiosas na categorização do conhecimento, é comum a necessidade de padronização do vocabulário utilizado no domínio. Nesse contexto, os tesouros surgem como instrumentos relevantes na área de BCI, desempenhando um papel fundamental na padronização do vocabulário. Os tesouros são vocabulários controlados empregados nos processos de indexação e busca, fornecendo informações úteis aos usuários que buscam identificar os termos mais apropriados em um determinado domínio (Lei Zeng, 2008).

Sob a perspectiva de sua estrutura, um tesouro é um vocabulário dinâmico que organiza conceitos e termos, estabelecendo relações semânticas entre eles. Essa

organização é composta por quatro elementos essenciais, conforme descrito por Maculan (2015): uma terminologia, que engloba descritores preferidos e não preferidos; uma estrutura gramatical, que define a forma de apresentação e composição dos descritores; uma rede paradigmática, indicando relações essenciais e estáveis entre conceitos; e uma rede sintagmática, estabelecendo relações contingentes entre descritores, válidas apenas em contextos específicos de uso.

Quanto às relações entre os termos de um tesauro, podem ser categorizadas em três tipos principais. As relações de equivalência abrangem diversos tipos de sinonímia (total e parcial), oposição, correspondência em outro idioma e variações linguísticas, indicando a presença de mais de um descritor para um mesmo conceito. As relações hierárquicas envolvem subordinação, subdividindo um assunto em subclasses (termos mais específicos); e superordenação, identificando as classes de nível mais alto (termos mais gerais). Por fim, as relações associativas, que não seguem hierarquias, ocorrem entre conceitos da mesma categoria e entre conceitos de categorias diferentes, estabelecendo ligações entre dois conceitos, independentemente de estarem relacionados hierarquicamente (Maculan, 2015).

Além dos Tesouros, um tipo adicional de Sistema de Organização do Conhecimento (SOC) amplamente adotado na área de BCI é a Ontologia, que integra uma estrutura classificatória utilizada tanto em Taxonomias quanto em Tesouros. A Ontologia possui propriedades específicas para cada classe dentro da estrutura classificatória, apresentando uma taxonomia abrangente e propriedades exaustivas. Funciona como um vocabulário conceitual e um modelo operacional que permite armazenar, pesquisar e raciocinar, baseado em instâncias e regras (Lei Zeng, 2008).

As Ontologias definem os termos utilizados para descrever e representar uma área do conhecimento, sendo aplicáveis tanto por pessoas quanto por máquinas, uma vez que são estruturadas conforme padrões e linguagens, como OWL, XML (*eXtensible Markup Language*), XML Schema, RDF e RDF Schema (Carlan; Brascher, 2011).

Semelhante ao Tesauro, uma Ontologia emprega os princípios de um SOC, priorizando funções, como a eliminação de ambiguidade, o controle de sinônimos e termos equivalentes, a explicitação dos relacionamentos e a apresentação tanto dos relacionamentos quanto das propriedades de cada conceito integrante do esquema (Lei Zeng, 2008).

A BCI possui um histórico de pesquisa bastante consolidado em termos de estratégias de organização e representação da informação e do conhecimento. Observa-se que princípios da BCI, como teorias, métodos, processos, técnicas e instrumentos, têm sido amplamente aplicados em diversos domínios do conhecimento, enfatizando o caráter e a interdisciplinaridade característica da área.

3.2 Ontologias

Na esfera da Ciência da Informação, as ontologias emergem como um elemento fundamental para a estruturação e a representação do conhecimento de maneira semântica. Tais estruturas conceituais desempenham uma função crucial na organização e no compartilhamento de informações em diversas disciplinas (Gruber, 1993).

De acordo com Gruber (1993), uma ontologia pode ser compreendida como uma especificação explícita de uma conceituação compartilhada. Formaliza o significado dos termos dentro de um domínio específico, delineando as relações entre esses termos e proporcionando uma estrutura comum para a representação de conhecimento.

No contexto da Ciência da Computação, ontologias são frequentemente construídas utilizando-se linguagens como OWL E RDF, oferecendo uma representação formal e estruturada dos conceitos e relações subjacentes a um determinado domínio (Guarino; Welty; Gruninger, 2009).

Ontologias, formalmente representando conceitos vinculados a um domínio específico e seus relacionamentos correspondentes, definem um vocabulário comum para compartilhamento de informações, incluindo interpretações por máquina dos conceitos de domínio (Hananto; Serdült; Kryssanov, 2021; Jalil; Nasir; Nasir, 2021).

As aplicações das ontologias são amplas e abrangem diversos campos, desde a biomedicina até a gestão do conhecimento e a Web Semântica. Em pesquisas biomédicas, por exemplo, ontologias são empregadas para padronizar terminologias médicas, facilitando a interoperabilidade entre sistemas de informação e promovendo avanços na pesquisa (Ruttenberg *et al.*, 2007).

O estudo das ontologias revela-se como uma área de crescente relevância, não apenas na academia, mas também na indústria e em diversas disciplinas científicas. A construção de ontologias eficazes e a adaptação contínua às mudanças no

conhecimento representam desafios que, quando superados, têm o potencial de revolucionar a organização e a disseminação do conhecimento em variados domínios.

Na literatura, as ontologias são amplamente adotadas como uma solução eficiente para a descrição de informações. Elas permitem a especificação formal dos componentes do domínio, introduzindo uma representação de padrões compartilháveis e reutilizáveis, facilitando, assim, a compreensão e a análise do conhecimento sobre o domínio em questão (Drissi *et al.*, 2021).

A crescente adoção de ontologias pela comunidade científica destaca seu papel promissor em diversas áreas, especialmente na sumarização de documentos. Isso contribui para a identificação de frases relevantes nos documentos, incorporando conhecimento ontológico para gerar resumos que mostram os conceitos críticos e sua correlação com a consulta do usuário, evitando ambiguidades (Jalil; Nasir; Nasir, 2021).

Apesar dos benefícios proporcionados pelas ontologias, os desafios persistem em sua criação e manutenção. A evolução contínua do conhecimento e a necessidade de atualizações frequentes demandam abordagens dinâmicas na construção de ontologias (Mizoguchi; Sunagawa; Kozaki, 2019). Além disso, a integração em larga escala de ontologias, a heterogeneidade de fontes de dados e a garantia da qualidade semântica constituem desafios contínuos a serem superados. Tendências recentes indicam o desenvolvimento de ontologias contextuais, capazes de considerar o contexto temporal e espacial na representação do conhecimento (Santos; Simão; Pinto, 2020). Essa abordagem dinâmica busca aprimorar a adaptabilidade das ontologias em ambientes complexos e em constante mutação.

A análise das metodologias destinadas ao desenvolvimento de ontologias revela que a maioria delas abrange etapas como análise de domínio, conceituação, implementação, avaliação, instanciação e apresenta um exemplo de ontologia de domínio. No entanto é observado que a maioria dessas metodologias não fornece todos os detalhes das técnicas e atividades envolvidas no processo. Apesar da abrangência dessas etapas, é evidente a lacuna nas metodologias estudadas em relação a áreas cruciais. Notadamente, a falta de suporte para manutenção, documentação, construção de ontologias colaborativas, reutilização, integração, interoperabilidade, localização de ontologias e estimativa de recursos humanos é uma limitação identificada (Sattar *et al.*, 2021).

A construção manual de ontologias é uma tarefa demorada e requer especialistas no domínio para validar cada etapa do processo, sendo limitada pelo chamado “gargalo de aquisição de conhecimento”. O desenvolvimento de ontologias é complexo, tedioso e demanda uma metodologia bem projetada. A construção manual de um modelo conceitual formal para uma ontologia é trabalhosa, e os desenvolvedores geralmente evitam essa abordagem. A demanda por tempo e conhecimento profundo do domínio é significativa, resultando em projetos de pesquisa focados na geração e/ou preenchimento automático ou semiautomático de ontologias (Deb *et al.*, 2018; Nguyen; Tettamanzi, 2019; Drissi *et al.*, 2021; Sattar *et al.*, 2021; Rajaonarivo; Arakawa, 2021).

A criação de ontologias ainda se apresenta como um processo demorado e desafiador, exigindo colaboração intensiva na área de engenharia ontológica. A dificuldade para especialistas do domínio dedicarem tempo ao desenvolvimento da ontologia é reconhecida, sendo o custo da utilização desses especialistas um fator limitante em muitos casos. Para superar esse desafio, foram propostos métodos, ferramentas e até estratégias de *crowdsourcing*, visando tornar o processo mais colaborativo e menos dispendioso (Yadav *et al.*, 2018; Enea *et al.*, 2018).

3.2.1 Metodologias para construção de ontologias

As ontologias são estruturas que organizam e relacionam informações de maneira lógica e padronizada, desempenhando um papel fundamental na Web Semântica, cujo objetivo é tornar a internet mais inteligente e eficiente. A criação dessas estruturas, quando realizada manualmente, é um processo complexo, especialmente considerando o aumento exponencial de dados não estruturados disponíveis na atualidade. Para lidar com esse desafio, surgiram as metodologias de aprendizagem de ontologias, que automatizam parte do processo, ampliando sua eficiência e escalabilidade.

Um *survey* conduzido por Asim *et al.* (2018a) apresenta uma revisão abrangente sobre as técnicas de aprendizagem de ontologias, abordando métodos, ferramentas, desafios e aplicações em diferentes domínios do conhecimento. O estudo destaca a relevância das ontologias tanto para a Web Semântica quanto para sistemas inteligentes, devido à sua capacidade de fornecer interoperabilidade e agregar semântica a grandes volumes de dados, especialmente os não estruturados.

Os autores classificaram as metodologias analisadas em três categorias principais: linguísticas, estatísticas e lógicas. As técnicas linguísticas exploram características da linguagem natural para identificar termos e relações em textos. As técnicas estatísticas, por sua vez, analisam padrões e associações em grandes volumes de dados, permitindo a organização eficiente das informações. Já as técnicas lógicas utilizam regras formais e raciocínio lógico, proporcionando ontologias robustas e reutilizáveis. Embora cada abordagem apresente vantagens e limitações, a combinação dessas técnicas pode resultar em processos mais eficazes e adaptáveis (Asim *et al.*, 2018a).

Apesar dos avanços alcançados, desafios ainda persistem, como a automação completa do processo de aprendizagem de ontologias e a avaliação da qualidade das ontologias geradas. A avaliação é essencial para garantir que as estruturas criadas atendam às necessidades específicas dos usuários. Diversos métodos têm sido explorados, incluindo a comparação com padrões existentes, a validação humana e testes em aplicações específicas. No entanto a escolha do método mais adequado depende do contexto de uso e das particularidades de cada aplicação (Asim *et al.*, 2018a).

Além das diversas propostas de metodologias identificadas na revisão de literatura, esta pesquisa destacou quatro metodologias elaboradas por pesquisadores do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Essas metodologias foram selecionadas com base na análise de elementos essenciais que as caracterizam. Entre esses elementos, estão a apresentação de um conjunto estruturado de procedimentos, técnicas, ferramentas e documentos, além da organização em fases bem definidas, subdivididas em subfases ou etapas com as respectivas descrições. Esse conjunto de elementos orienta os usuários na aplicação da metodologia e oferece mais clareza quanto às possibilidades, limitações e aos cuidados necessários para o sucesso na aplicação da metodologia em questão. A seguir, uma breve apresentação de cada uma das quatro metodologias selecionadas.

A primeira metodologia identificada, denominada OntoForInfoScience, foi desenvolvida por Mendonça (2015) e concebida para oferecer um modelo detalhado e sistemático de construção de ontologias, voltado para especialistas em Organização do Conhecimento na área de Ciência da Informação. A metodologia proposta busca superar algumas das limitações de metodologias existentes, como o pouco

detalhamento das etapas, a linguagem excessivamente técnica e a dificuldade de aplicação para profissionais sem formação em computação. Inspirada nas metodologias NeOn, Methontology e Método 101, a OntoForInfoScience adapta e detalha elementos dessas abordagens, tornando-as mais acessíveis e úteis para cientistas da informação.

A metodologia OntoForInfoScience é composta por fases estruturadas que abrangem todo o ciclo de desenvolvimento de uma ontologia. O processo se inicia com a avaliação da necessidade da ontologia, momento em que são identificadas as demandas do projeto e analisada sua viabilidade. Na sequência, define-se o escopo e os requisitos, determinando os conceitos e as relações que devem ser representados. A aquisição e extração de conhecimento ocorre em seguida, envolvendo a coleta de informações de fontes variadas, como literatura científica, bases de dados e especialistas do domínio, utilizando técnicas como entrevistas e análise documental. O material coletado é então organizado na fase de conceitualização, que inclui a criação de taxonomias e diagramas de classes para representar as relações entre os conceitos de forma preliminar.

A etapa de fundamentação ontológica é fundamental, pois utiliza bases teóricas sólidas, como princípios filosóficos e formalismos ontológicos, para garantir a coerência e a consistência do modelo. Na fase de formalização, a ontologia é codificada em linguagens computacionais, o que permite sua interpretação por sistemas automatizados. Posteriormente, a qualidade da ontologia é avaliada com base em métricas, como completude, consistência e reusabilidade, complementadas por validações de especialistas e testes práticos. A documentação gerada fornece uma explicação detalhada sobre a estrutura e os conceitos da ontologia, utilizando artefatos textuais e gráficos que facilitam sua compreensão e reutilização. Por fim, a ontologia é disponibilizada em repositórios ou plataformas online, assegurando seu acesso público e promovendo sua integração em diversas aplicações.

A eficácia da OntoForInfoScience foi comprovada por meio da criação da ontologia HEMONTO, que aborda os componentes do sangue humano utilizados em tratamentos terapêuticos. A HEMONTO integra conceitos biomédicos, propriedades formais e axiomas lógicos, oferecendo uma representação detalhada e consistente dos aspectos fisiológicos, hemocomponentes e hemoderivados. Essa aplicação prática demonstra a utilidade da metodologia, especialmente ao atender às demandas

de cientistas da informação e de outros profissionais que enfrentam desafios técnicos no desenvolvimento de ontologias.

A segunda metodologia identificada foi desenvolvida por Farinelli (2017), sendo apresentada em sua tese de doutorado intitulada "Realismo Ontológico Aplicado à Interoperabilidade Semântica entre Sistemas de Informação: um Estudo de Caso do Domínio Obstétrico e Neonatal". A metodologia propõe uma abordagem inovadora para resolver desafios na troca de informações entre sistemas de saúde. A fragmentação dos registros eletrônicos de saúde (EHRs) em diferentes sistemas, muitas vezes incompatíveis, é um grande obstáculo para a continuidade do cuidado, especialmente nos processos relacionados ao pré-natal, parto e pós-parto. A proposta metodológica visa promover a interoperabilidade semântica, ou seja, garantir que os sistemas "compreendam" os dados de forma mais consistente e com menos ambiguidades.

A metodologia proposta por Farinelli (2017) é estruturada em duas fases principais: exploratória e empírica. Na fase exploratória, a autora realizou um levantamento detalhado sobre o domínio obstétrico e neonatal, mapeando os conceitos e processos envolvidos. Além disso, foram analisadas metodologias existentes para a construção de ontologias, com o objetivo de identificar elementos que pudessem ser integrados a uma abordagem mais adequada para o contexto dos sistemas de saúde brasileiros. Essa fase resultou na escolha de uma metodologia fundamentada no Realismo Ontológico e nos princípios da *OBO Foundry*, um conjunto de diretrizes reconhecido por promover ontologias de alta qualidade.

Na fase empírica, a metodologia foi implementada em cinco etapas. Na primeira, chamada de conceitualização, o conhecimento do domínio foi organizado, identificando os conceitos mais relevantes, suas relações e restrições. Na etapa seguinte, de concepção, os requisitos para a ontologia foram detalhados em um documento estruturado, especificando os critérios que a ontologia deveria atender. Na fase de *design*, foi elaborada a arquitetura da ontologia, definindo hierarquias de classes e relações entre os conceitos. Em seguida, na fase de implementação, a ontologia foi formalizada em uma linguagem computacional, utilizando ferramentas como o Protégé. Por fim, na etapa de entrega, a ontologia foi avaliada quanto à sua qualidade e validada em um cenário prático de troca de dados entre sistemas de saúde.

Essa metodologia resultou na criação da ontologia OntONEo, um modelo que organiza e integra informações do domínio obstétrico e neonatal. A OntONEo foi projetada para unificar dados provenientes de diversos sistemas de saúde, facilitando a representação de processos, como diagnósticos, intervenções e resultados de cuidado materno e neonatal. A eficácia da ontologia foi demonstrada por meio de sua capacidade de promover consistência, reusabilidade e integração semântica entre sistemas heterogêneos, evidenciando seu potencial para melhorar os processos relacionados ao cuidado.

A terceira metodologia identificada foi desenvolvida por Fonseca (2023), também como parte de sua tese de Doutorado, intitulada "Modelagem de Conhecimento Baseada em Ontologia e Dados Abertos: a Representação do Domínio da Tomada e Prestação de Contas Anual dos Gestores Públicos Federais". A metodologia apresenta uma abordagem inovadora para modelar o conhecimento no domínio da gestão pública, utilizando ontologias e dados abertos. O objetivo central é reduzir a lacuna entre a vasta quantidade de dados gerados nos processos de tomada e prestação de contas e sua efetiva utilização por cidadãos e desenvolvedores de *softwares*, promovendo maior transparência e controle social.

A metodologia criada por Fonseca (2023) baseou-se na adaptação da metodologia OntoForInfoScience, desenvolvida por Mendonça (2015), visando atender às especificidades do domínio de conhecimento estudado, a tomada e prestação de contas anual dos gestores públicos. O processo começa pela especificação, que define o escopo da ontologia e os requisitos necessários para representar adequadamente o domínio da tomada e prestação de contas. Em seguida, realiza-se a aquisição de conhecimento, na qual dados e informações são coletados de documentos como relatórios de gestão e normas emitidas pelo Tribunal de Contas da União (TCU). Essa etapa inclui a identificação de conceitos, termos e relações relevantes para o contexto.

Na fase de conceitualização, o conhecimento coletado é estruturado em taxonomias e diagramas conceituais, representando entidades como órgãos públicos, processos e documentos. Essa estruturação inicial é formalizada utilizando a linguagem *Web Ontology Language* (OWL), com suporte da ferramenta Protégé, permitindo uma representação formal e semântica dos conceitos, propriedades e restrições. Após a formalização, a ontologia é avaliada quanto à consistência, completude e reusabilidade, garantindo sua conformidade com padrões semânticos e

a possibilidade de reutilização em diferentes aplicações. A documentação detalhada do modelo foi elaborada para facilitar seu uso e compreensão.

Como resultado dessa metodologia, foi desenvolvida a ontologia *OntoAccount*, que organiza e estrutura os conceitos e relações no domínio da tomada e prestação de contas anual dos gestores públicos federais. Essa ontologia permite transformar dados frequentemente armazenados em formatos não estruturados, como PDF, em um modelo semântico que facilita a interoperabilidade e o reuso de informações. A aplicação prática do modelo foi demonstrada em cenários que simulam o uso de dados estruturados para o desenvolvimento de *softwares* cívicos, capazes de monitorar políticas públicas e gerar informações sobre a eficácia, a eficiência e a efetividade da gestão pública.

Por fim, a quarta metodologia identificada foi desenvolvida por Almeida (2022), compilada no primeiro volume de um livro que reúne, segundo o autor, mais de 20 anos de experiência em pesquisa, ensino e consultoria sobre ontologias, no âmbito da metodologia denominada Projeto de Recursos Ontológicos (PRONTO). A metodologia tem como objetivo auxiliar estudantes e especialistas em gestão do conhecimento a construir ontologias, independentemente do tamanho ou nível de complexidade.

A metodologia PRONTO é composta por etapas bem definidas que guiam o processo desde a concepção inicial até a formalização e documentação. O objetivo é fornecer um roteiro estruturado e adaptável para criar ontologias consistentes e robustas, alinhadas às necessidades do domínio de aplicação. O processo começa com a especificação, etapa em que se define o propósito, o escopo, os usuários e os usos pretendidos da ontologia. Além disso, são detalhados os requisitos funcionais e não funcionais, assegurando que o projeto atenda aos objetivos estabelecidos. Nessa fase, é fundamental descrever claramente o objetivo principal e a abrangência esperada da ontologia.

Em seguida, parte para a aquisição de conhecimento que consiste na coleta de termos candidatos que integrarão a ontologia. Esses termos são obtidos a partir de diferentes fontes, como artefatos de representação (tesauros, glossários e dicionários), documentos relacionados ao domínio e interações com especialistas. Técnicas específicas são empregadas, incluindo inspeção manual, ferramentas como *AntCont* e *Sketch Engine*, protocolos verbais, entrevistas estruturadas ou semiestruturadas, e abordagens colaborativas como *card sorting* e *design thinking*.

Na etapa de organização de termos representativos de entidades, os termos coletados são estruturados em uma taxonomia preliminar, formando a base hierárquica da ontologia. Ferramentas como o Protégé são utilizadas para organizar os termos em classificações apropriadas, assegurando coerência e clareza na representação das entidades.

A organização de termos representativos de relacionamentos estabelece as conexões entre as classes e entidades definidas, abarcando tanto relacionamentos taxonômicos quanto não taxonômicos. Essa etapa é essencial para modelar as interações e as dependências entre os conceitos do domínio em foco.

Em seguida, na definição de entidades, as classes e os elementos da ontologia são descritos detalhadamente. O método aristotélico é empregado para criar definições precisas que associam a entidade ao seu gênero próximo e às suas diferenças específicas. Essas definições são registradas em linguagem natural, promovendo maior acessibilidade e compreensão.

A formalização transforma as definições criadas em axiomas formais no formato OWL. Por meio da sintaxe de Manchester no editor Protégé, esses axiomas permitem o uso de raciocinadores automáticos para validar a lógica da ontologia e inferir novos conhecimentos.

Na etapa de avaliação, a ontologia é revisada para garantir consistência e adequação aos requisitos especificados. Ferramentas como o Protégé auxiliam na análise de visualizações inferidas e no uso da funcionalidade *reasoner* para identificar possíveis erros ou inconsistências.

A documentação reúne todo o material utilizado ao longo do processo, consolidando as decisões e as atividades realizadas. Um repositório de documentos é criado para consultas futuras, e um resumo da ontologia é produzido, facilitando sua aplicação e manutenção.

A metodologia apresentada proporciona um processo integrado e abrangente, cobrindo todas as etapas necessárias para o desenvolvimento de ontologias robustas e bem estruturadas. A adoção de ferramentas especializadas e técnicas variadas assegura a adaptabilidade e a eficiência do processo, promovendo resultados consistentes e alinhados ao domínio de aplicação.

3.2.2 Aprendizagem de Ontologias

Atualmente, as ontologias emergem como elementos essenciais, especialmente no contexto da Web Semântica, destacando-se por sua ampla aplicabilidade em aplicativos Web. Além desse âmbito, as ontologias demonstram relevância em áreas como Gestão do Conhecimento, Inteligência Artificial e Processamento de Linguagem Natural, desempenhando um papel central na integração, questionamento e estruturação de conceitos e relações específicas de determinados domínios (Figueres-Esteban; Hughes; Van Gulijk, 2016; Asim *et al.*, 2018a; Belhoucine; Mourchid, 2018).

O aumento exponencial de dados não estruturados na Web impõe desafios significativos ao processo de construção de ontologias, tornando-o mais complexo, demorado e, em muitos casos, oneroso. Para enfrentar essa realidade, é necessário adotar estratégias que permitam a automação total ou parcial desse processo (Zahra *et al.*, 2014; Asim *et al.*, 2018a; Desul, 2019). Nesse contexto, surgem abordagens de geração automática ou semiautomática de ontologias, conhecidas como "*ontology learning*" ou aprendizagem de ontologias (Salatino *et al.*, 2020; Khemiri *et al.*, 2021).

A aprendizagem de ontologias configura-se como um processo de aquisição de conhecimento a partir de texto, empregando métodos e técnicas provenientes de diversas áreas, como Processamento de Linguagem Natural (PNL), mineração de dados e aprendizado de máquina (Drissi *et al.*, 2021). Seu principal objetivo é compartilhar e reutilizar eficientemente o conhecimento de um domínio. Para gerar uma ontologia em um campo específico, é frequentemente necessário criar um modelo de conhecimento na forma de uma taxonomia, que organiza hierarquicamente os conceitos e as relações (Alharbi; Tamma; Grasso, 2021; Sadeghianasl *et al.*, 2021).

Ao longo das últimas duas décadas, a construção de ontologias tem sido amplamente estudada, resultando na proposição de diversos métodos e ferramentas. No início dos anos 2000, esforços concentraram-se no desenvolvimento de abordagens semiautomáticas, voltadas para a extração de ontologias a partir de textos brutos, como artigos de notícias, publicações científicas e análises de produtos. Apesar de reduzirem parcialmente o esforço humano, essas abordagens ainda dependem significativamente de especialistas, limitando sua escalabilidade e eficiência em cenários de mudanças frequentes (Santosa; Miyazaki; Han, 2021).

A aprendizagem de ontologias surgiu, então, como uma disciplina fundamental na engenharia de ontologias, buscando automatizar a geração de ontologias a partir de fontes variadas, como texto, bancos de dados e dicionários (Aggoune, 2018). Essa abordagem torna-se cada vez mais relevante diante da necessidade de dar significado à vasta quantidade de dados disponíveis na Web. Contudo, ainda enfrenta desafios relacionados à qualidade da informação extraída e incorporada na ontologia, o que demanda esforços contínuos de pesquisa para aprimorar os resultados (Enea *et al.*, 2018).

O aumento do volume de documentos, tanto na Web quanto em bibliotecas digitais, tem impulsionado o interesse em aprendizagem de ontologias nos últimos anos. Esse cenário estimulou o desenvolvimento de métodos e ferramentas computacionais para organizar, representar e explorar automaticamente grandes *corpora* de informações (Khemiri *et al.*, 2021). Como campo de pesquisa emergente, a aprendizagem de ontologias visa superar o desafio da aquisição de conhecimento, especialmente na Web Semântica, gerando ontologias a partir de entradas como texto em linguagem natural, ontologias existentes expressas em OWL e dados de instância em RDF (Malchiodi; Tettamanzi, 2018).

Esse campo permite a construção, a ampliação ou a adaptação semiautomática de ontologias por meio da coleta e estruturação de conhecimento extraído de textos disponíveis na Web (Gómez-Suta; Echeverry-Correa; Soto-Mejía, 2020). A aprendizagem de ontologias envolve a criação de estruturas fundamentais que compõem uma ontologia: termos, conceitos e relações. Termos são representações textuais de conceitos, sendo unidades lexicais explícitas no *corpus* analisado. Conceitos, por sua vez, são expressões mentais cuja interpretação emerge do agrupamento contextual de termos relacionados. Relações estabelecem conexões entre conceitos e podem ser categorizadas como taxonômicas, que definem hierarquias, ou não taxonômicas, que descrevem interações adicionais relevantes (Gómez-Suta; Echeverry-Correa; Soto-Mejía, 2020).

Essas metodologias possibilitam uma extração eficiente de termos e relacionamentos, especialmente por meio de técnicas de mineração de dados. A aprendizagem de ontologias é amplamente reconhecida como uma tecnologia-chave para a construção de bases de conhecimento mais efetivas, superando limitações das abordagens tradicionais (Liao; Xiong; Chen, 2018; Liao; Wu; King, 2019). Técnicas como mineração de texto, PNL e aprendizado de máquina têm se destacado no

processo de extração de termos e conceitos em domínios específicos, demonstrando sua aplicabilidade e importância na construção automatizada de ontologias (Santosa; Miyazaki; Han, 2021).

A aprendizagem de ontologias se vale de um conjunto heterogêneo de métodos oriundos de distintas disciplinas, como a linguística computacional, a estatística aplicada, a lógica formal e o aprendizado de máquina. A literatura especializada propõe distintas taxonomias para a categorização dessas técnicas, sendo a proposta por Asim *et al.* (2018) amplamente referenciada, ao agrupá-las em três grandes classes: linguísticas, estatísticas e lógico-formais:

a) Técnicas Linguísticas

As técnicas linguísticas concentram-se na análise da estrutura da linguagem escrita ou falada, procurando compreender como os termos aparecem e se organizam nos textos. A ideia central é que a linguagem carrega pistas importantes sobre os conceitos e suas relações.

Uma das técnicas mais utilizadas nesse grupo é a análise gramatical, que busca identificar, por exemplo, quais palavras em uma frase são substantivos, verbos, adjetivos, entre outras categorias. Essa classificação ajuda a localizar candidatos a conceitos (geralmente substantivos) e ações ou relações (geralmente verbos).

Outra técnica relevante é o uso de padrões linguísticos recorrentes, também chamados de padrões léxico-sintáticos. Um exemplo clássico são expressões como "animais como gatos e cachorros", que indicam que "gatos" e "cachorros" são tipos do conceito mais geral "animais". Esses padrões ajudam a descobrir relações hierárquicas, como "é um tipo de".

Além disso, a identificação de nomes próprios ou entidades específicas (como nomes de pessoas, organizações ou locais) é útil para reconhecer exemplos concretos de conceitos, o que pode ser valioso na fase de instanciamento de uma ontologia.

Em resumo, as técnicas linguísticas operam por meio da leitura cuidadosa e sistemática dos textos, com apoio de ferramentas automatizadas que imitam a forma como humanos interpretam a linguagem, extraindo estrutura e significado.

b) Técnicas Estatísticas

As técnicas estatísticas baseiam-se na frequência e no padrão de aparecimento das palavras nos textos. A lógica é simples: palavras que aparecem com muita

frequência, ou que costumam aparecer juntas, provavelmente têm importância conceitual no domínio tratado.

Uma técnica comum é o cálculo da frequência de palavras, que serve para identificar quais termos são mais relevantes em um conjunto de documentos. No entanto essa contagem bruta nem sempre é suficiente, pois algumas palavras aparecem frequentemente em todos os textos, mas não dizem muito sobre o tema específico (como “o”, “a”, “é”, etc.). Por isso, essas técnicas costumam aplicar fórmulas que valorizam palavras que aparecem muitas vezes em um texto, mas não em todos, destacando aquelas mais representativas.

Outra técnica estatística é a análise de coocorrência, que observa quais palavras costumam aparecer próximas umas das outras. Por exemplo, se “médico” e “hospital” aparecem juntos com frequência, é provável que haja uma relação entre os conceitos que representam. Essa informação pode ser usada para sugerir relações semânticas entre conceitos.

As técnicas estatísticas não interpretam o conteúdo do texto diretamente, mas ajudam a identificar padrões numéricos que podem apontar para conceitos e relações importantes, especialmente em grandes volumes de dados.

c) Técnicas Lógico-Formais

As técnicas lógico-formais operam em um nível mais abstrato e conceitual. Elas são utilizadas principalmente para organizar os conceitos extraídos de forma coerente e precisa, por meio de regras e estruturas lógicas. Essas técnicas buscam responder a perguntas como: um conceito pode pertencer a mais de uma categoria? Quais são os limites e restrições de determinada relação?

Por exemplo, se uma ontologia afirma que “todo professor ensina alguma disciplina”, essa afirmação precisa ser formalizada e verificada para garantir que ela não entre em contradição com outras afirmações.

A formalização é feita por meio de representações que seguem regras da lógica matemática. Esse processo permite que sistemas computacionais realizem inferências, ou seja, tirem conclusões a partir do que foi previamente modelado.

As técnicas lógico-formais são fundamentais para conferir consistência e integridade à ontologia, especialmente quando ela será utilizada em sistemas de informação mais complexos, como assistentes virtuais, mecanismos de busca semânticos ou plataformas de integração de dados.

d) Técnicas Híbridas

Na prática, é comum que as abordagens de aprendizagem de ontologias combinem duas ou mais das técnicas anteriores, formando o que se chama de abordagens híbridas. Isso ocorre porque cada técnica, isoladamente, tem limitações que podem ser superadas por outras.

Por exemplo, uma técnica linguística pode identificar corretamente um termo, mas não entender sua importância no contexto. Uma técnica estatística, por sua vez, pode indicar que o termo é relevante com base em sua frequência. Ao combinar as duas, é possível ter mais segurança na identificação de conceitos-chave. Além disso, em fases mais avançadas da construção de ontologias, como a validação e a formalização dos conceitos e relações, entram em cena as técnicas lógico-formais, que ajudam a estruturar os elementos extraídos de forma coerente. Essa integração de técnicas permite construir ontologias mais completas, consistentes e adequadas a diferentes contextos de aplicação.

Em suma, a aprendizagem de ontologias destaca-se como um campo essencial para lidar com os desafios crescentes da Web Semântica, possibilitando a extração e a organização de conhecimento a partir de grandes volumes de dados não estruturados. Por meio de métodos automatizados e semiautomatizados, essa área tem promovido avanços significativos na construção de ontologias, integrando tecnologias como mineração de texto, processamento de linguagem natural e aprendizado de máquina. Apesar dos desafios relacionados à qualidade da informação extraída e à necessidade de infraestrutura robusta, os progressos recentes apontam para um futuro promissor, em que as ontologias desempenharão um papel central na representação e interoperabilidade do conhecimento em diversos domínios. Assim, a contínua pesquisa e o desenvolvimento nesse campo são fundamentais para potencializar o impacto das ontologias na organização e interpretação de dados complexos.

3.3 Governo Eletrônico

O advento da era digital transformou significativamente a prestação de serviços governamentais, impulsionando o governo eletrônico como um fenômeno global. O termo “governo eletrônico” refere-se ao uso das Tecnologias da Informação e Comunicação (TIC) para fornecer serviços públicos, interagir com cidadãos e melhorar

a eficiência e a transparência do aparato estatal (Cruz, 2001). O governo eletrônico representa uma evolução significativa na forma como os governos interagem com os cidadãos e oferecem serviços, sendo um marco na modernização da administração pública. Pesquisas recentes apontam para o potencial significativo da digitalização dos serviços governamentais, sugerindo uma possível economia de aproximadamente 1 trilhão de dólares anualmente em escala global (McGuinness; O'Neill; Cross, 2020). Esse fenômeno tem impulsionado a transformação digital em inúmeros países ao redor do mundo, indo além da mera introdução de tecnologias ou automação de processos.

A transformação digital no setor público demanda a implementação de ações estratégicas que facilitem o lançamento e a adoção efetiva de serviços públicos digitais (Dilmegani; Korkmaz; Lundqvist, 2014; Anderson; Fyall; Hoffman, 2018; Leão, 2018; Stern, 2018). Algumas medidas específicas, como o fornecimento de plataformas de Tecnologia da Informação (TI) comuns e a definição de padrões técnicos, promovem a interoperabilidade, reduzem custos operacionais e melhoram a qualidade dos serviços prestados aos cidadãos (Gil-Garcia; Helbig; Ojo, 2020). Além disso, a efetiva transformação digital requer uma abordagem estratégica para lidar com o conhecimento e a informação de maneira significativa, otimizando a entrega de serviços digitais (Scholl, 2018). A adoção de ontologias pode melhorar a interoperabilidade semântica entre diferentes sistemas de informação, permitindo uma comunicação mais eficaz e precisa entre órgãos governamentais (Wang; Wang; Wang, 2019).

Entretanto o governo eletrônico não está isento de desafios. A segurança cibernética, a inclusão digital e a proteção da privacidade emergem como preocupações cruciais para sua implementação (Castro; Santos; Pereira, 2017). Além disso, a resistência cultural e a falta de infraestrutura tecnológica podem limitar a eficácia dessas iniciativas (Chadwick, 2006). Superar essas barreiras requer investimentos contínuos em infraestrutura tecnológica, definição de padrões claros e promoção de uma cultura de inovação. Além disso, é essencial garantir que todos os cidadãos possam participar plenamente na sociedade digital, priorizando a inclusão digital e a proteção da privacidade.

Apesar desses desafios, as perspectivas futuras do governo eletrônico são promissoras. A integração de tecnologias emergentes, como inteligência artificial e *blockchain*, têm o potencial de potencializar a eficiência e a confiabilidade dos serviços

oferecidos (Silva; Santos; Ferreira, 2020). A contínua colaboração internacional e a troca de boas práticas também são fundamentais para o aprimoramento contínuo do governo eletrônico (United Nations, 2018). Inovações tecnológicas oferecem oportunidades únicas para aprimorar ainda mais a eficiência e a transparência do setor público, fortalecendo o impacto do governo eletrônico em escala global.

Diante desse cenário, destaca-se a necessidade de uma abordagem holística na implementação da transformação digital no setor público. Essa abordagem deve integrar o fornecimento de plataformas de TI comuns, a definição de padrões técnicos e o aprimoramento dos processos de organização e representação do conhecimento, garantindo a interoperabilidade e a eficiência dos serviços digitais. Além disso, é essencial alinhar esses esforços a estratégias que promovam a inclusão digital, a segurança cibernética e a proteção da privacidade, de modo a assegurar que todos os cidadãos possam usufruir dos benefícios do governo eletrônico. Com a integração de tecnologias emergentes e a colaboração internacional, o governo eletrônico tem o potencial de transformar a administração pública, aprimorando a transparência, a eficiência e a qualidade dos serviços prestados. Assim, o êxito do governo eletrônico depende de avanços tecnológicos e de um compromisso contínuo com a inovação, a equidade e o bem-estar da sociedade.

3.4 Plataforma de serviços públicos Gov.br

Desenvolvido pela Secretaria de Governo Digital do Ministério da Economia, o GOV.BR tem como objetivo principal proporcionar aos cidadãos um ponto único de acesso aos serviços públicos federais. A plataforma abrange uma ampla gama de temas, como benefícios sociais e guias para abertura de empresas, consolidando informações que antes estavam dispersas em diferentes plataformas. Com uma interface amigável e navegação intuitiva, o portal busca atender às demandas de uma população cada vez mais conectada e exigente.

O GOV.BR está alinhado aos princípios do governo aberto, sendo fundamentado pela Lei de Acesso à Informação (Lei n. 12.527/2011). Essa lei preconiza a transparência e a facilidade de acesso à informação, objetivos incorporados pela plataforma ao disponibilizar dados governamentais e relatórios. Além disso, o portal reflete as diretrizes da Parceria para Governo Aberto (OGP), promovendo a participação cidadã e o monitoramento das ações governamentais.

Esse esforço é importante para construir confiança pública no governo eletrônico, por meio da acessibilidade e abertura dos dados governamentais (Bertot; Jaeger; Grimes, 2010).

O GOV.BR centraliza serviços e informações em uma única plataforma, simplificando a interação dos cidadãos com o governo. A emissão de certidões, as consultas a benefícios sociais e os agendamentos de atendimentos presenciais são exemplos de serviços integrados que economizam tempo e recursos tanto para os cidadãos quanto para a administração pública. Essa abordagem representa um passo significativo na redução da burocracia e na melhoria da eficiência administrativa (Jaeger, 2003).

Apesar dos avanços, o GOV.BR enfrenta desafios substanciais. A inclusão digital e a acessibilidade para diferentes grupos sociais continuam sendo questões prioritárias (Castro; Santos; Pereira, 2017). No Brasil, as disparidades no acesso à internet dificultam a utilização plena do portal por todos os cidadãos (Freire; Rocha; Oliveira, 2020). Além disso, a segurança cibernética e a proteção de dados pessoais são preocupações recorrentes, especialmente sob o contexto da Lei Geral de Proteção de Dados (LGPD) (Brasil, 2018). Esses desafios demandam esforços contínuos para garantir que a plataforma seja eficaz e inclusiva.

A contínua evolução do GOV.BR passa pela integração de tecnologias emergentes, como inteligência artificial e aprendizado de máquina, que podem personalizar e aprimorar os serviços oferecidos (Silva; Santos; Ferreira, 2020). *Chatbots* e análise de grandes volumes de dados (*big data*) têm o potencial de prever demandas e otimizar recursos, enquanto o *blockchain* pode fortalecer a segurança e a transparência das transações eletrônicas governamentais (Tang; Zhang, 2020).

A troca de boas práticas e a colaboração internacional são fundamentais para o aprimoramento do governo eletrônico. Por meio de iniciativas, como a Parceria para Governo Aberto (OGP), países compartilham experiências e desenvolvem estratégias eficazes de governo digital (OECD, 2020). A participação do Brasil nessas iniciativas pode acelerar a adoção de práticas inovadoras e fornecer diretrizes para a melhoria contínua do GOV.BR.

O Portal GOV.BR representa uma iniciativa estratégica para a modernização do governo eletrônico no Brasil. Sua abordagem centralizada reflete um compromisso com a transparência, a eficiência e a acessibilidade. Entretanto é essencial enfrentar desafios como a inclusão digital e a segurança cibernética, além de acompanhar a

evolução tecnológica. A colaboração internacional e o uso de tecnologias emergentes serão fatores determinantes para o sucesso sustentável do GOV.BR, consolidando-o como uma ferramenta inclusiva e inovadora para todos os cidadãos brasileiros.

3.5 Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI)

O *E-Government Survey 2022*, elaborado pelo Departamento de Assuntos Econômicos e Sociais (UN DESA), oferece uma análise do desenvolvimento do governo eletrônico em nível global, destacando tendências, avanços e desafios enfrentados pelos chamados Estados-Membros. Esse documento é um recurso estratégico para formuladores de políticas, acadêmicos e profissionais, fornecendo uma ferramenta de *benchmarking* que facilita a comparação e a aprendizagem entre países.

O *E-Government Survey 2022* utiliza o Índice de Desenvolvimento de *E-Government* (EGDI), uma métrica composta que avalia o desenvolvimento do governo eletrônico em três dimensões: Infraestrutura de Telecomunicações, Capital Humano e Serviços Online. A metodologia do EGDI foi refinada ao longo dos anos para incorporar melhorias baseadas em *feedback* dos Estados-Membros, avanços tecnológicos e políticas emergentes. A edição de 2022 introduziu uma avaliação mais detalhada dos serviços online, dividida em cinco subíndices: quadro institucional, provisão de serviços, provisão de conteúdo, tecnologia e participação eletrônica.

A coleta de dados foi realizada por 227 pesquisadores voluntários da ONU de 130 países, que avaliaram os portais nacionais utilizando um questionário detalhado que cobria uma ampla gama de critérios. Além disso, os Estados-Membros foram solicitados a fornecer informações sobre seus portais governamentais, o que complementou a verificação dos dados. Esse processo permitiu uma avaliação abrangente e precisa dos portais nacionais e locais. A expansão da avaliação para incluir as cidades mais populosas de cada país, utilizando a mesma metodologia dos portais nacionais, permitiu uma análise mais granular e abrangente do desenvolvimento local no âmbito do governo eletrônico.

O EGDI é composto de três índices principais: (1) o Índice de Serviços Online (OSI), que avalia a qualidade e o alcance dos serviços eletrônicos oferecidos pelos governos; (2) o Índice de Infraestrutura de Telecomunicações (TII), que mede a infraestrutura tecnológica disponível para suportar serviços de e-government,

incluindo acesso à internet e telefonia móvel; e (3) o Índice de Capital Humano (HCI), que avalia a capacidade da população de utilizar serviços de e-government, medido por indicadores como níveis de alfabetização e escolaridade. Em 2022, a avaliação do OSI foi expandida para incluir 180 perguntas, divididas em cinco subíndices detalhados, permitindo uma análise mais precisa e melhor compreensão das áreas específicas nas quais os governos precisam melhorar.

O *E-Government Survey* serve como uma ferramenta estratégica para os países avaliarem seu progresso relativo em governo eletrônico. Ao identificar áreas de força e desafio, os Estados-Membros podem aprender uns com os outros, adotando melhores práticas. A transparência proporcionada pelo *ranking* pode estimular uma competição saudável e a colaboração entre nações. Países com desempenho destacado podem servir de modelo, compartilhando suas estratégias e soluções eficazes.

Além disso, o *Survey* fornece uma base de dados extensa que pode ser utilizada para pesquisas acadêmicas e para a formulação de políticas. As informações permitem análises detalhadas e comparativas, essenciais para entender as tendências globais e regionais do e-government. Os formuladores de políticas podem usar esses dados para desenvolver estratégias mais precisas, focando em áreas críticas que necessitam de melhorias, como, por exemplo, direcionar investimentos em infraestrutura de telecomunicações, melhorar a capacitação humana e aprimorar a entrega de serviços online.

O *Survey* destaca ainda a importância de uma abordagem integrada e holística para o desenvolvimento do e-government. Segundo o estudo, políticas eficazes devem considerar a implementação de tecnologias avançadas e a necessidade de desenvolver capacidades humanas e institucionais. A criação de um ambiente regulatório favorável e o fortalecimento das infraestruturas de TI são passos essenciais para garantir que os serviços eletrônicos sejam acessíveis e eficazes.

Em 2020, com a emergência da pandemia de covid-19, a necessidade de adaptação rápida dos governos a novos modos de operação destacou a importância crítica dos serviços digitais governamentais. A transformação digital já não é uma opção, mas uma necessidade para garantir a resiliência e a capacidade de resposta dos serviços públicos. O *E-Government Survey 2022* capta esse momento estratégico, documentando o progresso realizado e os desafios persistentes no caminho para a governança digital.

O estudo demonstra como os governos que já possuíam uma base sólida de *e-government* foram mais capazes de se adaptar rapidamente às necessidades emergentes, garantindo a continuidade dos serviços públicos e a comunicação eficaz com seus cidadãos. A digitalização tornou-se, portanto, um elemento essencial para a gestão de crises futuras, incluindo mudanças climáticas e outros desafios globais. Governos que investiram em tecnologias digitais e em capacitação foram capazes de oferecer serviços essenciais, como saúde e educação, de maneira remota e eficaz. A pandemia serviu como um catalisador para a transformação digital, incentivando inovações que permaneceram relevantes mesmo após a crise. O *Survey* documenta essas inovações e fornece uma base para que outros governos possam aprender e adotar estratégias semelhantes.

O *Survey* reforça também a necessidade de incluir todos os segmentos da sociedade na transformação digital, garantindo que ninguém seja deixado para trás. As lacunas no desenvolvimento de *e-government* são evidenciadas, especialmente entre países de baixa renda e grupos vulneráveis. Ao abordar essas desigualdades, os governos podem promover uma sociedade mais justa e inclusiva. A inclusão digital é um dos maiores desafios para muitos países, especialmente aqueles com recursos limitados. O *Survey* oferece informações sobre como diferentes países estão abordando essas questões, desde a implementação de programas de alfabetização digital até a criação de infraestruturas acessíveis. A promoção da inclusão digital é essencial para garantir que todos os cidadãos possam se beneficiar dos avanços tecnológicos e participar plenamente da sociedade digital.

Enfim, o *E-Government Survey 2022* pode ser considerado uma contribuição importante para o campo do governo eletrônico, oferecendo orientações para a melhoria contínua dos serviços públicos digitais. Ao destacar tanto os sucessos quanto os desafios, o *Survey* incentiva os Estados-Membros a reforçar suas infraestruturas digitais, desenvolver capacidades humanas e adotar abordagens inovadoras para enfrentar crises e promover o desenvolvimento sustentável. Em última análise, o *Survey* mede o progresso e inspira ações concretas para a construção de um futuro digital inclusivo e resiliente.

4 METODOLOGIA

Neste capítulo, é descrita a metodologia de pesquisa, contemplando sua caracterização, a delimitação do universo, o método adotado e os procedimentos metodológicos.

4.1 Caracterização da pesquisa

Adotou-se uma abordagem **mista**, utilizando os métodos **indutivo** e **comparativo**. O método indutivo sustenta-se na revisão sistemática da literatura (Bardin, 2011), enquanto o comparativo busca mapear, inventariar e comparar as fases e etapas de metodologias de construção de ontologias no contexto do governo eletrônico.

A pesquisa classifica-se como **aplicada**, pois visa gerar conhecimento prático sobre a automação na construção de ontologias nesse domínio (Gil, 2009). A **abordagem qualitativa** foi escolhida para identificar e compreender estratégias de organização e representação do conhecimento adotadas no governo eletrônico. Quanto aos objetivos, a **natureza exploratória** da pesquisa é justificada pelo levantamento de elementos metodológicos pouco explorados, a partir da análise documental e da literatura (Pádua, 2012; Marconi; Lakatos, 2016).

4.2 Delimitação do universo da pesquisa

A **delimitação da pesquisa**, conforme Marconi e Lakatos (2016), é fundamental para estabelecer limites que garantam a viabilidade e a aplicabilidade da investigação. O recorte inclui as seguintes especificações: (1) metodologias detalhadas de construção de ontologias vinculadas ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento da UFMG, as quais oferecem um contexto acadêmico robusto para a análise; (2) prova de conceito no domínio de Compras Públicas, selecionado devido à acessibilidade de dados, regulamentações amplamente documentadas e à experiência prática da pesquisadora nesse tema. A seleção das quatro metodologias analisadas nesta pesquisa foi guiada por critérios de relevância científica, disponibilidade de documentação completa, clareza na estrutura

metodológica e viabilidade de aplicação dentro do tempo disponível para o estudo. Essas metodologias foram escolhidas por apresentarem um conjunto coeso e detalhado de etapas, técnicas, ferramentas e documentos organizados em fases bem definidas, o que facilitou a análise crítica de seus componentes. Além disso, a facilidade de acesso a materiais descritivos permitiu uma compreensão aprofundada das propostas, algo essencial para os objetivos da pesquisa. A limitação temporal também foi um fator determinante: ampliar o escopo para incluir mais metodologias exigiria maior tempo de análise e validação, o que poderia comprometer a consistência e a profundidade desejadas. Dessa forma, a escolha por um número restrito, mas bem documentado e representativo, garantiu equilíbrio entre abrangência analítica, rigor metodológico e viabilidade prática.

4.3 Método

Com vistas a alcançar os objetivos propostos nesta pesquisa, foi adotada como método principal a análise de conteúdo, conforme as etapas indicadas por Bardin (2011): (1) pré-análise, que consiste na organização preliminar dos materiais coletados, identificando as categorias iniciais de análise; (2) exploração do material, com identificação de padrões e relações nos dados obtidos; e (3) interpretação, na qual os resultados são sistematizados e categorizados, formando a base para a proposta metodológica. Os materiais selecionados para análise incluem: (1) insumos obtidos na revisão sistemática da literatura, com foco na identificação de metodologias de aprendizagem de ontologias; (2) documentos sobre práticas de organização do conhecimento nos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da Organização da ONU, que adotam os idiomas inglês, português ou espanhol; (3) ontologias desenvolvidas no contexto do Governo Brasileiro, mapeando sua aplicação prática e possíveis lacunas metodológicas; e (4) quatro metodologias específicas de construção de ontologias: OntoForInfoScience, OntONeo, PRONTO e OntoAccount.

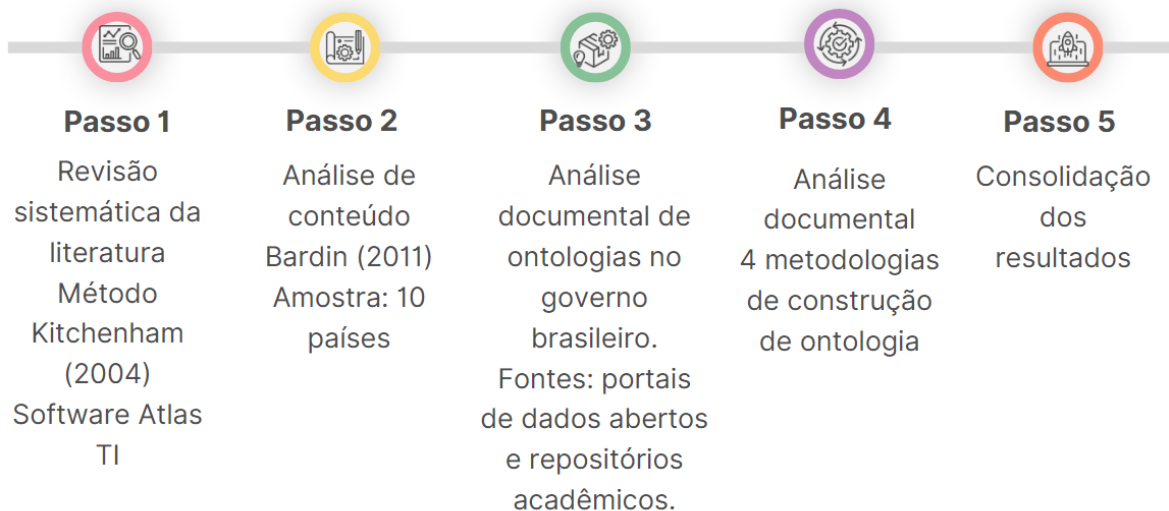
Essas fontes foram selecionadas pela relevância e alinhamento ao objetivo de propor uma metodologia de aprendizagem de ontologias no contexto do governo eletrônico. Uma vez codificados e categorizados, os conteúdos analisados servirão como insumos para a construção de fases, etapas e processos que comporão a metodologia final.

4.4 Visão geral dos procedimentos metodológicos

Nesta seção, são descritos os procedimentos metodológicos em cada etapa da pesquisa, incluindo uma visão geral dos procedimentos adotados na revisão sistemática de literatura (etapa 1), a análise documental acerca das práticas de organização e representação do conhecimento e da informação nos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico (etapa 2), a análise documental de ontologias desenvolvidas no contexto do governo brasileiro (etapa 3) e a análise documental de quatro metodologias de construção de ontologias (etapa 4).

A seguir, na Figura 1, tem-se um resumo dos passos adotados na revisão de literatura. O detalhamento dos procedimentos adotados nesta etapa 1 pode ser identificado na seção [2.2 Procedimentos metodológicos adotados na Revisão de Literatura](#).

Figura 1 – Passos da Etapa 1



Fonte: elaborado pela autora (2024).

4.4.1 Procedimentos metodológicos adotados na Etapa 2

Com vistas a alcançar os objetivos propostos nesta pesquisa, foram adotadas, para análise do conteúdo, as etapas indicadas por Bardin (2011), ou seja, pré-análise, exploração do material e interpretação. Primeiramente, foi realizada a seleção dos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU, com base nos dados fornecidos pelo *site* oficial do Survey. Foi realizada uma consulta no *ranking* geral, buscando-se identificar os dez primeiros países que tivessem como idioma oficial inglês, espanhol ou português. O Brasil⁴, apesar de ocupar a 49.^a posição na análise, foi também incluído para fins de comparação, visando identificar melhorias que possam ser adotadas pelo governo brasileiro.

Tabela 1 – Relação dos países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU

PAÍS	IDIOMA OFICIAL	PORTAL DE DADOS ABERTOS	EGDI 2022	POSIÇÃO NO RANKING DE E-GOVERNMENT 2022
New Zealand	Inglês	https://data.govt.nz	0.9432	4
Australia	Inglês	https://data.gov.au	0.9405	7
United States of America	Inglês	https://www.data.gov	0.9151	10
United Kingdom of Great Britain and Northern Ireland	Inglês	https://data.gov.uk	0.9138	11
Singapore	Inglês	https://data.gov.sg	0.9133	12
Malta	Maltês, Inglês	https://data.gov.mt	0.8943	15
Spain	Espanhol Irlandês,	https://datos.gob.es	0.8842	18
Ireland	Inglês	https://data.gov.ie	0.8539	31
Canada	Inglês, Francês	https://open.canada.ca	0.8511	32
Uruguay	Espanhol	https://datos.gub.uy	0.8388	35

Fonte: United Nations, 2022b.

⁴ O Brasil ocupa a 49.^a posição no *ranking*, com um índice de 0.7910. O portal de dados abertos do Brasil pode ser acessado por meio do seguinte endereço: <https://dados.gov.br>.

Em seguida, foram levantados os termos de busca, com base na Classificação de SOC proposta por Zeng (2008), nos idiomas inglês, português e espanhol, conforme Quadro 9, a seguir.

Quadro 9 – Termos de busca Etapa 2

INGLÊS	ESPAÑHOL	PORTUGUÊS
ontology	ontología	ontologia
semantic network	red semántica	rede semântica
thesauri	tesauros	tesauros
classification schemes	esquemas de clasificación	esquemas de classificação
taxonomies	taxonomías	taxonomias
categorization schemes	esquemas de categorización	esquemas de categorização
subject headings	encabezamientos de materia	cabeçalhos de assunto
gazetteers	gaceteros	gazetteers (ou gazeteiros)
directories	directorios	diretórios
authority files	archivos de autoridad	arquivos de autoridade
synonym rings	anillos de sinónimos	anéis de sinônimos
glossaries	glosarios	glossários
dictionaries	diccionarios	dicionários
pick lists	listas de selección	listas de seleção

Fonte: elaborado pela autora (2024).

Destaca-se que os termos de busca foram levantados nos três idiomas, considerando o idioma oficial dos países selecionados.

Com base na seleção dos países a serem analisados e já dispondo dos *sites* de dados abertos, bem como dos termos de busca selecionados para cada idioma, partiu-se para a coleta dos dados em cada um dos *sites*.

Dando prosseguimento a esta etapa, foi realizada uma leitura preliminar dos documentos para familiarização com o conteúdo e identificação de tópicos relevantes. Esta etapa envolveu a leitura exploratória dos documentos, a identificação do modelo de dados adotado em cada documento identificado e a classificação inicial dos

documentos por temas gerais, como tipo de SOC, domínio do conhecimento, formato do arquivo, entre outras classificações iniciais.

Na sequência, foi realizada a criação de categorias no *software* utilizado para análise qualitativa dos documentos, o Atlas TI. Nesta etapa, os documentos foram codificados utilizando-se as categorias previamente definidas.

Por fim, foi feita uma análise dos conteúdos codificados, buscando-se identificar padrões e variações nos SOC adotados pelos diferentes países. Esta etapa incluiu uma descrição dos SOC identificados em cada país; análise comparativa entre os países, destacando similaridades e diferenças; identificação de práticas inovadoras e eficazes que podem servir como modelo para outros países. A seguir, um resumo das etapas adotadas nos procedimentos metodológicos.

4.4.2 Procedimentos metodológicos adotados na Etapa 3

A metodologia adotada neste estudo é uma análise documental de ontologias no contexto governamental brasileiro, utilizando fontes de dados coletadas em portais de dados abertos e repositórios acadêmicos. Para garantir uma coleta de dados minimamente detalhada, foi utilizado um conjunto abrangente de termos de busca, organizados em diferentes categorias temáticas, visando ampliar as possibilidades de recuperação de ontologias. Realizou-se um levantamento sistemático utilizando-se os seguintes termos de busca, dispostos no Quadro 10, a seguir:

Quadro 10 – Termos de busca Etapa 3

CATEGORIAS TEMÁTICAS	TERMOS DE BUSCA
Ontologias e Governo Brasileiro	"ontologia" AND "governo brasileiro" "ontologias" AND "administração pública brasileira" "desenvolvimento de ontologias" AND "setor público brasileiro" "uso de ontologias" AND "governo do Brasil" "ontologia" AND "gestão pública" AND Brasil "ontologia" AND "serviços públicos" AND Brasil
Governo eletrônico e Ontologias no Brasil	"governo eletrônico" AND ontologia AND Brasil "governo digital" AND ontologias AND Brasil
Organização do Conhecimento e Ontologias no Setor Público Brasileiro	"organização do conhecimento" AND ontologia AND "setor público brasileiro" "organização da informação" AND ontologia AND "administração pública brasileira"
Desenvolvimento e Aplicação de Ontologias	"desenvolvimento de ontologias" AND "administração pública" AND Brasil "aplicação de ontologias" AND "governo brasileiro" "construção de ontologias" AND "setor público" AND Brasil

Fonte: elaborado pela autora (2024).

Esses termos de busca foram aplicados em bases de dados governamentais e acadêmicas, buscando-se diversificar as fontes, de maneira que fosse possível obter maior cobertura na identificação de possíveis ontologias desenvolvidas no contexto do governo brasileiro. As fontes selecionadas são compostas de seis *sites* de repositórios governamentais, uma biblioteca, dois repositórios acadêmicos e anais de um evento específico sobre ontologias, que foram utilizadas neste estudo, conforme apresentado no Quadro 11 a seguir:

Quadro 11 – Fontes de pesquisa Etapa 3

CATEGORIAS	SITES CONSULTADOS	ENDEREÇOS
Repositórios Governamentais	Portal de dados abertos do governo brasileiro	https://dados.gov.br/home
	Portal de serviços do governo brasileiro	https://www.gov.br/pt-br
	Portal da Transparência	https://portaldatransparencia.gov.br/
	Biblioteca Digital do Senado	https://www2.senado.leg.br/bdsf/
	Biblioteca do Instituto Brasileiro de Geografia e Estatística (IBGE)	https://biblioteca.ibge.gov.br/
	Repositório Institucional da ENAP	https://repositorio.enap.gov.br/
Bibliotecas, Repositórios Acadêmicos e Eventos	Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	https://bdtb.ibict.br/vufind/
	SciELO – Scientific Electronic Library Online	https://www.scielo.org/
	Anais Seminário de Pesquisa em Ontologia no Brasil – Ontobras	https://ceur-ws.org/iaoa.html
	Google Scholar	https://scholar.google.com/

Fonte: elaborado pela autora (2024).

Para a seleção dos documentos relevantes, foram definidos critérios de inclusão, a saber: (1) documentos de ontologias publicadas entre 2021 e 2024, visando garantir a atualidade; (2) documentos disponíveis em portais e repositórios oficiais. Documentos fora do escopo, como relatórios não relacionados a ontologias ou à organização do conhecimento foram excluídos da análise.

Os documentos coletados foram submetidos a uma análise qualitativa, que permitiu identificar padrões, lacunas e características das ontologias documentadas. Essa etapa envolveu a leitura detalhada dos documentos e a codificação de informações relevantes, como a descrição de conceitos, relacionamentos e a metodologia empregada para construção das ontologias.

4.4.3 Procedimentos metodológicos adotados na Etapa 4

Os procedimentos metodológicos adotados para a análise documental das quatro metodologias de construção de ontologias seguiram um processo sistemático, com o objetivo de identificar, comparar e sintetizar os principais elementos que caracterizam cada abordagem. Essa análise permitiu compreender as etapas, as técnicas e as ferramentas recomendadas em cada metodologia, bem como suas aplicações em diferentes contextos.

Inicialmente, foi realizada a identificação e a seleção das fontes documentais. Cada metodologia foi analisada com base em documentos primários, que incluíram três teses de Doutorado e um volume de um livro. Esses estudos foram escolhidos considerando-se a relevância e o alinhamento com o objetivo de construir ontologias em diferentes domínios do conhecimento. Em seguida, conduziu-se a leitura exploratória e analítica dos textos, com foco na estrutura geral e nas etapas descritas em cada metodologia. Durante essa fase, buscou-se identificar as similaridades e as diferenças entre os processos, detalhando as práticas recomendadas e os conceitos fundamentais adotados em cada uma delas.

A sistematização dos dados foi realizada por meio de uma matriz comparativa, na qual as fases e as etapas de cada metodologia foram organizadas e descritas. Essa categorização permitiu mapear os pontos de convergência e divergência entre as abordagens. Além disso, foi aplicada uma análise qualitativa para identificar os métodos e as ferramentas sugeridos por cada uma das metodologias, como o uso de *softwares* especializados e técnicas de elicitación com especialistas e procedimentos de formalização e validação.

Por fim, os resultados da análise foram organizados em um texto descritivo e integrado, apresentando as características específicas de cada metodologia, suas contribuições para o desenvolvimento de ontologias e sua aplicabilidade em diferentes contextos, com o intuito de fornecer uma visão abrangente e detalhada das práticas metodológicas, destacando a importância da adaptação das etapas às necessidades do domínio e do objetivo final da ontologia.

5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Este capítulo apresenta os resultados da pesquisa com foco em três perspectivas complementares que embasam o desenvolvimento da metodologia proposta. A análise documental inicial aborda os Sistemas de Organização do Conhecimento (SOC) adotados nos países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI) da ONU, permitindo identificar boas práticas e lacunas no uso dessas ferramentas em contextos internacionais. Em seguida, examinam-se as ontologias desenvolvidas no contexto do governo brasileiro, considerando repositórios acadêmicos e plataformas governamentais, o que fornece uma visão detalhada das iniciativas nacionais nesse campo. Por fim, realiza-se uma análise comparativa de quatro metodologias de construção de ontologias, mapeando fases, etapas, ferramentas e práticas fundamentais. Esses resultados integram-se para formar uma base sólida de referência, culminando na proposta detalhada da *Onto Merge Methodology (OMM)*, apresentada no [Capítulo 6](#).

5.1 Documentos recuperados em países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da Organização das Nações Unidas (EGDI)

Nesta seção, buscou-se examinar as práticas de organização e representação do conhecimento (SOC) adotadas pelos dez países que lideram o Índice Geral de Desenvolvimento de Governo Eletrônico da ONU, por meio de uma análise documental fundamentada na avaliação de como esses países estruturam e disponibilizam seus dados públicos por meio de suas plataformas de dados abertos. A escolha desses países teve como propósito identificar e compreender as melhores práticas internacionais em termos de organização e representação do conhecimento e da informação, proporcionando um referencial comparativo que poderá ser utilizado para aprimorar as estratégias de governo eletrônico de outros países, em especial o Brasil.

Os resultados a seguir oferecem uma visão parcial dos Sistemas de Organização e Representação do Conhecimento e Informação (SOC) adotados nos dez países selecionados, e também no Brasil, evidenciando lacunas e oportunidades de melhoria na gestão de dados abertos. Inicialmente, é apresentado o resultado consolidado da pesquisa realizada em cada um dos portais de dados abertos dos

países analisados, seguido, nas subseções, pela exposição detalhada dos dados por país.

Cabe destacar que, tanto na Tabela 2, a seguir, quanto nas subseções, são mencionados os totais de ocorrências dos termos relativos a cada tipo de SOC. A análise posterior desses documentos permitiu verificar se, de fato, esses termos correspondiam a um tipo específico de SOC ou se apenas faziam referência ao termo nos mecanismos de busca utilizados. Além disso, é importante também esclarecer que, em razão do volume de documentos recuperados, optou-se por realizar um recorte dos 10 primeiros documentos recuperados, nos casos em que houve número superior de documentos recuperados, de maneira que a análise se tornasse viável.

Tabela 2 – Documentos recuperados nos *sites* de dados abertos

SOC	NOVA ZELÂNDIA	AUSTRÁLIA	EUA	REINO UNIDO	SINGAPURA	MALTA	ESPANHA	IRLANDA	CANADÁ	URUGUAI	TOTAL
Arquivos de Autoridade	2077	0	2	0	0	0	0	0	30	1	2110
Esquemas de Categorização	0	0	3	0	0	0	0	0	1	0	4
Esquemas de Classificação	10	86	810	36	0	0	0	6	284	7	1239
Dicionários	4476	2135	3976	35	0	0	76	2	2804	20	13524
Diretórios	19	596	65833	392	4	0	509	19	7891	50	75313
<i>Gazetteers</i>	36	36	194	4	0	0	1	10	518	47	846
Glossários	6	127	318	4	0	0	18	6	20095	40	20614
Ontologia	11	17	44	20	0	0	17	1	130	20	260
Listas de Seleção	0	0	0	0	0	0	0	0	46	12	58
Rede Semântica	0	0	2	0	0	0	0	0	0	4	6
Cabeçalhos de Assunto	9	0	20	3	0	0	10	0	184	30	256
Anéis de Sinônimos	0	0	0	0	0	0	0	0	0	3	3
Taxonomias	177	154	727	12	0	0	3	1	2144	41	3259
Tesauros	1	3	29912	2	0	0	15	0	125	16	30074
Total	6822	3154	101841	508	4	0	649	45	34252	291	147566

Fonte: elaborado pela autora (2024).

5.1.1 Documentos recuperados no *site* de dados abertos do governo da Nova Zelândia

O *site* data.govt.nz oferece acesso a uma variedade de conjuntos de dados públicos da Nova Zelândia, organizados em categorias como terra, meio ambiente, saúde, transporte e educação. A página inicial destaca recursos para encontrar e usar dados, incluindo um guia do catálogo e um *kit* de ferramentas de dados. O *site* também apresenta artigos sobre inteligência artificial e a estratégia de dados do governo. Além disso, há seções para solicitar novos conjuntos de dados que porventura não tenham sido identificados e verificar os *datasets* mais recentemente atualizados. Na *home*, é possível realizar buscas tanto no conteúdo do *site* como um todo ou em *datasets* disponíveis. É informado que a página possui cerca de 34 mil *datasets* de 188 organizações. Esses *datasets* podem ser consultados por grupos temáticos.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 3, a seguir.

Tabela 3 – Menções aos SOC – *site* de dados abertos da Nova Zelândia

TERMOS	RESULTADOS
authority files	2077
categorization schemes	0
classification scheme	10
dictionary	4476
directory	19
gazetteer	36
glossary	6
ontology	11
pick list	0
semantic network	0
subject headings	9
synonym rings	0
taxonomy	177
thesaurus	1
Total	6822

Fonte: elaborada pela autora (2024).

Os Arquivos de Autoridade (*authority files*) têm a função de padronizar a forma de referenciar pessoas, lugares, coisas ou conceitos, promovendo consistência e controle de autoridade. Entre os documentos analisados, apenas três deles (*Geographic Areas File 2017; Meshblock 2008 Generalised Version; NZ Suburbs and Localities*) podem ser considerados exemplos de arquivos de autoridade, pois padronizam a referência a áreas geográficas, garantindo consistência e permitindo análises comparativas ao longo do tempo. Em contraste, os outros documentos recuperados, como gráficos náuticos, mapas geológicos e registros empresariais, embora úteis em seus contextos específicos, não representam exemplos desse tipo de SOC.

Em relação aos Esquemas de Classificação recuperados, avalia-se que vários podem ser classificados como este tipo de SOC, especificamente os documentos relacionados à classificação costeira da Nova Zelândia, como "*NZ Coastal Classification – Coastal Land Form Type*", "*NZ Coastal Classification – Beach Exposure*", "*NZ Coastal Classification – Foreshore Sediment*" e "*NZ Coastal Classification – Hinterland Characteristics*", pois organizam e definem diversos atributos costeiros, como tipos de formas de relevo, exposição das praias à energia das ondas, tipos de sedimentos, entre outros atributos. Por outro lado, os documentos "*TMS Daily Traffic Counts API*" e "*TMS Daily Traffic Counts CSV*" fornecem dados estruturados sobre a contagem diária de tráfego, mas não se enquadram como esquemas de classificação, pois não definem categorias ou termos específicos com o objetivo de padronizar a descrição de objetos ou conceitos, como é típico em um esquema de classificação.

Sobre os SOC do tipo Dicionários (*dictionary*), conclui-se que nenhum dos documentos recuperados corresponde a um “dicionário” conforme definição proposta por Zeng (2014). Embora vários documentos utilizem o termo “dicionário de dados”, os documentos representam recursos técnicos destinados a descrever e padronizar elementos de dados específicos em contextos como estudos clínicos, regulamentos de pesca e dados geológicos. Esses dicionários de dados fornecem definições e explicações sobre campos específicos de dados, mas não organizam palavras alfabeticamente nem fornecem informações detalhadas sobre formas, pronúncias, etimologias e usos idiomáticos das palavras.

A análise dos SOC recuperados do tipo Diretórios (*directory*) revelou que muitos deles são de fato diretórios completos e informativos, cada um servindo a

diferentes necessidades e audiências. O "*Directory of New Zealand Libraries*" e o "*Family Services Directory*" são exemplos de diretórios, contendo listas de organizações e seus contatos detalhados, facilitando a busca por serviços específicos como bibliotecas ou serviços de apoio familiar. Já os diretórios como o "*National Environmental Data Centre Online Directory*" e o "*Iwi Boundaries*" demonstram a diversidade de aplicações dessas listas organizadas. Enquanto o primeiro fornece acesso a conjuntos de dados ambientais, o segundo detalha as áreas de interesse das comunidades Māori, destacando a importância cultural e administrativa dessas informações.

Ao analisar os chamados Dicionários Geográficos (*gazetteer*), conclui-se que dois deles se destacam como exemplos desse tipo de SOC. O "*NZ Localities – NZGB Compliant Names – Dataset*" e o "*Suburb Locality – NZ Gazetteer – Dataset*" cumprem a definição de um dicionário de nomes de lugares georreferenciados, fornecendo nomes oficiais de localidades conforme o *New Zealand Geographic Board* (NZGB) e oferecendo coordenadas para a localização geográfica desses lugares. Esses *datasets* são essenciais para garantir a padronização e a precisão dos nomes geográficos na Nova Zelândia. Os outros documentos analisados, embora contenham elementos de localização e informações relevantes, não se encaixam completamente na definição de *gazetteer*. Por exemplo, o "*NZ Airsheds Gazetted – Dataset*" mapeia áreas específicas notificadas na *New Zealand Gazette*, mas seu foco é ambiental, e o "*Independent Marriage & Civil Union Celebrants – Number of Ceremonies Performed*" fornece dados geográficos vinculados a eventos e pessoas, não a nomes de lugares.

Após a análise dos documentos recuperados do tipo Glossário (*glossary*), constatou-se que, entre os seis arquivos examinados, apenas um corresponde a um glossário. O "*GeoNet Aotearoa New Zealand Glossary of Data-related terms*" apresenta uma lista de termos e suas respectivas definições, oferecendo uma referência clara e concisa de terminologias. Os outros documentos analisados fornecem descrições de dados técnicos e metodologias, mas não se configuram como glossários. Eles incluem definições de termos técnicos dentro de contextos específicos, porém seu objetivo principal é a apresentação de dados e referências bibliográficas e não a definição geral de termos.

Seguindo para análise das Ontologias (*ontology*), identificou-se um uso abrangente de ontologias para descrever e organizar informações complexas em diferentes domínios. Os documentos analisados empregam ontologias para mapear

relações entre genes, processos biológicos, metabólitos e achados radiológicos. Por exemplo, o documento “*Supplementary Table 6*” detalha a análise de uma ontologia gênica, identificando que esses eGenes estão envolvidos em processos relacionados ao sistema imunológico. De forma semelhante, o documento “*Supplementary Tables for Metabolite Data*” utiliza uma integração sistemática de dados, incluindo ontologia gênica (KEGG) e classificação de proteínas (*The Human Protein Atlas*). Além disso, os conjuntos de dados “CANDID-II” e “CANDID-III” exemplificam o uso de ontologias para anotar achados radiológicos em grandes conjuntos de dados de radiografias de tórax, provenientes do Hospital de Dunedin. Esses conjuntos de dados facilitam a integração e a interpretação dos dados e promovem a padronização e a interoperabilidade entre diferentes sistemas e estudos.

Sobre os documentos recuperados como Cabeçalho de Assunto (*subject headings*), conclui-se que nenhum deles corresponde a esse tipo de SOC. Todos os documentos analisados são conjuntos de dados específicos que abordam uma variedade de tópicos, como riscos naturais relacionados à subsidência de minas, monitoramento de espécies vegetais vulneráveis e detalhes técnicos sobre o registro de veículos a motor. Esses conjuntos de dados são voltados para finalidades de pesquisa científica, regulamentação ambiental e administração pública, fornecendo informações detalhadas e técnicas sobre cada tópico específico, mas nenhum dos documentos examinados se encaixa na definição de cabeçalho de assunto, visto que não possuem a estrutura hierárquica nem as regras de coordenação de termos típicas desse tipo de SOC, sendo mais apropriados para consultas de dados específicos em vez de servir como uma ferramenta de indexação bibliográfica.

Com relação aos documentos recuperados do tipo Taxonomia (*taxonomy*), foi possível identificar que seis deles correspondem ao conceito de taxonomia como um sistema de organização hierárquica de termos preferidos. Esses documentos são: “*GEA-NZ Reference Taxonomies*”, “*GEA-NZ Reference Taxonomies Modelling Files*”, “*World List of Isopod Species and Their Authorities.xlsx*”, “*World List of Isopod Species and Their Authorities_Updated_2021.xlsx*”, “*World List of Isopod Species and Their Authorities_Updated_2023.xlsx*” e “*WoRMS_Amphipoda_RawData.xlsx*”. Esses documentos fornecem estruturas taxonômicas detalhadas que facilitam a classificação e a organização de informações e espécies, atendendo ao propósito fundamental de uma taxonomia. Os demais documentos analisados, embora contenham dados relevantes para estudos científicos e taxonômicos, não constituem

taxonomias completas. Documentos como "*Insect-Plant Associations*", "*Notogaster Measurements*", "*Phaeohelotium Phylogeny and Notes*", e "*Threatened Species Biomass*" oferecem conjuntos de dados observacionais, medições e informações filogenéticas que são úteis para análises e pesquisas, mas não seguem a estrutura hierárquica característica de uma taxonomia.

Por fim, sobre a recuperação do único documento classificado como uma Tesouro (*thesaurus*), observou-se que, embora o documento "*National Paleontological Collection (NPC)*" represente uma coleção extensa e bem-organizada, abrangendo fósseis macro e micro de regiões como Nova Zelândia, Antártica e Nova Caledônia, ele não se configura explicitamente como um tesouro conforme os padrões NISO e ISO. Esses padrões exigem uma estruturação rigorosa de termos, com relações hierárquicas, de equivalência e associativas bem definidas, algo que não foi evidenciado na descrição do referido documento. Portanto, mesmo a NPC sendo uma coleção paleontológica fundamental para pesquisa e referência, ela não pode ser classificada diretamente como um tesouro formal. Para que a NPC se alinhe aos padrões de tesouro, seria necessário implementar um sistema de organização que incluía termos preferenciais, não preferenciais e suas relações, otimizando a indexação e recuperação de informações.

5.1.2 Documentos recuperados no *site* de dados abertos do governo da Austrália

O site data.gov.au é uma plataforma dedicada ao acesso a dados públicos da Austrália, oferecendo um catálogo extenso de conjuntos de dados categorizados por temas como saúde, meio ambiente, economia, educação e transporte. Destaca conjuntos de dados populares e recentemente atualizados, além de permitir solicitações de novos dados. O *site* fornece tutoriais e guias do usuário para ajudar na utilização dos dados, com uma seção de *blog* contendo artigos sobre práticas e iniciativas de dados abertos. Possui um motor de busca para facilitar a localização de dados específicos, bem como *links* para comunidades e eventos relacionados a dados abertos na Austrália.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 4, a seguir.

Tabela 4 – Menções aos SOC – site de dados abertos da Austrália

TERMOS	RESULTADOS
authority files	0
categorization	
schemes	0
classification scheme	86
dictionary	2135
directory	596
gazetteer	36
glossary	127
ontology	17
pick list	0
semantic network	0
subject headings	0
synonym rings	0
taxonomy	154
thesaurus	3
Total	3154

Fonte: elaborada pela autora (2024).

Dos documentos recuperados pelo termo Esquemas de Classificação (*classification scheme*), vários podem ser claramente identificados como esquemas de classificação. Documentos como "*Estuary Geomorphological Classification*", "*A Nationally Consistent Geomorphic Classification of the Australian Coastal Zone*" e "*National Coastal Geomorphology Information Framework: Discovery and Distribution*" podem ser considerados exemplos de esquemas de classificação, pois organizam e definem de maneira sistemática e padronizada os atributos geomorfológicos e ambientais de diversas áreas costeiras e marinhas na Austrália, além de fornecer uma estrutura clara e consistente para categorizar diferentes formas de relevo costeiro, tipos de sedimentos e características ambientais. Além desses, o "*Structural Vegetation (1995)*" e o "*Statewide Land Use 1996 to 2005*" também utilizam esquemas de classificação para organizar informações sobre a vegetação e o uso da terra, respectivamente. Esses documentos aplicam classificações padronizadas, como o esquema da *Land Conservation Council* e o *Australian Land Use Mapping*, para categorizar dados complexos de maneira acessível e consistente. Portanto a maioria dos documentos analisados pode ser considerada como esquemas de classificação, pois cumprem a função de organizar informações detalhadas e específicas de forma

estruturada, promovendo a clareza e a precisão na comunicação e a análise de dados ambientais e geoespaciais.

A análise dos documentos recuperados que utilizam o termo Dicionário (*dictionary*) revela que todos os documentos são dicionários de dados técnicos ou inventários específicos, destinados a padronizar e explicar elementos de dados em contextos técnicos ou científicos. Exemplos incluem os documentos relativos aos subsídios para primeiros compradores de imóveis no estado de Queensland, que fornecem dicionários de dados para interpretar corretamente os conjuntos de dados estatísticos, e o "*Botanical Dictionary Queensland 2022*", que serve para checagem ortográfica de termos botânicos. Embora os documentos possam ser úteis em seus respectivos contextos, não contemplam informações linguísticas típicas de um dicionário tradicional, como pronúncias, etimologias ou usos idiomáticos.

Em relação aos documentos que mencionava o termo Diretórios (*directory*), observou-se que todos os documentos recuperados se enquadram na definição de diretórios, oferecendo informações organizadas sobre uma variedade de serviços e recursos. Esses diretórios cobrem desde negócios que aceitam o chamado *Companion Card*, capacidades científicas e organizações comunitárias, até a estrutura detalhada do governo australiano. Servem ao público geral e são cruciais para profissionais que necessitam de informações específicas para suas atividades diárias. A presença de diretórios detalhados e atualizados ajuda a conectar indivíduos com serviços relevantes, apoia a tomada de decisões informada e melhora a comunicação dentro das organizações e entre elas.

Sobre os documentos recuperados que mencionam os Dicionários Geográficos (*gazetteer*), observou-se que vários se enquadravam na definição de *gazetteer*, fornecendo dados georreferenciados e autorizados sobre nomes de lugares. Exemplos incluem o "*Gazetteer of Australia 2010*" e o "*Gazetteer of Australia 2012 Release*", que oferecem informações detalhadas sobre a localização e a grafia de mais de 370.000 nomes geográficos. Esses documentos são complementados por serviços Web, como o MapServer, WMS e WFS, que permitem acesso programático a esses dados, facilitando sua integração em sistemas de informações geográficas e outras aplicações. Além disso, há também *gazetteers* estaduais, como o "*Place names gazetteer – Queensland*" e o "*Place names (State Gazetteer)*" de South Australia, que reforçam a padronização e a precisão dos nomes geográficos em níveis regionais. Esses *datasets* fornecem listas de nomes de lugares aprovados com coordenadas

geográficas, cumprindo a função de um *gazetteer*. Os documentos analisados atendem aos critérios tradicionais de um *gazetteer* e demonstram a evolução desse conceito por meio da integração de tecnologias Web, ampliando seu alcance e aplicabilidade.

Após a análise dos documentos recuperados que mencionavam o termo Glossário (*glossary*), constatou-se que apenas um deles, o "*Glossary – NSW Department of Education*", configura-se como esse tipo de SOC, visto que apresenta uma lista consolidada de termos com definições específicas, cumprindo assim a função de esclarecer e padronizar a terminologia dentro do contexto educacional. Os demais documentos, embora mencionem glossários ou façam referência a páginas de glossários nos seus respectivos *sites*, são predominantemente conjunto de dados que servem mais como referências a dados e informações contextuais, em que a presença de glossários é secundária e dependente de *links* externos para acesso a essas definições.

Sobre os documentos que mencionavam Ontologia (*ontology*), observou-se o uso desse tipo de SOC em diferentes contextos pelo governo australiano. Por exemplo, a "*GeoSPARQL Extensions Ontology*" e a "*Geological Timescale Ontology*" são utilizadas para representar dados geoespaciais e geológicos com alta precisão, facilitando a interoperabilidade e a utilização eficiente desses dados por diferentes agências e pesquisadores. O governo australiano também adapta ontologias para atender a necessidades específicas, como evidenciado pela "*Modifications to W3C OWL – Time Ontology*", que suporta sistemas de referência temporal não gregorianos, e pela "*Ontology for Observations and Sampling Features*", que padroniza dados observacionais científicos. Além disso, ontologias como a "*Agreements Ontology*" e a "*data.gov.au Dataset Ontology*" mostram como o governo da Austrália está promovendo a governança de dados e a transparência. Essas ontologias modelam acordos legais e descrevem conjuntos de dados, respectivamente, facilitando a reutilização de informações como *Linked Data* e melhorando a acessibilidade e a gestão de dados. Em suma, o uso de ontologias pelo governo australiano destaca seu papel central na padronização, integração e gestão eficaz de dados, impulsionando avanços em diversas áreas científicas e administrativas.

Em relação aos documentos recuperados que faziam menção às Taxonomias (*taxonomy*), identificou-se que apenas o documento "*Taxonomy and phylogenetics of marine pill bugs*" alinha-se completamente ao conceito de uma taxonomia. Esse

estudo organiza grupos sub-familiares de crustáceos marinhos, proporcionando uma estrutura detalhada de termos preferidos, conforme a definição de taxonomia. Os demais documentos, incluindo as diversas coleções do *South Australian Museum* e os estudos sobre biodiversidade de esponjas e morfologia de plantas, podem ser considerados catálogos descritivos de coleções biológicas. Embora esses documentos ofereçam classificações básicas e sejam úteis para a pesquisa científica e taxonômica, eles não fornecem a estrutura hierárquica detalhada necessária para serem considerados taxonomias.

Quanto ao último tipo de SOC identificado, o Tesouro (*thesaurus*), observou-se que dos quatro documentos recuperados, o "*Australian Sport Thesaurus*" é o único que se alinha à definição tradicional de um tesouro. Esse documento contém uma vasta coleção de termos e definições relacionadas ao esporte, organizados de forma a reduzir ambiguidades terminológicas e promover a consistência no uso dos termos. Ele estabelece relações hierárquicas e associativas, incluindo termos preferenciais e não preferenciais, tornando-se uma ferramenta útil para a interoperabilidade de dados e a padronização terminológica no setor esportivo. Os outros três documentos – "*Austlang dataset*", "*Acid Sulfate Soil Risk Data*" e o registro de resultados de busca "*Datasets search*" – não correspondem à definição formal de tesouro. Embora sejam recursos importantes em seus respectivos contextos, eles não organizam termos com base em relações hierárquicas ou associativas padronizadas.

5.1.3 Documentos recuperados no *site* de dados abertos do governo dos Estados Unidos da América

A página Data.gov é dedicada ao acesso a dados públicos dos Estados Unidos, oferecendo mais de 300 mil conjuntos de dados de várias agências governamentais. A plataforma facilita o uso de dados para conduzir pesquisas, desenvolver aplicativos e criar visualizações de dados. Ela destaca conjuntos de dados populares e recentemente adicionados, organizados por categorias como geoespacial, saúde e transporte. O Data.gov implementa a Lei de Dados Abertos do Governo (*OPEN Government Data Act*), que exige que as agências federais publiquem suas informações online em formatos legíveis por máquina e com metadados incluídos no catálogo do Data.gov. A página também oferece um repositório online de ferramentas, melhores práticas e padrões de esquemas para facilitar a adoção de práticas de dados

abertos em todo o governo federal. O catálogo de dados do Data.gov organiza informações em várias categorias, como clima, saúde, transporte e mais. Ele oferece filtros por tipo de *dataset*, *tags*, formatos e organizações, facilitando a busca por informações específicas. Os dados disponíveis incluem formatos como CSV, JSON, XML, entre outros, fornecendo flexibilidade para diferentes usos.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 5, a seguir.

Tabela 5 – Menções aos SOC – *site* de dados abertos dos Estados Unidos

TERMOS	RESULTADOS
authority files	2
categorization	
schemes	3
classification scheme	810
dictionary	3976
directory	65833
gazetteer	194
glossary	318
ontology	44
pick list	0
semantic network	2
subject headings	20
synonym rings	0
taxonomy	727
thesaurus	29912
Total	101841

Fonte: elaborada pela autora (2024).

Iniciando a análise dos documentos recuperados que faziam menção aos Arquivos de Autoridade (*authority files*), identificou-se que o documento "*Organization Descriptions from the National Archives Catalog*" enquadra-se na definição de "*authority files*". Esse arquivo padroniza a evolução dos nomes e das histórias administrativas de organizações federais e não federais, proporcionando pontos de acesso para a indexação de descrições arquivísticas. A padronização promovida por esse documento é essencial para garantir a consistência na nomenclatura e facilitar a busca e o acesso a informações arquivísticas, demonstrando seu papel na organização e no controle de autoridade de dados governamentais. Em contraste, o

documento "*Bathymetric surveys of the Neosho River, Spring River, and Elk River*" fornece dados técnicos detalhados para a modelagem hidráulica, mas não estabelece padrões de referência universal para nomes ou conceitos. Portanto avalia-se que ele não se qualifica como um Arquivo de autoridade.

A análise dos documentos recuperados que faziam menção ao termo Esquemas de Categorização (*categorization schemes*) revela que todos eles utilizam esquemas de categorização para organizar e interpretar dados em diferentes contextos científicos. O primeiro documento, "*Acoustic backscatter – Data and Python Code*", aplica um modelo probabilístico para classificar sedimentos de leitos de rios, categorizando-os em cinco classes distintas. O segundo documento, "*Evaluation of Existing QSAR Models and Structural Alerts and Development of New Ensemble Models for Genotoxicity Using a Newly Compiled Experimental Dataset*", categoriza substâncias com base em sua genotoxicidade. O terceiro documento, "*Limacina helicina shell dissolution due to ocean acidification in the California Current Ecosystem*", emprega um esquema de categorização para avaliar a dissolução de conchas sob diferentes condições de acidificação do oceano. Esses esquemas de categorização demonstram ser ferramentas essenciais para a sistematização e a análise de grandes volumes de dados complexos. A aplicação desses esquemas facilita a interpretação dos dados e a tomada de decisões, evidenciando a importância de métodos estruturados para a análise científica.

Sobre os documentos que faziam menção aos Esquemas de Classificação (*classification scheme*), observou-se que vários deles podem ser considerados esquemas de classificação, enquanto outros utilizam classificações para organizar dados complexos. Documentos como "*CDC WONDER: Mortality – Underlying Cause of Death*" e "*NCHS – Potentially Excess Deaths from the Five Leading Causes of Death*" utilizam a CID-9 e CID-10 para categorizar causas de morte, fornecendo uma estrutura padronizada essencial para a análise epidemiológica e a formulação de políticas de saúde pública. De forma semelhante, os documentos relacionados à cobertura do solo, como "*MODIS_Terra+Aqua Land Cover Type Yearly*" e "*USGS Land Cover (NLCD)*", aplicam esquemas de classificação internacionais para monitoramento ambiental e estudos climáticos. Por outro lado, documentos como "*Extreme Significant Wave Heights for US Coastal Waters*" e "*Multi-Temporal Remote Sensing Image Classification – A Multi-View Approach*" utilizam classificações para estruturar dados específicos, mas não se enquadram como esquemas de

classificação. Eles fornecem metodologias e estruturas de dados que auxiliam na análise e compreensão de fenômenos específicos, como alturas de ondas extremas e classificações de imagens de sensoriamento remoto. A maioria dos documentos analisados utiliza esquemas de classificação para organizar e padronizar informações, desempenhando um importante papel na gestão e análise de dados em diversas áreas, incluindo saúde pública, monitoramento ambiental e gestão de recursos naturais.

A análise dos documentos recuperados classificados como Dicionários (*dictionary*) revela que todos os documentos representam dicionários de dados técnicos destinados a padronizar e explicar elementos de dados em contextos específicos, como saúde pública, educação e segurança. Exemplos incluem o "*College Scorecard*", que fornece dados sobre conclusão de cursos e dívidas estudantis, e o "*Chemicals in Cosmetics*", que detalha ingredientes perigosos em produtos cosméticos. Esses documentos são essenciais para a correta interpretação e utilização dos dados apresentados, mas não se enquadram na definição de um dicionário, que organiza palavras alfabeticamente e fornece informações detalhadas sobre formas, pronúncias, etimologias e usos idiomáticos.

Em relação aos documentos recuperados do tipo Diretórios (*directories*), observou-se que uma parte dos documentos recuperados funciona como diretórios detalhados e organizados, oferecendo informações essenciais sobre locais, serviços e eventos. Documentos como "*Electric Power Substations*", "*Farmers Markets Directory and Geographic Data*", "*NCDC Storm Events Database*", "*NGS Survey Control Map*" e "*Public School Characteristics – Current*" são claros exemplos de diretórios, pois fornecem listas estruturadas de pontos de interesse, contatos e dados específicos que facilitam a navegação e o acesso a informações cruciais para diversos públicos, desde consumidores e agricultores até engenheiros e administradores de educação. Por outro lado, alguns documentos analisados, como aqueles relacionados a dados climáticos e meteorológicos, funcionam mais como bases de dados científicas, oferecendo informações valiosas para pesquisadores e profissionais, mas não se enquadram na definição tradicional de diretórios.

A análise dos documentos recuperados que faziam menção aos Dicionários Geográficos (*gazetteer*) demonstrou que alguns documentos se encaixam na definição de *gazetteer*, enquanto outros não. Documentos como o "*Gazetteer of Planetary Nomenclature*" e suas variantes específicas (Europa, Ganymede, Venus)

podem ser considerados exemplos de *gazetteers*, fornecendo informações detalhadas e georreferenciadas sobre características topográficas em corpos celestes. O "GNIS" (*Geographic Names Information System*) também é um exemplo de *gazetteer*, oferecendo um banco de dados abrangente de nomes geográficos nos Estados Unidos, essencial para a padronização e referência geográfica precisa. Por outro lado, documentos como o "*Gateway to Astronaut Photography of Earth*" e os focados em propriedade intelectual, como o "*IP Marketplace Platform API*" e o "*Trademark Official Gazettes*", não se enquadram na definição tradicional de *gazetteer*. Eles fornecem informações visuais e legais, respectivamente, mas não estruturam nomes e localizações geográficas de maneira sistemática.

A respeito dos documentos recuperados que faziam menção ao Glossário (*glossary*), conclui-se que nenhum deles se configura como um glossário, ou seja, uma lista de termos acompanhados de suas definições diretamente no próprio documento. Embora vários mencionem a existência de glossários ou forneçam *links* para glossários externos, os documentos em si são predominantemente conjuntos de dados que oferecem informações detalhadas sobre diversos temas, como produção de energia, disponibilidade de alimentos, estatísticas de saúde e dados financeiros.

Em relação aos documentos relacionados ao termo Ontologia (*ontology*), observou-se que documentos como o "*MedGen*" e o "*Immunology Database and Analysis Portal (ImmPort)*" mostram como as ontologias são essenciais para a centralização e a padronização de informações genéticas e clínicas, facilitando a interoperabilidade e a pesquisa colaborativa. Essas ontologias permitem que dados de múltiplas fontes sejam agregados e utilizados de maneira eficiente, promovendo avanços significativos na medicina e na biologia. No setor ambiental, o "*NaKnowBase Interoperability Tools*" e o estudo "*Prior knowledge-based approach for associating contaminants with biological effects*" ilustram como as ontologias podem ser usadas para integrar e analisar dados ambientais e de saúde pública. Da mesma forma, a aplicação de ontologias na manufatura inteligente, como demonstrado no documento "*Reference Knowledge Graphs of STEP and QIF Data for a Three-Part Box Assembly*", destaca a importância dessas ferramentas para a rastreabilidade e a qualidade dos produtos. Por fim, o uso de ontologias em bancos de dados agrícolas, como o "*CottonGen: Cotton Database Resources*", evidencia seu papel na facilitação da pesquisa e inovação em setores específicos. Em resumo, os documentos analisados demonstram que o governo dos EUA utiliza ontologias de maneira eficaz para

promover a padronização, a interoperabilidade e o avanço científico em diversas áreas.

Outro tipo de SOC pesquisado foi o chamado Redes Semânticas (*semantic network*). Analisando os dois documentos recuperados, observou-se que nenhum dos dois representa uma rede semântica. O primeiro documento "*Unified Medical Language System (UMLS) – Catalog*" propõe um sistema unificado de linguagem médica, padronizando terminologias chave, padrões de classificação e codificação e recursos para promover a criação de sistemas de informação biomédica mais eficazes e interoperáveis; e serviços, incluindo registros eletrônicos de saúde. Já o segundo documento, "*A knowledge-based system approach for sensor fault modeling, detection and mitigation*", é apenas um artigo que menciona redes semânticas em seu resumo.

Sobre os documentos que faziam menção ao Cabeçalho de Assunto (*subject headings*), a análise revelou que, entre eles, apenas o "*Medical Subject Headings (MeSH)*" corresponde a um Cabeçalho de Assunto. O MeSH é uma terminologia hierarquicamente organizada utilizada para a indexação e catalogação de informações biomédicas, sendo essencial para a classificação precisa e a recuperação eficiente de dados em bases, contemplando a estrutura necessária para fornecer uma lista controlada de termos, facilitando a organização e a pesquisa de informações. Os demais documentos analisados são conjuntos de dados técnicos e científicos específicos, abrangendo áreas como geologia, geoquímica e gestão de recursos naturais. Embora sejam documentos úteis para finalidades de pesquisa e regulamentação, esses documentos não possuem a estrutura hierárquica e as regras de coordenação de termos características dos cabeçalhos de assunto.

Em análise dos documentos recuperados quanto ao termo Taxonomia (*taxonomy*), apenas quatro deles podem ser classificados como taxonomias. Os documentos "*Medicare Provider and Supplier Taxonomy Crosswalk*", "*Restructured BETOS Classification System*", "*Taxonomy Database*" e "*Plant List of Accepted Nomenclature, Taxonomy, and Symbols (PLANTS) Database*" fornecem estruturas hierárquicas detalhadas para a classificação de informações em seus respectivos campos. Esses documentos organizam dados de maneira sistemática e hierárquica, atendendo ao conceito de taxonomia. Os demais documentos, como o "*2019 California Water Quality Status Report*", "*Official Soil Series Descriptions (OSDs)*", "*National Transportation Safety Board Data*", "*Data from: A century of wild bee sampling*", "*Bee-Gap: Ecology, Life-History, and Distribution of Bee Species in the*

United States 2017" e "*Tax Exempt Organizations (Extracted from the Internal Revenue Service)*", embora forneçam informações detalhadas em suas respectivas áreas, não seguem uma estrutura taxonômica hierárquica. Esses documentos servem mais como catálogos descritivos ou conjuntos de dados analíticos, fornecendo informações úteis para a pesquisa e análise.

Por fim, em relação aos documentos recuperados que mencionavam o termo Tesouro (*thesaurus*), nenhum dos documentos se configura como um tesouro conforme os padrões que envolvem a organização terminológica com relações hierárquicas, de equivalência e associativas bem definidas. A maioria dos documentos analisados são relatórios técnicos e metadados de levantamentos hidrográficos realizados pela NOAA e estudos paleoclimatológicos. O único documento que menciona um tesouro é uma referência ao "*Getty Thesaurus of Geographic Names Online Page*", presente nos metadados de alguns relatórios hidrográficos. No entanto essa menção é a um recurso externo e não faz parte do conteúdo principal dos documentos analisados. Portanto, embora os documentos sejam úteis para suas respectivas áreas de pesquisa, eles não atendem aos critérios específicos que caracterizam um tesouro.

5.1.4 Documentos recuperados no *site* de dados abertos do governo do Reino Unido

A página Data.gov.uk oferece acesso a dados abertos publicados por governos central, locais e órgãos públicos do Reino Unido, organizados por tópicos como negócios, economia, crime, justiça, defesa, educação, meio ambiente, saúde e transporte. O *site* facilita a busca e o uso de dados para desenvolver produtos e serviços, fornecendo ferramentas e orientações para a publicação e gestão de dados. As informações publicadas no Data.gov.uk não são hospedadas diretamente pelo *site*, mas sim nos *sites* das respectivas organizações, com descrições e *links* disponibilizados no catálogo.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 6, a seguir.

Tabela 6 – Menções aos SOC – *site* de dados abertos do Reino Unido

TERMOS	RESULTADOS
authority files	0
categorization	
schemes	0
classification scheme	36
dictionary	35
directory	392
gazetteer	4
glossary	4
ontology	20
pick list	0
semantic network	0
subject headings	3
synonym rings	0
taxonomy	12
thesaurus	2
Total	508

Fonte: elaborada pela autora (2024).

Em relação à análise dos documentos recuperados que mencionavam Esquemas de Classificação (*classification scheme*), observou-se que vários dos documentos recuperados são exemplos desse tipo de SOC, essenciais para a organização e a padronização de dados complexos. Documentos como "*Compliance Classification Scheme*", "*Historic River Quality Objectives*" e "*The BGS Rock Classification Scheme*" exemplificam esquemas de classificação que categorizam informações ambientais e geológicas de maneira sistemática e estruturada. Esses esquemas fornecem uma base consistente para a avaliação da conformidade ambiental, qualidade da água e estudos geológicos, facilitando a comunicação e a tomada de decisões informadas em diversas áreas. Além disso, documentos como "*Digital Geological Map Data of Great Britain*" e "*Species and Habitat Point Records from 1987 AWA Breydon Water Littoral Survey*" aplicam classificações existentes para organizar e apresentar dados de forma estruturada, contribuindo para a conservação ambiental e a gestão de riscos geológicos. Embora nem todos os documentos sejam esquemas de classificação por definição, eles utilizam classificações para estruturar dados, demonstrando a importância dessas ferramentas na gestão e análise de informações.

Os documentos recuperados mencionam o termo Dicionário (*dictionary*) em um contexto técnico específico, como dicionários de dados médicos, ambientais e estatísticos. Exemplos incluem o "*NHS Data Model and Dictionary*" e o "*UK Marine Noise Registry Outputs*", que fornecem definições e formatos de dados para padronização e análise correta em seus respectivos campos. Esses documentos podem ser considerados essenciais para a interpretação precisa de dados técnicos, mas não se enquadram na definição de um dicionário, que organiza palavras alfabeticamente e fornece informações detalhadas sobre formas, pronúncias, etimologias e usos idiomáticos. Nesse sentido, não foi identificado nenhum documento que pudesse ser classificado como esse tipo de SOC.

A análise dos documentos recuperados que mencionavam o termo Diretório (*directories*) revela que muitos deles funcionam como diretórios abrangentes, fornecendo informações organizadas sobre locais, serviços e contatos em diversos setores. Diretórios como o "*Points of Interest*", "*Department for Transport Central Staff Directory*", "*Directory of Business*", e os documentos sobre corpos públicos de 2010, 2012 e 2013 oferecem uma visão detalhada e atualizada de entidades governamentais, negócios e serviços públicos e privados. Além disso, diretórios como o "*Care Quality Commission Care Directory*" desempenham um papel importante na área de saúde, fornecendo listas abrangentes de locais onde são realizados cuidados regulamentados, ajudando pacientes e profissionais a localizar e acessar serviços de saúde. Embora alguns documentos, como os registros de espécies e habitats, sirvam mais como bases de dados científicas, a maioria dos documentos analisados podem ser classificados como diretórios.

A respeito dos documentos recuperados que mencionavam o termo Dicionário Geográfico (*gazetteer*), observou-se que todos os quatro documentos se encaixam na definição de *gazetteers*, fornecendo bases de dados georreferenciadas de endereços e nomes de lugares. Documentos como o "*Local Land and Property Gazetteers*" e o "*One Scotland Gazetteer*" exemplificam a função de *gazetteers* ao consolidar informações de endereços de diversas autoridades locais em bases de dados padronizadas e continuamente atualizadas. Esses recursos são essenciais para o planejamento urbano, gestão de propriedades e serviços públicos, garantindo precisão e consistência nos dados geográficos. O "*Place Name Gazetteer – Scotland*" e o "*National Address Gazetteer*" destacam-se por seu valor cultural e utilidade para pesquisas acadêmicas, projetos de preservação histórica e aplicação em políticas

públicas. O "*Place Name Gazetteer – Scotland*" integra nomes de lugares com importância cultural, histórica e linguística, enquanto o "*National Address Gazetteer*" unifica dados de endereços de várias fontes, oferecendo uma referência única e definitiva para o setor público.

Dos quatro documentos recuperados que mencionavam o termo Glossário (*glossary*), apenas o "*Glossary of Export Control*" pode ser classificado como esse tipo de SOC. Esse documento, publicado pelo Departamento de Negócios, Inovação e Competências, fornece uma lista detalhada de termos especializados e acrônimos usados no controle de exportações, acompanhados de suas definições. Sua estrutura e propósito atendem à definição de um glossário, tornando-o um recurso essencial para a compreensão da terminologia específica utilizada nesse contexto. Os documentos "*Spend over £25,000 in HM Revenue & Customs*" e "*Materials collected for preparation for reuse and recycling*" fazem referência a glossários externos, enquanto "*2021 Camden School Places Planning Reporting*" inclui um glossário de abreviações como anexo.

A análise dos documentos recuperados que mencionavam o termo Ontologia (*ontology*) revela que a maioria utiliza ontologias para estruturar e organizar dados. Documentos como "*Libraries in East Sussex*" e "*Waste and Recycling Sites in East Sussex*" aplicam a ontologia *schema.org*, enquanto os diversos "*Indices of Multiple Deprivation 2010*" utilizam a ontologia "*Data Cube*" para representar os dados como *Linked Data*. No entanto dois dos documentos analisados, "*Indices of Multiple Deprivation 2010, Crime Rank*" e "*Indices of Multiple Deprivation 2010, Housing Rank*", apesar de estruturarem dados de privação de maneira detalhada, não fazem uso explícito de ontologias. Isso indica que, embora a aplicação de ontologias seja uma prática prevalente e benéfica, ainda há espaço para expandir seu uso em outros conjuntos de dados.

A análise dos documentos recuperados que citavam o termo Cabeçalho de Assunto (*subject headings*) revelou que nenhum deles corresponde a esse tipo de SOC. Os documentos, como "*General and Personal Medical Services, England December 2016, Provisional Experimental statistics*" e "*NHS Staff Earnings Estimates*", são conjuntos de dados estatísticos focados em informações específicas sobre serviços médicos e ganhos dos funcionários do NHS. Embora forneçam dados detalhados e úteis para suas áreas de atuação, não possuem a estrutura necessária para indexação e recuperação sistemática de informações.

Após a análise dos dez documentos recuperados que mencionavam o termo Taxonomia (*taxonomy*), identificou-se que apenas um deles, "*Taxonomy for macroinvertebrates in Welsh upland rivers (2012-2013)*", alinha-se ao conceito de uma taxonomia. Esse documento fornece uma estrutura detalhada para a classificação dos macroinvertebrados, incluindo dados sobre sua distribuição e abundância, o que o torna um exemplo de taxonomia organizada hierarquicamente. Os demais documentos analisados, como "*Biodiversity Action Plan (BAP) Gap – Bryophytes*" e "*All BGS Raw Biostratigraphy Collections And Data*", embora contenham informações detalhadas, servem como relatórios descritivos ou catálogos de coleções. Eles não apresentam a estrutura hierárquica detalhada que caracteriza uma taxonomia.

Sobre os documentos recuperados que mencionavam o termo Tesouro (*thesaurus*), identificou-se que os dois documentos podem ser classificados como tesouros. O "*BGS Geoscience Thesaurus*" contém aproximadamente 6000 termos relacionados à geociência, organizados com relações hierárquicas, sinônimos e notas explicativas, atendendo aos critérios dos padrões NISO e ISO. O "*UK Read Code*" é um tesouro de termos clínicos que fornece um vocabulário padronizado para o registro de informações de saúde. Ambos os tesouros mencionados demonstram as características definidoras de um tesouro, promovendo a padronização e a interoperabilidade em seus respectivos campos.

5.1.5 Documentos recuperados no *site* de dados abertos do governo de Singapura

A página Data.gov.sg é a principal plataforma de dados abertos de Singapura, oferecendo acesso a mais de 4.000 conjuntos de dados de 70 agências governamentais. Os dados são organizados em categorias como economia, educação, saúde, transporte, meio ambiente e cultura. A plataforma facilita a busca, acesso e uso dos dados, fornecendo APIs para desenvolvimento de aplicativos e serviços, além de *dashboards* interativos, gráficos e tabelas compartilháveis. O *site* também apresenta *blogs* que explicam tendências sociais e políticas através de análises de dados, promovendo transparência e inovação.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 7, a seguir.

Tabela 7 – Menções aos SOC – *site* de dados abertos de Singapura

TERMOS	RESULTADOS
authority files	0
categorization	
schemes	0
classification scheme	0
dictionary	0
directory	4
gazetteer	0
glossary	0
ontology	0
pick list	0
semantic network	0
subject headings	0
synonym rings	0
taxonomy	0
thesaurus	0
Total	4

Fonte: elaborada pela autora (2024).

Mesmo considerando a informação do *site* de dados abertos da existência de mais de 4 mil conjuntos de dados, em relação aos SOC pesquisados, houve registro apenas do tipo Diretório (*directory*). Dos cinco documentos analisados desse tipo de SOC, todos foram classificados como diretórios, pois fornecem listas organizadas e detalhadas de escolas, serviços para pessoas com deficiência, postos de bombeiros, agências governamentais e projetos de habitação pública, demonstrando a utilidade desses diretórios na promoção da transparência, acessibilidade e eficiência na busca e uso de informações.

5.1.6 Documentos recuperados no *site* de dados abertos do governo de Malta

A página Data.gov.mt é a principal plataforma de dados abertos de Malta, oferecendo acesso a conjuntos de dados publicados por vários ministérios e organizações governamentais. As categorias incluem agricultura, energia, transporte, economia, justiça, meio ambiente, educação e saúde. O *site* facilita a busca e a filtragem de dados por categoria, ministério, organização, setor e função. A plataforma tem como objetivo promover a transparência e a inovação, incentivando o uso dos

dados para pesquisa, desenvolvimento de aplicativos e serviços, e a criação de visualizações de dados.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 8, a seguir.

Tabela 8 – Menções aos SOC – *site* de dados abertos de Malta

TERMOS	RESULTADOS
authority files	0
categorization schemes	0
classification scheme	0
dictionary	0
directory	0
gazetteer	0
glossary	0
ontology	0
pick list	0
semantic network	0
subject headings	0
synonym rings	0
taxonomy	0
thesaurus	0
Total	0

Fonte: elaborada pela autora (2024).

A análise no *site* de dados abertos do governo de Malta revelou uma ausência completa de referências explícitas a qualquer tipo de SOC. Essa ausência sugere que a plataforma pode não utilizar ou referenciar explicitamente SOCs formais na organização e disponibilização de seus dados. Avalia-se que a falta de menções a SOCs específicos apresenta uma oportunidade para o governo de Malta aprimorar a estruturação e a transparência de suas práticas de organização de dados. Adotar e explicitar SOCs reconhecidos internacionalmente poderia melhorar a interoperabilidade, a acessibilidade e a reutilização dos dados públicos, trazendo benefícios especialmente para pesquisadores e desenvolvedores, facilitando a busca, a filtragem e o uso dos dados, e potencialmente promovendo maior inovação e desenvolvimento de serviços baseados em dados.

5.1.7 Documentos recuperados no *site* de dados abertos do governo da Espanha

A página Data.gob.es é a principal plataforma de dados abertos do governo da Espanha, oferecendo acesso a mais de 80 mil conjuntos de dados publicados por diversas administrações e entidades públicas. Os dados estão organizados em categorias como meio ambiente, cultura, educação, transporte, saúde, turismo e justiça. A plataforma facilita a busca e o uso dos dados por meio de um sistema de filtragem eficiente e ferramentas como APIs, tutoriais e guias para publicação e gestão de dados. Além de promover a transparência, o Data.gob.es incentiva a inovação e o uso dos dados para pesquisa, desenvolvimento de aplicativos e serviços. A plataforma inclui seções dedicadas a eventos, notícias e projetos relacionados ao uso de dados abertos, oferecendo suporte e inspiração para cidadãos, pesquisadores e desenvolvedores. O *site* também apresenta casos de sucesso e exemplos de como os dados abertos estão sendo utilizados para criar valor e resolver problemas sociais.

Partindo-se para a consulta dos documentos recuperados no idioma espanhol, têm-se os resultados dispostos na Tabela 9, a seguir.

Tabela 9 – Menções aos SOC – *site* de dados abertos da Espanha

TERMOS	RESULTADOS
anillos de sinónimos	0
archivos de autoridad	0
diccionario geográfico	1
diccionarios	76
directorios	509
encabezamientos de materia	10
esquemas de categorización	0
esquemas de clasificación	0
glosarios	18
listas de selección	0
ontología	17
red semántica	0
taxonomías	3
tesauros	15
Total	649

Fonte: elaborada pela autora (2024).

A análise do documento recuperado do tipo Dicionário Geográfico (*diccionario geográfico*) indica que o "*Diccionario Geográfico – Conjunto de datos*" disponibilizado pelo Ministério de Cultura y Deporte da Espanha pode ser considerado um exemplo de dicionário geográfico, pois dispõe de uma base de dados detalhada e georreferenciada de mais de 46.000 termos geográficos, essencial para a catalogação e contextualização de bens culturais.

Dos documentos analisados do tipo Dicionários (*diccionarios*), três podem ser classificados como esse tipo de SOC. Os documentos "*Diccionario de Contextos Culturales*", "*Diccionario Enciclopédico Lu*" e "*Diccionario Geográfico*" podem ser considerados mais próximos do conceito de dicionário, pois organizam termos alfabeticamente e oferecem informações amplas e contextuais. Os demais documentos identificados fornecem definições e contextos detalhados, podendo ser classificados como dicionários técnicos, mas não organizam as palavras alfabeticamente com informações sobre pronúncia ou etimologia.

A análise dos documentos recuperados que mencionavam o termo Diretório (*directorios*) revela que todos se enquadram na definição de diretórios, fornecendo informações detalhadas sobre uma variedade de serviços e entidades. Documentos como "*Directorio de Empresas*", "*Directorio de Asociaciones*", "*Directorio de Centros Religiosos*", "*Directorio de Centros Sanitarios*" e "*Directorio de Farmacias*" organizam e disponibilizam informações essenciais sobre locais, contatos e funções de diferentes organizações e serviços. Esses diretórios são fundamentais para promover a transparência, a acessibilidade e a eficiência na busca e utilização de informações por cidadãos e profissionais.

A análise dos documentos recuperados que mencionavam o termo Cabeçalho de Assunto (*encabezamientos de materia*) revelou que apenas o documento "*Lista de Encabezamientos de Materia para las Bibliotecas Públicas en SKOS*" pode ser classificado como esse tipo de SOC. Esse documento fornece uma lista estruturada e controlada de termos utilizados para representar assuntos em bibliotecas públicas, abrangendo vários idiomas e utilizando padrões de dados abertos vinculados, como *Dublin Core* e SKOS. Os demais documentos analisados, incluindo diversos dicionários especializados e catálogos de autoridades, embora ricos em terminologia e organização de dados específicos, não possuem a estrutura hierárquica e a funcionalidade de um cabeçalho de assunto. Esses documentos são mais adequados

como recursos terminológicos especializados e ferramentas de catalogação em áreas específicas, como patrimônio cultural, técnicas, mobiliário e geografia.

Dos documentos recuperados que mencionavam o termo Glossário (*glosarios*), seis podem ser claramente classificados como glossários. Os documentos "*Diccionarios terminológicos de Euskalterm*", "*Glosario del Gobierno Vasco (2020-2024)*", "*Estadística tributaria autonómica 2008*", "*Estadística tributaria autonómica 2009*", "*Estadística tributaria autonómica 2011*" e "*Sistema Integrado de Metadatos (SIM)*" fornecem listas estruturadas de termos e suas definições, servindo como ferramentas essenciais para padronização terminológica e facilitando a compreensão de termos técnicos e administrativos em seus respectivos contextos. Os demais documentos, como o "*Directorio de Servicios Públicos Avanzados ZERPA 2021*", "*Ejes de las carreteras de titularidad provincial*" e "*Esquema de conceptos transversales de SDMX*", apesar de fornecerem informações detalhadas e úteis em suas áreas de foco, não incluem glossários ou listas de termos com definições.

A análise dos documentos recuperados que mencionavam o termo Ontologia (*ontología*) revela um uso diversificado desse tipo de SOC para a organização e gestão de dados em diversas áreas. Quatro dos documentos analisados, incluindo "*Ontología CAP IACS – Open IACS*", "*Ontología SIGPAC – Open IACS*", "*Diccionario de Denominaciones de Bienes Culturales*" e "*Oferta de Empleo*", demonstram a aplicação de ontologias para estruturar dados, facilitando a interoperabilidade e a integração das informações. Esses documentos utilizam vocabulários RDF/OWL e outras tecnologias semânticas, mostrando como as ontologias podem melhorar significativamente a acessibilidade e a utilidade dos dados em contextos específicos, como agricultura, cultura e emprego. Por outro lado, cinco dos documentos, como "*Alojamientos*", "*Restaurantes*", "*Monumentos*", "*Catálogo de datos*" e "*Impresos*", apesar de apresentarem dados estruturados e em formatos compatíveis com *Linked Data*, não mencionam explicitamente o uso de ontologias.

Após a análise dos documentos recuperados que mencionavam o termo Taxonomia (*taxonomías*), conclui-se que apenas dois deles, "*Agrupación temática nivel dominio*" e "*Agrupación temática nivel subdominio*", podem ser classificados como taxonomias. Ambos os documentos apresentam estruturas hierárquicas detalhadas para a classificação de serviços, conceitos e dados temáticos, atendendo ao conceito de taxonomia. Esses documentos organizam termos em domínios e subdomínios, facilitando a categorização e o acesso às informações. Os outros dois

documentos, "*Conjunto de datos*" e "*Registro de Reglas de Conservación*", embora valiosos e informativos, não se enquadram como taxonomias. O "*Conjunto de datos*" é um catálogo geral de dados disponíveis, sem uma estrutura hierárquica detalhada, enquanto o "*Registro de Reglas de Conservación*" concentra-se na gestão e conservação de documentos, estabelecendo prazos de transferência e eliminação, mas sem organizar termos de forma hierárquica.

Por fim, em relação aos documentos recuperados que mencionavam o termo Tesouro, identificou-se que sete desses documentos podem ser classificados como tesouros. Esses documentos incluem os catálogos de autoridades em áreas específicas como "Literatura e Cômico", "Dibujos y Grabado", "Fotografía, Cine e Cartografía", além do "Tesouro del Patrimonio Histórico Andaluz" e do "Diccionario de Mobiliario". Todos esses vocabulários controlados apresentam termos normalizados, definições detalhadas, e relações hierárquicas e associativas, cumprindo assim os critérios para serem considerados tesouros, conforme os padrões NISO e ISO. Os outros três documentos ("*Diccionario de Contextos culturales*", "*Diccionario de Cerámica*" e "*Diccionario de Técnicas*") não atendem à definição de tesouro, pois carecem das relações hierárquicas e associativas características desses sistemas de organização do conhecimento.

5.1.8 Documentos recuperados no *site* de dados abertos do governo da Irlanda

O data.gov.ie é a principal plataforma de dados abertos da Irlanda, lançada em 2014 como parte da Iniciativa de Dados Abertos da Irlanda. Ela oferece acesso a uma vasta gama de conjuntos de dados fornecidos por departamentos governamentais e organismos públicos, promovendo a transparência, a inovação e a reutilização de dados. A plataforma organiza os dados em várias categorias, como agricultura, saúde, educação, meio ambiente e transporte, facilitando a busca e o acesso por meio de ferramentas como APIs e tutoriais. Além de disponibilizar dados para *download* e uso, Data.gov.ie incentiva a participação dos usuários ao permitir que sugiram novos conjuntos de dados e compartilhem suas experiências e aplicações dos dados. A plataforma também apoia o desenvolvimento econômico, social e ambiental por meio do uso de dados abertos, destacando casos de sucesso e exemplos práticos de como os dados estão sendo utilizados para resolver problemas reais.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 10, a seguir.

Tabela 10 – Menções aos SOC – *site* de dados abertos da Irlanda

TERMOS	RESULTADOS
authority files	0
categorization schemes	0
classification scheme	6
dictionary	2
directory	19
gazetteer	10
glossary	6
ontology	1
pick list	0
semantic network	0
subject headings	0
synonym rings	0
taxonomy	1
thesaurus	0
Total	45

Fonte: elaborada pela autora (2024).

A análise dos documentos recuperados referentes ao termo Esquemas de classificação (*classification scheme*) revela que todos os seis documentos recuperados podem ser considerados esse tipo de SOC. Documentos como "*GS/ Physiographic Units Level 1, Level 2, e Level 3*" e "*INFOMAR Seabed Substrate*" utilizam classificações detalhadas e padronizadas para organizar dados geográficos, geomorfológicos e sedimentares. Esses esquemas são essenciais para a compreensão e gestão dos recursos naturais da Irlanda, fornecendo uma base consistente para a análise de paisagens, tipos de sedimentos e características geomorfológicas do leito marinho. Além disso, o "*Irish Shelf Seabed Geomorphological Map v2023*" aplica um esquema de classificação internacionalmente padronizado para categorizar características do leito marinho da plataforma continental irlandesa.

Os documentos analisados que utilizam o termo Dicionário (*dictionary*) referem-se ao uso do termo no contexto de dicionários de dados técnicos, que servem para padronizar e explicar elementos de dados específicos em contextos como pesca e saúde. Eles são essenciais para a interpretação correta e a utilização dos dados

apresentados, mas não correspondem à definição tradicional de um dicionário, que organiza palavras alfabeticamente e fornece informações detalhadas sobre formas, pronúncias, etimologias e usos idiomáticos. Portanto nenhum dos documentos analisados pode ser classificado como um dicionário.

Em relação aos documentos analisados que mencionavam o termo Diretório (*directory*), identificou-se que oito documentos podem ser classificados como esse tipo de SOC. Eles incluem listas organizadas e detalhadas de locais, contatos e serviços para práticas médicas familiares, edifícios do conselho, estações de bombeiros, lares de idosos, hospitais, serviços de saúde mental e pedreiras. Apenas dois documentos, que fornecem dados geoespaciais sobre pequenas áreas estatísticas, não se enquadram na definição de diretório.

Dos documentos analisados que mencionavam o termo Dicionário Geográfico (*gazetteer*), nove podem ser classificados como esse tipo de SOC, fornecendo bases de dados georreferenciadas e estruturadas sobre nomes de lugares na Irlanda. Esses *gazetteers* são essenciais para diversas aplicações, incluindo planejamento urbano, transporte público e pesquisa histórica. Apenas o documento "*Lacewings (Neuroptera) of Ireland*" não se enquadra como um *gazetteer*, pois trata de informações taxonômicas sobre insetos.

Sobre documentos que citavam o termo Glossário (*glossary*), dos seis documentos analisados, cinco podem ser claramente classificados como esse tipo de SOC. Os documentos "*Field Alias Glossary (Multi-Storey Developments)*", "*Field Alias Glossary (PRA State Assets)*", "*Field Alias Glossary (State Assets Sourced by LDA)*", "*Public Transport Infrastructure Management System (PTIMS)*" e "*Quarterly Key Data Reports (QDKR) ComReg*" fornecem listas estruturadas de termos e suas definições, servindo como ferramentas de padronização terminológica. Já o documento "*Hydro Energy Potential*", embora inclua algumas definições de termos técnicos, não se configura como um glossário abrangente e, portanto, não pode ser classificado como tal.

A análise do documento recuperado que mencionava o termo Ontologia (*ontology*) revela uma aplicação clara de ontologia no contexto da proteção de dados. O documento intitulado "GDPRtEXT" é uma ontologia que estrutura referências específicas do Regulamento Geral sobre a Proteção de Dados (GDPR) utilizando a ontologia do *European Legislation Identifier* e o *Simple Knowledge Organization System* (SKOS). Essa ontologia visa facilitar a referência e o entendimento dos artigos

e pontos do GDPR de forma consistente e interoperável, melhorando a conformidade regulatória e a análise jurídica. Assim, o GDPRtEXT exemplifica como as ontologias podem organizar informações legais complexas, tornando-as mais acessíveis e compreensíveis para sistemas automatizados e usuários humanos.

Por fim, em relação à análise do documento recuperado que fazia menção ao termo Taxonomia (*taxonomy*), observou-se que o documento "*Groundfish Survey Invertebrate Data*" apresenta dados detalhados sobre invertebrados capturados durante pesquisas anuais de peixes, coordenadas pelo *Marine Institute* como parte do *International Bottom Trawl Survey* (IBTS). Embora inclua informações sobre espécies identificadas e quantificadas, o foco principal é a coleta e a análise de dados sobre a abundância e distribuição dessas espécies, sem organizar os termos em uma estrutura hierárquica detalhada. Portanto, apesar de ser um recurso valioso para a pesquisa e gestão pesqueira, não se qualifica como uma taxonomia.

5.1.9 Documentos recuperados no *site* de dados abertos do governo do Canadá

A página [Open.Canada.ca](https://open.canada.ca) é a principal plataforma de dados abertos do governo do Canadá, oferecendo acesso a uma vasta gama de informações e conjuntos de dados de várias instituições governamentais. A plataforma facilita a busca e o uso de dados abertos para promover a transparência e a inovação, com categorias como saúde, meio ambiente, economia, transporte e cultura. Além disso, inclui recursos como APIs, tutoriais e diretrizes para publicação e gestão de dados. A plataforma também destaca conjuntos de dados sobre gastos governamentais e operações, pedidos de acesso à informação e contratos públicos. A plataforma visa aumentar a responsabilização e a transparência do governo, fortalecendo a democracia e impulsionando a inovação por meio do acesso a dados. A página inclui seções dedicadas ao uso e publicação de dados abertos, orientações para instituições e informações sobre o compromisso do Canadá com a abertura governamental. O *site* também oferece exemplos de como os dados abertos estão sendo utilizados para criar valor e resolver problemas sociais, incentivando cidadãos, pesquisadores e desenvolvedores a explorar e utilizar os dados disponíveis.

Partindo-se para a consulta dos documentos recuperados no idioma inglês, têm-se os resultados dispostos na Tabela 11, a seguir.

Tabela 11 – Menções aos SOC – *site* de dados abertos do Canadá

TERMOS	RESULTADOS
authority files	30
categorization schemes	1
classification scheme	284
dictionary	2804
directory	7891
gazetteer	518
glossary	20095
ontology	130
pick list	46
semantic network	0
subject headings	184
synonym rings	0
taxonomy	2144
thesaurus	125
Total	34252

Fonte: elaborada pela autora (2024).

Ao analisar os documentos recuperados que faziam menção ao termo Arquivo de Autoridade (*authority files*), conclui-se que a maioria deles não se enquadra na definição desse tipo de SOC. Documentos como "*Guidance on Prescribing Metadata Reference Standards*", "*Metadata (data structure standards)*" e os perfis de *software* "*TMS Collections and eMuseum*", "*PastPerfect Museum Software Version 5.0*" e "*VITA by Our Digital World*" fornecem diretrizes e descrições detalhadas sobre padrões de metadados e sistemas de gestão de coleções, mas não padronizam a referência a entidades específicas. Em contraste, o documento "*Vocabulary (Data Value Standards)*" pode ser considerado um exemplo de "*authority file*", pois padroniza termos e nomes usados para criar metadados, promovendo consistência nos registros de dados. Esse documento assegura que informações semelhantes sejam referenciadas de maneira uniforme, o que é fundamental para a integração e o compartilhamento de dados.

Sobre o documento recuperado que fazia menção ao termo Esquema de Categorização (*categorization schemes*), observou-se que o documento "*Health Portfolio Sex- and Gender-Based Analysis Plus Policy: Advancing Equity, Diversity and Inclusion*" utiliza um esquema de categorização detalhado, denominado SGBA Plus, para integrar diversos fatores de identidade (como sexo, gênero, raça e *status*

socioeconômico) em todas as atividades do portfólio de saúde. Esse esquema interseccional permite uma análise abrangente de como essas identidades interagem e influenciam o acesso e os resultados de saúde, resultando em políticas e programas mais inclusivos e equitativos. Assim, o documento pode ser classificado como um esquema de categorização no contexto da política de saúde pública.

Em relação aos documentos que referenciavam o termo Esquemas de Classificação (*classification scheme*), identificou-se que seis deles podem ser classificados como esse tipo de SOC. Documentos como o "*Classification Scheme*" do *Transport Canada*, "*The Canadian Bikeway Comfort and Safety (Can-BICS) Classification System*", e "*0XX Codes – Classification – MARC Standards*" organizam e padronizam informações de maneira sistemática e estruturada. Esses esquemas são essenciais para garantir a consistência e a clareza na comunicação e na gestão de dados em áreas como transporte de mercadorias perigosas, planejamento urbano de infraestrutura cicloviária e catalogação bibliográfica em bibliotecas. Os demais documentos, embora não sejam esquemas de classificação formais, ainda desempenham um papel importante na organização e análise de informações específicas. Eles fornecem diretrizes, relatórios de pesquisa e estratégias que utilizam classificações para estruturar dados de maneira eficiente.

Dos documentos analisados que faziam menção ao termo Dicionário (*dictionary*), apenas o "*About the Medical Dictionary for Regulatory Activities (MedDRA)*" pode ser considerado um dicionário no sentido mais próximo do tradicional, devido à sua abrangência e organização detalhada de termos médicos. Os demais documentos são dicionários de dados técnicos ou guias de competências, úteis em seus contextos específicos, mas não se enquadram na definição tradicional de um dicionário. Portanto, dos dez documentos analisados, apenas um pode ser classificado como um SOC do tipo "dicionário" no sentido mais amplo, enquanto os outros nove são dicionários de dados técnicos ou guias de competências que não correspondem à definição tradicional de um dicionário.

Com relação aos documentos relacionados ao termo Diretórios (*directory*), observou-se que, dos documentos analisados, seis podem ser classificados como esse tipo de SOC, pois incluem listas organizadas de contatos e serviços para departamentos governamentais, cientistas e profissionais de pesquisa, funcionários públicos, museus militares, instalações correcionais e serviços de apoio a vítimas. Quatro documentos não se enquadram na definição de diretórios, pois fornecem

especificações técnicas ou informações sobre programas sem listar contatos específicos.

A respeito dos documentos que referenciavam Dicionários geográficos (*gazetteer*), conclui-se que, dos documentos analisados, cinco podem ser classificados como esse tipo de SOC: "*Gazetteer of Canada. Prince Edward Island*", "*Gazetteer of Canada. Alberta*", "*Gazetteer of Canada. Saskatchewan*", "*Gazetteer of Canada. Manitoba*" e "*Concise Gazetteer of Canada*". Esses documentos fornecem bases de dados georreferenciadas e detalhadas sobre nomes de lugares e suas localizações geográficas, cumprindo os critérios tradicionais de um *gazetteer*. Os outros cinco documentos, embora contenham informações sobre nomes geográficos e serviços associados, não se enquadram na definição de dicionário geográfico.

Após a análise dos dez documentos recuperados que faziam menção ao termo Glossário (*glossary*), conclui-se que todos eles podem ser classificados como esse tipo de SOC. Cada documento fornece uma lista estruturada de termos acompanhados de suas definições específicas no contexto de seus respectivos assuntos, como práticas ambientais, segurança de medicamentos veterinários, resposta a condutas sexuais inadequadas, e serviços digitais para empresas. Esses glossários são ferramentas essenciais para padronizar a terminologia técnica e facilitar a compreensão dos conceitos abordados nos documentos.

Sobre os documentos que faziam menção a Ontologias (*ontology*), apenas dois deles podem ser classificados como esse tipo de SOC: Os documentos "*Linked Open Data*" e "*Digital Health and Geospatial Analytics*" utilizam princípios e tecnologias de ontologia para estruturar e integrar dados de maneira eficiente, promovendo a interoperabilidade e a acessibilidade das informações. A aplicação dessas ontologias facilita a gestão de dados culturais e históricos, bem como a análise avançada de dados de saúde e geoespaciais, destacando a importância dessas ferramentas na melhoria da qualidade da informação e na tomada de decisões informadas. Os outros oito documentos analisados, apesar de fornecerem informações estruturadas em diversas áreas, como programas educacionais, revisões de literatura e diretrizes de metadados, não mencionam explicitamente o uso de ontologias.

Após análise dos documentos recuperados que mencionavam o termo Listas de Seleção (*Pick List*), observou-se que documentos como "*Help with Enquiries Service*" e os guias de relatórios do "*National Pollutant Release Inventory*" (NPRI) destacam o uso desse tipo de SOC para padronizar a entrada de dados, minimizar

erros e garantir que as informações sejam uniformes. Esses documentos demonstram que as listas de opções são fundamentais para a identificação precisa de corpos d'água, instalações fora do local e substâncias químicas, assegurando que todos os usuários sigam os mesmos critérios e nomenclaturas, o que facilita a análise e o processamento posterior dos dados. Além disso, documentos como o "*Natural Health Product Licence Application Form User Manual*" utilizam listas de opções para selecionar ingredientes e rotas de administração, alinhando-se com normas regulatórias e garantindo a uniformidade dos dados. Embora alguns documentos, como os FAQs sobre solicitações online e revisão de nomes de marcas de medicamentos não mencionem diretamente o uso de "Pick Lists", a maioria dos documentos analisados enfatiza a sua relevância na manutenção da consistência e precisão dos dados governamentais.

A análise dos documentos recuperados que faziam menção a Cabeçalhos de Assuntos (*subject headings*) indicou que apenas dois deles, "*Canadian Subject Headings*" e "*Retrieving Canadian Subject Headings*", podem ser classificados como esse tipo de SOC. Esses documentos fornecem uma lista estruturada e controlada de termos, destinada a facilitar a indexação e a recuperação de informações sobre o Canadá, sendo essenciais para a organização eficiente do conhecimento e da informação em bibliotecas e outras instituições. Os demais documentos analisados, incluindo diretrizes técnicas, programas de financiamento e avaliações específicas, não correspondem a esquemas do tipo Cabeçalhos de Assunto.

Após a análise dos documentos recuperados que faziam menção a Taxonomia (*taxonomy*), conclui-se que apenas três deles podem ser classificados como esse tipo de SOC. Os documentos "*Taxonomy*" do *Agriculture and Agri-Food Canada*, "*Taxonomy of wood products*" do *Natural Resources Canada*, e "*Chapter 16: Correlation of Canadian Soil Taxonomy with Other Systems*" fornecem estruturas hierárquicas detalhadas para a classificação de solos e produtos de madeira. Esses documentos organizam termos de maneira sistemática e hierárquica, facilitando a categorização e a pesquisa de informações, alinhando-se ao conceito clássico de taxonomia. Os demais documentos, como "*Taxonomy Roadmap Report*", "*Skill components and proficiency levels*" e os relatórios de progresso relacionados ao plano de redução de emissões e financiamento climático não atendem aos critérios de uma taxonomia.

Dos documentos recuperados que faziam menção ao termo Tesouro (*thesaurus*), sete podem ser classificados como esse tipo de SOC. O "*Government of Canada Core Subject Thesaurus*", o "*Thesaurus of Information Sciences and Technologies*" e o "*Thesaurus | CRTC*" contêm termos controlados organizados hierarquicamente com relações associativas, atendendo aos critérios estabelecidos pelos padrões NISO e ISO para a definição de tesouro. Os outros três documentos, "*Core Standards for Canadian Museums*", "*Collections Management System Criteria Checklist*" e "*Making Up the Rules: New Documentation Standards for Canadian Museums*", embora sejam recursos úteis para a gestão e documentação de coleções museológicas, não apresentam a estrutura hierárquica e as relações entre termos características de um tesouro.

5.1.10 Documentos recuperados no *site* de dados abertos do governo do Uruguai

A página *Datos Abiertos* do Uruguai oferece acesso a uma ampla gama de dados abertos publicados por organismos públicos, organizações da sociedade civil, academia e empresas privadas. Os dados estão organizados em diversas categorias, como agricultura, saúde, educação, meio ambiente, transporte e economia. A plataforma facilita a busca e o uso dos dados, fornecendo recursos como APIs, guias e tutoriais para a publicação e gestão de dados. Além disso, promove a transparência e a inovação, incentivando o uso dos dados para pesquisa, desenvolvimento de aplicativos e serviços. A página também destaca eventos e novidades relacionados a dados abertos, como lançamentos de comunidades e consultas públicas sobre recomendações técnicas. Há seções dedicadas a aplicações práticas dos dados, como informações sobre violência de gênero, prestação de serviços de saúde, compras e contratações do Estado, e indicadores de desenvolvimento sustentável. A plataforma incentiva a participação cidadã e oferece recursos para desenvolvedores criarem visualizações e aplicações baseadas em dados abertos.

Partindo-se para a consulta dos documentos recuperados no idioma espanhol, têm-se os resultados dispostos na Tabela 12, a seguir.

Tabela 12 – Menções aos SOC – *site* de dados abertos do Uruguai

TERMOS	RESULTADOS
anillos de sinónimos	3
archivos de autoridad	1
diccionario geográfico	47
diccionarios	20
directorios	50
encabezamientos de materia	30
esquemas de categorización	0
esquemas de clasificación	7
glosarios	40
listas de selección	12
ontología	20
red semántica	4
taxonomías	41
tesauros	16
Total	291

Fonte: elaborada pela autora (2024).

Os documentos recuperados que faziam menção a Anéis de Sinônimos (*anillos de sinónimos*) tratam, em sua maioria, de especificações técnicas e processos de licitação para a implementação de sistemas de governo eletrônico. O documento "pliego_159685.pdf" detalha uma licitação pública para a aquisição de componentes tecnológicos, incluindo sistemas de autenticação, autorização, *gateways/firewalls* XML, auditoria e serviços de consultoria e capacitação. Esse documento é abrangente e técnico, focado na infraestrutura e segurança da plataforma de governo eletrônico, mas não aborda o conceito de "anéis de sinônimos". O segundo documento, "investigacion_bibliotecologica_2030.pdf", concentra-se em uma investigação bibliotecológica com foco na interoperabilidade, metadados e integração de sistemas de informação para o futuro. Embora discuta a padronização terminológica e a importância da interoperabilidade, não menciona explicitamente os "*anillos de sinónimos*". O terceiro documento, "pliego_159531.pdf", similar ao primeiro, também detalha especificações técnicas e processos de licitação para componentes tecnológicos, sem referência a conjuntos de termos equivalentes. Em resumo, os

documentos são ricos em detalhes técnicos e processos de implementação, mas não correspondem diretamente ao conceito de Anéis de Sinônimos.

Ao analisar o documento "Pedido 401586", que fazia referência ao termo Arquivos de Autoridade (*archivos de autoridad*), observou-se que ele detalha um projeto para melhorar a infraestrutura de TI da Biblioteca e Publicações da Junta Departamental de Montevideo, incluindo a organização de bases de dados e a implementação de arquivos de autoridade para padronizar e controlar a qualidade dos registros bibliográficos, por meio de listas autoritativas de autores, instituições e outros elementos descritivos. Ou seja, o documento fazia menção ao SOC, mas não pode ser classificado como tal.

Sobre os documentos que referenciam o termo Dicionário Geográfico (*diccionario geográfico*), dos oito documentos analisados, seis podem ser classificados como esse tipo de SOC: "*Diccionario Geográfico Universal*", "*Diccionario Geográfico del Uruguay*", "*Diccionario Geográfico-Postal de la República Oriental del Uruguay*", "*Novísimo Diccionario Geográfico, Histórico, Pintoresco Universal*", "*Diccionario Geográfico de la República de Chile*" e "*Diccionario Geográfico-Estadístico-Histórico de España y sus Posesiones de Ultramar*". Esses documentos fornecem informações detalhadas e georreferenciadas sobre localidades, cumprindo os critérios de dicionários geográficos. Os outros dois documentos, "*Catálogo en línea Biblioteca del Poder Legislativo*" e "*Parque 'Artigas'*", não se enquadram nessa definição, pois funcionam como um índice de referência e uma descrição de um local específico, respectivamente.

Dos documentos analisados que mencionavam o termo Dicionário (*diccionarios*), três podem ser classificados como esse tipo de SOC. O "*Catálogo en línea Biblioteca José Artigas*", o "*Catálogo en línea Biblioteca del Ministerio de Ambiente*" e o "PORTUGUES – DICCIONARIOS – ESPAÑOL" organizam palavras alfabeticamente e fornecem traduções e definições, correspondendo ao conceito clássico de um dicionário. Os demais documentos são dicionários técnicos ou discussões sobre a evolução dos dicionários no contexto digital. Eles fornecem terminologias padronizadas e classificações específicas para contextos médicos, jurídicos e ambientais, sendo essenciais para a padronização e a análise de dados especializados, mas não se enquadram na definição de um dicionário.

Sobre os documentos que mencionavam o termo Diretórios (*directorios*), três podem ser classificados como esse tipo de SOC, que seriam os documentos

"*Directorios de Representaciones Uruguayas*", "*Directorios del Partido Nacional y del Partido Nacional Independiente 1887-1994*" e "*Directorio Estados Unidos – Catálogo en Línea*". Os demais documentos são registros administrativos, normativos ou manuais que não se enquadram na definição tradicional de diretórios.

A análise dos documentos recuperados que se referiam ao termo Cabeçalho de Assunto (*encabezamientos de materia*) indicou que seis deles podem ser classificados como esse tipo de SOC. Documentos como "*Lista de Encabezamientos de Materia para Bibliotecas*", "*Encabezamientos de Materia*", "*Los epígrafes en el catálogo diccionario*", "*Procultura*", "*Unión Panamericana*" e "*Acme*" fornecem listas estruturadas e controladas de termos para indexação e recuperação de informações em bibliotecas. Os outros quatro documentos, incluindo "*Catalogación sencilla para bibliotecas*", "*Catálogo en línea Biblioteca del Poder Legislativo*", "*Clasificación Decimal de Dewey*" e "*Catalogación – Normas*" são guias técnicos e manuais de catalogação que, apesar de mencionarem o termo "*encabezamientos de materia*", não possuem a estrutura hierárquica e abrangente necessária para serem considerados esquemas completos de cabeçalhos de assunto.

Dos documentos analisados que remetiam ao termo Esquemas de Classificação (*esquemas de clasificación*), dois deles podem ser considerados como esse tipo de SOC. O documento "*S2201035_es.pdf*" e o "*Rojido et al 2024 Tipología de los Homicidios en Uruguay vF.pdf*" fornecem categorias e definições claras para a organização de dados, cumprindo os critérios de um esquema de classificação. Os outros documentos analisados, embora úteis e informativos, não se enquadram como esquemas de classificação, pois não organizam informações de maneira sistemática e padronizada.

Em relação aos documentos que faziam menção ao termo Glossário (*glosarios*), identificou-se que três podem ser classificados como glossários: "*Glosarios | SINAÉ*", "*Decreto 150/000 Aprobación del documento de definiciones y glosarios para productos domisanitarios*", e "*Glosarios | SINAÉ-2*". Esses documentos fornecem listas estruturadas de termos e suas definições, servindo como ferramentas de padronização terminológica. Os outros documentos, embora mencionem o termo "glosarios" ou façam referência a materiais que os contêm, não fornecem diretamente uma lista de termos com definições e, portanto, não podem ser classificados como esse tipo de SOC.

A análise dos documentos que remetiam ao termo Listas de Seleção (*listas de selección*) indicou que documentos como o "*Manual para Registradores Agosto 2019*", o "*AG-SAE-DA-01-ES Documento de Arquitectura V1.7*", o "*MA-13210-002 Manual Buenas Prácticas Ciudadano*" e o "*Registro de Actividades Condiciones de Trabajo y Personas*" demonstram a utilização explícita de listas de seleção para facilitar a escolha de opções predefinidas, minimizando erros e garantindo a uniformidade dos dados coletados. Essas listas são integradas em formulários eletrônicos, em que os usuários podem selecionar informações específicas, como tipos de envase, materiais, departamentos e países, assegurando que as entradas de dados estejam em conformidade com os requisitos regulamentares e sejam consistentes em toda a base de dados. Por outro lado, o documento "*LPN_10018.12 – Pliego*" não menciona diretamente o uso de listas de seleção, mas sugere a possibilidade de sua utilização implícita em seus procedimentos técnicos.

A análise dos documentos que faziam referência ao termo Ontologia (*ontología*) indicou que apenas um deles, "*Guía Técnica Ontología de Documentos – Versión 5.0.0*", pode ser classificado como esse tipo de SOC. Esse documento detalha a utilização de padrões para classificar e mapear diferentes tipos de documentos clínicos no contexto da História Clínica Electrónica Nacional (HCEN) do Uruguai. A aplicação dessa ontologia facilita a interoperabilidade e a consistência dos dados clínicos, evidenciando um uso claro e estruturado de ontologias para a organização e gestão de informações de saúde. Os outros nove documentos, incluindo ensaios filosóficos, catálogos bibliográficos e recursos informativos, tratam o conceito de ontologia de forma mais teórica ou descritiva, sem aplicá-lo tecnicamente como um sistema de organização do conhecimento.

Após a análise dos documentos recuperados que remetiam ao termo Rede Semântica (*red semántica*), conclui-se que nenhum dos arquivos corresponde à definição desse tipo de SOC. Embora os documentos abranjam uma ampla gama de temas, como envelhecimento da população, práticas sociais e relatos sobre a morte, todos são estruturados de maneira analítica e descritiva sem a organização de conceitos e termos em uma rede inter-relacionada típica das redes semânticas.

Em relação aos documentos recuperados que faziam menção ao termo Taxonomia (*taxonomías*), conclui-se que quatro deles podem ser classificados como esse tipo de SOC. Os documentos "*Taxonomía por tipo, públicos y temas para sitios web e intranets*", "*Normativa*", "*Descarga de plantillas*" e "*Nueva contenido para*

informes de gestión" apresentam estruturas hierárquicas detalhadas para a classificação de informações, sejam elas dados financeiros, sejam conteúdos Web ou relatórios de gestão. Os demais documentos, incluindo "*Sobre pérdidas y paraísos*", "*Llamado para la contratación de Consultor Informático*", "*Drupal para el sitio web*", e "*Cómo usar Drupal*" são predominantemente manuais de instrução, chamados de contratação ou críticas culturais, que mencionam ou utilizam conceitos de taxonomia, mas não apresentam uma estrutura hierárquica completa de termos.

Por fim, em relação aos documentos recuperados que mencionavam o termo Tesouro (*tesauros*), três podem ser classificados como esse tipo de SOC. Esses documentos, incluindo "*Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*", "*Administración – Tesauros Catálogo en Línea*" e "*Tesauros Catálogo en línea*", apresentam termos organizados hierarquicamente e relações associativas, atendendo aos critérios estabelecidos para tesauros. Eles fornecem uma estrutura robusta para a organização e a recuperação de informações, promovendo a consistência e a padronização terminológica. Os outros sete documentos incluem manuais, relatórios técnicos, bases de dados e descrições de serviços terminológicos que não atendem aos critérios de organização hierárquica e relações associativas entre termos.

5.1.11 Documentos recuperados no *site* de dados abertos do governo do Brasil

O portal dados.gov.br é a principal plataforma de dados abertos do governo brasileiro, oferecendo acesso a uma vasta gama de conjuntos de dados de diversos órgãos e entidades públicas. Esses dados são organizados em categorias como saúde, educação, transporte, meio ambiente, economia e mais. A plataforma visa promover a transparência e a inovação, facilitando a busca e o uso dos dados para pesquisa, desenvolvimento de aplicativos e serviços, fornecendo APIs, tutoriais e guias para a publicação e gestão de dados. Além de disponibilizar dados para *download* e uso, o portal Dados.gov.br incentiva a participação cidadã e a colaboração entre governo, sociedade civil e setor privado. A página destaca iniciativas e eventos relacionados a dados abertos, apresentando casos de sucesso e exemplos práticos de como os dados estão sendo utilizados para criar valor e resolver problemas sociais. A plataforma também oferece recursos para desenvolvedores criarem visualizações e aplicações baseadas nos dados abertos disponibilizados.

Partindo-se para a consulta dos documentos recuperados no idioma português, têm-se os resultados dispostos na Tabela 13, a seguir.

Tabela 13 – Menções aos SOC – *site* de dados abertos do Brasil

TERMOS	RESULTADOS
anéis de sinônimos	0
arquivos de autoridade	0
cabeçalhos de assunto	0
dicionários	2618
diretórios	18
esquemas de categorização	0
esquemas de classificação	0
<i>gazetteers</i> (ou gazeteiros)	0
glossários	132
listas de seleção	0
ontologia	1
rede semântica	0
taxonomias	11
tesauros	0
Total	2780

Fonte: elaborada pela autora (2024).

Iniciando a análise pelos documentos recuperados que faziam menção ao termo Dicionários, observou-se que nenhum dos documentos analisados pode ser classificado como esse tipo de SOC. Todos os documentos fornecem conjuntos de dados técnicos e específicos, úteis para análises em suas respectivas áreas, mas não organizam palavras alfabeticamente nem fornecem informações detalhadas sobre formas, pronúncias, etimologias e usos idiomáticos.

Dos sete documentos analisados que faziam menção ao termo Diretório, três podem ser classificados como esse tipo de SOC, quais sejam "Grupos de Pesquisa", "Ciência Aberta é Vida" e "Representação Estudantil do IFSP". Os demais documentos são bases de dados ou ferramentas técnicas que não se enquadram na definição tradicional de diretórios.

Após a análise dos documentos recuperados que faziam menção ao termo Glossário, dos dez documentos analisados, três podem ser classificados como esse tipo de SOC "Glossário da Saúde Suplementar", "Glossário de campos dos relatórios de Pessoal" e "Glossário de Segurança da Informação". Esses documentos fornecem

listas estruturadas de termos e suas definições, servindo como ferramentas de padronização terminológica. Os outros sete documentos, apesar de fornecerem informações e dados detalhados, não incluem listas de termos com definições e, portanto, não podem ser classificados como glossários.

Em relação ao documento recuperado que fazia menção ao termo ontologia, observou-se que ele fornece dados detalhados sobre a execução orçamentária e financeira, incluindo despesas, receitas, empenhos, liquidações e pagamentos, em formatos como HTML, OWL e RDF, que facilitam a interoperabilidade. No entanto, embora utilize tecnologias compatíveis com *Linked Data*, o documento não aplica uma ontologia formal para organizar e representar o conhecimento de maneira estruturada e semântica, classificando-o mais como um conjunto de dados estruturado do que como uma ontologia.

Sobre os documentos recuperados que faziam menção ao termo Taxonomias, conclui-se que quatro deles podem ser classificados como esse tipo de SOC. Os documentos "Solos da Amazônia Legal", "Relevo do Estado de Tocantins", "Relevo do Brasil" e "Lista de Espécies da Flora do Brasil 2015 – *Brazilian Flora Checklist*" apresentam estruturas hierárquicas detalhadas para a classificação de solos, relevo e espécies de plantas, respectivamente. Esses documentos organizam termos de maneira sistemática, facilitando a categorização e o acesso eficiente às informações, atendendo plenamente ao conceito de taxonomia. Os demais documentos, como "Fiscalização – auto de infração", "Base de Dados das avaliações de risco de extinção da flora brasileira" e "Produção Intelectual" são predominantemente descritivos ou analíticos e não apresentam a estrutura hierárquica de classificação de termos necessária para serem considerados taxonomias.

5.2 Discussão dos resultados

A presente fase da pesquisa teve como objetivo identificar os SOCs adotados nos portais de dados abertos dos dez países com melhor desempenho no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU (EGDI), com o intuito de compreender em que medida esses países fazem uso de instrumentos semânticos – notadamente ontologias – na estruturação e disponibilização da informação pública. O pressuposto norteador dessa etapa da investigação considerava que a liderança desses países no referido índice estaria associada à adoção sistemática de

ontologias, compreendidas como ferramentas catalisadoras da interoperabilidade, integração de serviços e foco na experiência do cidadão.

A análise documental dos portais de dados abertos confirmou, de maneira geral, o pressuposto de que há uma presença significativa de SOC, em especial ontologias, nos países mais bem classificados no EGDI, como Estados Unidos, Canadá, Reino Unido e Austrália. Nesses países, constatou-se não apenas a menção recorrente a ontologias, mas também seu emprego em domínios específicos, como saúde, meio ambiente, transporte e serviços públicos, integrando modelos de dados com vocabulários padronizados, frequentemente publicados em formatos compatíveis com dados abertos vinculados (*Linked Data*). As ontologias nesses contextos são frequentemente utilizadas com vistas à interoperabilidade semântica entre sistemas, à integração de bases heterogêneas de dados e à melhoria da governança da informação.

Todavia, nem todos os países do grupo analisado demonstraram a mesma maturidade ou presença sistemática de SOC. Em Malta, por exemplo, não foram identificadas menções a qualquer tipo de SOC. Em Singapura, embora a plataforma seja estruturada e funcional, a ocorrência de SOC se limitou à presença de diretórios, com ausência de ontologias ou estruturas semanticamente robustas. Esses casos indicam que a alta posição no EGDI pode ser alcançada por meio de outros fatores – como infraestrutura tecnológica e serviços digitais responsivos – sem que, necessariamente, haja uma adoção expressiva de ontologias.

A investigação também evidenciou que, mesmo nos países em que ontologias são adotadas, a distribuição dos tipos de SOC é heterogênea. Diretórios, dicionários técnicos e glossários são mais comuns, possivelmente por demandarem menor esforço de formalização e especialização. Já os SOCs com maior expressividade semântica – como taxonomias, tesouros e ontologias – aparecem com maior frequência em países com tradição em políticas públicas de dados abertos e interoperabilidade, como Estados Unidos, Reino Unido e Canadá.

Importa destacar, ainda, que a análise foi deliberadamente não probabilística: dada a elevada quantidade de documentos retornados para cada termo pesquisado, limitou-se a análise aos dez primeiros documentos para cada tipo de SOC em cada país. Esse recorte metodológico visa à viabilidade da análise, mas impõe limites à generalização dos resultados, os quais devem ser interpretados como indícios e não como representações exaustivas do uso de SOC nos países estudados.

No caso do Brasil, observou-se uma presença tímida de SOC, com destaque para dicionários técnicos e glossários. A única ocorrência de menção ao termo "ontologia" não correspondeu a uma ontologia formal aplicada. Isso indica um cenário de baixa maturidade em termos de uso de instrumentos de representação do conhecimento, sobretudo no que tange às ontologias e demais SOCs com alto grau de formalismo e expressividade semântica.

À luz das evidências coletadas nos países mais bem posicionados no EGDI, torna-se evidente que a adoção de SOCs mais robustos – especialmente ontologias – está associada a práticas mais maduras de gestão da informação pública e interoperabilidade de sistemas. Nesse sentido, o Brasil, apesar de apresentar avanços pontuais, ainda revela uma adoção tímida de SOC com maior expressividade semântica, o que limita o potencial de integração e reutilização de dados governamentais.

Considerando esse cenário, é recomendável que o país invista na construção de uma estratégia nacional voltada à institucionalização do uso de ontologias na administração pública, especialmente em domínios prioritários como saúde, educação, segurança e meio ambiente. Essa estratégia deve contemplar diretrizes para o desenvolvimento, a documentação e a publicação de ontologias, preferencialmente alinhadas a padrões internacionais como OWL, SKOS e RDF, promovendo assim a interoperabilidade semântica entre sistemas de informação governamentais.

Além disso, a implementação de programas de capacitação voltados a servidores públicos e profissionais da informação é imprescindível para consolidar uma cultura institucional de governança semântica. A formação continuada em modelagem ontológica, dados abertos e vocabulários controlados pode potencializar a apropriação dessas ferramentas e favorecer a articulação de esforços interinstitucionais.

Por fim, recomenda-se o fortalecimento de repositórios públicos de vocabulários e ontologias, mantidos por instituições oficiais, como bibliotecas nacionais, ministérios e universidades públicas. Tais repositórios devem servir como referência para iniciativas locais e regionais, promovendo a padronização terminológica e incentivando o reuso de modelos ontológicos já consolidados. Ao seguir essas diretrizes, o Brasil poderá alinhar-se às melhores práticas internacionais

e avançar significativamente na promoção de um ecossistema informacional mais integrado, transparente e orientado ao cidadão.

5.3 Ontologias desenvolvidas no contexto do governo brasileiro

Esta seção apresenta uma análise das ontologias desenvolvidas no âmbito do governo brasileiro, com foco na identificação de iniciativas voltadas à organização e representação do conhecimento em plataformas públicas. Partindo de um levantamento sistemático em portais governamentais e repositórios acadêmicos, busca-se mapear as práticas existentes, apontar lacunas e avaliar o uso desses SOCs no fortalecimento da eficiência administrativa, da transparência e da acessibilidade no governo eletrônico. A análise documental realizada fundamenta-se na importância estratégica das ontologias para a interoperabilidade e a integração de dados, oferecendo subsídios para reflexões críticas sobre os desafios e as oportunidades relacionadas à aplicação dessas estruturas no setor público brasileiro. A partir dos resultados obtidos, esta seção visa contribuir com propostas que fomentem a padronização e a ampliação do uso de ontologias, promovendo avanços significativos na gestão do conhecimento na administração pública.

5.3.1 Apresentação dos resultados

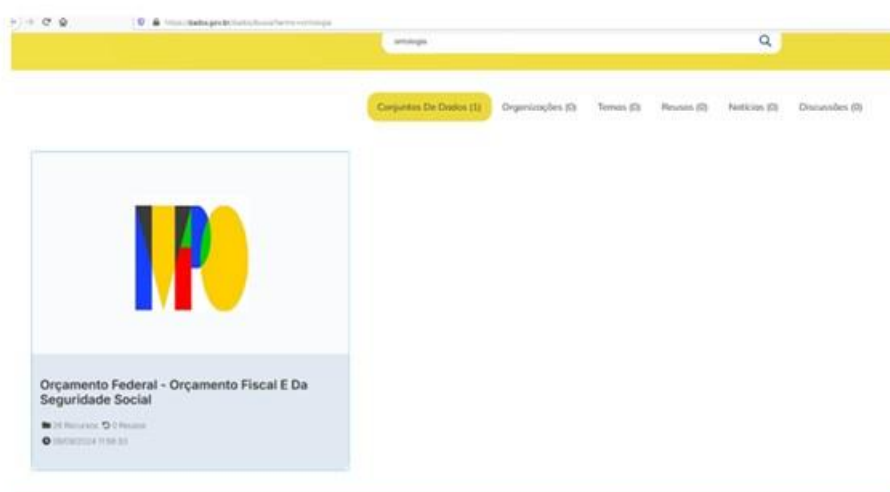
Nesta seção, têm-se os resultados da coleta de dados realizada nas fontes de pesquisa apresentadas. A coleta foi feita manualmente nos dez *sites* priorizados nos procedimentos metodológicos e ocorreu no mês de outubro de 2024.

5.3.1.1 Portal de dados abertos do governo brasileiro

O portal de dados abertos do governo brasileiro centraliza informações e *datasets* disponibilizados por órgãos federais, promovendo a transparência e o acesso público aos dados governamentais. Esse portal é essencial para pesquisadores e cidadãos interessados em explorar dados sobre orçamentos, políticas públicas e atividades administrativas, fornecendo um acervo estruturado, que favorece o uso em projetos de ciência de dados e análise governamental.

Adotando-se os termos de busca selecionados, não foram encontrados resultados para nenhuma das expressões de busca consultadas. Buscando-se aplicar pesquisa complementar apenas com o termo “ontologia”, foi identificado um conjunto de arquivos denominado “Orçamento Federal – Orçamento Fiscal e da Seguridade Social” que contém dados da execução orçamentária e financeira da despesa orçamentária de órgãos da União para os exercícios financeiros desde 2000 até o ano corrente, conforme informações dispostas na página do recurso.

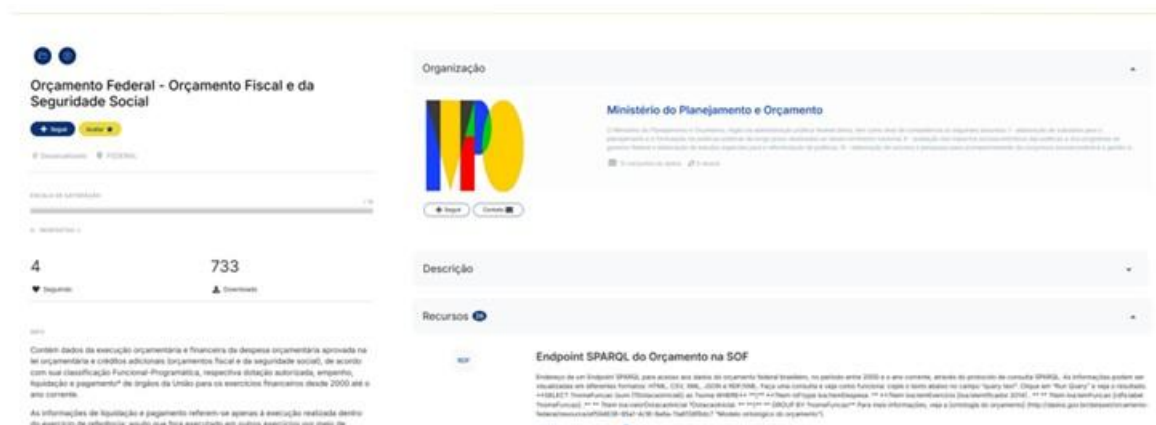
Figura 2 – Página inicial do documento Orçamento Federal – Orçamento Fiscal e da Seguridade Social



Fonte: elaborada pela autora (2024).

Em um dos diretórios disponíveis, foram identificados arquivos do Orçamento anual, desde 2014, em formato RDF. Além dos arquivos RDF, foram identificados alguns metadados, como “Conceito”, que traz a definição do Orçamento Federal; “Legislação”, que lista a legislação aplicável ao tema, incluindo portarias, decretos e leis; e “Veja Também”, que traz o *link* de uma ferramenta para consulta pública aos orçamentos fiscal e de seguridade social da União, chamado Painel do Orçamento, mas o *link* se encontrava inativo no momento da consulta.

Figura 3 – Página de detalhamento do documento Orçamento Federal – Orçamento Fiscal e da Seguridade Social



Fonte: elaborada pela autora (2024).

Na página do recurso, também é possível identificar o tipo de licença adotado, os formatos dos arquivos, periodicidade de atualização, data da última alteração e algumas *tags*.

Ao final da página, é possível identificar uma aba com “Ações”, que permite ao usuário sugerir *tags*, reportar problemas e sugerir melhoria, além de um “Conjunto de dados similares”, que prevê dois arquivos, sendo um de “Acompanhamento Físico-Financeiro das Ações Orçamentárias” e outro de “Emendas Parlamentares Individuais de Execução Obrigatória”, ambos vinculados à categoria “Orçamento Federal”.

5.3.1.2 Portal de serviços do governo brasileiro – GovBR

O GovBR é a principal plataforma de serviços digitais do governo brasileiro, na qual os cidadãos podem acessar informações sobre serviços públicos e realizar consultas ou solicitações online. A plataforma agrega um vasto acervo de dados, integrando diferentes serviços públicos que podem ser acionados pela própria plataforma.

Identificou-se que a funcionalidade de busca da plataforma GovBR não considera a busca pelo termo composto absoluto, mesmo utilizando-se aspas duplas. Foram recuperados 408 registros, porém não foi possível confirmar quais parâmetros foram considerados na recuperação. Analisando-se os documentos recuperados, identificou-se que nenhum deles se tratava de arquivos de ontologias e sim que faziam menção ao termo ontologia, seja em notícias, seja em títulos de contratações de

curso, registros de pagamentos de disciplinas, entre documentos de outras naturezas.

Figura 4 – Tela de busca da plataforma GovBR

The screenshot shows the GovBR search interface. At the top, there is a navigation bar with the GovBR logo, links for 'Órgãos do Governo', 'Acesso à Informação', 'Legislação', and 'Acessibilidade', along with a language dropdown set to 'PT' and a user profile 'Olá, PATRICIA'. Below the navigation bar, there is a search bar containing the query "'ontologia" AND "governo brasileiro'". The search results page displays the title "'ontologia" AND "governo brasileiro'" and indicates 408 results. There are tabs for 'Notícias (1)' and 'Todos (408)'. A filter icon is visible. The main content area features a news item titled 'Seminário internacional sobre "A Representação do Conhecimento e as Novas Formas de Comunicação Científica: Interlocuções Necessárias" é promovido pelo grupo BRIET do IbiCT' with a date of 29/05/2023 and a brief description of the event.

Fonte: elaborada pela autora (2024).

Buscando-se realizar a pesquisa complementar com o termo “ontologia”, foram recuperados um total de 1610 registros, sendo que 1 (um) foi classificado na categoria “Serviço”, 57 na categoria “Notícias” e 1610 na categoria “Todos”.

Figura 5 – Tela de pesquisa complementar na plataforma GovBR

The screenshot shows the GovBR search interface for the query "ontologia". The search bar contains the query "ontologia". The search results page displays the title "'ontologia'" and indicates 1610 results. There are tabs for 'Serviços (1)', 'Notícias (57)', and 'Todos (1610)'. The main content area features a service item titled 'Consultar o Portal da Ontologia Brasileira de Medicamentos (Portal OBM)' with a sub-header 'Conhecido por: "Ontologia Brasileira de Medicamentos"'. Below the title, there is a description of the portal and a blue button labeled 'Iniciar'. There are also links for 'Etapas para a realização deste serviço' and 'Quem pode utilizar este serviço?'. A section for 'Conteúdos Relacionados' is partially visible at the bottom.

Fonte: elaborada pela autora (2024).

Primeiramente, analisando-se o arquivo “Portal da Ontologia Brasileira de Medicamentos (Portal OBM)”, recuperado na aba serviços, são identificados alguns metadados comuns a outros serviços da plataforma GovBR: “O que é”, que traz uma apresentação do serviço, não necessariamente o conceito; “Quem pode utilizar esse serviço”, que indica o público-alvo do serviço; “Etapas para a realização deste serviço”, que traz instruções sobre como acessar o serviço, incluindo o *link* de acesso; “Outras informações”, que apresenta informações adicionais sobre o serviço, como tempo estimado, canais para dúvidas, condições para tratamento prioritário, entre outras; e, por fim, é prevista uma aba da “Lei Geral de Proteção de Dados Pessoais – LGPD”, onde constam as informações quanto ao tratamento de dados realizado pelo serviço. Além dos metadados, a página principal do serviço também indica os serviços relacionados, e possui um botão “Iniciar”, por meio do qual o usuário acessa o serviço diretamente.

Ao acessar o serviço, pode ser identificada uma lista de 430 medicamentos, em formato de tabela, contendo “Identificador do AMPP”, “Nome do AMPP”, “Código EAN” e “Registro Sanitário”. A lista de medicamentos não possibilita acesso a cada item, sendo possível apenas navegar pelas páginas ou então utilizar o mecanismo de busca (Pesquisar).

Partindo-se para a análise dos arquivos recuperados nas abas “Notícias” e “Todos”, novamente o termo “ontologia” foi identificado em uma variedade de documentos, como ementas de cursos, editais, programas de provas de concursos públicos, títulos de livros, trabalhos de conclusão de cursos, entre outros. Destaca-se que os documentos não possuem quaisquer formas de categorização, contendo apenas o nome dado ao arquivo, alguns com a extensão, em sua maioria, arquivos em formato pdf.

5.3.1.3 Portal da transparência

Este portal promove a transparência administrativa, oferecendo dados detalhados sobre despesas, contratações, viagens e demais atividades financeiras do governo. O acesso a esses dados é fundamental para a análise da aplicação de ontologias na área da prestação de contas e auditoria governamental, proporcionando uma interface de pesquisa que permite filtrar e explorar informações com alto nível de granularidade.

Em consulta ao Portal da Transparência, novamente nenhum dos termos de busca priorizados foi suficiente para recuperação de quaisquer registros. Ampliando-se a busca com o uso do termo “ontologia”, foi possível a recuperação de 301 registros, englobando informações sobre pauta de reuniões de serviço de servidores públicos, contratações de cursos e eventos sobre o tema, aquisição de livros, pagamento de bolsas de pesquisa, entre outros documentos de prestação de contas que citam o termo ontologia.

Figura 6 – Página de pesquisa do Portal da Transparência

The screenshot shows the search results page for the term "ontologia" on the Portal da Transparência. The page header includes navigation links and a search bar. The main content area displays a search bar with "ontologia" entered and a search icon. Below the search bar, it states "Aproximadamente 301 resultados encontrados para ontologia". The results are listed in a table-like format, showing details for each record, including the title, the organization (Órgão), and the object (Objeto). A sidebar on the right titled "FILTROS APLICADOS" offers various categories to refine the search results, such as Despesas, Documentos, Viagens, Receitas públicas, Servidores, Imóveis Funcionais, Sanções, Benefícios, Órgãos / entidades, and Pessoas físicas e jurídicas.

Resultado da busca	FILTROS APLICADOS
<p>Viagem a serviço de DALTON LOPES MARTINS com destino a Brasília/DF</p> <p>Motivo: Como principais temas, a reunião tratará de: Definir o escopo da ontologia. 1 - Qual o domínio a ontologia Motivo: temas deve cobrir e que funções na gestão e operação cultural deve cobrir 2 - Para qual vamos usar a ontologia Motivo: Construir casos objetivos e exemplos de formas de uso da ontologia. Motivo: Ex: gerar relatórios, integrar com sistemas, produzir análises, etc. 3 - Que tipos de questões a ontologia Motivo: Construir perguntas objetivas que a ontologia deveria responder</p>	<p>Utilize as categorias abaixo para refinar o resultado da busca</p> <p><input type="checkbox"/> Despesas</p> <p><input type="checkbox"/> Documentos</p> <p><input type="checkbox"/> Viagens</p> <p><input type="checkbox"/> Receitas públicas</p> <p><input type="checkbox"/> Servidores</p> <p><input type="checkbox"/> Imóveis Funcionais</p> <p><input type="checkbox"/> Sanções</p> <p><input type="checkbox"/> Benefícios</p> <p><input type="checkbox"/> Órgãos / entidades</p> <p><input type="checkbox"/> Pessoas físicas e jurídicas</p>
<p>Licitação 00022/2018 - Órgão: Universidade Federal de Minas Gerais</p> <p>Objeto: Objeto: Inscrição em Seminário de Pesquisa em Ontologia no Brasil - Oncobras</p>	
<p>Licitação 00036/2018 - Órgão: Instituto Federal do Espírito Santo</p> <p>Objeto: Objeto: Pagamento de inscrição de servidor em evento denominado "Ontobras - Seminário de Pesquisa em Ontologias</p>	
<p>Documento Pagamento 2020OB800030</p> <p>Observação do documento: DESPESAS RELATIVAS A PARTICIPACAO E APRESENTACAO DE TRABALHO NO WORKSHOP DE TESES E DISSERTACOES EM ONTOLOGIAS</p>	
<p>Viagem a serviço de VICTOR XIMENES MARQUES com destino a Rio de Janeiro/RJ</p> <p>Motivo: Encontro do GT Ontologias Contemporâneas</p>	

Fonte: elaborada pela autora (2024).

O Portal da Transparência possui um conjunto de categorias que permitem ao usuário filtrar os resultados recuperados. Cada um dos registros disponíveis conta com um conjunto de atributos específicos. Por exemplo, em um registro de viagem de um servidor público, constam informações como nome do servidor, período da viagem, motivação, valor total da viagem, itinerário, entre outros.

Já no registro de uma aquisição por meio de licitação, constam atributos como modalidade da licitação, quantidade de itens licitados, valor da licitação, objeto, entre outros.

Atributos específicos também podem ser visualizados em um registro denominado “Empenho”, instrumento utilizado pela administração pública para formalizar o pagamento de alguma aquisição. Na Figura 7, detalha-se uma ajuda de custo a um estudante para participação em um evento que engloba o tema ontologia.

Figura 7 – Detalhamento de um documento de empenho do Portal da Transparência

The screenshot shows the 'Portal da Transparência' website interface. At the top, there is a search bar and navigation links. The main heading is 'Portal da Transparência' with the subtitle 'CONTROLADORIA-GERAL DA UNIÃO'. Below this, there is a breadcrumb trail: 'VOCÊ ESTÁ AQUI: INÍCIO > DESPESAS PÚBLICAS > CONSULTA > DOCUMENTOS DE EXECUÇÃO DA DESPESA PÚBLICA > DOCUMENTO'. The title of the page is 'Detalhamento do documento de Empenho' with a button for 'ORIGEM DOS DADOS'.

Nº do documento 2020NE000014	Última atualização 10/11/2020	Descrição NOTA DE EMPENHO (NE)
Fase EMPENHO	Espécie/tipo de documento ORIGINAL	Valor atual do documento R\$ 80,00

Observação do documento
PAGAMENTO DE AUXILIO FINANCEIRO AO ESTUDANTE PARA CUSTEAR SUA PARTICIPACAO COMAPRESENTACAO DE TRABALHO NO WORKSHOP DE TESES E DISSERTACOES EM ONTOLOGIAS NA ONTOBRAS 2020.

DADOS DO FAVORECIDO

CPF/CNPJ/Outros *** 532.656.**	Nome RAFAEL ROCHA
--	-----------------------------

DADOS DO ÓRGÃO EMITENTE

Órgão Superior 26000 MINISTÉRIO DA EDUCAÇÃO	Órgão / Entidade Vinculada 26238 UNIVERSIDADE FEDERAL DE MINAS GERAIS	Unidade Gestora 153277 ESCOLA DE CIENCIA DA INFORMACAO/UFGM	Gestão 15229 UNIVERSIDADE FEDERAL DE MINAS GERAIS
--	--	--	--

Fonte: elaborada pela autora (2024).

5.3.1.4 Biblioteca Digital do Senado

A Biblioteca Digital do Senado Federal disponibiliza publicações, legislações, estudos e teses de interesse público e parlamentar. Esse repositório é útil para pesquisas acadêmicas e documentais, apresentando um sistema de metadados

detalhado que segue o padrão *Dublin Core*, o que facilita a categorização e a busca por documentos relevantes ao estudo de ontologias na administração pública.

A pesquisa pelos termos de busca na Biblioteca Digital do Senado também não retornou resultados satisfatórios, visto que a funcionalidade de busca não considera o termo composto absoluto, recuperando registros com o termo “ontologia”, por exemplo.

Figura 8 – Resultado de busca na Biblioteca Digital do Senado

Escolha a Coleção: Toda a Biblioteca Digital

"ontologia" AND "governo brasileiro"

Filtros
Utilize filtros para refinar o resultado de busca.

Filtros correntes:
Título Diferentes Artigo de jornal

Novo Filtro:
Assunto Contém
Data de publicação Contém

Aplicar

Apresentando 10 de um total de 56 resultados. (0,213 segundos)

1 2 3 4 ... 6 Próxima página

Impacto fiscal de proposições legislativas : unindo LexML e uma web semântica fiscal
Silva, Fábio Vinicius Pinto e
Senado Federal, Instituto Legislativo Brasileiro, 2017.
O trabalho analisa meios para que proposições legislativas, suas estimativas de impacto fiscal e correspondentes medidas de compensação, sejam expressas em formato de dados abertos através da definição de uma **ontologia** orçamentária-financeira...

Revista de informação legislativa : v. 36, n. 143 (jul./set. 1999)
Senado Federal, Subsecretaria de Edições Técnicas, 07/1999.
nacionalidade no Direito Grego 309 A filosofia do controle concentrado de constitucionalidade das leis na ordem jurídica brasileira pós-88 313 Usucapião: doutrina e jurisprudência 321 **Ontologia** jurídica e realidade – o problema da “ética da tolerância” 335 Da...

Fonte: elaborada pela autora (2024).

O número de registros recuperados variou conforme o termo de busca utilizado, porém a recuperação manteve as características, ou seja, trazendo registros diferentes daqueles esperados na pesquisa realizada.

Buscando-se então a pesquisa complementar utilizando apenas o termo “ontologia”, houve a recuperação de 152 documentos. Entre eles, podem ser identificados artigos científicos, trabalhos de conclusão de curso, anais de eventos, livros, dicionários, periódicos, revistas técnicas etc.

Observou-se que todos os documentos possuem um conjunto detalhado de metadados, além de notas explicativas, os arquivos para *download*, tipo de licença adotada, bem como coleção à qual o documento possa estar vinculado. Além disso, identificou-se também que todos os documentos são catalogados por meio do padrão *Dublin Core*.

Figura 9 – Detalhamento dos metadados no padrão *Dublin Core*

Mostrar registro simples

dc.contributor.author	Maia, Alexandre da	
dc.date.accessioned	2007-07-06T20:53:56Z	pt_BR
dc.date.accessioned	2013-05-15T14:33:10Z	
dc.date.issued	1999-07	
dc.identifier.citation	Revista de informação legislativa, v. 36, n. 143, p. 335-345, jul./set. 1999	pt_BR
dc.identifier.uri	http://www2.senado.leg.br/bdsf/handle/id/525	
dc.description	Inclui notas explicativas e bibliográficas.	pt_BR
dc.description.statemntofresponsibility	Alexandre da Maia	pt_BR
dc.relation.ispartof	Revista de informação legislativa	
dc.rights	Attribution-NonCommercial-ShareAlike 3.0 Brazil	
dc.rights.uri	https://creativecommons.org/licenses/by-nc-sa/3.0/br/	
dc.subject.other	Filosofia do direito	pt_BR
dc.subject.other	Ética forense	pt_BR
dc.subject.other	Ontologia jurídica	pt_BR
dc.title	Ontologia jurídica e realidade : o problema da ética da tolerância	pt_BR
dc.type	Artigo de revista	pt_BR
dc.type	Texto	pt_BR
local.citation.volume	36	
local.citation.issue	143	
local.citation.spage	335	
local.citation.epage	345	
local.relation.ispartof	id/496875	

Fonte: elaborada pela autora (2024).

5.3.1.5 Biblioteca do Instituto Brasileiro de Geografia e Estatística (IBGE)

A Biblioteca do IBGE reúne publicações estatísticas, relatórios técnicos e documentos sobre geografia e estatísticas nacionais. Esse repositório é relevante para o levantamento de dados que podem ser organizados por ontologias voltadas para a representação de informações demográficas e econômicas, fornecendo uma base rica para a modelagem de conhecimento no contexto governamental.

A pesquisa realizada na Biblioteca do IBGE também não retornou registros de ontologias a partir dos termos de busca priorizados. Partindo-se para a pesquisa complementar utilizando apenas o termo “ontologia”, houve a recuperação de apenas dois arquivos.

Figura 10 – Resultado de busca na página da biblioteca do IBGE

Busca rápida Busca combinada

A **Busca Rápida** permite a localização de materiais de maneira mais simples, com duas possibilidades de combinação por tipo de material e por apenas um campo específico.

Para realizar a busca, selecione o tipo de material na aba correspondente. Em seguida, selecione o campo desejado (as opções variam de acordo com o tipo de material selecionado), digite a palavra-chave (termo que se deseja localizar) e execute a busca. Selecione "buscar frase exata" para refinar os resultados da busca.

TIPO DE MATERIAL
 Todos os materiais

CAMPO
 Todos os campos

SUA BUSCA
 ontologia

apenas itens digitalizados buscar frase exata

Buscar

[Clique aqui para pesquisar o acervo físico de todas as bibliotecas do IBGE no Brasil](#)

Sua busca retornou **2** resultados:

Tipo	Título/Estado	Autor/Município	Ano	ID	Download
Publicação	Ter ou ser?	Fromm, Erich, 1900-1980	1980	6817	
Publicação	The Courage to be	Tillich, Paul, 1888-1965	1952	159	

Fonte: elaborada pela autora (2024).

Observou-se que há alguns metadados relacionados a cada registro recuperado, sendo ambos do tipo “Publicação”, porém sem acesso direto ao arquivo.

Figura 11 – Detalhamento de documento recuperado na Biblioteca do IBGE

gob.br

ACESSO À INFORMAÇÃO PARTICIPE LEGISLAÇÃO ÓRGÃOS DO GOVERNO

Biblioteca IBGE

Catálogo Sobre a Biblioteca Atendimento Serviços Links

IBGE

Id: 6817
Autor: Fromm, Erich, 1900-1980
Título: Ter ou ser?
Local: Rio de Janeiro
Editor: Zahar Ed.
Ano: 1980
Descrição física: 202p.
Notas: Bibliografia: p. 196-202.
Assuntos: Humanismo; Ontologia; Personalidade

Download
 Download não disponível.

© 2024 IBGE - Instituto Brasileiro de Geografia e Estatística

IBGE

Fonte: elaborada pela autora (2024).

5.3.1.6 Repositório Institucional da ENAP

O Repositório Institucional da Escola Nacional de Administração Pública (ENAP) armazena artigos, dissertações e materiais técnicos voltados à gestão pública e à capacitação de servidores. Esse repositório oferece documentos essenciais para compreender as práticas e diretrizes que orientam a administração pública, sendo uma fonte relevante para estudos que buscam mapear o desenvolvimento e a aplicação de ontologias neste setor.

A pesquisa no Repositório da Enap novamente não recuperou arquivos de ontologias. Para a maioria dos termos de busca priorizados, não houve recuperação de nenhum arquivo. Para dois dos termos de busca (“ontologia” AND “governo brasileiro” e “ontologias” AND “administração pública brasileira”), houve a recuperação de alguns documentos, entre eles apostilas de cursos, documentos técnicos, editais de concursos públicos, artigos de revistas técnicas, trabalhos de conclusão de curso. Porém não foram identificados documentos que contemplassem propostas de ontologias no contexto da administração pública.

Figura 12 – Resultado de busca no Repositório Institucional da ENAP

The screenshot displays the search interface of the ENAP Institutional Repository. At the top, the search bar contains the query "ontologia" AND "governo brasileiro". Below the search bar, there are filters for "Adicionar filtros" and options to refine the search. The results are displayed in a table with columns for "Pre-visualização", "Data do documento", "Título", and "Autor(es)".

Pre-visualização	Data do documento	Título	Autor(es)
	2016	Gestão de políticas públicas de direitos humanos: costâneas	Delegado Ana Luiza de Mendez (Org.); Rec. Maria Stela (Org.); Alves, Pedro Assumpção (Org.)
	2018-07	Revista do Serviço Público (RSP), vol. 69, n. 3	Escola Nacional de Administração Pública (Brasil)
	2013	Uma proposta de aperfeiçoamento do planejamento público a partir do paradigma do Governo Aberto - como superar as lacunas do	Rec. Sérgio Roberto Cuedes

On the right side of the page, there is a sidebar titled "Explore o Repositório" which lists authors and titles. The authors listed include: Escola Nacional de Administração..., Alves, Pedro Assumpção (Org.), Carvalho, Marique Demis de, Delegado Ana Luiza de Mendez (Org.), Duernas, Rogério Allan, Elias de Oliveira, Vanessa (Org...), Escola de Administração Fazendas, Espejo, Marcia Maria dos Santos B., Fogaça, Vanessa de Souza (Conteúdo...), Fogaça, Vanessa de Souza (Conteúdo...), and Título: Governança de Dados na Transforma....

Fonte: elaborada pela autora (2024).

Todos os documentos disponíveis no repositório contam com uma variedade de metadados, identificador persistente, acesso direto ao arquivo, além de estatísticas de acesso ao documento. Os documentos são catalogados conforme o padrão *Dublin Core*, sendo possível detalhar os metadados em uma segunda tela.

Figura 13 – Detalhamento de documento recuperado no Repositório Institucional da ENAP

The screenshot shows the ENAP Institutional Repository interface. At the top, there is a navigation bar with the ENAP logo and links for 'INÍCIO', 'COMUNIDADES', 'NAVEGAR', and 'SOBRE A BIBLIOTECA DIGITAL'. Below the navigation bar, the breadcrumb trail reads: 'Cursos da Enap > Escola Virtual > Escola Virtual: Governo e Transformação Digital > Módulo 3: a arquitetura ePING'. The main heading is 'Módulo 3: a arquitetura ePING'. Below the heading, there is a social media sharing section with icons for WhatsApp, Facebook, Twitter, and LinkedIn. A button labeled 'Visualizar arquivos' is visible. The document details are presented in a table-like format:

Título:	Módulo 3: a arquitetura ePING
Autor(es):	Mello, Ana Paula Pessoa (Conteudista, 2015) Mesquita, Hudson (Conteudista, 2015) Vieira, Carlos Eduardo (Conteudista, 2016)
Editor:	Escola Nacional de Administração Pública (Enap)
Endereço Eletrônico:	http://www.enap.gov.br/web/pt-br/sobre-curso?p_p_id=enapvisualizardetalhescurso_WAR_enapinformacoescursoportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_564233524_0Curso=2706
Idioma:	Português
País:	Brasil
Tipo:	Módulo Completo

Fonte: elaborada pela autora (2024).

Mesmo realizando as buscas com os termos compostos entre aspas duplas, a pesquisa retornou registros que apenas faziam menção ao termo ontologia, conforme identificado no exemplo a seguir, que se trata da única página que faz menção ao termo ontologia no documento recuperado pela pesquisa.

Figura 14 – Recorte de documento recuperado no Repositório Institucional da ENAP

The screenshot shows a snippet of a document with the ENAP logo on the left. The text on the right is as follows:

governo eletrônico. São abordadas as políticas e as especificações para estações de trabalho, televisão digital e mobilidade.

Organização e Intercâmbio de Informações

Aborda os aspectos relativos ao tratamento e à transferência de informações nos serviços de governo eletrônico. Inclui padrão de vocabulários controlados, taxonomias, **ontologias** e outros métodos de organização e recuperação de informações.

Áreas de Integração para Governo Eletrônico

Fonte: elaborada pela autora (2024).

5.3.1.7 Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)

A BDTD reúne teses e dissertações de diversas áreas de conhecimento, oferecendo uma visão abrangente da pesquisa acadêmica no Brasil. Esse acervo é essencial para identificar estudos que exploram ontologias no contexto do governo brasileiro, além de possibilitar a análise de abordagens teóricas e metodológicas para a construção e aplicação de ontologias na organização do conhecimento.

A pesquisa com o termo de busca “ontologia” AND "governo brasileiro" retornou um total de 13 trabalhos acadêmicos. Boa parte dos trabalhos aborda o tema ontologia de maneira subjacente, não sendo o tema central dos trabalhos recuperados. Além disso, nenhum dos trabalhos recuperados contempla uma ontologia, mesmo que em nível de proposta.

Figura 15 – Primeiro resultado de busca na Biblioteca de Teses e Dissertações

A imagem mostra a interface de busca da BDTD. No topo, há o logo da BDTD e links para Institucional, Rede, Faq e Contato. O campo de busca contém o texto "ontologia" AND "governo brasileiro". Abaixo do campo de busca, há uma seção de resultados da busca. O primeiro resultado é "Ontologia e avaliação da educação básica no Brasil (1990-2007)" por Zanardini, João Batista, publicado em 2008. O segundo resultado é "Redes sociotécnicas, práticas de conhecimento e ontologias na Amazônia: tradução de saberes no campo da biodiversidade" por Silveira, Diego Soares da, publicado em 2011. A interface também inclui opções de refinamento de busca e botões para acessar documentos e exportar.

Fonte: elaborada pela autora (2024).

Adotando-se o termo “ontologias” AND "administração pública brasileira", houve a recuperação de quatro documentos, sendo três teses e uma dissertação. Novamente, os documentos recuperados contemplam o termo ontologia como assunto subjacente, sendo que alguns estudos abordam a ontologia do ponto de vista filosófico e não como artefato de organização do conhecimento e da informação.

Figura 16 – Documento recuperado na Biblioteca de Teses e Dissertações

The screenshot displays the BDTD website interface. At the top, there is a navigation menu with links for 'Institucional', 'Rede', 'Faq', and 'Contato'. Below the navigation is a search bar containing the query "ontologias" AND "administração pública brasileira". To the right of the search bar are options for "Todos os campos" and a "Buscar" button, along with a link to "Busca avançada".

Below the search bar, the page indicates "Resultados da busca: 'ontologias' AND 'administração pública brasileira'". There is a section for "Buscas alternativas" with a link to "publica brasileira + publica brasil (Expandir a busca)".

The search results section shows "Mostrando 1 - 4 resultados de 4 para a busca 'ontologias' AND 'administração pública brasileira', tempo de busca: 0.27s". There are options to "Ordenar" by "Relevância" and a button to "Exportar".

On the left side, there is a "Refinar a Busca" section with several filters: "Instituição de defesa", "Bases coletadas", "Programa de Pós-Graduação", "Autor", "Orientador(a)", "Tipo de documento", and "Tipo de acesso".

The search results are listed as follows:

- Assuntos: "Ontologias"**
Um modelo para implementação de aplicações da Argument Web integradas com bases de dados abertos e ligados
 Por [Niche, Roberto](#)
 Publicado em 2015
[Acessar documento](#)
 Dissertação
- A hibridiz material das Cortes de Contas como atributo determinante de sua organicidade e a metamorfose institucional dos Tribunais da Governança Pública**
 Por [Ferreira Junior, Adilson de Moraes](#)
 Publicado em 2021
[Acessar documento](#)

Fonte: elaborada pelas autoras (2024).

Na pesquisa com o termo "ontologia" AND "gestão pública" AND Brasil, foram recuperados 12 trabalhos acadêmicos. Identificou-se que dois deles, apesar de não apresentarem ontologias enquanto artefatos, apresentam propostas de uso desse SOC no contexto da administração pública: a dissertação “Validação de dados em sistemas de data *warehouse* através de índice de similaridade no processo de ETL e mapeamento de trilhas de auditoria utilizando indexação ontológica” e a tese “Padrão de projeto de ontologias para inclusão de referências do novo serviço público em plataformas de governo aberto”.

Figura 17 – Segundo resultado de busca na Biblioteca de Teses e Dissertações

BDTD

Institucional Rede Faq Contato

"ontologia" AND "gestão pública" AND Brasil

Todos os campos

Buscar

Busca avançada

Resultados da busca: "ontologia" AND "gestão pública" AND Brasil

Mostrando 1 - 12 resultados de 12 para a busca "ontologia" AND "gestão pública" AND Brasil, tempo de busca: 0,20s

Ordenar: Relevância

Exportar

Refinar a Busca

Instituição de defesa

Bases coletadas

Programa de Pós-Graduação

Autor

Orientador(a)

Tipo de documento

Tipo de acesso

openAccess 12

1 | Assuntos: "...Brasil", "Fome..."

O processo de **gestão pública** da miséria

Por [Barbosa, Maria Geralda \(UNESP\)](#)
Publicado em 2007

[Acessar documento](#)

Tese

2 | Assuntos: "...Ontologia..."

Validação de dados em sistemas de data warehouse através de índice de similaridade no processo de ETL e mapeamento de trilhas de auditoria utilizando indexação **ontológica**

Por [Campos, Sander Rodrigues](#)
Publicado em 2013

[Acessar documento](#)

Dissertação

Fonte: elaborada pelas autoras (2024).

Por fim, a pesquisa com o termo "desenvolvimento de ontologias" AND "administração pública" AND Brasil retornou 1 (um) registro que poderá ser analisado em um segundo momento, visto que o resumo apresenta a proposta de desenvolvimento de uma ontologia para um domínio específico da administração pública.

Figura 18 – Terceiro resultado de busca na Biblioteca de Teses e Dissertações

The screenshot displays the BDTD search interface. At the top, there are navigation links: Institucional, Rede, Faq, and Contato. The search bar contains the query: "desenvolvimento de ontologias" AND "administração pública" AND Brasil. Below the search bar, the results section shows "Resultados da busca: 'desenvolvimento de ontologias' AND 'administração pública' AND Brasil". It lists "Buscas alternativas" and "Mostrando 1 - 1 resultados de 1 para a busca 'desenvolvimento de ontologias' AND 'administração pública' AND Brasil". The search time is 0,25s. On the left, there is a "Refinar a Busca" sidebar with filters for "Instituição de defesa", "Bases coletadas", "Programa de Pós-Graduação", "Autor", "Orientador(a)", and "Tipo de documento". The main result is a document titled "PSO: uma ontologia de domínio sobre servidores públicos federais" by GONCALVES, Victor, published in 2017. The document is available as a "Dissertação".

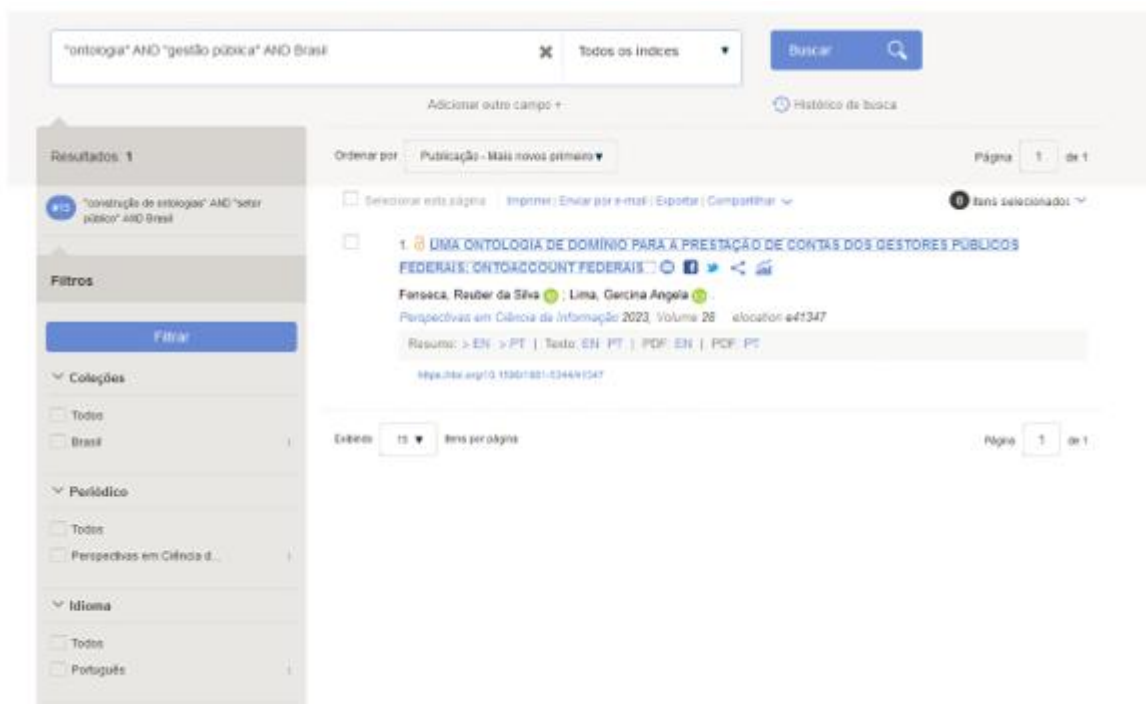
Fonte: elaborada pela autora (2024).

5.3.1.8 SciELO – *Scientific Electronic Library Online*

SciELO é uma biblioteca digital de acesso aberto que publica periódicos acadêmicos brasileiros e latino-americanos. A plataforma é uma fonte importante de estudos científicos que abordam temáticas diversas.

Entre os termos de busca priorizados, houve recuperação apenas por meio do termo "ontologia" AND "gestão pública" AND Brasil de um estudo intitulado "Uma Ontologia de Domínio para a Prestação de Contas Dos Gestores Públicos Federais: *Ontoaccount*", que apresenta uma proposta de ontologia, desenvolvida no *software Protégé*. No caso do Scielo, a busca complementar apenas com o termo "ontologia" não foi realizada, partindo-se do pressuposto de que os documentos recuperados não abordariam o tema no contexto da administração pública, foco desta etapa da pesquisa.

Figura 19 – Resultado de busca no SciELO



Fonte: elaborada pela autora (2024).

5.3.1.9 Anais Seminário de Pesquisa em Ontologia no Brasil – Ontobras

Os Anais do Ontobras, seminário dedicado à pesquisa em ontologias no Brasil, oferecem acesso a estudos sobre a aplicação e o desenvolvimento de ontologias em diversos contextos, incluindo a administração pública. Esse repositório especializado é essencial para identificar tendências e desafios específicos do campo de ontologia no Brasil e no mundo.

A busca nos Anais dos Seminários Ontobras deu-se nos anos de 2021 a 2024, nas páginas dos eventos, procurando-se identificar algum estudo que relacionasse aplicação de ontologias na administração pública ou mesmo geração automática ou semiautomática de ontologias. Foi identificado apenas o trabalho “*Ontology Requirements Elicitation in Brazilian Public Administration: Methodologies and Challenges*” que, em tese, estaria dentro do contexto desta pesquisa.

5.3.1.10 Google Acadêmico

O *Google Scholar* é um mecanismo de busca acadêmica que indexa publicações científicas, incluindo artigos, teses, livros e conferências. É amplamente

utilizado para explorar a literatura acadêmica sobre diversos temas, incluindo ontologias no governo eletrônico, permitindo o acesso a uma grande variedade de estudos relevantes para a pesquisa em curso.

A última pesquisa deu-se no Google Acadêmico, buscando-se complementar as pesquisas anteriores. Foi necessário realizar uma adaptação dos termos de busca, visto que os termos selecionados inicialmente não recuperaram nenhum documento. Após adaptações, com o termo de busca "ontologia na administração pública", recuperou-se a tese "Modelagem de conhecimento baseada em ontologia e dados abertos: a representação do domínio da tomada e prestação de contas anual dos gestores públicos federais", trabalho este relacionado ao artigo recuperado na busca realizada no Scielo.

5.3.2 Discussão dos resultados

Em relação ao Portal de dados abertos do governo brasileiro, onde foi identificada uma possível ontologia do Orçamento Federal, avalia-se que o conjunto de metadados pode auxiliar o usuário a compreender melhor o recurso disponível, trazendo informações com o conceito de orçamento, legislação relacionada, tipo de licença adotada, formatos dos arquivos, *tags*, entre outros metadados. Entretanto, apesar de a busca ter sido realizada utilizando-se o termo "ontologia", não há nenhuma menção na página do recurso de que ele se trata de uma ontologia. Além disso, a página não dispõe de informações básicas de uma ontologia como, por exemplo, descrição, classes, relações, entidades e outros elementos. Ao tentar realizar o *download* do arquivo para uma verificação mais aprofundada por meio do *software Protégé*, um editor de ontologias, não foi possível descompactar o suposto arquivo da ontologia. Dessa forma, não houve condições técnicas para confirmar se os arquivos disponibilizados são de fato arquivos de ontologias do orçamento federal.

Passando para a análise dos documentos recuperados no Portal de serviços do governo brasileiro – GovBR, foi identificado o "Portal da Ontologia Brasileira de Medicamentos (Portal OBM)", que também apresenta alguns metadados comuns a outros serviços da plataforma GovBR. Avalia-se que o recurso recuperado se trata de uma tabela com uma relação de medicamentos, não podendo ser considerado uma ontologia, por não conter informações básicas desse tipo de SOC. Além da ausência de elementos básicos de uma ontologia, a página não permite acesso aos itens,

impossibilitando quaisquer consultas a conceitos ou atributos relacionados. Considerando a informação em destaque na página do serviço de que a base de dados ainda está sendo consolidada, com ressalva de constar apenas produtos de saúde da Farmácia Popular, avalia-se que há potencial para que a lista de medicamentos se torne futuramente uma ontologia. Porém, da forma que se apresenta no momento dessa consulta, entende-se que a lista de medicamentos não pode ser considerada uma ontologia, por não prever características como definição clara dos itens (conceitos), relacionamentos, atributos, hierarquia de conceitos, restrições (axiomas), entre outras. Além disso, a lista não está estruturada em uma linguagem formal de ontologias.

Na pesquisa realizada no Portal da Transparência, apesar de não ter havido a recuperação de nenhum arquivo que pudesse se aproximar de uma ontologia, destaca-se que houve a recuperação de 301 registros que mencionavam o termo ontologia. Entre esses registros, documentos de prestação de contas de reuniões de trabalho para tratar sobre o tema, participação em eventos e cursos de ontologias, além de registros de pagamentos de bolsas de pesquisa sobre o tema. Com base nos documentos recuperados, pode-se inferir que há um interesse pelo tema no contexto da administração pública, mas, ao mesmo tempo, identifica-se ainda incipiência na temática, visto que era esperado que, em um portal de transparência de dados públicos, houvesse pelo menos a menção a ontologias já desenvolvidas no âmbito da administração pública brasileira.

O *site* da Biblioteca do Senado Federal também não recuperou documentos coerentes com os termos de busca. O número de arquivos recuperados variava conforme o termo de busca utilizado, porém identificou-se que os arquivos recuperados não estavam coerentes com tais termos. Boa parte dos documentos recuperados eram artigos, trabalhos de conclusão de curso, livros, entre outros. Destaca-se que, diferentemente dos *sites* anteriores analisados, a Biblioteca do Senado adota o padrão *Dublin Core* para representação dos metadados de todos os documentos disponíveis, o que pode contribuir para a recuperação dos arquivos e para a otimização da análise de cada documento com base nos metadados registrados.

Da mesma forma, a Biblioteca do IBGE também não recuperou documentos que atendessem aos termos de busca utilizados. Mesmo com a utilização apenas do termo ontologia, houve a recuperação de somente dois arquivos, que davam acesso

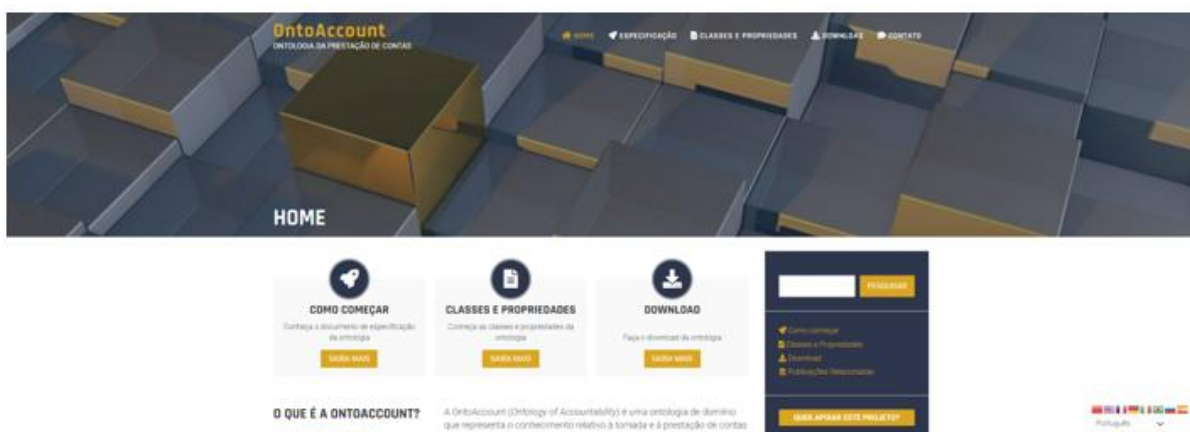
ao conjunto bastante limitado de metadados, sem possibilidade de acesso direto ao documento.

A pesquisa ao Repositório Institucional da Enap recuperou uma variedade de documentos que faziam menção ao termo ontologia, mas não recuperou de forma adequada documentos relacionados aos termos de busca utilizados. O repositório da Enap também adota o padrão *Dublin Core* para catalogação dos documentos, além de disponibilizar outras informações, como identificador persistente, filtros por título, por autor, funcionalidade para ordenar os documentos recuperados, entre outras.

A busca realizada na Biblioteca Digital Brasileira de Teses e Dissertações demonstrou maior coerência em relação aos termos de busca utilizados e os documentos recuperados. Boa parte dos documentos recuperados abordava o tema ontologia de maneira subjacente, sem um foco específico no contexto da administração pública. Entre os arquivos recuperados, identificou-se a Dissertação intitulada “PSO: uma ontologia de domínio sobre servidores públicos federais”. Além de ser uma proposta de desenvolvimento de uma ontologia para o contexto da administração pública, o estudo apresenta a proposta da ontologia, e disponibiliza o modelo conceitual, exemplos de consultas SQL, bem como o arquivo da ontologia desenvolvido nos formatos RDF e XML.

Por meio da pesquisa realizada no site SciELO – *Scientific Electronic Library Online*, foi possível identificar outro trabalho acadêmico, neste caso o resultado de uma tese, que propôs o desenvolvimento de uma ontologia denominada “*Ontoaccount*”, voltada para o domínio do conhecimento da tomada de decisão e prestação de contas anual dos gestores públicos federais. A ontologia foi desenvolvida com base na metodologia de construção de ontologias denominada *OntoForInforScience* (Mendonça, 2015), utilizando-se o *software* Protégé. Em pesquisa complementar, identificou-se que a ontologia possui um *site* específico (<https://ontoaccount.org/>), conforme disposto na Figura 20, por meio do qual é possível acessar a descrição da ontologia, objetivos, tecnologias utilizadas, documento de especificação, taxonomia da ontologia, relação de classes e propriedades, versão disponível, *link* para *download* do código em formato OWL e publicações relacionadas à ontologia.

Figura 20 – Página inicial da ontologia OntoAccount



Fonte: elaborada pela autora (2024).

Outra publicação relacionada à “*Ontoaccount*” foi identificada na busca realizada no Google Acadêmico – a tese de doutorado, em que o pesquisador traz mais detalhes acerca da construção da ontologia.

Por fim, na busca realizada nos Anais do Seminário de Pesquisa em Ontologia no Brasil – Ontobras, foi possível identificar um estudo que, em tese, teria relevância para esta pesquisa: “*Ontology Requirements Elicitation in Brazilian Public Administration: Methodologies and Challenges*”. Tratou-se de uma sessão de pôster realizada na edição 2024 da Ontobras, entretanto não foi disponibilizado acesso ao arquivo do estudo no *site* do evento.

Analisando-se as pesquisas realizadas de forma geral, um aspecto recorrente em praticamente todos os *sites* consultados foi a dificuldade em relação aos mecanismos de busca que não orientam o usuário acerca dos operadores lógicos que poderiam ser utilizados, ausência de mecanismo de busca avançada, não disponibilização de filtros de maneira a restringir os resultados, além de falhas claras na recuperação, pois mesmo nos *sites* que permitiam o uso dos operadores lógicos, os documentos recuperadores não refletiam a busca realizada.

Avalia-se que a variedade de tipos de documentos no contexto da administração pública, bem como a quantidade de atributos específicos relacionados a cada tipo de documento demonstram o potencial que as ontologias têm para organização e representação desse vasto conjunto de entidades. A falta de padronização de metadados entre portais de informação estratégicos do governo brasileiro também é um indicativo de que há uma lacuna importante no que diz respeito à gestão do conhecimento e da informação nesse contexto.

Diante dos resultados apresentados, cumpre resgatar o objetivo central desta etapa da pesquisa: inventariar as ontologias desenvolvidas no contexto do governo brasileiro. A análise documental realizada em diferentes plataformas institucionais e repositórios de dados públicos revelou a escassez de ontologias formalizadas e amplamente acessíveis, bem como a dificuldade de localizá-las por meio dos mecanismos de busca disponíveis. Embora tenham sido identificadas menções pontuais ao termo “ontologia” em serviços, notícias e documentos administrativos, os dados não evidenciam a existência de um repositório integrado de ontologias governamentais, tampouco de iniciativas sistemáticas voltadas à sua promoção enquanto instrumentos de organização do conhecimento.

Tais achados permitem confrontar o segundo pressuposto teórico desta pesquisa, segundo o qual as iniciativas de construção de ontologias no contexto do governo brasileiro constituem ações isoladas, desvinculadas de uma metodologia unificada. A análise sustenta esse pressuposto, na medida em que as ontologias identificadas aparecem como experiências localizadas, sem articulação evidente com diretrizes comuns de desenvolvimento, documentação e interoperabilidade. Essa fragmentação compromete a padronização terminológica, dificulta a integração entre sistemas e enfraquece o potencial das ontologias enquanto instrumentos de interoperabilidade semântica no governo eletrônico.

Em consequência, observa-se que a ausência de um modelo metodológico compartilhado e amplamente adotado para a construção de ontologias governamentais limita a criação de uma infraestrutura de conhecimento sólida e reutilizável. Essa lacuna pode dificultar o alinhamento semântico entre órgãos e esferas de governo, afetando diretamente a consistência dos dados públicos e as possibilidades de integração informacional. Para superar esse desafio, torna-se necessário fomentar políticas e iniciativas voltadas à adoção de metodologias sistemáticas que favoreçam a reutilização, a interoperabilidade e o alinhamento ontológico em ambientes governamentais.

5.4 Metodologias de construção de ontologias: uma análise comparativa

Esta seção apresenta uma análise comparativa de quatro metodologias desenvolvidas no âmbito do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPGGOC) da UFMG: OntoForInfoScience, OntONeo,

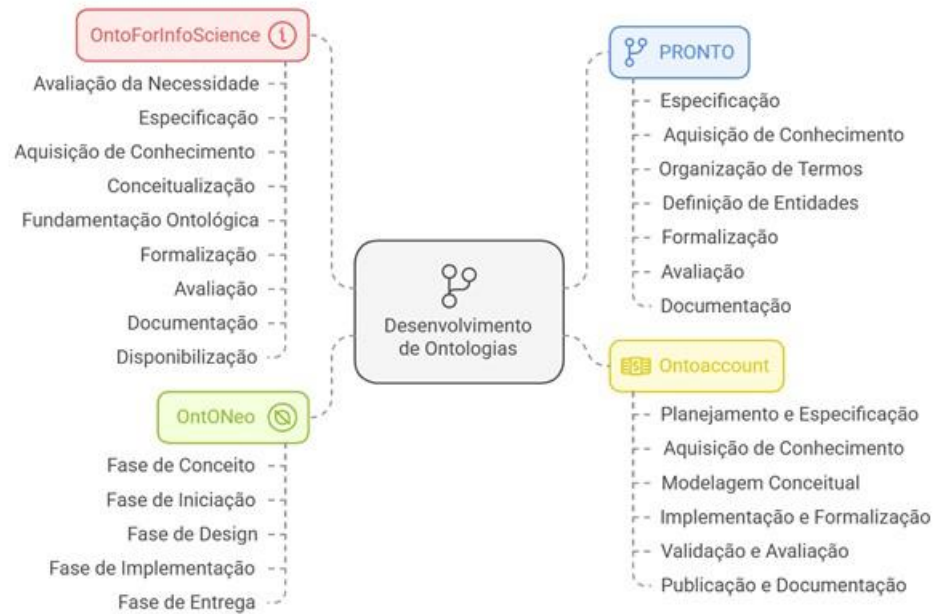
OntoAccount e PRONTO. A seleção dessas metodologias foi orientada por critérios de relevância científica, completude documental, clareza metodológica e aplicabilidade prática. Tais critérios visam assegurar não apenas a representatividade das abordagens no campo da Ciência da Informação, mas também a viabilidade analítica dentro dos limites temporais e operacionais da pesquisa. Além disso, por serem metodologias originadas em um mesmo contexto institucional e epistemológico, parte-se do pressuposto de que compartilham premissas teóricas comuns, o que favorece uma comparação mais equilibrada entre seus componentes. A análise que se segue busca descrever suas fases, etapas e objetivos, evidenciando os pontos de convergência e distinção, bem como as potencialidades e limitações de cada proposta no processo de desenvolvimento ontológico.

Embora a OntONeo e a OntoAccount tenham sido originalmente concebidas como ontologias específicas para os domínios obstétrico-neonatal e de prestação de contas públicas, respectivamente, compreende-se que ambas incorporam em seus processos de construção um conjunto sistemático de fases, etapas e procedimentos suficientemente estruturados para que possam ser analisadas também sob a perspectiva metodológica. Ainda que esses processos derivem de metodologias consolidadas, como NeOn, Methontology e Método 101, as adaptações realizadas pelos autores configuram instâncias metodológicas particulares, adaptadas às especificidades dos domínios tratados. Assim, para fins desta análise comparativa, considerou-se pertinente tratar OntONeo e OntoAccount não apenas como produtos ontológicos finais, mas como proposições metodológicas aplicadas, com potencial para orientar novos processos de modelagem em contextos similares.

5.4.1 Descrição das metodologias

A partir da descrição das metodologias, buscou-se compreender como essas propostas são estruturadas, permitindo identificar similaridades e diferenças. Na Figura 21, a seguir, é apresentado um consolidado das fases e etapas identificadas de cada uma das metodologias que são descritas na sequência.

Figura 21 – Visão consolidada das fases e etapas das metodologias analisadas



Fonte: elaborada pela autora (2024).

5.4.1.1 Fases e Etapas da Metodologia OntoForInfoScience

O desenvolvimento de uma ontologia sob a perspectiva da Metodologia OntoForInfoScience é estruturado em fases e etapas bem definidas, garantindo que o processo seja conduzido de forma sistemática, abrangente e alinhada aos objetivos específicos do domínio de aplicação. Cada fase desempenha um papel estratégico, desde a identificação das necessidades até a disponibilização prática da ontologia.

A Fase 1 – Avaliação da Necessidade da Ontologia (Etapa 0), consiste na identificação da pertinência de uma ontologia para um domínio específico. Durante essa fase, são realizadas atividades como análise de lacunas, levantamento de requisitos e entendimento do domínio. Esse diagnóstico inicial é essencial para justificar o esforço de construção da ontologia, estabelecendo as bases conceituais para as etapas seguintes.

Na Fase 2 – Especificação da Ontologia (Etapa 1), o propósito, o escopo e os usuários-alvo da ontologia são definidos. Essa fase inclui a elaboração de *templates*, a identificação de objetivos e uma análise detalhada dos requisitos do domínio. O objetivo é assegurar que a ontologia seja desenvolvida para atender de maneira precisa às necessidades de seus usuários.

A Fase 3 – Aquisição e Extração de Conhecimento (Etapa 2) compreende a coleta e a extração de informações relevantes. Atividades como entrevistas com especialistas, revisão bibliográfica e análise de dados disponíveis são realizadas para consolidar um entendimento aprofundado do domínio. Essa fase é fundamental para reunir os insumos necessários à estruturação da ontologia.

Durante a Fase 4 – Conceitualização (Etapa 3), os conceitos e as relações do domínio são estruturados de forma inicial. São criados glossários, diagramas conceituais e taxonomias que organizam o conhecimento coletado. Essa fase estabelece a base para o desenvolvimento posterior, assegurando a coerência e a clareza na organização das informações.

Na Fase 5 – Fundamentação Ontológica (Etapa 4), os conceitos definidos são alinhados a ontologias de base já existentes. Para isso, são selecionadas ontologias de referência, e as estruturas conceituais são ajustadas para garantir consistência e interoperabilidade semântica. Essa fase conecta a ontologia desenvolvida a padrões estabelecidos, aumentando sua aplicabilidade.

A Fase 6 – Formalização da Ontologia (Etapa 5) traduz os conceitos e as relações definidos em representações formais, utilizando linguagens ontológicas, como OWL. Essa formalização inclui a criação de axiomas, a definição de classes e propriedades, conferindo à ontologia um formato que pode ser processado por sistemas computacionais e aplicado em diferentes contextos.

Na Fase 7 – Avaliação da Ontologia (Etapa 6), a qualidade, a consistência e a aplicabilidade da ontologia são verificadas. Atividades como testes com especialistas, análise de consistência e refinamento são conduzidas para assegurar que a ontologia atende aos requisitos identificados e que está livre de inconsistências lógicas.

A Fase 8 – Documentação da Ontologia (Etapa 7) é responsável pelo registro detalhado de todo o processo de desenvolvimento e da estrutura final da ontologia. São produzidos relatórios, diagramas e documentação técnica que tornam a ontologia compreensível e acessível a diferentes públicos.

Por fim, a Fase 9 – Disponibilização da Ontologia (Etapa 8) conclui o processo, publicando e integrando a ontologia para uso prático. Nessa fase, são realizadas atividades como *upload* em repositórios, testes em sistemas e integração com ferramentas existentes. O objetivo é garantir que a ontologia esteja disponível e funcional para seu domínio de aplicação.

5.4.1.2 Fases e Etapas da Metodologia OntONEo

A Metodologia OntONEo organiza o desenvolvimento de ontologias em fases estruturadas, cada uma com objetivos específicos e etapas bem definidas, assegurando um processo sistemático e alinhado às demandas do domínio de aplicação. Essa abordagem fornece um guia metodológico robusto para criar ontologias eficazes e práticas.

A Fase de Conceito tem como objetivo estabelecer a base conceitual da ontologia. Durante essa fase, são realizadas atividades fundamentais, como o levantamento de requisitos e o escopo do domínio, o mapeamento do conhecimento existente por meio de revisões de literatura e análise de sistemas, além da definição inicial de conceitos e relações. Essa etapa inicial é importante para garantir que a ontologia seja desenvolvida com uma compreensão clara e fundamentada do domínio.

A Fase de Iniciação busca formalizar os requisitos iniciais da ontologia. Nessa fase, é elaborado o Documento de Especificação de Requisitos da Ontologia (*Ontology Requirements Specification Document – ORSD*), que detalha os objetivos e os limites do projeto. Também são selecionadas ontologias e padrões existentes para reutilização, com a identificação de lacunas semânticas e estruturais que precisam ser abordadas. Essa etapa assegura que a ontologia será desenvolvida de forma eficiente, aproveitando recursos já existentes.

A Fase de *Design* é voltada para estruturar a ontologia de forma que ela reflita com precisão o conhecimento do domínio. As atividades incluem a modelagem conceitual utilizando ferramentas como o Protégé, a criação de hierarquias taxonômicas e relações entre conceitos, além da definição de axiomas e restrições semânticas baseadas no realismo ontológico. Essa fase estabelece a estrutura lógica e organizacional da ontologia.

Na Fase de Implementação, o foco é formalizar e integrar a ontologia. Durante essa etapa, a ontologia é codificada em linguagens formais, como OWL, garantindo sua interoperabilidade com sistemas computacionais. Além disso, são realizados testes iniciais de consistência e funcionalidade e a integração com sistemas existentes e bancos de dados, assegurando sua aplicação prática.

Por fim, a Fase de Entrega tem como objetivo disponibilizar a ontologia e garantir sua utilização prática. Nessa fase, a ontologia é publicada em repositórios como BioPortal e OBO *Foundry*, tornando-se acessível à comunidade. Além disso, a

validação prática é realizada por meio de estudos de caso e integração em sistemas reais. Com base no *feedback* de usuários e especialistas, são realizados ajustes para melhorar ainda mais a funcionalidade e a aplicabilidade da ontologia.

5.4.1.3 Fases e Etapas da Metodologia OntoAccount

A Metodologia OntoAccount apresenta um conjunto estruturado de fases e etapas para o desenvolvimento de ontologias, com foco específico no domínio da prestação de contas. Cada fase é organizada de forma a garantir a precisão, a consistência e a aplicabilidade prática da ontologia, abordando desde o planejamento inicial até a publicação e a promoção de seu uso.

A Fase de Planejamento e Especificação tem como objetivo definir o escopo e os requisitos da ontologia. Durante essa fase, são levantados os objetivos do domínio e as necessidades dos usuários. Em seguida, é elaborado um Documento de Especificação de Requisitos da Ontologia (ORSD), que formaliza os parâmetros do projeto. Também são identificadas ontologias de base reutilizáveis e padrões relevantes, como a *Basic Formal Ontology* (BFO), assegurando que o desenvolvimento seja baseado em referências confiáveis.

Na Fase de Aquisição e Extração de Conhecimento, o foco é a coleta de informações detalhadas sobre o domínio. Para isso, são realizadas revisões de literatura e análises de documentos de referência, como relatórios de gestão pública e normas do Tribunal de Contas da União (TCU). A extração de conceitos, verbos e relações relevantes dessas fontes documentais é acompanhada da construção de glossários iniciais, que ajudam a consolidar o conhecimento necessário para a próxima etapa.

A Fase de Modelagem Conceitual visa estruturar os conceitos e as relações do domínio. Essa estruturação inclui a criação de diagramas conceituais e taxonomias para organizar o conhecimento coletado, bem como a elaboração de hierarquias taxonômicas e definições formais de classes, propriedades e relações. Além disso, ocorre a integração com ontologias reutilizáveis e o alinhamento com padrões semânticos, garantindo consistência e interoperabilidade no modelo conceitual.

Na Fase de Implementação e Formalização, a ontologia é codificada em um formato formal e interoperável. O Protégé é utilizado para implementar a ontologia em OWL, uma linguagem amplamente reconhecida para representação formal. Durante

essa fase, são definidos axiomas e restrições que asseguram consistência e validade semântica. Testes iniciais são realizados para verificar a consistência lógica e estrutural, garantindo que a ontologia atenda aos requisitos estabelecidos.

A Fase de Validação e Avaliação busca garantir a qualidade e a aplicabilidade prática da ontologia. Nessa etapa, são conduzidos testes de usabilidade com especialistas e usuários finais, permitindo uma análise prática do modelo desenvolvido. Também é verificada a interoperabilidade da ontologia com sistemas de dados existentes, e ajustes são realizados com base no *feedback* recebido, refinando o modelo para atender às necessidades reais do domínio.

Por fim, a Fase de Publicação e Documentação conclui o processo, tornando a ontologia acessível e promovendo seu uso. Durante essa fase, a ontologia é publicada em repositórios de dados abertos e plataformas ontológicas, como o BioPortal. Além disso, é criada uma documentação detalhada que inclui definições, exemplos e aplicações da ontologia, facilitando sua compreensão e utilização por diferentes públicos. Por último, ocorre a divulgação da ontologia para a comunidade de pesquisa e outros interessados no domínio, promovendo sua adoção e aplicabilidade.

5.4.1.4 Fases e Etapas da Metodologia PRONTO

A Metodologia PRONTO organiza o desenvolvimento de ontologias em um conjunto de fases sequenciais, cada uma composta por etapas específicas e objetivos claros. Esse modelo garante que o processo de construção da ontologia seja abrangente, sistemático e adaptado às demandas do domínio de aplicação, desde a definição de requisitos até a documentação final.

A Fase I – Especificação tem como objetivo descrever e delimitar os requisitos e os objetivos da ontologia. Nessa etapa inicial, são identificados o propósito principal da ontologia, o escopo (cobertura e nível de detalhamento necessário) e a linguagem de implementação que será utilizada. Além disso, são definidos os usuários e os usos pretendidos da ontologia. Essa fase também contempla a identificação dos requisitos da ontologia, que podem ser classificados como não funcionais (requisitos gerais e restrições) ou funcionais (questões de competência e respostas esperadas).

A Fase II – Aquisição de Conhecimento concentra-se na identificação e registro de termos relevantes para o domínio da ontologia. Durante essa fase, termos são

extraídos de artefatos como tesouros, glossários e ontologias existentes, bem como de documentos relevantes analisados manualmente ou por métodos automáticos. Além disso, entrevistas, *brainstorms* e outros métodos de elicitación de conhecimento com especialistas são utilizados. Esses termos e relações são organizados em ferramentas como o Protégé, e um modelo preliminar é criado para estruturar as informações coletadas.

Na Fase III – Organização de Termos Representativos de Entidades, os termos identificados são estruturados de maneira preliminar. Isso envolve a importação de termos genéricos a partir de ontologias de alto nível, a classificação de termos específicos do domínio e a criação de hierarquias taxonômicas no Protégé. Essa fase é essencial para estabelecer uma base sólida para a organização conceitual da ontologia.

A Fase IV – Organização de Termos Representativos de Relacionamentos é voltada para a conexão das entidades estruturadas na ontologia. Durante essa fase, relacionamentos padrão são selecionados e inseridos, utilizando ontologias de alto nível como referência. Além disso, relações adicionais específicas do domínio são criadas e integradas, garantindo que as conexões entre as classes reflitam o conhecimento real do domínio.

Na Fase V – Definição de Entidades, o significado das entidades e sua classificação são determinados. O método de definições aristotélico (“S é um G que D”) é aplicado para descrever as entidades em detalhes, e essas definições são registradas em linguagem natural, criando metadados úteis para consultas e documentação.

A Fase VI – Formalização converte as definições criadas em axiomas formais que podem ser utilizados por *reasoners*. Durante essa etapa, axiomas são produzidos em OWL usando a sintaxe Manchester, e a ontologia é classificada por meio de *reasoners* para validar as inferências geradas e assegurar sua consistência lógica.

A Fase VII – Avaliação garante que a ontologia esteja consistente e alinhada aos objetivos estabelecidos. A avaliação para ajustes verifica erros e inconsistências utilizando o *reasoner* no Protégé, enquanto a avaliação para resultados assegura que a ontologia responda adequadamente às questões de competência do domínio.

Por fim, a Fase VIII – Documentação organiza e registra todos os materiais utilizados e decisões tomadas durante o processo. Nessa etapa, um repositório com os documentos de cada fase é criado, e um documento resumo da ontologia é gerado,

contendo as principais características, definições e aplicações. Esse documento facilita consultas futuras e promove a disseminação do trabalho realizado.

5.4.2 Análise comparativa das quatro metodologias selecionadas

As quatro metodologias analisadas apresentam diferentes abordagens para o desenvolvimento de ontologias, cada uma com características que refletem suas prioridades e contextos de aplicação. Apesar de suas peculiaridades, elas compartilham elementos fundamentais, como a identificação inicial da necessidade da ontologia, a estruturação de conceitos e relações, a formalização em linguagens específicas e a documentação do processo. Esses pontos comuns evidenciam a importância de passos básicos e estruturados para garantir que uma ontologia seja coerente, funcional e adaptada ao seu domínio de aplicação.

No entanto foram observadas algumas divergências entre as metodologias, tanto na nomenclatura quanto no nível de detalhamento de suas fases e etapas. A metodologia OntoForInfoScience, por exemplo, adota uma abordagem mais detalhada e sistemática, destacando etapas específicas, como a fundamentação ontológica, que envolve o alinhamento com ontologias de base. Esse nível de especificidade da OntoForInfoScience ressalta uma preocupação com a fundamentação teórica robusta, mas pode torná-la menos flexível para aplicações em contextos mais dinâmicos ou com recursos limitados.

A OntONeo, por sua vez, adota uma estrutura mais agregada, agrupando atividades relacionadas em fases amplas, como conceituação, *design* e implementação. Essa abordagem a torna eficiente para projetos que demandam rapidez ou integração com sistemas já existentes, mas pode deixar lacunas em questões mais teóricas ou estruturais. Já a metodologia OntoAccount é fortemente orientada para domínios específicos, como prestação de contas, incorporando etapas que enfatizam a análise de documentos e padrões regulatórios. Essa especificidade proporciona maior aderência a domínios complexos, mas pode limitar sua aplicabilidade em áreas mais amplas ou diversificadas.

A metodologia PRONTO se destaca por sua granularidade e detalhamento nas subetapas. Ao organizar as atividades em pequenos blocos, como a organização separada de entidades e relacionamentos, ela facilita a gestão do processo e promove

maior clareza para não especialistas. Contudo esse detalhamento pode ser excessivo em projetos menores ou em fases que exigem mais agilidade.

Outro ponto de divergência importante é a etapa de disponibilização da ontologia. Enquanto ela é tratada como uma fase separada e final nas metodologias OntoForInfoScience e OntONEo, sendo explícita quanto à publicação e integração, está implícita em OntoAccount e PRONTO. Isso pode gerar lacunas no planejamento, especialmente em contextos que demandam acessibilidade ampla e documentação pública.

Adicionalmente, as abordagens para conceitualização e avaliação variam. A OntoForInfoScience detalha a conceitualização como uma etapa autônoma, enquanto OntONEo e OntoAccount a integram em fases mais abrangentes, como *design* ou modelagem. No que diz respeito à avaliação, todas as metodologias reconhecem sua relevância, mas a profundidade e o foco variam. A PRONTO, por exemplo, combina avaliação técnica e funcional em fases distintas, demonstrando preocupação com a adaptação da ontologia aos objetivos finais.

De maneira geral, as metodologias revelam diferentes forças e fraquezas, dependendo do contexto de aplicação. Entende-se que a OntoForInfoScience é ideal para projetos academicamente rigorosos; já a OntONEo oferece flexibilidade e eficiência para aplicações práticas; a OntoAccount é indicada para domínios regulamentados; e, por fim, a metodologia PRONTO destaca-se em cenários que exigem clareza e rastreabilidade do processo. Contudo entende-se que nenhuma delas é totalmente abrangente, o que sugere a necessidade de combinações adaptativas ou ajustes para atender às especificidades de projetos em domínios variados.

Na próxima seção, apresenta-se uma proposta de nova metodologia para o desenvolvimento de ontologias, construída a partir da análise das fases, etapas, objetivos e descrições das metodologias previamente examinadas. A análise comparativa realizada destacou as semelhanças e as diferenças entre as abordagens, bem como lacunas e oportunidades de integração que podem ser exploradas. Essa proposta busca consolidar os pontos fortes das metodologias analisadas, ao mesmo tempo que aborda limitações identificadas, visando oferecer uma abordagem mais abrangente e adaptada às necessidades específicas do domínio de aplicação. O Quadro 12 apresenta resumidamente o resultado da comparação entre as quatro metodologias.

Quadro 12 – Comparação entre as quatro metodologias

FASES/ETAPAS	OntoForInfoScience	OntONeo	OntoAccount	PRONTO
1. Avaliação/ Planejamento	Identifica lacunas e requisitos	Define escopo e requisitos	Levanta objetivos e reutiliza padrões	É determinada por propósito
2. Especificação	Define escopo, objetivos e usuários-alvo	Formaliza requisitos no ORSD	ORSD para padrões regulatórios	Inclui requisitos funcionais e não funcionais
3. Aquisição de Conhecimento	Entrevistas, revisão e análise documental	Organiza modelos de conhecimento	Extraí conceitos de documentos e normas	Organiza termos e relações preliminares
4. Conceitualização	Cria glossários, diagramas e taxonomias	Modelagem no Protégé e hierarquia	Estrutura conceitos e relações específicas	Separa organização de entidades e relações
5. Fundamentação	Alinha com as ontologias de base	Integra padrões em <i>design</i> /implementação	Alinha com padrões semânticos	Importa termos de ontologia de alto nível
6. Formalização	Tradução para OWL e criação de axiomas	Codifica formalmente e testa consistência	Formaliza no Protégé com axiomas definidos	Usa OWL e valida lógica com <i>reasoners</i>
7. Avaliação	Testa qualidade e refina com especialistas	Válida com estudo de caso	Verifica usabilidade e testes	Avaliação técnica e funcional separada
8. Documentação	Relatórios técnicos e diagramas	Publicação e documentação técnica	Documentação prática e detalhada	Repositório completo de processo
9. Disponibilização	Publica em sistemas práticos	Disponibiliza com ajustes via <i>feedback</i>	Publica em plataformas abertas	Repositórios e divulgação acessível

Fonte: elaborado pela autora (2024).

Os achados desta etapa da pesquisa permitem retomar o objetivo específico de identificar possibilidades de automatização de fases e etapas das metodologias analisadas. Embora todas as metodologias apresentem processos bem estruturados, observou-se que a automatização — seja por meio de ferramentas computacionais

integradas, seja de mecanismos formais de apoio às decisões — ainda não constitui um elemento central nos modelos examinados. Atividades como aquisição de conhecimento, formalização e avaliação, por exemplo, são predominantemente realizadas de forma manual, exigindo *expertise* técnica e dedicação intensiva de tempo. Esse cenário corrobora o quarto pressuposto da pesquisa, ao indicar que a ausência de metodologias que ofereçam suporte automatizado ou semiautomatizado pode contribuir para o baixo uso de ontologias no governo eletrônico. Avalia-se que a complexidade inerente à construção manual de ontologias, somada à carência de processos escaláveis e replicáveis, tende a desestimular sua adoção em ambientes institucionais que operam sob pressões de tempo, orçamento e interoperabilidade. Portanto, torna-se evidente a necessidade de metodologias que incorporem, de forma mais explícita e integrada, recursos automatizados que favoreçam não apenas a eficiência operacional, mas também a difusão e a sustentabilidade das ontologias no setor público.

Os resultados apresentados até aqui evidenciam as práticas e as lacunas associadas à organização e à representação do conhecimento, com foco nos SOCs adotados no contexto do governo eletrônico. A análise das ontologias existentes e das metodologias de construção comparadas permite identificar elementos fundamentais para o avanço teórico e prático nesse domínio. A partir dessas bases, a próxima seção detalha a *Onto Merge Methodology* (OMM), proposta que consolida as principais fases e etapas identificadas, promovendo uma abordagem integrada para a aprendizagem de ontologias. A OMM apresenta as fases e as etapas de construção, e as sugestões de execução e exemplos práticos, com foco na automatização de processos e na validação de sua aplicabilidade, conforme ilustrado na prova de conceito descrita na subseção 6.4.

6 ONTO MERGE METHODOLOGY (OMM)

A *Onto Merge Methodology* (OMM) foi concebida como uma abordagem que integra fases e etapas das quatro metodologias analisadas, buscando aproveitar os pontos fortes de cada uma. Dada a complexidade inerente ao processo de automatização, a metodologia concentrou seus esforços na fase de aquisição de conhecimento, considerada crucial para a construção de ontologias eficientes e adaptadas ao contexto do governo eletrônico. Além disso, a OMM apresenta uma descrição detalhada das etapas, oferece sugestões de alternativas para a execução de cada fase e inclui exemplos práticos para facilitar a aplicação. Os resultados desse processo de fusão e sua aplicação prática são apresentados na prova de conceito descrita na subseção 6.4, ilustrando como os elementos da OMM foram operacionalizados para alcançar os objetivos propostos.

6.1 Resumo do Percorso Metodológico

Inicialmente, foi feita uma revisão de literatura, buscando-se identificar possíveis técnicas de automatização utilizadas na construção de ontologias, mesmo que essas técnicas fossem aplicadas em parte das etapas de construção. A partir dos resultados da revisão de literatura, foi possível identificar diferentes técnicas para automatização de etapas da construção de ontologias, principalmente em fases consideradas críticas e dispendiosas, como é o caso da fase de aquisição de conhecimento.

Com o objetivo de avaliar a aplicação de ontologias em outros países bem posicionados em implementação de práticas de Governo Eletrônico, foi realizada uma análise dos principais SOCs adotados nos dez países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU; também do Brasil, para fins comparativos, apesar de ele ocupar a 49.^a posição nesse índice. Por meio da análise documental, observou-se um uso intensivo dos mais variados tipos de SOC pelos países analisados, incluindo as ontologias. Em contraste, na análise dos dados abertos do governo brasileiro, pressupõe-se baixo uso de SOC ou mesmo ausência de plataformas que integrem esses artefatos de maneira organizada e acessível.

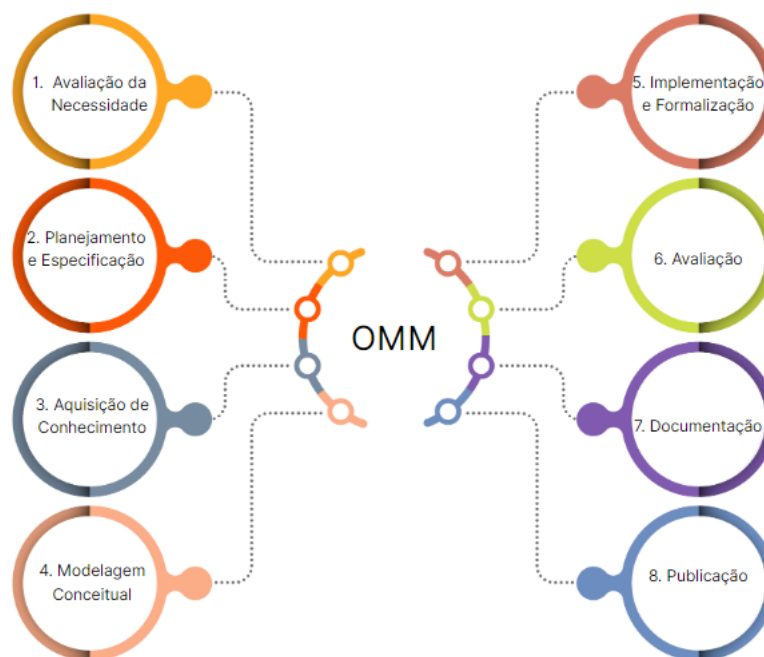
Essa percepção de baixo uso de SOC para organização e representação do conhecimento e informação pelo governo brasileiro pode ser confirmada por meio de uma análise documental complementar, focada especificamente na identificação de ontologias. A partir da pesquisa em *sites* de dados abertos e *sites* acadêmicos, constatou-se que não há uma integração das poucas iniciativas de construção de ontologias identificadas, apesar de recuperados estudos e documentos técnicos referenciando esse tipo de SOC. A análise documental ratificou, portanto, a percepção de que as ontologias ainda são pouco exploradas pelo governo brasileiro, no que diz respeito à organização e representação do conhecimento e informação.

Dando sequência ao processo investigativo, buscou-se então identificar na literatura metodologias de construção de ontologias que pudessem ser utilizadas como base para estruturação da nova metodologia proposta nesta pesquisa. Foram então identificadas quatro metodologias produzidas por pesquisadores do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da UFMG. As quatro metodologias foram selecionadas por contemplarem elementos essenciais a uma proposta de metodologia, como previsão de fases e etapas e suas respectivas descrições, com linguagem acessível e clara e passível de replicação. As quatro metodologias selecionadas (OntoForinfoscience OntONeo; OntoAccount e PRONTO), incluindo fases, objetivos e etapas estão descritas na seção 5.3. Nas seções seguintes, são apresentados os resultados desse percurso metodológico, que desencadearam um novo conjunto de Fases e uma versão integrada das etapas, ainda dentro de uma perspectiva de construção de ontologias não automatizada.

6.2 Fases propostas da OMM

A partir da análise das fases e etapas descritas nas quatro metodologias de referência, foi realizado um processo de análise, comparação e integração das fases com seus respectivos objetivos. A integração das fases resultou em oito fases, conforme ilustrado na Figura 22.

Figura 22 – Visão Geral da Metodologia Onto Merge



Fonte: elaborada pela autora (2024).

No Quadro 13, descrevem-se os objetivos das oito fases da proposta.

Quadro 13 – Fases da Onto Merge Methodology (OMM)

(Continua...)

FASES	OBJETIVOS
Avaliação da Necessidade	Analisar e justificar a necessidade de uma ontologia no domínio específico, considerando lacunas e demandas existentes.
Planejamento e Especificação	Delimitar propósito, escopo, requisitos funcionais e não funcionais, usuários-alvo e objetivos da ontologia, criando uma base clara e formal para seu desenvolvimento.
Aquisição de Conhecimento	Realizar coleta, extração e registro sistemático de informações e termos relevantes do domínio, abrangendo dados, conceitos e relações fundamentais para a ontologia.

(Conclusão.)

FASES	OBJETIVOS
Modelagem Conceitual	Estruturar e organizar conceitos, entidades e relações do domínio, alinhando-os a ontologias de base e padrões semânticos para construir uma representação conceitual fundamentada e coerente.
Implementação e Formalização	Implementar e formalizar a ontologia em uma linguagem formal, desenvolvendo axiomas e garantindo sua integração em sistemas computacionais de maneira interoperável.
Avaliação	Avaliar a ontologia quanto à consistência lógica, qualidade, adequação aos objetivos e aplicabilidade prática, garantindo que os requisitos e propósitos sejam plenamente atendidos.
Documentação	Registrar detalhadamente os materiais, as decisões, os processos e a estrutura da ontologia, criando documentação clara e acessível para uso e manutenção futura.
Publicação	Publicar e integrar a ontologia para uso prático, garantindo acessibilidade, aplicabilidade e promoção junto aos usuários e sistemas relevantes.

Fonte: elaborado pela autora (2024).

Observou-se que a fase de Avaliação da Necessidade foi explicitamente prevista apenas na OntoForInfoScience, considerada essencial para a sequência da fase de planejamento, justamente para que se justifique a necessidade de construção de uma ontologia, visto que, conforme a demanda apresentada, a solução poderia ser alcançada por meio de um outro tipo de SOC.

A fase de Planejamento e Especificação foi prevista nas quatro metodologias analisadas, apesar de pequenas variações terminológicas. Avaliou-se que etapas previstas nas fases denominadas de “Conceito” e “Iniciação” na metodologia OntONEo eram equivalentes à fase de Planejamento e Especificação, sendo então agrupadas nessa fase.

A Aquisição de Conhecimento não foi prevista explicitamente apenas na metodologia OntONEo. Entretanto observou-se que algumas das etapas previstas nas fases de “Conceito” e “Iniciação” atendiam ao objetivo dessa fase do processo. Nela, houve também uma variação terminológica, com a inclusão do termo “Extração” no título das fases das metodologias OntoForInfoScience e OntoAccount, entretanto avaliou-se que o processo de extração estaria subentendido e contemplado no objetivo e nas respectivas etapas dessa fase.

Para a fase de Modelagem Conceitual, observou-se maior divergência entre as metodologias. A OntoForInfoScience denominou esta fase como “Conceitualização”, enquanto a OntONeo a denominou de “*Design*”. Em relação à metodologia PRONTO, avaliou-se que a fase foi devidamente contemplada, porém em um nível de granularidade que estaria mais adequado a etapas e não a uma fase como, por exemplo, a fase “Organização de Termos Representativos de Entidades”, compreendida na proposta desta pesquisa como uma etapa. Da mesma forma, observou-se que a OntoForInfoScience previu uma fase específica denominada “Fundamentação Ontológica”, também compreendida enquanto etapa da fase de Modelagem Conceitual.

Com relação à Implementação e Formalização, observou-se que as quatro metodologias previram essa fase, com pequenas diferenças de terminologia tanto nos rótulos adotados para as fases quanto nas etapas vinculadas. A partir da análise das etapas vinculadas, foi possível confirmar a equivalência entre os diferentes termos usados e contemplar etapas que não foram previstas de forma explícita nas quatro metodologias.

A fase de Avaliação não foi explicitamente prevista apenas na OntONeo, entretanto a metodologia contemplou essa fase nas etapas previstas nas fases de “Implementação” e “Entrega”. Nas demais metodologias, a fase é explicitada com foco na verificação da consistência, acurácia, aplicabilidade e atendimento aos objetivos e às questões de competências levantados nos requisitos.

Na sequência, sobre a etapa de Documentação, observou-se que não foi prevista na OntONeo, mas contemplada nas demais metodologias, com menor ou maior detalhamento das etapas. Observou-se que, na OntoAccount, essa fase foi prevista de forma integrada com a fase de Publicação, mas avaliou-se que, em razão da importância de cada uma dessas fases, seria mais assertiva a abordagem em separado.

Por fim, a fase de Publicação, não explicitamente prevista apenas na metodologia PRONTO, sendo contemplada nas demais metodologias. Destaca-se a importância dessa fase, principalmente nas etapas de disponibilização em repositórios de ontologias ou portais de dados abertos, de maneira que outros pesquisadores possam utilizar o artefato produzido e até mesmo evoluir a proposta.

6.3 Apresentação e descrição das etapas de construção de Ontologias não automatizadas

Para se chegar a uma estrutura consolidada, foi feita uma análise detalhada de cada uma das etapas previstas nas metodologias de referência. O maior desafio para realizar a integração foi lidar com as diferenças de terminologia e assegurar que os autores estavam se referindo aos mesmos conceitos. Em razão do volume de redundâncias, foram necessários diversos refinamentos, tendo o cuidado para não deixar de contemplar nenhuma das etapas previstas pelos autores e, ao mesmo tempo, buscar equivalências entre as etapas com terminologias diferentes. Após análise das redundâncias, equivalência de termos, com consulta à metodologia de origem sempre que necessário, chegou-se a uma proposta de integração das etapas, conforme Quadro 14, a seguir. Essa proposta não é exaustiva, entretanto avalia-se que é capaz de representar um processo lógico e abrangente de construção manual de ontologias.

Quadro 14 – Integração das Etapas para construção de ontologias com base nas Metodologias de Referência

(Continua...)

FASES	ETAPAS
Avaliação da Necessidade	Proceder ao entendimento do domínio
	Analisar lacunas e demandas existentes
Planejamento e Especificação	Definir o objetivo principal da ontologia
	Definir usuários-alvo e formas de aplicação da ontologia
	Identificar os objetivos do domínio e as necessidades dos usuários
	Definir requisitos funcionais e não funcionais da ontologia
	Descrever a cobertura geral e o nível de detalhamento necessário
	Identificar lacunas semânticas e estruturais
	Identificar ontologias de base reutilizáveis e padrões relevantes
	Elaborar o Documento de Especificação de Requisitos da Ontologia (ORSO)
	Criar <i>templates</i> dos requisitos

(Continua...)

FASES	ETAPAS
Aquisição de Conhecimento	Realizar revisão de literatura e análise de documentos
	Elicitar conhecimento com especialistas
	Identificar e extrair conceitos, verbos e relações relevantes
	Examinar e interpretar dados do domínio
	Desenvolver glossários iniciais
	Documentar sistematicamente os resultados
	Criar e validar um modelo preliminar
Modelagem Conceitual	Importar termos genéricos e inserir termos do domínio
	Organizar termos representativos de entidades e relacionamentos
	Elaborar hierarquias taxonômicas
	Adicionar relacionamentos predefinidos e adicionais
	Criar diagramas conceituais e taxonomias
	Definir axiomas e restrições semânticas
	Registrar definições em linguagem natural
	Realizar ajustes conceituais
Implementação e Formalização	Criar axiomas e restrições semânticas
	Implementar a ontologia
	Empregar um editor de ontologias
	Refinar as definições formais
	Utilizar um <i>reasoner</i> para validação automática
	Executar testes para identificar inconsistências
	Garantir a interoperabilidade da ontologia
Avaliação	Verificar consistência lógica
	Realizar testes práticos
	Avaliar a compatibilidade com sistemas e padrões
	Identificar possíveis erros e lacunas na ontologia
	Implementar correções e otimizações
	Realizar ajustes finais na ontologia
	Certificar-se de que a ontologia alcança os resultados esperados

(Conclusão.)

FASES	ETAPAS
Documentação	Organizar todos os materiais e decisões tomadas
	Elaborar relatórios técnicos
	Desenvolver representações visuais
	Produzir um documento abrangente da ontologia
	Preparar materiais técnicos
	Elaborar um resumo executivo
Publicação	Realizar testes em ambientes reais
	Avaliar funcionalidade em sistemas específicos
	Incorporar sugestões e observações recebidas
	Configurar a ontologia para ferramentas específicas
	Disponibilizar a ontologia em repositórios públicos
	Divulgar resultados do trabalho

Fonte: elaborado pela autora (2024).

Para que fosse possível identificar etapas passíveis de automatização, avaliou-se a necessidade de incluir uma breve descrição, possíveis alternativas de execução, bem como exemplos práticos de aplicação de cada etapa.

6.3.1 Etapas detalhadas da Fase Avaliação da Necessidade

Nesta fase, identifica-se e justifica-se a necessidade de uma ontologia no domínio específico, avaliando lacunas e demandas existentes. Busca-se compreender os problemas a serem resolvidos e identificar oportunidades para melhorar a organização e o uso do conhecimento no domínio.

Quadro 15 – Etapas detalhadas da Fase Avaliação da Necessidade

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
<p>Proceder ao entendimento do domínio.</p>	<p>Compreender o contexto e os requisitos do domínio alvo.</p>	<p>Mapear o contexto por meio de entrevistas com especialistas e análise documental. Usar a análise SWOT para identificar forças, fraquezas, oportunidades e ameaças ou técnicas de análise de cenários.</p>	<p>Para desenvolver uma ontologia na área médica, revisar documentos de políticas públicas de saúde, como o Sistema Único de Saúde (SUS) no Brasil, e identificar os termos mais comuns relacionados ao atendimento primário.</p>
<p>Analisar lacunas e demandas existentes.</p>	<p>Identificar problemas e necessidades específicas no domínio.</p>	<p>Realizar revisões de literatura e utilizar ferramentas de análise de dados para mapear lacunas e desafios.</p>	<p>Durante a análise de registros de pacientes em um hospital, identificar que os dados de diagnósticos não possuem padronização, causando problemas em relatórios de epidemiologia.</p>

Fonte: elaborado pela autora (2024).

6.3.2 Etapas detalhadas da Fase Planejamento e Especificação

Nesta fase, objetivos, requisitos e escopo da ontologia são estruturados, criando uma base clara e formal para o desenvolvimento.

Quadro 16 – Etapas detalhadas da Fase Planejamento e Especificação

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Definir o objetivo principal da ontologia.	Estabelecer as metas da ontologia de forma clara e concisa.	Conduzir sessões de <i>brainstorming</i> com patrocinadores para alinhar expectativas e objetivos.	Em um projeto de <i>e-commerce</i> , o objetivo da ontologia é padronizar a classificação de produtos e melhorar a busca por itens no <i>site</i> .
Definir usuários-alvo e formas de aplicação da ontologia.	Determinar os usuários-alvo e as formas como a ontologia será aplicada.	Realizar análise de patrocinadores e cenários de aplicação. Adotar a técnica de " <i>Personas</i> " para mapeamento dos usuários.	Para uma ontologia educacional, definir os usuários como professores e instituições acadêmicas.
Identificar os objetivos do domínio e as necessidades dos usuários.	Garantir alinhamento entre os requisitos do domínio e as expectativas dos usuários.	Entrevistar usuários e analisar processos existentes no domínio.	Em um banco, os usuários precisam de uma ontologia que padronize conceitos financeiros, como taxas de juros e tipos de empréstimos, para garantir consistência nos relatórios.
Definir requisitos funcionais e não funcionais da ontologia.	Estabelecer critérios técnicos e operacionais para o funcionamento da ontologia. Para os requisitos funcionais, deve-se levantar as chamadas "Questões de Competências", que são todas as possíveis questões para as quais a ontologia desenvolvida apresentará as respostas.	Documentar requisitos como suporte a consultas SPARQL (funcional) ou tempo máximo de resposta para inferências (não funcional).	Em uma ontologia financeira, definir como requisito funcional a integração com sistemas de ERP.

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Descrever cobertura geral e nível de detalhamento necessário.	Definir o escopo de conceitos e detalhes esperados da ontologia.	Usar <i>benchmarks</i> ou padrões existentes para guiar o nível de detalhamento.	Em uma ontologia de saúde, limitar a cobertura a doenças infecciosas específicas.
Identificar lacunas semânticas e estruturais.	Localizar áreas do domínio onde falta cobertura ou precisão semântica.	Comparar os resultados das análises anteriores com padrões existentes no domínio.	Identificar que a ontologia de biodiversidade existente não cobre espécies de regiões tropicais.
Identificar ontologias de base reutilizáveis e padrões relevantes.	Identificar ontologias de base reutilizáveis e padrões relevantes, como a <i>Basic Formal Ontology</i> (BFO) ou DOLCE.	Consultar bibliotecas como OntoBee ⁵ e BioPortal ⁶ .	Adotar o padrão BFO para uma ontologia de processos industriais.
Elaborar o Documento de Especificação de Requisitos da Ontologia (ORSD).	Consolidar os requisitos, objetivos e escopo em um único documento.	Usar templates baseados em padrões como o NeOn.	Criar o ORSD de uma ontologia de transporte público, detalhando conceitos como rotas e horários.
Criar <i>templates</i> dos requisitos.	Criar <i>templates</i> para organizar os requisitos levantados para facilitar a rastreabilidade e a organização dos requisitos.	Utilizar ferramentas como Google Sheets ou Excel.	Desenvolver um <i>template</i> que organize requisitos por prioridade e categoria.

Fonte: elaborado pela autora (2024).

⁵ Disponível em: <https://ontobee.org/>. Acesso em: 12 abr. 2025.

⁶ Disponível em: <https://www.bioontology.org/>. Acesso em: 12 abr. 2025.

6.3.3 Etapas detalhadas da Fase Aquisição de Conhecimento

Nesta fase, é feita a coleta e o registro sistemático de informações do domínio, utilizando fontes documentais, revisão de literatura e elicitación com especialistas.

Quadro 17 – Etapas detalhadas da Fase Aquisição de Conhecimento

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Realizar revisão de literatura e análise de documentos.	Realizar revisão de literatura e análise de documentos de referência relevantes no domínio do conhecimento para identificar informações relevantes.	Utilizar <i>softwares</i> específicos para gerar termos com maior incidência.	Para criar uma ontologia de biodiversidade, revisar artigos científicos sobre fauna e flora amazônica.
Elicitar conhecimento com especialistas.	Elicitar conhecimento com especialistas, por meio de entrevistas, questionários ou outras técnicas apropriadas de elicitación visando obter conhecimento especializado.	Aplicar questionários ou entrevistas semiestruturadas com especialistas.	Realizar entrevistas com biólogos para entender como eles classificam espécies e habitats.
Identificar e extrair conceitos, verbos e relações relevantes.	Identificar e extrair conceitos, verbos e relações relevantes das fontes documentais, visando mapear elementos-chave do domínio.	Usar ferramentas de processamento de linguagem natural para automatizar a extração de termos e relações.	Extrair termos como "habitat", "espécie" e "bioma" de relatórios ambientais.
Examinar e interpretar dados do domínio	Examinar e interpretar dados do domínio, com base nas fontes documentais e conhecimento de especialistas para compreender o contexto usando fontes documentais.	Usar mapas mentais ou ferramentas de modelagem conceitual para organizar o conhecimento.	Analisar dados de sensores climáticos para identificar padrões de relações entre "temperatura" e "chuvas".

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Desenvolver glossários iniciais.	Desenvolver glossários iniciais contendo termos e conceitos do domínio, assegurando definições claras, consistentes e alinhadas ao contexto de aplicação.	Organizar definições iniciais e validá-las com especialistas.	Criar um glossário inicial contendo termos como "ecossistema", "predador" e "presa", com definições claras validadas por especialistas.
Documentar sistematicamente os resultados.	Documentar sistematicamente os resultados da aquisição de conhecimento em um formato padronizado, incluindo justificativas para escolhas realizadas.	Usar ferramentas como Miro ⁷ ou Lucidchart ⁸ para criar diagramas colaborativos.	Registrar todas as informações coletadas em um formato padrão, incluindo as fontes usadas e as justificativas para cada escolha.
Criar e validar um modelo preliminar.	Criar e validar um modelo preliminar para organização e representação do conhecimento adquirido, utilizando <i>feedback</i> iterativo de especialistas e gerando entregáveis como diagramas conceituais ou taxonomias iniciais.	Usar ferramentas como o Protégé ⁹ ou Lucidchart para criar diagramas conceituais iniciais.	Desenvolver diagramas conceituais no Lucidchart representando as relações entre "espécies", "biomas" e "ameaças ambientais".

Fonte: elaborado pela autora, 2024.

⁷ Disponível em: <https://miro.com/pt/>. Acesso em: 14 abr. 2025.

⁸ Disponível em: <https://www.lucidchart.com/pages>. Acesso em: 14 abr. 2025.

⁹ Disponível em: <https://protege.stanford.edu/>. Acesso em: 14 abr. 2025.

6.3.4 Etapas detalhadas da Fase Modelagem Conceitual

Nesta fase, conceitos e relações são coletados e organizados em representações estruturadas, alinhadas a padrões ontológicos.

Quadro 18 – Etapas detalhadas da Fase Modelagem Conceitual

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Importar termos genéricos e inserir termos do domínio.	Expandir o vocabulário especializado, importando termos genéricos e inserindo termos do domínio.	Utilizar APIs de repositórios ontológicos.	Importar termos como "cliente" e "fornecedor" de uma ontologia de negócios existente para um sistema ERP.
Organizar termos representativos de entidades e relacionamentos.	Organizar termos representativos de entidades e relacionamentos, criando uma taxonomia preliminar com base nos termos identificados visando estruturar relações hierárquicas.	Criar taxonomias preliminares no Protégé.	Estruturar termos em uma hierarquia, como "Animal > Vertebrado > Mamífero > Carnívoro".
Elaborar hierarquias taxonômicas.	Elaborar hierarquias taxonômicas e definir classes, propriedades e relações.	Criar hierarquias detalhadas no Protégé.	Desenvolver uma hierarquia detalhada no Protégé para representar categorias de equipamentos médicos.
Adicionar relacionamentos predefinidos e adicionais.	Selecionar e inserir relacionamentos padrão, bem como criar e inserir relacionamentos adicionais.	Validar padrões com especialistas.	Adicionar relações como "é parte de" e "é usado por" em uma ontologia de veículos.
Criar diagramas conceituais e taxonomias.	Criar diagramas conceituais e taxonomias para representar visualmente o conhecimento.	Desenvolver diagramas no Lucidchart ou Protégé.	Desenvolver diagramas no Lucidchart para representar conceitos de "fornecimento" e "armazenamento".

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Definir axiomas e restrições semânticas.	Definir axiomas e restrições semânticas, integrando validação com especialistas durante o processo para garantir consistência.	Formalizar restrições em editores de ontologias como o Protégé.	Criar restrições como "Todo veículo deve ter pelo menos 4 rodas".
Registrar definições em linguagem natural.	Documentar definições de classes, propriedades e relações em linguagem natural para assegurar entendimento claro e alinhamento ao domínio.	Redigir definições concisas e precisas para cada elemento da ontologia, revisando com especialistas para evitar ambiguidades. Utilizar ferramentas como Google Docs ou Protégé para centralizar as definições.	Em uma ontologia de turismo, registrar a definição de "Atração Turística" como: "Um local, evento ou objeto que atrai visitantes por seu valor cultural, natural ou recreativo."
Realizar ajustes conceituais.	Realizar ajustes conceituais e validar estruturas criadas com especialistas, refinando os conceitos e as relações do modelo com base em feedbacks, garantindo a validade e o alinhamento com os objetivos do domínio.	Submeter o modelo preliminar à revisão de especialistas e patrocinadores por meio de <i>workshops</i> ou entrevistas, coletar <i>feedbacks</i> e realizar ajustes necessários. Utilizar ferramentas como Protégé para aplicar as revisões e manter rastreabilidade.	Em uma ontologia de saúde pública, ajustar a hierarquia de "Doenças Infecciosas" para incluir subclasses específicas como "Viroses" e "Infecções Bacterianas", conforme orientação de epidemiologistas.

Fonte: elaborado pela autora (2024).

6.3.5 Etapas detalhadas da Fase Implementação e Formalização

Nesta fase, os conceitos modelados são traduzidos para linguagens formais, garantindo precisão lógica e semântica.

Quadro 19 – Etapas detalhadas da Fase Implementação e formalização

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Criar axiomas e restrições semânticas.	Criar formalmente os axiomas necessários para representar as relações, restrições e regras lógicas entre as classes e propriedades, garantindo uma base sólida para a inferência.	Usar Protégé para definir restrições e regras que garantam consistência lógica.	Definir no Protégé que "uma entrega só pode estar associada a uma transportadora".
Implementar a ontologia.	Implementar a ontologia utilizando linguagens formais compatíveis, como OWL, que permitem a representação precisa e interoperável do conhecimento.	Utilizar editores como Protégé ou WebProtégé para estruturar e documentar a ontologia.	Codificar uma ontologia de saúde em OWL para representar doenças, sintomas e tratamentos. Conectar com bancos de dados de registros clínicos.
Empregar um editor de ontologias.	Empregar um editor de ontologias para a estruturação, edição e refinamento, aproveitando seus recursos para simplificar o processo de modelagem.	Usar funcionalidades avançadas do Protégé, como edição de propriedades e definição de classes derivadas.	Usar o WebProtégé para organizar classes e propriedades em uma ontologia de gerenciamento de projetos, incluindo "tarefas", "recursos" e "cronogramas".
Refinar definições formais.	Refinar as definições formais por meio da criação de axiomas e restrições, com vistas a garantir consistência e validade semântica.	Realizar ciclos iterativos de revisão com especialistas.	Revisar os axiomas de uma ontologia financeira para garantir que "taxa de juros" esteja restrita a valores numéricos positivos.

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Utilizar um <i>reasoner</i> para validação automática.	Utilizar um <i>reasoner</i> para organizar e validar as inferências automáticas, classificando as classes e ajustando a hierarquia conforme necessário para garantir a precisão.	Usar <i>reasoners</i> para detectar inconsistências e garantir a precisão lógica.	Validar uma ontologia jurídica no Protégé utilizando o Hermit para verificar se "contratos inválidos" não possuem relações incorretas, como "assinado por múltiplos clientes".
Executar testes para identificar inconsistências.	Executar testes para identificar e corrigir inconsistências lógicas e estruturais, visando validar a estrutura lógica da ontologia e corrigir erros.	Testar cenários hipotéticos no Protégé para verificar a consistência das relações.	Simular cenários em uma ontologia de ensino, como verificar se "um professor" deve estar associado apenas a "uma disciplina específica".
Garantir a interoperabilidade da ontologia.	Garantir a interoperabilidade da ontologia ao conectá-la com sistemas existentes e bancos de dados, possibilitando sua aplicação prática e funcional em cenários reais.	Integrar a ontologia com APIs e ferramentas de integração semântica.	Integrar uma ontologia de produtos agrícolas com APIs de sistemas de rastreamento de transporte, permitindo consultas em SPARQL.

Fonte: elaborado pela autora (2024).

6.3.6 Etapas detalhadas da Fase Avaliação

Nesta fase, avalia-se a consistência lógica, a qualidade, a adequação aos objetivos e a aplicabilidade da ontologia.

Quadro 20 – Etapas detalhadas da Fase Avaliação

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Verificar consistência lógica.	Verificar a consistência lógica da ontologia, avaliando sua estrutura e relações.	Utilizar <i>reasoners</i> no Protégé para validar a consistência lógica das inferências.	Usar um <i>reasoner</i> para verificar se todas as subclasses de "móveis" em uma ontologia de <i>design</i> de interiores possuem a propriedade "fabricado em algum material".
Realizar testes práticos.	Realizar testes práticos para identificar problemas e garantir que a ontologia atende aos requisitos do domínio.	Testar a ontologia em ambientes simulados ou reais, verificando se as inferências são consistentes com os dados.	Em um sistema de gerenciamento de estoque, testar se a ontologia permite buscar "produtos com validade próxima do vencimento".
Avaliar a compatibilidade com sistemas e padrões.	Avaliar se a ontologia é compatível com os sistemas e padrões já utilizados no domínio.	Realizar testes de interoperabilidade com sistemas que utilizem OWL ou RDF.	Validar se uma ontologia de engenharia civil é compatível com o padrão BIM (<i>Building Information Modeling</i>).
Identificar possíveis erros e lacunas na ontologia.	Identificar possíveis erros e lacunas na ontologia, com foco na melhoria contínua do modelo ontológico.	Revisar o modelo com especialistas do domínio e documentar os ajustes necessários.	Revisar com especialistas de recursos humanos uma ontologia de carreiras para identificar conceitos ausentes, como "progressão de cargos".
Implementar correções e otimizações.	Implementar correções e otimizações a partir das análises anteriores e de <i>feedbacks</i> recebidos.	Atualizar a ontologia no Protégé, garantindo que os ajustes atendam às necessidades levantadas.	Ajustar definições na ontologia de um aplicativo de <i>e-commerce</i> , corrigindo a relação entre "descontos" e "produtos em promoção".

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Realizar ajustes finais na ontologia.	Realizar ajustes finais na ontologia para garantir qualidade e alinhamento com os objetivos iniciais.	Fazer uma revisão geral com todos os envolvidos e validar o modelo final.	Após testes em uma ontologia de ensino a distância, adicionar a propriedade "modalidade" para diferenciar aulas síncronas e assíncronas.
Certificar-se de que a ontologia alcança os resultados esperados.	Certificar-se de que a ontologia alcança os resultados esperados e responde às questões de competência definidas e demais requisitos previstos no ORSD.	Usar métricas de avaliação definidas no ORSD para verificar a completude e a aplicabilidade da ontologia.	Verificar que uma ontologia de turismo responde corretamente a consultas como "lugares para visitar na região nordeste com atrações culturais".

Fonte: elaborado pela autora (2024).

6.3.7 Etapas detalhadas da Fase Documentação

Nesta fase, decisões, processos e estrutura da ontologia são registrados para facilitar a manutenção futura, com vistas a criar uma documentação clara e acessível para uso e manutenção.

Quadro 21 – Etapas detalhadas da Fase Documentação

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Organizar todos os materiais e decisões tomadas.	Organizar todos os materiais, fontes e decisões utilizados ao longo do desenvolvimento da ontologia, garantindo a rastreabilidade e o suporte para futuras atualizações.	Usar ferramentas de gestão de conhecimento para armazenar decisões e materiais.	No desenvolvimento de uma ontologia de produção industrial, registrar as justificativas para cada escolha de classe e propriedade em um documento compartilhado.

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Elaborar relatórios técnicos.	Elaborar relatórios técnicos que detalhem os resultados obtidos em cada etapa do processo, incluindo decisões tomadas e justificativas.	Criar relatórios formais utilizando ferramentas como LaTeX ou Word.	Gerar um relatório técnico detalhado com gráficos de hierarquias e exemplos de consultas SPARQL de uma ontologia de gestão ambiental.
Desenvolver representações visuais.	Desenvolver representações visuais que ilustrem a estrutura da ontologia, como hierarquias de classes, relações e axiomas.	Usar <i>Lucidchart</i> ou <i>PowerPoint</i> para criar representações visuais acessíveis.	Criar um diagrama que ilustre as relações entre "profissões", "habilidades" e "setores de trabalho" para uma ontologia de mercado de trabalho.
Produzir um documento abrangente da ontologia.	Produzir um documento abrangente contendo definições, exemplos e aplicações práticas da ontologia, para facilitar seu uso e entendimento.	Estruturar as informações de forma técnica e prática, incluindo exemplos de uso.	Compilar um documento contendo exemplos de uso, como consultas práticas em uma ontologia de seguros, para auxiliar novos usuários.
Preparar materiais técnicos.	Preparar materiais técnicos, incluindo especificações formais e instruções de uso, para apoiar desenvolvedores e usuários finais.	Consolidar os principais resultados e informações em um manual de referência.	Escrever um guia técnico com instruções para carregar e consultar uma ontologia jurídica em um sistema de busca semântica.
Elaborar um resumo executivo.	Sintetizar as informações principais da ontologia em um formato resumido, oferecendo uma visão geral acessível para novos usuários e partes interessadas.	Desenvolver um guia prático com os principais resultados.	Elaborar um resumo executivo de uma ontologia de gestão hospitalar, destacando classes principais como "médicos", "pacientes" e "procedimentos".

Fonte: elaborado pela autora (2024).

6.3.8 Etapas detalhadas da Fase Publicação

Nesta fase, a ontologia é disponibilizada e divulgada, garantindo acessibilidade e visibilidade.

Quadro 22 – Etapas detalhadas da Fase Publicação

(Continua...)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Realizar testes em ambientes reais.	Realizar testes em ambientes reais para avaliar a aplicabilidade e o desempenho da ontologia em diferentes cenários.	Implementar a ontologia em sistemas reais e monitorar seu uso.	Testar uma ontologia de gestão de transporte público em um sistema de bilhetagem eletrônica, verificando sua funcionalidade em consultas sobre "linhas e horários de ônibus".
Avaliar funcionalidade em sistemas específicos.	Avaliar a funcionalidade da ontologia em sistemas específicos, identificando possíveis incompatibilidades ou melhorias necessárias.	Realizar integrações experimentais e documentar os resultados.	Integrar uma ontologia de gerenciamento agrícola em um sistema de IoT, validando sensores que monitoram a umidade do solo.
Incorporar sugestões e observações recebidas.	Incorporar as sugestões e observações recebidas durante os testes práticos e avaliações, refinando a ontologia com base em <i>feedbacks</i> recebidos.	Realizar revisões iterativas e ajustar conforme necessário.	Ajustar uma ontologia de gerenciamento acadêmico com base no <i>feedback</i> de professores que solicitam novas categorias de "cursos híbridos".
Configurar a ontologia para ferramentas específicas	Configurar a ontologia para funcionar de maneira eficiente com ferramentas e plataformas relevantes ao domínio.	Criar conexões com APIs e sistemas de suporte semântico.	Configurar uma ontologia de recursos humanos para funcionar com ferramentas de análise de dados.

(Conclusão.)

ETAPA	DESCRIÇÃO	ALTERNATIVAS PARA EXECUÇÃO	EXEMPLO
Disponibilizar a ontologia em repositórios públicos.	Disponibilizar a ontologia em repositórios específicos, como BioPortal e OBO Foundry, garantindo o acesso público e a disseminação do trabalho.	Disponibilizar a ontologia em ferramentas como GitHub, de maneira que ela possa ser acessada, compartilhada e reutilizada por outros desenvolvedores e pesquisadores.	Publicar uma ontologia de biodiversidade no BioPortal, incluindo metadados e exemplos de uso.
Divulgar resultados do trabalho.	Divulgar a ontologia e os resultados do trabalho para comunidades científicas e técnicas, ampliando o alcance e incentivando colaborações futuras.	Publicar artigos científicos e apresentar em conferências técnicas.	Apresentar uma ontologia de saúde em um congresso científico, destacando sua aplicação em prontuários eletrônicos.

Fonte: elaborado pela autora (2024).

6.4 Prova de conceito para Aquisição de Conhecimento no Domínio de Compras Públicas

Com o objetivo de validar parcialmente a metodologia proposta, realizou-se uma prova de conceito, aplicando a *Onto Merge Methodology* (OMM) no domínio de compras públicas. A escolha desse domínio decorre de sua relevância estratégica para o governo eletrônico, da disponibilidade de dados públicos e da experiência prática da pesquisadora na área.

A prova de conceito concentrou-se na fase de Aquisição de Conhecimento da OMM, mais especificamente na identificação automática de conceitos e relações semânticas presentes em documentos normativos. Para tanto, empregaram-se técnicas de mineração de texto e de processamento de linguagem natural, aliadas a modelos de linguagem de grande escala (LLM), visando à automação parcial das tarefas.

O processo adotado concentrou-se na exploração da viabilidade de cada tarefa de forma individual, permitindo a validação de múltiplas etapas em escala reduzida, em vez de um procedimento exaustivo em etapa única.

Para a coleta e processamento dos documentos, foi desenvolvido um *webcrawler*¹⁰, responsável por extrair dados textuais relevantes. Esses documentos foram convertidos em texto simples, submetidos a pré-processamento e, posteriormente, enviados para análise por meio de uma LLM.

Em relação à fonte das informações coletadas na prova de conceito, foi considerado o *site* oficial do governo de compras públicas. Dessa maneira, em um cenário de produção, a página principal poderia ser considerada como um ponto de partida para o *webcrawler*. Para orientar de forma mais clara o *crawler*, este poderia ser direcionado a explorar subpáginas tematicamente relevantes, como “Legislação” e “Manuais” pulando um nível de profundidade de navegação.

Como o foco recaiu sobre documentos de natureza legislativa, identificou-se a necessidade de uma etapa específica de processamento textual, aplicável a todos os textos provenientes dessa mesma fonte, de modo a abranger os diversos casos pertinentes à prova de conceito.

Inicialmente, as passagens relevantes foram extraídas do arquivo HTML, mantendo-se apenas os trechos que integravam diretamente a peça legislativa analisada. Trechos referentes a alterações em outras normas foram desconsiderados, por extrapolarem o escopo da análise e exigirem verificação posterior do texto completo.

Com base no resultado da etapa anterior, cada segmento textual foi submetido a um processo de limpeza, removendo-se marcadores e caracteres que pudessem comprometer a análise automatizada. Caracteres especiais, como “§”, foram substituídos por “parágrafo”; abreviações, como “arts.”, por “artigos”; e sinais cuja função não era semântica, como “.”, foram eliminados. Trechos autorreferenciais, como “desta Lei”, foram substituídos por referências genéricas, por exemplo, “Lei de Licitações e Contratos Administrativos”. Finalmente, artefatos de ordenação (numerais romanos, letras etc.) foram convertidos em contagem numérica simples.

Ao término desse processamento, obteve-se uma lista de trechos, cada qual correspondendo a um elemento único da legislação (artigo, parágrafo, inciso ou alínea). Em razão da estrutura hierárquica do texto jurídico, esses trechos estão vinculados aos seus níveis superiores — alíneas a incisos, incisos a parágrafos —,

¹⁰ *Webcrawler* trata-se de um programa automatizado que percorre a *World Wide Web* de maneira sistemática, seguindo *hiperlinks* para coletar em larga escala páginas, metadados e outros objetos digitais, criando assim repositórios ou índices que alimentam mecanismos de busca, mineração ou arquivamento.

motivo pelo qual cada unidade de análise foi acompanhada de sua referência hierárquica imediata. Essa abordagem apresenta desafios, como estruturas ordenadas complexas e listagens incompletas, que deverão ser consideradas em implementações de maior escala.

Para a escolha do modelo de linguagem, optou-se por ferramentas gratuitas e de código aberto, capazes de processar texto em português e passíveis de ajuste via instruções. Um requisito central era a viabilidade de execução em ambiente local. Assim, diante das limitações técnicas da prova de conceito, priorizaram-se modelos que pudessem ser executados com até 12 GB de VRAM, assegurando o melhor desempenho possível no contexto.

Foram feitos testes com modelos como o *google/gemma-7b*, *google/gemma-2-9b*, *meta-llama/Llama-3.1-8B*. Entretanto, com a publicação dos modelos da empresa chinesa *Deepseek*, em especial, o modelo “raciocinador” R1, avaliou-se que esse modelo seria mais adequado para utilização na prova de conceito, visto que poderia possibilitar maior auditabilidade dos resultados com os trechos de “raciocínio”. Portanto, foi escolhida a versão “destilada” do modelo R1, baseada no modelo *llama*, “*deepseek-ai/DeepSeek-R1-Distill-Llama-8B*”. Além disso, a boa qualidade de resultados de forma generalista do modelo possibilitava mais flexibilidade para seguir as instruções específicas e detalhadas envolvidas no processo de extração e sumarização.

O modelo foi utilizado de forma quantizada em 4 bits, adotando-se mecanismo de atenção de “*Scaled Dot-Product Attention*”. Foram usados também modelos de transformação para geração de *embeddings*¹¹ para pesquisa semântica, “*ricardo-filho/bert-base-portuguese-cased-nli-assin-2*”, e modelos para processamento de linguagem natural, “*pt_core_news_lg*”.

Para a demanda específica de extração e sumarização de conceitos, foi estruturada uma arquitetura de *Retrieval-augmented Generation* (RAG), com encadeamento de *prompts* de funções distintas. A partir de um termo ou expressão, realizava-se busca na base de *embeddings* obtida do texto analisado. Em seguida,

¹¹ *Embedding* trata-se de uma representação compacta em forma de vetor numérico atribuída a cada palavra, frase ou documento. Nesse espaço vetorial, elementos com significados semelhantes ocupam posições próximas, permitindo que algoritmos comparem e recuperem conteúdos relacionados com rapidez e precisão.

um número “n” de trechos era retornado e fornecido ao modelo junto a um *prompt*, conforme ilustrado no Quadro 23, a seguir:

Quadro 23 – *Prompt 1*

<p>### Quem é você:</p> <p>- Você é um modelo de linguagem avançado especializado em extrair informações de texto para modelagem conceitual.</p> <p>- Sua tarefa é analisar o texto fornecido e extrair as informações disponíveis **exclusivamente do contexto apresentado**.</p> <p>### Instruções:</p> <p>1. Identifique todas as informações **explícitas ou implicitamente relacionadas** ao termo "{query}" no texto abaixo.</p> <p>2. Liste cada informação relevante no formato:</p> <p>- {query} - [Informação específica]</p> <p>3. **Não inclua suposições ou conhecimentos externos**.</p> <p>4. Caso o termo não seja mencionado ou não haja informações relacionadas, retorne exatamente: "sem informações encontradas".</p> <p>### Exemplo Orientativo:</p> <p>Texto: "Os contratos devem seguir normas específicas e são supervisionados pelos gestores."</p> <p>Termo: "contratos"</p> <p>Resposta:</p> <p>- contratos - Devem seguir normas específicas</p> <p>- contratos - São supervisionados pelos gestores</p> <p>### Sua tarefa</p> <p>Com base no texto abaixo, faça a extração de informações no formato solicitado.</p> <p>### Texto para análise:</p> <p>{context}</p> <p>### Termo analisado:</p> <p>"{query}"</p>
--

Fonte: elaborado pela autora (2024).

Em seguida, o resultado desse *prompt*, sem os trechos de “raciocínio”, foi novamente encaminhado ao mesmo modelo, segundo o *prompt* apresentado no Quadro 24, a seguir:

Quadro 24 – *Prompt 2*

<p>### Quem é você:</p> <p>- Você é um especialista em criação de glossários, focado em extrair informações específicas de textos especiais.</p> <p>Dessa forma, seu processo de extração precisa analisar somente as informações presentes no texto fornecido e a relação entre elas.</p> <p>### Instruções:</p> <ol style="list-style-type: none"> 1. Analise o texto fornecido para identificar todas as informações relacionadas ao termo "{query}". 2. Resuma e organize as informações disponibilizadas em um único trecho com as informações coletadas: <p>- {query} - [Definição]</p> <ol style="list-style-type: none"> 3. Use apenas dados presentes no texto; **não adicione conhecimento externo**. 4. Se não houver informações, retorne exatamente: "sem informações encontradas". <p>### Exemplo Orientativo:</p> <p>#### Texto para análise:</p> <ul style="list-style-type: none"> - Contratos - São acordos legais - Contratos - Devem seguir normas específicas - Contratos - Exigem conformidade <p>#### Termo analisado: "Contratos"</p> <p>#### Resposta:</p> <p>[INICIO DA RESPOSTA]</p> <p>Contratos - Contratos são acordos legais que exigem conformidade e devem seguir normas específicas.</p> <p>[FIM DA RESPOSTA]</p> <p>### Sua tarefa</p> <p>Com base no texto abaixo, crie uma entrada no glossário para o termo {query}.</p>

Texto para análise:

{context}

Fonte: elaborado pela autora (2024).

Esse processo foi repetido para os termos-chave, definidos em etapas anteriores como mais relevantes, viabilizando a extração de uma lista de atributos e a elaboração de uma “definição” pautada nas informações contidas no material analisado. Abaixo, segue exemplo de resultado desse processo para o termo “Licitação”:

Licitação – São modalidades de licitação: pregão; concorrência; concurso; leilão; diálogo competitivo; procedimentos auxiliares: credenciamento; processo de licitação observa fases: divulgação do edital; homologação; Lei de Licitações e Contratos Administrativos entra em vigor na data de sua publicação; aplica-se a contratações de tecnologia da informação e de comunicação.

Considerando esse fluxo integrado de processamento textual, análise semântica e aprendizado, testou-se também a utilização direta do modelo para extração de relações, conforme ilustrado no Quadro 25. Os resultados, entretanto, foram considerados insatisfatórios, dada a dificuldade de isolar entidades dos qualificadores e de computar múltiplas entradas para a mesma entidade.

Quadro 25 – Prompt 3

Quem é você:

- Você é um modelo de linguagem avançado especializado em extrair informações de texto para modelagem conceitual.

- Sua tarefa é analisar o texto fornecido e identificar as entidades mencionadas, as relações entre elas e os conectores semânticos que as unem.

- Você vai se basear exclusivamente no contexto apresentado. Não inclua conhecimentos externos. Siga estas instruções:

Suas Instruções:

1. Identifique **todas** as **entidades** mencionadas no texto.

2. Determine **todas** as **relações explícitas ou implícitas** que essas entidades possuem com outras entidades, conforme descrito no texto.

3. Liste as informações extraídas no seguinte formato:

- [Entidade 1] - [conector semântico] - [Entidade 2]

4. Caso não encontre informações relevantes sobre as entidades ou suas relações, retorne "sem informações".

Exemplo de como deve ser feito:

Texto de exemplo:

"Os gestores têm a responsabilidade de supervisionar os contratos, e cada contrato pode ser considerado uma obrigação vinculada à administração pública. Os prazos contratuais previstos nesta Lei não excluem nem revogam os prazos contratuais previstos em lei especial."

Resposta esperada:

[INICIO DA RESPOSTA]

- gestores - tem - responsabilidade

- supervisionar - é uma - responsabilidade

- supervisionar - é aplicado sobre - contratos

- contrato - pode ser considerado uma - obrigação

- obrigação - pode estar vinculada a - administração pública.

- Prazos contratuais previstos nesta lei - não excluem - prazos contratuais previstos em lei especial.

- Prazos contratuais previstos nesta lei - não revogam - prazos contratuais previstos em lei especial.

[FIM DA RESPOSTA]

Sua tarefa

Com base no texto abaixo, faça a extração de informações no formato solicitado.

{context}

Fonte: elaborado pela autora (2024).

A aplicação prática da OMM, nesse contexto, possibilitou identificar e estruturar um conjunto de conceitos relevantes para o domínio de compras públicas. Como resultado, foram extraídas triplas semânticas e elaborado um glossário inicial a partir dos dados textuais normativos.

Fatores como a escolha do modelo, formulação do *prompt*, treinamento adicional, ou mesmo melhorias na implementação e arquitetura do sistema podem contribuir com um melhor desempenho individual de cada *prompt*, garantindo uma extração mais detalhada e com menos casos de resultados indevidos. Entretanto, concluiu-se que o uso e a articulação de estratégias combinadas — associando

análises de PLN mais simples, como extração de substantivos, ao modelo de linguagem — têm potencial ainda maior para gerar resultados consistentes. Tal estratégia é mais interessante, pois alavanca os resultados mais determinísticos das ferramentas de PLN com a flexibilidade de interpretação e manipulação de texto dos LLM.

Apesar dos avanços, algumas limitações permanecem, por exemplo, em relação à etapa de avaliação e curadoria dos conceitos extraídos que ainda demandam validação humana, restringindo, em parte, os ganhos de uma automatização integral. Mas, de forma geral, para as etapas selecionadas da fase de aquisição de conhecimento para a prova de conceito, quais sejam “Realizar revisão de literatura e análise de documentos”, “Identificar e extrair conceitos, verbos e relações relevantes”, “Examinar e interpretar dados do domínio” e “Desenvolver glossários iniciais”, pode-se considerar que o processo foi bem sucedido. O código e arquivos pertinentes à prova de conceito podem ser acessados por meio do link: <https://github.com/afflz/ontomerge>.

Conclui-se, por fim, que a aplicação prática da OMM demonstrou viabilidade técnica e ganhos de eficiência na etapa de aquisição de conhecimento, validando parcialmente o modelo proposto. Os resultados indicam que a automatização, ainda que parcial, pode tornar o processo de aprendizagem de ontologias mais escalável, replicável e adequado à realidade do governo eletrônico brasileiro. Todavia são necessárias validações adicionais em outros domínios e fases da metodologia, bem como aprimoramentos técnicos para superar os desafios identificados.

7 CONSIDERAÇÕES FINAIS

Esta tese teve como objetivo principal propor uma metodologia de aprendizagem de ontologias voltada à organização e representação do conhecimento no contexto do governo eletrônico brasileiro. Para atingir esse objetivo, foram definidos quatro objetivos específicos: (i) mapear o estado da arte da aprendizagem de ontologias; (ii) identificar Sistemas de Organização do Conhecimento (SOC) utilizados nos dez países mais bem colocados no Índice Geral de Desenvolvimento de Governo Eletrônico (EGDI) da ONU; (iii) inventariar ontologias desenvolvidas no âmbito do governo brasileiro; e (iv) identificar possibilidades de automatização de fases e etapas de quatro metodologias específicas de construção de ontologias. Ao longo do trabalho, os pressupostos formulados orientaram criticamente a análise dos dados e subsidiaram a elaboração da proposta metodológica apresentada.

O primeiro objetivo foi alcançado por meio de uma Revisão Sistemática da Literatura (RSL), que permitiu evidenciar o estágio atual das pesquisas em *Ontology Learning*. A análise dos estudos selecionados revelou a predominância de abordagens centradas na extração automática de conhecimento a partir de dados textuais e estruturados, com ênfase em técnicas de PLN e mineração de texto. Contudo observou-se uma carência de metodologias sistematizadas e aplicáveis a contextos institucionais e governamentais, particularmente na realidade brasileira. Assim, o primeiro pressuposto — de que a aprendizagem de ontologias constitui um campo emergente com fragilidades metodológicas — foi confirmado. A literatura carece de propostas que integrem de maneira robusta etapas formais, recursos automatizados e aderência a demandas organizacionais reais.

O segundo objetivo foi atendido por meio de uma análise documental dos portais e repositórios de dados abertos dos países mais bem avaliados no EGDI. A investigação demonstrou que a adoção de SOC, incluindo ontologias, é prática consolidada em diversos contextos internacionais, estando frequentemente associada à interoperabilidade semântica, à transparência de dados e à prestação de serviços digitais. No entanto, a aplicação das ontologias ocorre de modo articulado a diretrizes estratégicas nacionais e a padrões técnicos compartilhados. A partir desse cenário, o segundo pressuposto — de que as iniciativas brasileiras são ações isoladas, sem a adoção de uma metodologia unificada — foi também confirmado. No Brasil, apesar de experiências pontuais em instituições públicas, não foram identificados esforços

coordenados de padronização ontológica, o que compromete a integração semântica e dificulta a criação de uma infraestrutura de conhecimento comum no governo eletrônico.

O terceiro objetivo da pesquisa consistiu em inventariar ontologias desenvolvidas no contexto do governo brasileiro. A análise documental de plataformas como o GovBR, o Portal de Dados Abertos e a Biblioteca Digital do Senado evidenciou uma baixa incidência de ontologias publicamente acessíveis e reutilizáveis. Os poucos registros encontrados demonstram a ausência de um ecossistema ontológico integrado, além da inexistência de diretrizes nacionais para desenvolvimento, catalogação e publicação desses artefatos. Com isso, o terceiro pressuposto — que propunha a escassez de ontologias formalizadas e acessíveis no setor público brasileiro — foi confirmado. A realidade observada indica não apenas a baixa adoção, mas também a invisibilidade e a falta de governança em relação às ontologias existentes.

O quarto objetivo foi abordado por meio de uma análise comparativa de quatro metodologias desenvolvidas no PPGOC/UFMG: OntoForInfoScience, OntONEo, OntoAccount e PRONTO. A comparação sistemática dessas metodologias permitiu identificar suas fases, etapas e técnicas, destacando semelhanças e diferenças estruturais. Verificou-se que, embora bem organizadas, essas metodologias ainda apresentam um baixo nível de automatização, principalmente nas etapas de aquisição de conhecimento, modelagem conceitual e avaliação. Com isso, o quarto pressuposto — segundo o qual a inexistência de uma metodologia com etapas automatizadas contribui para o baixo uso de ontologias no governo eletrônico — foi igualmente confirmado. A complexidade do processo de construção de ontologias, quando inteiramente manual, torna-se um fator limitante à sua adoção em ambientes que demandam escalabilidade e agilidade, como a administração pública.

Como principal contribuição, esta tese propôs a Onto Merge Methodology (OMM), uma metodologia híbrida e adaptável, que combina elementos das metodologias analisadas com sugestões práticas de automatização. A OMM estrutura-se em oito fases e múltiplas etapas, integrando processos manuais, semiautomatizados e recursos de apoio ao processo de construção de ontologias. A proposta foi validada por meio de uma prova de conceito no domínio das compras públicas, demonstrando sua aplicabilidade, clareza metodológica e potencial de replicabilidade em outros contextos.

Ainda assim, é necessário reconhecer algumas limitações da pesquisa. Em primeiro lugar, a validação da OMM foi restrita a um único domínio. Embora o campo das compras públicas tenha oferecido um ambiente controlado e com dados acessíveis, outras áreas — como saúde, educação ou segurança — podem demandar adaptações nas etapas de aquisição de conhecimento, tratamento terminológico e validação. Em segundo lugar, nem todas as fases da OMM são passíveis de automatização integral. Etapas como a formalização de axiomas e a avaliação ontológica ainda exigem julgamento humano, interpretação contextual e refinamento semântico, o que limita o escopo de automação. Dessa forma, entende-se que a automatização total da OMM não é viável em curto prazo, mas é possível avançar significativamente por meio da incorporação progressiva de recursos de inteligência artificial e aprendizado de máquina.

A implementação da OMM em ambientes reais de governo eletrônico pode encontrar obstáculos técnicos (infraestrutura precária e sistemas heterogêneos), organizacionais (falta de capacitação e de cultura de interoperabilidade) e políticos (descontinuidade administrativa e ausência de diretrizes nacionais). Enfrentar esses desafios exigirá investimentos em capacitação, políticas públicas de gestão do conhecimento e ações normativas que incentivem a adoção de ontologias interoperáveis e reutilizáveis.

A realização desta pesquisa também proporcionou amadurecimento acadêmico e profissional à autora, sobretudo no enfrentamento de desafios interdisciplinares e na articulação entre teoria, prática e inovação. A construção da OMM exigiu aprofundamento técnico em engenharia ontológica, experimentação com ferramentas, domínio de procedimentos metodológicos e sensibilidade para traduzir necessidades institucionais em soluções estruturadas. Entre os principais aprendizados, destacam-se a importância da clareza conceitual, da modularidade metodológica e da escuta ativa ao contexto de aplicação.

Os resultados obtidos ao longo desta pesquisa também proporcionaram duas publicações científicas que ampliam sua contribuição para a comunidade acadêmica. A primeira, intitulada “Metodologias sobre aprendizagem de ontologias: uma revisão sistemática de literatura”¹², apresenta uma análise abrangente das abordagens de

¹² Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/aprendizagem-de-ontologias>. Acesso em: 12 abr. 2025.

geração automática e semiautomática de ontologias, com ênfase nos desafios metodológicos enfrentados por propostas que carecem de etapas bem definidas e replicáveis. O estudo reforça a necessidade de metodologias mais detalhadas e sistemáticas, alinhando-se aos objetivos desta tese e antecipando discussões aprofundadas na formulação da OMM. A segunda publicação, “Sistemas de Organização do Conhecimento (SOC) adotados em países mais bem posicionados no Índice Geral de Desenvolvimento de Governo Eletrônico da ONU (EGDI)”¹³, oferece uma análise documental comparativa sobre como nações líderes em governo eletrônico estruturam seus dados públicos por meio de SOC, com destaque para o uso de ontologias e vocabulários controlados em plataformas digitais. Essa investigação forneceu subsídios empíricos essenciais para a análise do contexto internacional e para a reflexão sobre os desafios da realidade brasileira frente à modernização da gestão do conhecimento no setor público.

Para pesquisas futuras, recomenda-se: (i) expandir a validação da OMM a diferentes domínios do setor público; (ii) desenvolver indicadores de maturidade ontológica para instituições governamentais; (iii) aprimorar os módulos automatizados da OMM com uso de técnicas de PLN e IA generativa; (iv) investigar sua integração a arquiteturas de governo digital, como plataformas interoperáveis e sistemas de serviços públicos; e (v) propor políticas de governança ontológica para uso federado, colaborativo e transparente das ontologias públicas.

Conclui-se, portanto, que a *Onto Merge Methodology* representa um avanço teórico, metodológico e prático para o campo da aprendizagem de ontologias e para a modernização da gestão da informação no setor público. Ao integrar rigor metodológico, modularidade, potencial de automatização e aplicabilidade real, a OMM oferece um caminho promissor para o fortalecimento da interoperabilidade semântica e da inteligência institucional no governo eletrônico brasileiro.

¹³ Disponível em: <https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/12630>. Acesso em: 12 abr. 2025.

REFERÊNCIAS

- AGANETTE, E. C. **Taxonomias Corporativas: Um Estudo Sobre Definições e Etapas de Construção Fundamentado na Literatura Publicada**. 2010. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.
- AGGOUNE, A. Automatic ontology learning from heterogeneous relational databases: Application in alimentation risks field. *In*: AMINE, A; MOUHOU, M.; MOHAMED, A.; DJEBBAR, B. (ed.) **Computational Intelligence and Its Applications**. Cham, Alemanha: Springer International Publishing, 2018. p. 199-210.
- ALHARBI, R.; TAMMA, V.; GRASSO, F. Characterising the gap between theory and practice of ontology reuse. *In*: PROCEEDINGS OF THE 11TH ON KNOWLEDGE CAPTURE CONFERENCE, 11., 2021, New York, NY, United States. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 217-224.
- ALMEIDA, M. B. **Ontologia na prática: projeto, metodologia e construção**. Curitiba: CRV, 2022. v. 01.
- ANDERSON, D. M.; FYALL, R.; HOFFMAN, S. P. Overcoming coordination challenges in the U.S. environmental policy arena: The case for a cross-sectoral approach. **Public Performance & Management Review**, United States, v. 41, n. 1, p. 1-29, 2018.
- ARAÚJO, W. J. **Proposta metodológica para enriquecimento de ontologias de domínio**. 2021. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2021.
- ASIM, M. *et al.* A survey of ontology learning techniques and applications. **Database**, United States, v. 2018, p. bay101, 2018a.
- ASIM, M., *et al.* Ontology Learning Techniques: A Review and Future Directions. **Information Processing & Management**, United Kingdom, v. 54, n. 6, p. 951-986, 2018b.
- ATLASTI. Disponível em: <https://atlasti.com>. Acesso em: 27 abr. 2025.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições, 2011. v. 70.
- BELHOUCINE, K.; MOURCHID, M. **A Survey of Ontology Learning from Text**, 2018. Disponível em: https://www.thinkmind.org/index.php?view=article&articleid=semapro_2018_1_40_30_027. Acesso em: 21 out. 2019.
- BERTOT, J. C.; JAEGER, P. T.; GRIMES, J. M. Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. **Government Information Quarterly**, United States, v. 27, n. 3, p. 264-271, 2010.

BIOONTOLOGY. Disponível em: <https://www.bioontology.org/>. Acesso em: 12 abr. 2025.

BRASCHER, M.; CAFÉ, L. Organização do Conhecimento: Subsídios para uma Disciplina da Ciência da Informação. **Em Questão**, Natal, RN, v. 14, n. 1, p. 31-43, 2008.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento. *In*: IX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: [s.n], 2008. p. 1-14. Disponível em: [http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89\(2008\)-1835.pdf](http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89(2008)-1835.pdf). Acesso em: 08 dez. 2015.

BRASIL. **Lei n. 12.527 de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5.º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei n. 8.112, de 11 de dezembro de 1990; revoga a Lei n. 11.111, de 5 de maio de 2005, e dispositivos da Lei n. 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília: Congresso Nacional, 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 24 jul. 2024.

BRASIL. **Lei n. 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília: Congresso Nacional, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 24 jul. 2024.

CAMPOS, M. L.; GOMES, H. E. Taxonomia e Classificação: a categorização como princípio. *In*: **VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação**, 8., Salvador, 2007. **Anais...** Salvador: [s.n], 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf>>. Acesso em: 21 maio 2016.

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

CARLAN, E.; MEDEIROS, M. B. B. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 4, n. 1, 2011.

CARLAN, G.; BRASCHER, M. Ontologies: What They Are, What They Do, How to Build One. **Knowledge Organization**, Germany, v.38, n. 2, p. 101-114, 2011.

CASTRO, C. A. F.; SANTOS, A. L. M.; PEREIRA, H. R. Segurança cibernética no governo eletrônico: desafios e perspectivas. **Revista de Administração Pública**, Rio de Janeiro, v. 51, n. 5, p. 667-687, 2017.

CHADWICK, A. **Internet Politics: States, Citizens, and New Communication Technologies**. Oxford: Oxford University Press, 2006.

CONTRIBUCIONES. Disponível em:

<https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/12630>. Acesso em: 12 abr. 2025.

CRUZ, M. **O governo eletrônico no Brasil: desafios e perspectivas**. Brasília: Editora UnB, 2001.

CRUZ, R. S. E-Government in Singapore. **International Journal of Electronic Government Research (IJEGR)**, United States, v. 1, n. 3, p. 1-20, 2001.

DADOS ABERTOS. Disponível em: <https://dados.gov.br>. Acesso em: 12 abr. 2025.

DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What is a knowledge representation?. **AI magazine**, United States, v. 14, n. 1, p. 17, 1993.

DEB, C. K. *et al.* A framework for ontology learning from taxonomic data. *In: BIG DATA ANALYTICS*, 2018, Singapore. **Proceedings of CSI 2015**. Singapore: Springer Singapore, 2018. p. 29-37.

DESUL, S. *et al.* Method for automatic key concepts extraction: Application to documents in the domain of nuclear reactors. **The Electronic Library**, United Kingdom, v. 37, n. 1, p. 2-15, 2019.

DILMEGANI, C.; KORKMAZ, B.; LUNDQVIST, M. Public-sector digitization: The trillion-dollar opportunity. **McKinsey & Company**, Germany, 2014. Disponível em: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/public-sector-digitization-the-trillion-dollar-opportunity>. Acesso em: 24 jul. 2024.

DRISSI, A. *et al.* A New Automatic Ontology Construction Method Based on Machine Learning Techniques: Application on Financial Corpus. *In: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON MANAGEMENT OF DIGITAL ECOSYSTEMS*, 13., 2021, New York, NY, USA. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 57-61.

ENEA, R. *et al.* How to Support Human Operator in "Uncertainty" Managing during the Ontology Learning Process. *In: COMPANION PROCEEDINGS OF THE THE WEB CONFERENCE 2018 (WWW '18)*, 18., 2018, Lyon, France. **Proceedings** [...] Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018. p. 1147-1154.

FARINELLI, F. **Realismo ontológico aplicado a interoperabilidade semântica entre sistemas de informação: um estudo de caso do domínio obstétrico e neonatal**. 2017. Tese (Doutorado em Gestão e Organização do Conhecimento) – Faculdade de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

FIGUERES-ESTEBAN, M.; HUGHES, P.; VAN GULIJK, C. Ontology network analysis for safety learning in the railway domain. *In: EUROPEAN SAFETY AND RELIABILITY CONFERENCE*, 26., 2016, Scotland. **Proceedings** [...] Scotland: ESREL 2016, 2016.

FONSECA, R. S. **Modelagem de conhecimento baseada em ontologia e dados abertos**: a representação do domínio da tomada e prestação de contas anual dos gestores públicos federais. 2023. Tese (Doutorado em Gestão e Organização do Conhecimento) – Faculdade de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2023.

FREIRE, M. P.; ROCHA, A. C. P.; OLIVEIRA, J. M. Inclusão digital no Brasil: desafios e políticas públicas. **Revista Brasileira de Políticas Públicas**, Brasília, v. 10, n. 2, p. 455-470, 2020.

GIL-GARCIA, J. R.; HELBIG, N.; OJO, A. **Enacting Electronic Government Success**: An Integrative Study of Government-wide Websites, Organizational Capabilities, and Institutions. Berlim, Alemanha: Springer, 2020.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. São Paulo: Editora Atlas, 2009.

GLUSHKO, R. J. Foundations for Information Organizing and retrieval, and Use. *In*: GLUSHKO, R. J. (ed.). **Rhe Discipline of Organizing**. London: MIT Press, 2013, p.1-36.

GÓMEZ-SUTA, M.; ECHEVERRY-CORREA, J. D.; SOTO-MEJÍA, J. A. Semi-automatic extraction and validation of concepts in ontology learning from texts in Spanish. *In*: PROCEEDINGS OF THE 10TH INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, MINING AND SEMANTICS. 10., 2020, New York, NY, United States. **Proceedings** [...]. New York: Association for Computing Machinery, 2020. p. 7-16.

GRUBER, T. R. A Translation Approach to Portable Ontology Specifications. **Knowledge Acquisition**, California, v.5, p. 199–220, 1993.

GUARINO, N.; WELTY, C.; GRUNINGER, M. Formal ontology and information systems. **Frontiers in Artificial Intelligence and Applications**, Netherlands, v. 183, p. 3-15, 2009.

GYRARD, A.; ZIMMERMANN, A.; SHETH, A. Building IoT-based applications for smart cities: How can ontology catalogs help?. **IEEE Internet of Things Journal**, United States, v. 5, n. 5, p. 3978-3990, 2018.

HANANTO, V. R.; SERDÜLT, U.; KRYSSANOV, V. A Tourism Knowledge Model through Topic Modeling from Online Reviews. *In*: INTERNATIONAL CONFERENCE ON COMPUTING AND DATA ENGINEERING, 7., 2021, Phuket, Thailand, 15–17 January 2021. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 87-93.

HJØRLAND, B. What is knowledge organization (KO)?. **Knowledge organization. International journal devoted to concept theory, classification, indexing and knowledge representation**, Germany, v. 35, n. 2, p. 86-101, 2008.

JACOB, E. Classification and categorization: a difference that makes a difference. **Library Trends**, United States, v. 52, n.3, p. 515-540, 2004.

JAEGER, P. T. The endless wire: E-government as global phenomenon. **Government Information Quarterly**, United States, v. 20, n. 4, p. 323-331, 2003.

JALIL, Z.; NASIR, J. A.; NASIR, M. Extractive Multi-Document Summarization: A Review of Progress in the Last Decade. **IEEE Access**, United States, v. 9, p. 130928-130946, 2021.

JANOWSKI, T. Digital government evolution: From transformation to contextualization. **Government Information Quarterly**, United States, v. 32, n. 3, p. 221-236, 2015.

KHEMIRI, A. *et al.* Learn2Construct: an automatic ontology construction based on LDA from textual data. *In: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON MANAGEMENT OF DIGITAL ECOSYSTEMS, 13.*, 2021, New York, NY. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 49-56.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, United Kingdom, v. 33, n. 2004, p. 1-26, 2004. Disponível em: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=29890a936639862f45cb9a987dd599dce9759bf5>. Acesso em: 12 jul. 2024.

LEÃO, D. S. O impacto da transformação digital nos serviços públicos: um estudo de caso. **Revista de Administração Pública e Gestão Social**, Viçosa, Minas Gerais, v. 10, n. 2, p. 150-165, 2018.

LEÃO, H. A. T. **Digitização de serviços públicos**: um modelo para administração pública federal. 2018. Dissertação (Mestrado Profissional em Computação Aplicada) – Universidade de Brasília, Brasília, 2018.

LEI ZENG, M. Knowledge organization systems (KOS). **Knowledge Organization**, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

LEI ZENG, M.; HODGE, G. Developing a Dublin Core Application Profile for the knowledge organization systems (KOS) resources. **Bulletin of the Association for Information Science and Technology**, United States, v. 37, n. 4, p. 30-34, 2011.

LEI ZENG, M. Knowledge organization systems (KOS). **Knowledge Organization**, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

LIAO, C. C.; WU, Y.F.; KING, G.H. Research on Learning OWL Ontology from Relational Database. **Journal of Physics: Conference Series**. **IOP Publishing**, United States, p. 022031, 2019.

LIAO, C. H.; XIONG, G.Y.; CHEN, C.L. Research on OWL Ontology Learning Method Based on Relational Schema. **DEStech Transactions on Computer Science and Engineering**, United States, n. msota, 2018.

LIAO, Y.; WU, Y.; KING, I. A Survey on Ontology Learning from Social Network Data. **Knowledge and Information Systems**, United Kingdom, v. 61, p. 1833-1865, 2019.

LIAO, Y., XIONG, Y.; CHEN, Y. Ontology Learning from Linked Data: A Review. **Journal of Network and Computer Applications**, United Kingdom, v. 103, p. 60-70, 2018.

LIMA, G. A. B. O. **MHTX: Modelagem hipertextual para organização de documentos**: princípios e aplicação. 1. ed. Rio de Janeiro: Interciência, 2015. 208p.

LIMA, G. A. B. O. Modelos de categorização: apresentando o modelo clássico e o modelo de protótipos. **Perspectivas em ciência da informação**, Belo Horizonte, v. 15, n. 2, p. 108-122, 2010.

LIMA, G. A. B. O.; MACULAN, B. C. M. S. Universo do Conhecimento: classificação e categorização sob o prisma da organização do conhecimento. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 22, p. e024017, 2024.

LIMA, G. A. B. O.; RAGHAVAN, K. S. Categories in Knowledge Organization. *In*: KNOWLEDGE ORGANIZATION IN THE 21ST CENTURY: BETWEEN HISTORICAL PATTERNS AND FUTURE PROSPECTS, 21., 2014, Kraków, Poland. **Proceedings** [...]. Poland: Ergon, 2014. p. 19-22.

LOURENÇO, C. A. **Análise do Padrão Brasileiro de Metadados de Teses e Dissertações segundo o Modelo Entidade-Relacionamento**. 2005. Tese. (Doutorado em Ciência da Informação) – Faculdade de Ciência da informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

LUCIDCHART. Disponível em: <https://www.lucidchart.com/pages>. Acesso em: 14 abr. 2025.

MACULAN, B. C. M. S. A Classification Schema for Ontologies and Ontology Applications. **International Journal of Software Engineering and Knowledge Engineering**, Singapore, v. 25, n. 2, p. 235-259, 2015.

MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesouro**: remodelagem do THESAGRO. 2015. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MAHMOUD, N.; ELBEH, H.; ABDLKADER, H. M. Ontology Learning Based on Word Embeddings for Text Big Data Extraction. *In*: **INTERNATIONAL COMPUTER ENGINEERING CONFERENCE (ICENCO)**, 14., 2018, Cairo, Egypt. Proceedings [...], Cairo: IEEE, 2018. p. 183-188.

MALCHIODI, D.; DA COSTA PEREIRA, C.; TETTAMANZI, A. G. B. Predicting the possibilistic score of OWL axioms through support vector regression. *In: SCALABLE UNCERTAINTY MANAGEMENT: 12TH INTERNATIONAL CONFERENCE, SUM 2018*, 12., 2018. Milan, Italy. **Proceedings** [...]. Milan, Italy: Springer International Publishing, October 3-5, 2018. p. 380-386.

MALCHIODI, D.; TETTAMANZI, A. G. B. Predicting the possibilistic score of OWL axioms through modified support vector clustering. *In: PROCEEDINGS OF THE 33RD ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING*, 33., 2018. Pau, France. **Proceedings** [...]. New York: Association for Computing Machinery, 2018. p. 1984-1991.

MARCONI, M.; LAKATOS, E. M. **Fundamentos de Metodologia científica**. 7. ed. São Paulo: Atlas, 2016.

MCGUINNESS, C.; O'NEILL, P.; CROSS, R. Digital government in the public sector: challenges and opportunities. **Journal of Information Technology**, United Kingdom, v. 35, n. 2, p. 130-145, 2020.

MENDONCA, 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). 2015. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MENDELEY. Disponível em: www.mendeley.com. Acesso em: 27 abr. 2025.

MGI. Ministério da Gestão e da Inovação em Serviços Públicos. O maior portal de serviços digitais do mundo. **Agência Gov.**, 4 de março de 2024. Disponível em: <https://agenciagov.ebc.com.br/noticias/202403/gov-br-e-a-pagina-de-governo-mais-acessada-do-mundo>. Acesso em: 24 jul. 2024.

MIRO. Disponível em: <https://miro.com/pt/>. Acesso em: 14 abr. 2025.

MIZOGUCHI, R., SUNAGAWA, E., KOZAKI, K. Evolution of ontological engineering: A brief personal reflection. **Applied Ontology**, Netherlands, v. 14, n. 4, p. 331-354, 2019.

NGUYEN, T. H.; TETTAMANZI, A. G. B. An evolutionary approach to class disjointness axiom discovery. *In: IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE*, 2019, Greece. **Proceedings** [...]. Greece: IEEE, 2019. p. 68-75.

NOURA, M. *et al.* Automatic knowledge extraction to build semantic web of things applications. **IEEE Internet of Things Journal**, United States, v. 6, n. 5, p. 8447-8454, 2019.

NÚCLEO DO CONHECIMENTO. Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/aprendizagem-de-ontologias>. Acesso em: 12 abr. 2025.

OA. Como o Portal GOV.BR pode facilitar sua vida: Entenda e crie sua conta de acesso. **O Antagonista**, 3 de julho de 2024. Disponível em: <https://oantagonista.com.br/brasil/como-o-portal-gov-br-pode-facilitar-sua-vida-entenda-e-crie-sua-conta-de-acesso/>. Acesso em: 24 jul. 2024.

OECD. The Path to Becoming a Data-Driven Public Sector. **OECD Publishing**, United States 2020. Disponível em: <https://www.oecd.org/gov/the-path-to-becoming-a-data-driven-public-sector-9789264910552-en.htm>. Acesso em: 24 jul. 2024.

ONTOBEE. Disponível em: <https://ontobee.org/>. Acesso em: 12 abr. 2025.

PÁDUA, C. I. P. S. **Engenharia de Usabilidade**. Material de Referência. Belo Horizonte: Departamento de Ciência da Computação da Universidade Federal de Minas Gerais, 2012. (Apostila Disciplina Engenharia de usabilidade). Disponível em: <<http://homepages.dcc.ufmg.br/~clarindo/arquivos/disciplinas/eu/material/referencias/apostila-usabilidade.pdf>> Acesso em: 21 out 2019.

PROTÉGÉ. Disponível em: <https://protege.stanford.edu/>. Acesso em: 14 abr. 2025.

RAJAONARIVO, L.; MINE, T.; ARAKAWA, Y. Automatic generation of event ontology from social network and mobile positioning data. *In*: IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2021, New York. **Proceedings** [...]. New York: IEEE, 2021. p. 87-94.

RODRIGUEZ, T.; AGUILAR, J. Knowledge extraction system from unstructured documents. **IEEE Latin America Transactions**, United States, v. 16, n. 2, p. 639-646, 2018.

RUTTENBERG, A.; CLARK, T.; BUG, W.; SAMWALD, M.; BODENREIDER, O.; CHEN, H. *et al.* Advancing translational research with the Semantic Web. **BMC Bioinformatics**, London, v. 8, Suppl 3, S2, 2007.

SADEGHIANASL, S. *et al.* Process activity ontology learning from event logs through gamification. **IEEE Access**, United States, v. 9, p. 165865-165880, 2021.

SALATINO, A. *et al.* The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. **Data Intelligence**, United States, v. 2, n. 3, p. 379-416, 2020.

SANTOS, A.; SIMÃO, A.; PINTO, H. Semantic Ontologies and Contextual Reasoning: A Systematic Review. **Information**, Suécia, v. 11, n. 7, p. 339, 2020.

SANTOSA, N. C.; MIYAZAKI, J.; HAN, H. Automating computer science ontology extension with classification techniques. **IEEE Access**, United States, v. 9, p. 161815-161833, 2021.

SATTAR, A. *et al.* An improved methodology for collaborative construction of reusable, localized, and shareable ontology. **IEEE Access**, United States, v. 9, p. 17463-17484, 2021.

SCHOLL, H. J. **E-Government: Information, Technology, and Transformation**. Abingdon, Inglaterra: Routledge, 2018.

SILVA, J. C. *et al.* Blockchain-Based E-Government: An Overview. **Sustainability**, United States, v. 12, n. 7, p. 3009, 2020.

SILVA, L. G.; SANTOS, M. A.; FERREIRA, J. R. A utilização de *blockchain* no governo eletrônico: um estudo exploratório. **Revista Tecnologia e Sociedade**, Curitiba, Paraná, v. 16, n. 4, p. 45-60, 2020.

SOERGEL, D. Information organization. *In*: BAINBRIDGE, D. (ed.) **Berkshire Encyclopedia of Human-Computer Interaction**, Estados Unidos: ACM Digital Library, 2004. p. 355-360.

SOERGEL, D. **Knowledge Organization Systems: Overview**. 2014. Disponível em: https://www.researchgate.net/profile/Dagobert_Soergel/publication/303610010_Knowledge_Organization_Systems_Overview/links/5749ef1608ae5bf2e63f1b17/Knowledge_Organization_Systems_Overview.pdf. Acesso em: 06 jan. 2018.

SOMODEVILLA GARCÍA, M.; VILARIÑO AYALA, D.; PINEDA, I. An Overview of Ontology Learning Tasks. **Computación y Sistemas**, México, v. 22, n. 1, p. 137-146, 2018.

STERN, E. Transformação digital no setor público: desafios e oportunidades. **Revista de Administração Pública e Gestão Social**, Viçosa, Minas Gerais, v. 12, n. 3, p. 45-63, 2018.

STERN, S. *et al.* **Government 4.0: the public sector in the digital age**. Disponível em: https://www.mckinsey.de/~media/McKinsey/Locations/Europe%20and%20Middle%20East/Deutschland/Publikationen/2018%20Compendium/Government%2040%20the%20public%20sector%20in%20the%20digital%20age/kompendium_04_ps.ashx. Acesso em: 03 set. 2019.

TANG, C.; ZHANG, J. Blockchain in government: applications and challenges. **Government Information Quarterly**, United States, v. 37, n. 1, p. 101-123, 2020.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistemas de classificação facetados e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004.

TUDHOPE, D.; NIELSEN, L. M. Introduction to knowledge organization systems and services. **New Review of Hypermedia and Multimedia**, United Kingdom, v. 12, n. 1, p. 3-9, 2006.

UNITED NATIONS. E-Government Survey 2018: Gearing e-government to support transformation towards sustainable and resilient societies. **United Nations Department of Economic and Social Affairs**, New York City, New York, 2018. Disponível em: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2018>. Acesso em: 24 jul. 2024.

UNITED NATIONS. E-Government Survey 2022: The future of digital government. **United Nations Department of Economic and Social Affairs**, New York City, New York, 2022a. Disponível em: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022>. Acesso em: 24 jul. 2024.

UNITED NATIONS. **UN E-Government Knowledgebase**. 2022b. Disponível em: <https://publicadministration.un.org/egovkb/en-us/Data-Center>. Acesso em: 12 jul. 2024.

VEDULA, N.; MANERIKER, P.; PARTHASARATHY, S. Bolt-k: Bootstrapping ontology learning via transfer of knowledge. *In: THE WORLD WIDE WEB CONFERENCE*, 19., 2019, New York. **Proceedings [...]**. New York: Association for Computing Machinery, 2019. p. 1897-1908.

WANG, H.; WANG, C.; WANG, S. Ontology-based framework for e-government systems interoperability. **Journal of Software**, United States, v. 14, n. 6, p. 257-271, 2019.

WHITE, H. **Theories of Knowledge Organization literature review chapter**. 2011. Disponível em: <http://sites.duke.edu/holliewhite/files/2011/10/Theory-of-Knowledge-Organization-lit-review-Mar-2010.pdf>. Acesso em: 06 jan. 2018.

YADAV, U. *et al.* EasyOnto: A Collaborative Semiformal Ontology Development Platform. *In: NATURE INSPIRED COMPUTING*, Singapore. **Proceedings [...]**. Singapore: Springer Singapore, 2018. p. 1-11.

ZAHRA, F. M. *et al.* Ferramentas para aprendizagem de ontologias a partir de textos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 1, p. 3-21, 2014.

ZENG, M. L.; HODGE, G. Knowledge Organization Systems: Research and Development. **Journal of the American Society for Information Science and Technology**, United States, v. 62, n. 8, p. 1387-1389, 2011.

ZENG, M. L. Knowledge Organization Systems (KOS). **Knowledge Organization**, Frankfurt, Alemanha, v. 35, n. 2-3, p. 160-182, 2008.

ZENG, M. L. **Kos Types Vocabulary**. 2014. Disponível em: <https://nkos.dublincore.org/nkos-type.html>. Acesso em: 06 jan. 2024.

APÊNDICE A – PROTOCOLO DE REVISÃO SISTEMÁTICA DE LITERATURA

TEMÁTICA: Geração automática de Ontologias

ÁREAS DE CONHECIMENTO ESPECÍFICAS: Ciência da Informação. Ciência da Computação. Educação.

OBJETIVO

Este Protocolo de Revisão Sistemática da Literatura (PRSL) apresenta a estrutura metodológica para a execução da etapa de revisão da literatura sobre métodos para o processo de enriquecimento de ontologias de domínio. Para a construção desse protocolo, respaldou-se no método proposto por Kitchenham (2004), bem como no trabalho de Araújo (2021). Para organização e gestão das referências, foi utilizado o *software Mendeley*. Para análise quanti e qualitativa do material selecionado, foi utilizado o software **Atlas TI**.

EQUIPE

Nome	Papel	Afiliação
Gercina Ângela de Lima	Orientadora	PPGGOC/UFMG
Patrícia Lopes Ferreira	Doutoranda	PPGGOC/UFMG

PERGUNTA DE PESQUISA

Principal:

- Quais trabalhos aplicaram métodos de aprendizagem de ontologias entre os anos de 2018 e 2023?

Secundárias

- É possível identificar a utilização de alguma metodologia de aprendizagem de ontologias?
- Quais as etapas/fases da metodologia ora adotada?
- Quais as vantagens e os diferenciais da metodologia adotada?
- Quais as lacunas existentes nessas metodologias?
- Quais os principais resultados evidenciados?

BASES DE DADOS

Considerando a temática da pesquisa, as seguintes bases de dados são utilizadas nesta RSL:

- ACM Digital Library
- IEEE Xplore Digital Library
- Scopus Preview
- Web of Science

TERMOS DE BUSCA

Termos em Inglês	Termos em Português
[Ontology Learning] OR [Ontologies Learning]	[Aprendizagem de Ontologia] OR [Aprendizagem de Ontologias]
[Ontology Engineering] OR [Ontologies Engineering]	[Engenharia de ontologia] OR [Engenharia de ontologias]
[Ontology Extraction] OR [Ontologies Extraction]	[Extração de Ontologia] OR [Extração de Ontologias]
[Automatic ontology Generation] OR [Automatic ontologies Generation]	[Geração automática de ontologia] OR [Geração automática de ontologias]

STRINGS DE BUSCA

Base de dados	String utilizada
ACM Digital Library	[Title: "ontology learning"] OR [Title: "ontologies learning"] OR [Title: "aprendizagem de ontologia"] OR [Title: "aprendizagem de ontologias"] OR [Title: "ontology engineering"] OR [Title: "ontologies engineering"] OR [Title: "engenharia de ontologia"] OR [Title: "engenharia de ontologias"] OR [Title: "ontology extraction"] OR [Title: "ontologies extraction"] OR [Title: "extração de ontologia"] OR [Title: "extração de ontologias"] OR [Title: "automatic ontology generation"] OR [Title: "automatic ontologies generation"] OR [Title: "geração automática de ontologia"] OR [Title: "geração automática de ontologias"] OR [Abstract: "ontology learning"] OR [Abstract: "ontologies learning"] OR [Abstract: "aprendizagem de ontologia"] OR [Abstract: "aprendizagem de ontologias"] OR [Abstract: "ontology engineering"] OR [Abstract: "ontologies engineering"] OR [Abstract: "engenharia de ontologia"] OR [Abstract: "engenharia de ontologias"] OR [Abstract: "ontology extraction"] OR [Abstract: "ontologies extraction"] OR [Abstract: "extração de ontologia"] OR [Abstract: "extração de ontologias"] OR [Abstract: "automatic ontology generation"] OR [Abstract: "automatic ontologies generation"] OR [Abstract: "geração automática de ontologia"] OR [Abstract: "geração automática de ontologias"] OR [Keywords: "ontology learning"] OR [Keywords: "ontologies learning"] OR [Keywords: "aprendizagem de ontologia"] OR [Keywords: "aprendizagem de ontologias"] OR [Keywords: "ontology engineering"] OR [Keywords: "ontologies engineering"] OR [Keywords: "engenharia de ontologia"] OR [Keywords: "engenharia de ontologias"] OR [Keywords: "ontology extraction"] OR [Keywords: "ontologies extraction"] OR [Keywords: "extração de ontologia"] OR [Keywords: "extração de ontologias"] OR [Keywords: "automatic ontology generation"] OR [Keywords: "automatic ontologies generation"] OR [Keywords: "geração automática de ontologia"] OR [Keywords: "geração automática de ontologias"] AND [E-Publication Date: (01/01/2018 TO 12/31/2023)]
IEEE	("Document Title": "Ontology Learning") OR ("Document Title": "Ontologies Learning") OR ("Document Title": "Aprendizagem de Ontologia") OR ("Document Title": "Aprendizagem de Ontologias") OR ("Document Title": "Ontology Engineering") OR ("Document Title": "Ontologies Engineering") OR ("Document Title": "Engenharia de ontologia") OR ("Document Title": "Engenharia de ontologies") OR ("Document Title": "Ontology Extraction") OR ("Document Title": "Ontologies Extraction") OR ("Document Title": "Extração de Ontologia") OR ("Document Title": "Extração de Ontologias") OR ("Document Title": "Automatic ontology Generation") OR ("Document Title": "Automatic ontologies Generation") OR ("Document Title": "Geração automática de ontologia") OR ("Document Title": "Geração

	<p>automática de ontologies”) OR (“Abstract”:“Ontology Learning” OR “Abstract”:“Ontologies Learning” OR “Abstract”:“Aprendizagem de Ontologia” OR “Abstract”:“Aprendizagem de Ontologias” OR “Abstract”:“Ontology Engineering” OR “Abstract”:“Ontologies Engineering” OR “Abstract”:“Engenharia de ontologia” OR “Abstract”:“Engenharia de ontologies” OR “Abstract”:“Ontology Extraction” OR “Abstract”:“Ontologies Extraction” OR “Abstract”:“Extração de Ontologia” OR “Abstract”:“Extração de Ontologias” OR “Abstract”:“Automatic ontology Generation” OR “Abstract”:“Automatic ontologies Generation” OR “Abstract”:“Geração automática de ontologia” OR “Abstract”:“Geração automática de ontologies”) OR (“Author Keywords”:“Ontology Learning” OR “Author Keywords”:“Ontologies Learning” OR “Author Keywords”:“Aprendizagem de Ontologia” OR “Author Keywords”:“Aprendizagem de Ontologias” OR “Author Keywords”:“Ontology Engineering” OR “Author Keywords”:“Ontologies Engineering” OR “Author Keywords”:“Engenharia de ontologia” OR “Author Keywords”:“Engenharia de ontologies” OR “Author Keywords”:“Ontology Extraction” OR “Author Keywords”:“Ontologies Extraction” OR “Author Keywords”:“Extração de Ontologia” OR “Author Keywords”:“Extração de Ontologias” OR “Author Keywords”:“Automatic ontology Generation” OR “Author Keywords”:“Automatic ontologies Generation” OR “Author Keywords”:“Geração automática de ontologia” OR “Author Keywords”:“Geração automática de ontologies”)</p> <p>Filters Applied: JournalsBooks2018 - 2023</p>
<i>Scopus</i>	<p>TITLE-ABS-KEY (“Ontology Learning” OR “Ontologies Learning” OR “Aprendizagem de Ontologia” OR “Aprendizagem de Ontologias” OR “Ontology Engineering” OR “Ontologies Engineering” OR “Engenharia de ontologia” OR “Engenharia de ontologies” OR “Ontology Extraction” OR “Ontologies Extraction” OR “Extração de Ontologia” OR “Extração de Ontologias” OR “Automatic ontology Generation” OR “Automatic ontologies Generation” OR “Geração automática de ontologia” OR “Geração automática de ontologies”) AND PUBYEAR > 2017 AND PUBYEAR < 2024</p>
<i>Web of Science</i>	<p>“Ontology Learning” OR “Ontologies Learning” OR “Aprendizagem de Ontologia” OR “Aprendizagem de Ontologias” OR “Ontology Engineering” OR “Ontologies Engineering” OR “Engenharia de ontologia” OR “Engenharia de ontologies” OR “Ontology Extraction” OR “Ontologies Extraction” OR “Extração de Ontologia” OR “Extração de Ontologias” OR “Automatic ontology Generation” OR “Automatic ontologies Generation” OR “Geração automática de ontologia” OR “Geração automática de ontologies”</p> <p>P.S. Utilizou-se como parâmetro de busca o item “Topic”, que realiza a busca nas seções Título, Resumo e Palavras-Chave. Utilizou o filtro Ano de publicação, adotando-se o intervalo entre 2018 e 2023.</p>

CRITÉRIOS DE INCLUSÃO

Os Critérios de Inclusão dos estudos recuperados são apresentados no quadro a seguir:

Critério	Descrição do Critério de Inclusão
CI1	Título possui relação com a temática aprendizagem de ontologias.
CI2	Resumo possui relação com a temática aprendizagem de ontologias.
CI3	Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas.
CI4	Serão incluídos trabalhos que tenham aplicado um método de geração automática ou semiautomática de ontologias.

CI5	Trabalhos nos idiomas inglês, português e espanhol.
CI6	Período de publicação compreendido entre 2018 e 2023.
CI7	Serão considerados trabalhos dos seguintes tipos: artigo de periódico, capítulo de livro, artigo de evento.

CRITÉRIOS DE EXCLUSÃO

Os Critérios de Exclusão dos estudos recuperados são apresentados no quadro a seguir:

Critério	Descrição do Critério de Inclusão
CE1	Título não possui relação com a temática aprendizagem de ontologias.
CE2	Resumo não possui relação com a temática aprendizagem de ontologias.
CE3	Serão excluídos trabalhos que não apresentem resumo/ <i>abstract</i> .
CE4	Serão excluídos <i>short papers</i> e trabalhos disponibilizados em bases preprint.
CE5	Serão excluídos documentos que não estejam em português, inglês ou espanhol.
CE6	Serão excluídos trabalhos que não estejam disponíveis em texto completo.
CE7	Serão excluídos os trabalhos que não apliquem o processo de geração automática ou semiautomática de ontologias.

PROCESSO DE SELEÇÃO DAS PUBLICAÇÕES

Inicialmente, executaram-se as expressões de busca adaptadas em cada uma das bases de dados. A partir da lista de estudos retornados, exportaram-se os resultados em formato RIS para importação na ferramenta Mendeley. Na primeira etapa de seleção dos textos, foram excluídas as duplicatas, resultando-se em **183 textos**. Considerando a exclusão dos trabalhos que não disponibilizavam o texto completo, chegou-se a um total de **170 textos**. As etapas 1 a 3 foram realizadas entre os meses de setembro e outubro de 2023. Em seguida, foram identificados os trabalhos aderentes aos critérios de inclusão e exclusão, refletindo o *corpus* final para a etapa de análise, chegando-se a **22 textos**. Esta etapa foi realizada entre novembro de 2023 e janeiro de 2024. Finalizada a etapa de extração dos dados, realizou-se a análise, a interpretação e a documentação dos resultados encontrados nos estudos, tendo em vista a pergunta principal e as perguntas secundárias previstas nesta RSL, com a redação do capítulo de revisão. Esta etapa foi realizada entre os meses de janeiro e fevereiro de 2024 com o apoio do *software* Atlas TI.