



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E  
GESTÃO DO CONHECIMENTO

Jeferson de Oliveira Mello

**Análise da Contribuição de Grafos de Conhecimento para a Engenharia do  
Conhecimento**

Florianópolis  
2024

Jeferson de Oliveira Mello

**Análise da Contribuição de Grafos de Conhecimento para a Engenharia do  
Conhecimento**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia do Conhecimento.

Orientador: Prof., Dr. Fernando Álvaro Ostuni Gauthier  
Coorientador: Prof., Dr. Marcelo Macedo

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.  
Dados inseridos pelo próprio autor.

Mello, Jeferson de Oliveira

Análise da contribuição de grafos de conhecimento para a Engenharia do Conhecimento / Jeferson de Oliveira Mello ; orientador, Fernando Álvaro Ostuni Gauthier, coorientador, Marcelo Macedo, 2024.

98 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2024.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Grafos de conhecimento. 3. Banco de dados em grafos. 4. Ferramentas de grafos. 5. Representação do conhecimento. I. Gauthier, Fernando Álvaro Ostuni . II. Macedo, Marcelo. III. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Jeferson de Oliveira Mello

**Análise da Contribuição de Grafos de Conhecimento para a Engenharia do  
Conhecimento**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 14 de agosto de 2024,  
pela banca examinadora composta pelos seguintes membros:

Prof. Rogerio Cid Bastos, Dr.  
Universidade Federal de Santa Catarina

Prof. Vinicius Faria Culmant Ramos, Dr.  
Universidade Federal de Santa Catarina

Prof. Antônio Pereira Cândido, Dr.  
Instituto Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado  
adequado para obtenção do título de Mestre em Engenharia do Conhecimento.

---

Prof.<sup>a</sup> Luciane Maria Fadel, Dr.<sup>a</sup>  
Coordenação do Programa de Pós-Graduação

---

Prof. Fernando Alvaro Ostuni Gauthier, Dr.  
Orientador

Florianópolis, 2024.

Dedico este trabalho às pessoas especiais que me apoiaram ao longo desta caminhada.

## AGRADECIMENTOS

No decorrer desta jornada, foram muitas as batalhas, infindáveis momentos difíceis, incontáveis períodos turbulentos, mas, como escreveu Carlos Drummond de Andrade, “só é lutador quem sabe lutar consigo mesmo”. É nesses instantes que o apoio da família e dos amigos são fundamentais, para manter a serenidade necessária e a autoestima, tão relevante para o ser humano no enfrentamento dos desafios diários.

Nesta oportunidade, expresso meus agradecimentos a todos que de forma direta ou indireta contribuíram para a conclusão deste estudo.

Aos meus pais, Ângelo e Claudete, e à minha irmã, Priscila, agradeço o apoio incondicional, pois sempre estiveram presentes em todas as circunstâncias da minha vida.

Ao meu orientador, Professor Dr. Fernando Álvaro Ostuni Gauthier, meu mais profundo agradecimento pela direção, aconselhamento, confiança e apoio para que este trabalho fosse concretizado.

Ao meu coorientador, Professor Dr. Marcelo Macedo pelas pontuais contribuições.

Agradeço a todos os professores do EGC por todo o conhecimento compartilhado e por ter me concedido a oportunidade de galgar mais um degrau em minha formação acadêmica.

Ao meu amigo, doutorando Ranieri Alves dos Santos, expresso minha gratidão pelas valiosas conversas, sugestões e apoio.

Agradeço aos demais funcionários do Programa pela acolhida e atendimento.

Agradeço a CAPES pelo apoio à pesquisa.

“Todo o conhecimento humano começou com intuições, passou daí aos conceitos e terminou  
com ideias.”

*Immanuel Kant*

## RESUMO

Com o exponencial aumento de dados no mundo, a organização e representação eficaz dessas informações tornaram-se cruciais. Nesse contexto, os grafos de conhecimento emergem como uma solução poderosa, aplicável em diversas áreas, desde sistemas de consumo até infraestruturas críticas. Para explorar essa potencialidade, este estudo se concentra no desenvolvimento de um grafo de conhecimento utilizando dados do Congresso Internacional de Conhecimento e Inovação – CIKI 2023 e do Repositório Institucional da UFSC, por meio de implementação na plataforma Neo4j. O processo de construção do grafo envolveu a extração, a modelagem e a inserção de dados para a análise das interconexões entre artigos, autores e instituições. Essa abordagem possibilitou identificar áreas emergentes de pesquisa, mapear colaborações e analisar a evolução de tópicos ao longo do tempo. Além disso, ferramentas de visualização e algoritmos de análise de grafos foram empregados para destacar padrões e áreas com potencial de inovação, demonstrando a eficácia dos grafos de conhecimento. Os grafos de conhecimento utilizam metadados para criar regras que organizam e interpretam dados, automatizando processos de otimização. Eles maximizam os benefícios das informações, melhoram previsões e facilitam respostas ágeis em contextos dinâmicos. O Neo4j é destaque em razão do uso intuitivo, da flexibilidade, da eficiência e da escalabilidade, sendo amplamente usado em áreas como sistemas de recomendação e detecção de fraudes. A integração de informações do CIKI 2023 e do repositório da UFSC revelou padrões de colaboração e contribuições acadêmicas, reforçando a importância de eventos como o CIKI na promoção da inovação e colaboração científica. Assim, este estudo não somente valida a utilidade prática dos grafos de conhecimento, mas também sublinha seu papel essencial na evolução da pesquisa científica e tecnológica. Por fim, neste estudo, destaca-se a importância da colaboração interdisciplinar e internacional para superar desafios e promover avanços significativos na Engenharia do Conhecimento.

Palavras-chave: grafos de conhecimento; representação do conhecimento; banco de dados em grafos; ferramentas de grafos; Engenharia do Conhecimento.

## ABSTRACT

With the exponential increase in data globally, the organization and effective representation of this information has become crucial. In this context, knowledge graphs have emerged as a powerful solution, applicable in a variety of areas, from consumer systems to critical infrastructures. To explore this potential, this study focuses on the development of a knowledge graph using data from the International Congress on Knowledge and Innovation (CIKI) 2023 and the UFSC Institutional Repository, implemented on the Neo4j platform. The process of building the graph involved extracting, modeling and inserting data, allowing analysis of the interconnections between articles, authors and institutions. This approach made it possible to identify emerging areas of research, map collaborations and analyze the evolution of topics over time. In addition, visualization tools and graph analysis algorithms were used to highlight patterns and areas with potential for innovation, demonstrating the effectiveness of knowledge graphs. Knowledge graphs use metadata to create rules that organize and interpret data, automating optimization processes. They maximize the benefits of information, improve forecasts and facilitate agile responses in dynamic contexts. Neo4j stands out due to its intuitive use, flexibility, efficiency and scalability, and is widely used in areas such as recommendation systems and fraud detection. Finally, the dissertation highlights the importance of interdisciplinary and international collaboration to overcome challenges and promote significant advances in knowledge engineering. The integration of information from CIKI 2023 and the UFSC repository revealed patterns of collaboration and academic contributions, reinforcing the importance of events such as CIKI in promoting innovation and scientific collaboration. Thus, this study not only validates the practical usefulness of knowledge graphs, but also underlines their essential role in the evolution of scientific and technological research.

**Keywords:** knowledge graphs; representation of knowledge; graphs database; graph tools; Knowledge Engineering.

## LISTA DE FIGURAS

Figura 1	– Camadas da <i>web</i> semântica.....	34
Figura 2	– Tripla RDF .....	36
Figura 3	– Grafo RDF sobre produtos.....	37
Figura 4	– Representação gráfica de Königsberg.....	40
Figura 5	– Uma representação do grafo de conhecimento de uma rede social .....	45
Figura 6	– Exemplo de grafo de propriedades .....	46
Figura 7	– Representação gráfica de um grafo de propriedade.....	47
Figura 8	– Um grafo que representa pessoas, suas amigas e suas localizações.....	48
Figura 9	– Exemplo de uma tripla .....	50
Figura 10	– Principais etapas do processo de construção de grafos de conhecimento .....	59
Figura 11	– Alguns grafos de conhecimento em grande escala bem conhecidos .....	64
Figura 12	– Visão Geral da estrutura de extração DBpedia .....	67
Figura 13	– Tela da página inicial do <i>site</i> CIKI 2023 .....	73
Figura 14	– Tela de visualização da lista de artigos do CIKI.....	74
Figura 15	– Tela página inicial Repositório Institucional Teses Dissertações/UFSC.....	74
Figura 16	– Tela do <i>web scraping</i> .....	75
Figura 17	– Tela do código-fonte HTML.....	76
Figura 18	– Tela metadado da página HTML .....	77
Figura 19	– Grafo de conhecimento do CIKI 2023 .....	78
Figura 20	– Informações completas de uma das dissertações dispostas no Repositório Institucional de Tese e Dissertações da UFSC de 2023.....	80
Figura 21	– Planilha eletrônica com informações de Teses e Dissertações .....	81
Figura 22	– <i>Script</i> na linguagem Cypher com as informações de Teses e Dissertações.....	81
Figura 23	– Grafo de conhecimento de Teses e Dissertações do PPGEGC/UFSC.....	82
Figura 24	– Visão do Neo4j com artigos CIKI 2023 e Teses e Dissertações do PPGEGC/UFSC .....	83
Figura 25	– Visão do Neo4j: pessoas que escreveram artigos no CIKI 2023 e defenderam Teses/Dissertações em 2023 .....	84

## LISTA DE QUADROS

Quadro 1 – Desafios da Engenharia do Conhecimento: aquisição e melhoria do conhecimento.....	17
Quadro 2 – Principais trabalhos relacionados ao tema.....	23
Quadro 3 – Descrição das etapas.....	26
Quadro 4 – Ferramentas para Engenharia do Conhecimento.....	31
Quadro 5 – Categorização das tecnologias da <i>web</i> semântica.....	35
Quadro 6 – Requisitos fundamentais do projeto para armazenar grafos de conhecimento.....	53
Quadro 7 – Processo de avaliação de dados e conhecimento do grafo de conhecimento.....	61
Quadro 8 – Comparativo de bases de conhecimento.....	66
Quadro 9 – Pessoas: Artigo(s) e Tese/Dissertação.....	85

## LISTA DE ABREVIATURAS E SIGLAS

ACID	Atomicity, Consistency, Isolation, Durability
API	Application Programming Interface
CIKI	Congresso Internacional de Conhecimento e Inovação
EGC	Engenharia e Gestão do Conhecimento
ER	Entity Resolution
FMD	Fashion Model Directory
FOAF	Friend of a Friend
GC	Grafo de conhecimento
HTML	Hyper Text Markup Language
IA	Inteligência artificial
IRI	Internationalised Resource Identifier
KDD	Knowledge Discovery in Databases
KG	Knowledge Graph
LOD	Linked Open Data
LPG	Labeled Property Graph
NELL	Never-Ending Language Learner
NLP	Natural Language Processing
NNDB	Notable Names Database
NoSQL	Not Only SQL
OWL	Web Ontology Language
PPGEGC	Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento
PROSPERA	PRospering knOWledge with Scalability, PRrecision, and RecAll
RDB2RDF	Relational Databases to RDF
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RIF	Requirements Interchange Format
SPARQL	Protocol and RDF Query Language
SQL	Structured Query Language
SWT	Semantic Web Technologies
SW	Semantic Web
TI	Tecnologia da Informação
UFSC	Universidade Federal de Santa Catarina

URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WWW	World Wide Web
WMDE	Wikimedia Deutschland
XML	Extensible Markup Language
YAGO	Yet Another Great Ontology

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	16
1.1	CONSIDERAÇÕES INICIAIS.....	16
1.2	PROBLEMA DE PESQUISA.....	18
1.3	OBJETIVOS .....	19
<b>1.3.1</b>	<b>Objetivo geral</b> .....	19
<b>1.3.2</b>	<b>Objetivos específicos</b> .....	19
1.4	JUSTIFICATIVA.....	19
1.5	ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO.....	22
1.6	PROCEDIMENTOS METODOLÓGICOS.....	24
1.7	LIMITAÇÕES DA PESQUISA.....	26
1.8	ESTRUTURA DO TRABALHO.....	26
<b>2</b>	<b>ENGENHARIA DO CONHECIMENTO</b> .....	28
2.1	REPRESENTAÇÃO DO CONHECIMENTO .....	29
2.2	CONCEITOS E ARQUITETURAS DA <i>WEB</i> .....	31
2.3	<i>WEB</i> SEMÂNTICA .....	32
<b>2.3.1</b>	<b>Resource Description Framework (RDF)</b> .....	35
<b>2.3.2</b>	<b>Resource Description Framework Schema (RDFS)</b> .....	37
<b>2.3.3</b>	<b>Web Ontology Language (OWL)</b> .....	38
<b>3</b>	<b>GRAFOS DE CONHECIMENTO</b> .....	40
3.1	UM BREVE HISTÓRICO SOBRE O SURGIMENTO DOS GRAFOS .....	40
3.2	GRAFOS DE CONHECIMENTO (GC): UMA BREVE DEFINIÇÃO .....	41
3.3	MODELO DE GRAFOS DE CONHECIMENTO .....	44
<b>3.3.1</b>	<b>Exemplo de grafo</b> .....	44
3.4	GRAFOS DE PROPRIEDADE .....	45
3.5	GRAFOS RDF .....	48
3.6	ARMAZENAMENTO E GERENCIAMENTO DE GRAFOS DE CONHECIMENTO.....	51
3.7	FERRAMENTAS PARA CRIAR GRAFOS.....	54
<b>3.7.1</b>	<b>Neo4j</b> .....	56
3.8	CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO .....	57
3.9	ALGUNS GRAFOS DE CONHECIMENTO .....	63

<b>3.9.1 DBpedia</b> .....	67
<b>3.9.2 YAGO</b> .....	68
<b>3.9.3 Freebase</b> .....	69
<b>3.9.4 Wikidata</b> .....	70
<b>3.9.5 Cyc</b> .....	71
<b>4 DESENVOLVIMENTO DE UM GRAFO DE CONHECIMENTO</b> .....	73
4.1 CONSIDERAÇÕES INICIAIS.....	73
4.2 GRAFO DE CONHECIMENTO SOBRE ARTIGOS DO CIKI 2023.....	75
4.3 GRAFO DE CONHECIMENTO SOBRE TESES E DISSERTAÇÕES DO EGC/UFSC EM 2023.....	79
4.4 GRAFO DE CONHECIMENTO AGREGADO .....	82
4.5 CONSIDERAÇÕES SOBRE O CAPÍTULO .....	86
<b>CONSIDERAÇÕES FINAIS</b> .....	88
<b>REFERÊNCIAS</b> .....	91

# 1 INTRODUÇÃO

## 1.1 CONSIDERAÇÕES INICIAIS

Sivarajah *et al.* (2017) afirmam que a Engenharia do Conhecimento enfrenta grandes desafios na integração, organização e análise de enormes volumes de dados provenientes de diversas fontes. De acordo com os autores, a complexidade e a interconectividade desses dados dificultam a extração de informações relevantes e a geração de *insights* valiosos, comprometendo a eficiência e a eficácia dos processos de gestão do conhecimento. Embora existam várias ferramentas e metodologias disponíveis para manipular esses dados, muitas delas falham em fornecer uma representação intuitiva e eficiente das relações entre os dados, limitando a capacidade de conduzir análises detalhadas e obter resultados precisos, concluem os autores.

Além disso, a Engenharia do Conhecimento se defronta com vários obstáculos tanto na aquisição quanto na melhoria do conhecimento. Esses desafios são fundamentais para o desenvolvimento e a manutenção de sistemas de conhecimento eficazes e eficientes. Alguns dos principais desafios são apresentados no Quadro 1 a seguir.

Quadro 1 – Desafios da Engenharia do Conhecimento: aquisição e melhoria do conhecimento

<b>Aquisição do conhecimento</b>	
1. Complexidade da captura de conhecimento	Capturar conhecimento tácito (não documentado) de especialistas e convertê-lo em uma forma que possa ser processada por sistemas computacionais é complexo e requer técnicas sofisticadas de elicitación.
2. Variedade e heterogeneidade de fontes	A diversidade de fontes de conhecimento, que podem variar de textos escritos e bases de dados a conhecimentos especializados humanos, apresenta desafios significativos para a coleta e unificação de informações.
3. Qualidade e confiança das informações	Avaliar a qualidade, precisão e confiabilidade das informações coletadas é crucial, mas desafiador, especialmente em domínios nos quais o conhecimento está em constante evolução.
4. Escalabilidade	À medida que a quantidade de conhecimento disponível cresce, torna-se cada vez mais difícil capturar, processar e organizar essas informações de maneira eficiente.
<b>Melhoria do conhecimento</b>	
1. Manutenção e atualização contínuas	Garantir que a base de conhecimento permaneça atualizada com as mais recentes descobertas e informações é um desafio constante, especialmente em campos que evoluem rapidamente.
2. Consistência e conflitos	Manter a consistência dentro da base de conhecimento, especialmente quando novas informações contradizem ou alteram o entendimento existente, pode ser difícil.
3. Representação e modelagem de conhecimento	Desenvolver representações de conhecimento que sejam ao mesmo tempo ricas, precisas e computacionalmente eficientes é um desafio significativo. Isso inclui a escolha de esquemas de representação adequados que possam capturar a complexidade do domínio.
4. Interoperabilidade	Facilitar a interoperabilidade entre diferentes sistemas e bases de conhecimento, cada um possivelmente usando diferentes esquemas de representação e ontologias, é um desafio técnico importante.
5. Acessibilidade e usabilidade	Garantir que o conhecimento possa ser facilmente acessado e utilizado por diferentes usuários, incluindo tanto especialistas no domínio quanto leigos, requer interfaces intuitivas e mecanismos eficientes de busca e recuperação.

Fonte: Adaptado de Sivarajah *et al.* (2017).

Os desafios na integração de dados, de acordo com Sivarajah *et al.* (2017), são especialmente graves, uma vez que geralmente provêm de fontes heterogêneas e apresentam distintos formatos e estruturas. A falta de padronização e a necessidade de unificar dados de diversas origens complicam o processo, aumentando a possibilidade de erros. Além disso, defendem os autores, organizar os dados de tal maneira que se torne fácil a sua recuperação e análise é uma tarefa complexa, sobretudo quando se maneja grandes volumes de informações que devem ser estruturadas de maneira coerente e acessível.

Nesse contexto, Fensel *et al.* (2020) destacam que os grafos de conhecimento desempenham um papel crucial na construção e visualização de redes de autores e coautores em artigos científicos, oferecendo uma representação visual e estrutural das relações e colaborações no universo acadêmico. Para os autores, essa abordagem não somente facilita a identificação de padrões de colaboração e áreas de pesquisa comuns, mas também promove

uma compreensão mais profunda das dinâmicas e tendências dentro de comunidades científicas específicas. Sivarajah *et al.* (2017) concordam com Fensel *et al.* (2020) e ainda complementam que, ao mapear as conexões entre autores e coautores, os grafos de conhecimento proporcionam *insights* valiosos sobre a estrutura de redes de pesquisa, possibilitando analisar como a colaboração influencia o desenvolvimento e a disseminação do conhecimento científico.

## 1.2 PROBLEMA DE PESQUISA

Os grafos de conhecimento ajudam a revelar a centralidade e a influência de determinados pesquisadores dentro de uma rede, identificando aqueles que atuam como pontos de conexão cruciais entre diferentes subgrupos ou áreas de pesquisa. Isso é especialmente útil para novos pesquisadores que buscam colaboradores potenciais ou desejam entender melhor as relações existentes dentro de seu campo.

A visualização de grafos de conhecimento pode destacar padrões de colaboração interdisciplinar, evidenciando como diferentes áreas do conhecimento se interconectam e contribuem para o avanço da ciência. Essa capacidade de mapear e analisar redes complexas de colaboração pode orientar estratégias de pesquisa, fomentar parcerias produtivas e estimular inovações científicas (Sivarajah *et al.*, 2017).

Por fim, no contexto da gestão do conhecimento e da análise bibliométrica, os grafos de conhecimento são ferramentas indispensáveis. Eles possibilitam a identificação de tendências emergentes, áreas de pesquisa em crescimento e lacunas no conhecimento existente, fornecendo uma base sólida para a tomada de decisões estratégicas em pesquisa e desenvolvimento. A visualização de redes de coautoria por intermédio de grafos de conhecimento não somente enriquece a compreensão das dinâmicas de colaboração científica, mas também amplia a capacidade de pesquisadores e instituições de navegarem efetivamente no vasto e complexo ecossistema da pesquisa científica. Em resumo, os grafos de conhecimento são fundamentais para desvendar a estrutura e a evolução das colaborações científicas, impulsionando a inovação e o progresso do conhecimento humano (Fensel *et al.*, 2020).

Nesse sentido, a questão norteadora desta pesquisa é: quais as contribuições dos grafos de conhecimento para a Engenharia do Conhecimento?

## 1.3 OBJETIVOS

### 1.3.1 Objetivo geral

Identificar as contribuições de grafos de conhecimento para a Engenharia do Conhecimento.

### 1.3.2 Objetivos específicos

Para alcançar o objetivo geral, foram delimitados os seguintes objetivos específicos:

- a) identificar definições, modelos e propriedades de grafos de conhecimento por meio de uma revisão de literatura;
- b) apresentar as características de representação e uso do conhecimento com grafos de conhecimento;
- c) reconhecer ferramentas de armazenamento, recuperação, representação visual e operação de grafos de conhecimento;
- d) desenvolver um grafo de conhecimento para expor as potencialidades da abordagem.

## 1.4 JUSTIFICATIVA

Por mais de cinco décadas, especialistas e cientistas têm se dedicado à descoberta e formalização de conhecimentos especializados para apoiar aplicações de Inteligência Artificial (IA), o que resultou em sistemas especializados e bases de conhecimento. Segundo Simsek *et al.* (2022), os primeiros passos focaram a criação de bases de conhecimento com ontologias detalhadas. Atualmente, os grafos de conhecimento representam a mais recente tentativa de atingir esse objetivo, visto que constituem redes semânticas extensas que combinam diversas fontes de informação heterogênea, destacando-se das tradicionais bases de conhecimento.

Os grafos de conhecimento são essenciais na Engenharia do Conhecimento por sua capacidade de estruturar e interligar dados de forma significativa, promovendo uma compreensão mais profunda e uma exploração eficiente do conhecimento. Em um cenário no qual o volume de dados cresce exponencialmente, eles possibilitam a representação intuitiva de conceitos e relações, o que facilita a descoberta, a recuperação e a aplicação do conhecimento. Além disso, servem como base para sistemas de recomendação, motores de busca semântica e

assistentes inteligentes, sendo indispensáveis para a análise de dados complexos, a tomada de decisões informadas e o avanço tecnológico e científico (Ji *et al.*, 2021).

Ainda, segundo Ji *et al.* (2021), os grafos de conhecimento possibilitam a representação explícita de relações entre entidades, o que melhora significativamente a capacidade de realizar inferências e análises avançadas.

Ehrlinger e Wöb (2016) destacam que a estrutura de grafos é particularmente eficaz para a modelagem de dados semânticos, o que proporciona uma melhor organização e recuperação de informações. Essas características tornam os grafos de conhecimento indispensáveis para a construção de sistemas inteligentes e a resolução de problemas complexos de extração e demonstração de conhecimento.

Hogan *et al.* (2021) e Paulheim (2017) apontam que os grafos de conhecimento facilitam a integração de dados heterogêneos, o que concede uma visão unificada e coerente que é essencial para a extração e demonstração de conhecimento. Além disso, autores como Ehrlinger e Wöb (2016) e Suchanek *et al.* (2007) ressaltam que a estrutura gráfica dos dados melhora a eficiência das consultas complexas, tornando possível a análise de grandes volumes de informações inter-relacionadas de maneira mais rápida e precisa.

Auer *et al.* (2007) também reforçam a importância dos grafos de conhecimento na criação de sistemas inteligentes, capazes de inferir novas informações com base nos dados existentes, de modo a promover uma maior precisão e eficiência na gestão do conhecimento. Essas características tornam os grafos de conhecimento uma ferramenta indispensável para resolver problemas complexos e extrair valor de grandes volumes de dados interconectados.

Ferramentas como Neo4j, amplamente discutidas por Robinson, Webber e Eifrem (2015), exemplificam a aplicação prática dos grafos de conhecimento, oferecendo uma linguagem de consulta intuitiva e uma *performance* robusta, mesmo em cenários de alta complexidade e volume de dados.

Conforme Simsek *et al.* (2022), o uso de dados em forma de grafos e o desenvolvimento dos grafos de conhecimento tornaram-se comuns, facilitando a exploração e integração de dados em um contexto com metadados que definem regras para estruturar e interpretar informações. Esses grafos potencializam a eficiência na gestão de dados, otimizam a automação, refinam operações e facilitam previsões mais acertadas, sendo especialmente úteis nos âmbitos empresarial e acadêmico. O crescente interesse por essa ferramenta é evidenciado pelo aumento de publicações científicas, análises de especialistas, grupos de discussão e conferências a respeito do tema, destacando sua ligação com novas tecnologias *web*,

interoperabilidade de dados, análise de grandes volumes de dados e processamento em *cloud computing*.

A organização de dados como grafos de conhecimento, conforme discutido por Hur, Janjua e Ahmed (2021), oferece diversas aplicações inteligentes, como sistemas de perguntas e respostas, mecanismos de recomendação e pesquisa semântica. Os autores afirmam que esses grafos representam uma inovação que facilita a inferência lógica e a descoberta de novos conhecimentos, considerando tanto o conteúdo quanto o contexto. Esse método possibilita compreender de forma mais aprofundada as relações entre dados, promovendo análises avançadas e detalhadas. Assim, os grafos de conhecimento têm potencial para impulsionar significativamente o desenvolvimento de soluções inovadoras em diversas áreas, o que contribui para avanços na (IA) e na ciência de dados.

Com o surgimento da internet e a massiva produção de dados em diversos formatos, como textos, gráficos, sons e vídeos, os grafos de conhecimento tornaram-se essenciais em aplicações de inteligência artificial, incluindo sistemas de recomendação, pesquisa semântica e assistentes virtuais, em razão da sua capacidade de representar conhecimento de forma estruturada e acessível. Além disso, a integração com tecnologias de processamento de linguagem natural é crucial para organizar essas informações em grandes grafos, visto que fornece respostas precisas no momento necessário e apoia o desejo dos usuários de adquirir conhecimento. A ontologia, que define um conjunto de conceitos e categorias em um sujeito ou domínio e apresenta suas propriedades e as relações entre eles, desempenha um papel fundamental na estruturação desses grafos de conhecimento, a fim de proporcionar uma base sólida para a interoperabilidade e a análise semântica dos dados (Fensel *et al.*, 2020).

No contexto acadêmico, pontuam Fensel *et al.* (2020), os grafos de conhecimento contribuem significativamente para a análise e visualização das redes de autores e coautores, fornecendo ideias vantajosas sobre as dinâmicas de colaboração e produção científica. De acordo com os autores, ao mapear as conexões entre pesquisadores e seus trabalhos, revelam padrões de colaboração, influência e especialização dentro de comunidades acadêmicas, ajudando a identificar líderes de pensamento e centros de excelência, além de facilitar a descoberta de áreas emergentes de pesquisa e potenciais lacunas no conhecimento. Fensel *et al.* (2020) afirmam que, para pesquisadores individuais e instituições acadêmicas, entender essas redes pode orientar estratégias de publicação, fomentar colaborações produtivas e maximizar o impacto científico.

Esta pesquisa, desafiadora pela sua abrangência e relevância, se alinha com os objetivos do programa de Pós-graduação em Engenharia e Gestão do Conhecimento

(PPGEGC). A integração de grafos de conhecimento a outras tecnologias, como ontologias, bancos de dados de grafos, ferramentas de análise de dados e inteligência artificial, é uma área crucial de pesquisa e desenvolvimento, que amplia suas funcionalidades e melhora a eficiência dessas tecnologias.

A temática dos grafos de conhecimento é ampla e relevante, com crescente importância em áreas como negócios, saúde, segurança e educação. Nas universidades, eles melhoram a busca e a análise de informações, visto que personalizam recomendações e experiências de aprendizado.

## 1.5 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO

Nesta dissertação, objetiva-se identificar as contribuições dos grafos de conhecimento para a Engenharia do Conhecimento ao explorar sua aplicação na organização e facilitação do uso do conhecimento.

No Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC) da Universidade Federal de Santa Catarina (UFSC), conforme descrito por Pacheco (2016), a Engenharia do Conhecimento é voltada para a criação e o aprimoramento de métodos, técnicas e instrumentos dedicados à construção de sistemas de conhecimento. Esses sistemas são projetados para fornecer suporte a organizações e indivíduos. Esta dissertação está associada à linha de pesquisa de Engenharia do Conhecimento do PPGEGC, alinhando-se à abordagem que visa ao aprimoramento de serviços de conhecimento.

No PPGEGC, esta pesquisa se insere no contexto do grupo de pesquisa “StudioKEM – Modelos e sistemas computacionais apoiados em engenharia e gestão do conhecimento”, que explora métodos, tecnologias e ferramentas para o desenvolvimento de modelos e sistemas computacionais apoiados em Engenharia e Gestão do Conhecimento. A utilização de grafos de conhecimento é relevante para esse grupo, pois facilita a agregação e a análise de informações de diversas fontes, promovendo a descoberta de novos conhecimentos.

Segundo Pacheco (2016), a especialização em Engenharia do Conhecimento no PPGEGC analisa o conhecimento sob a perspectiva cognitiva, considerando-o como “processo e produto tangível ou intangível efetivado na relação entre pessoas e agentes não humanos para a geração de valor”. Os grafos de conhecimento, ao facilitar a integração e a interpretação de grandes volumes de informações, apoiam essa visão, uma vez que possibilitam que as máquinas forneçam informações e conhecimentos mais precisos.

A Engenharia do Conhecimento visa desenvolver ferramentas que facilitem a aquisição, o armazenamento e a disseminação do conhecimento em organizações centradas em atividades de alta intensidade intelectual. Conforme apontam Rochadel, Souza e Dandolini (2016), essa disciplina busca gerar valor tecnológico, científico e econômico por meio da padronização terminológica, como o uso de ontologias. Para isso, considera a semântica das informações e a interoperabilidade dos processos, assegurando a utilização eficiente e eficaz do conhecimento. Tal abordagem oportuniza às organizações aprimorar sua capacidade de inovação, solucionar problemas complexos e tomar decisões informadas, fortalecendo, assim, seu potencial intelectual e fomentando o desenvolvimento sustentável.

Na revisão de literatura baseada no Repositório Institucional da UFSC e nos dados do PPGEGC, não foram encontrados estudos anteriores que abordem diretamente o tema deste trabalho. Entretanto, conforme o Quadro 2 a seguir, há pesquisas que exploram técnicas semelhantes, como *web* semântica, ontologia, agentes e modelagem de conhecimento, o que destaca a pertinência e o potencial dos grafos de conhecimento. Esses grafos contribuem significativamente para a Engenharia do Conhecimento ao organizar e utilizar grandes volumes de informação, estimulando a inovação e a resolução de problemas complexos.

Quadro 2 – Principais trabalhos relacionados ao tema

Ano	Título	Autor	Orientador	Tipo
2014	Uso da <i>web</i> de dados como fonte de informação no processo de inteligência competitiva setorial	Dal Pizzol, Leandro	Prof. Dr. José Leomar Todesco	Dissertação
2015	Um modelo baseado em ontologia para suporte a tarefa intensiva em conhecimento de recomendação	Silva, Thales do Nascimento da	Prof. Dr. Alexandre Leopoldo Gonçalves	Dissertação
2015	Aplicação de ontologias na organização de conteúdos para apoio a equipes de desenvolvimento de <i>software</i>	Botelho, Maurício	Prof. Dr. Denilson Sell	Dissertação
2016	Modelo para descoberta de conhecimento baseado em associação semântica e temporal entre elementos textuais	Woszezenki, Cristiane Raquel	Prof. Dr. Alexandre Leopoldo Gonçalves	Tese
2017	Sistema baseado em conhecimento (SBC) de apoio à capacitação organizacional	Martínez, Diego Jessie	Prof. <sup>a</sup> Dra. Lia Caetano Bastos	Dissertação
2017	Modelo de conhecimento para representação semântica de <i>smart cities</i> com foco nas pessoas	Anderle, Daniel Fernando	Prof. Dr. Marcelo Macedo	Tese
2019	KE-IoT: uma proposta de modelo baseado em conhecimento para ambientes de internet das coisas (IoT)	Quingerski, Leandro	Prof. Dr. Mário Antônio Ribeiro Dantas	Dissertação

Fonte: Elaborado pelo autor (2024).

A aplicação específica de grafos de conhecimento facilita a integração de dados complexos e promove a descoberta de novas conexões. O foco na aplicação teórica e prática desses grafos para a representação de conhecimento oferece suporte a pesquisadores e organizações.

Em síntese, esta dissertação busca contribuir para a área da Engenharia e Gestão do Conhecimento, apresentando uma abordagem inovadora e prática para a gestão e análise de informações complexas, em consonância com os objetivos do PPGEGC e suas linhas de pesquisa.

## 1.6 PROCEDIMENTOS METODOLÓGICOS

Dresch, Lacerda e Antunes Júnior (2015) afirmam que a ciência e a educação sustentam o conhecimento, e que o avanço do conhecimento científico se baseia na pesquisa com a finalidade de solucionar problemas e comparar teorias. Nesse contexto, a ciência desempenha um papel crucial ao fornecer métodos e ferramentas para a investigação sistemática, enquanto a educação dissemina esses conhecimentos, formando novas gerações de pesquisadores e profissionais capacitados.

Os grafos de conhecimento, segundo Peng *et al.* (2023), têm se mostrado uma ferramenta poderosa na produção e gestão do conhecimento científico. De acordo com os autores, eles possibilitam a integração de dados de diversas fontes e áreas do conhecimento, o que facilita a compreensão e análise de informações complexas. Essa capacidade de conectar diferentes domínios do conhecimento, completam os autores, é essencial para a pesquisa científica, pois assim é possível que haja a identificação de novas relações e padrões que podem levar a descobertas inovadoras. Além disso, concluem, os grafos de conhecimento auxiliam na organização e representação do conhecimento, tornando-o mais acessível e utilizável para pesquisadores e profissionais de diversas áreas.

De acordo com Peng *et al.* (2023), a aplicação de grafos de conhecimento na pesquisa científica também contribui para a melhoria da precisão e relevância das informações fornecidas pelas máquinas. Os autores pontuam que, ao estruturar os dados de maneira que as máquinas possam entender melhor o seu significado, os grafos possibilitam que sistemas de inteligência artificial e aprendizado de máquina façam inferências mais precisas e forneçam recomendações mais relevantes. Isso não só aumenta a eficiência da pesquisa científica, mas também facilita a tomada de decisões informadas, promovendo avanços significativos em diversas áreas do conhecimento.

A pesquisa, sendo uma investigação sistemática, segundo Dresch, Lacerda e Antunes Júnior (2015), pode ser de natureza teórica, cujo limite é o progresso científico, ou prática, aplicada à solução de problemas. Os autores enfatizam que a escolha de um método de pesquisa adequado torna fundamentado o conhecimento científico. Para os autores, o avanço do conhecimento científico parte da pesquisa com o objetivo de solucionar problemas específicos ou comprovar teorias.

Este estudo, qualitativo e descritivo em natureza, visa explorar a aplicação teórica de grafos de conhecimento por meio de pesquisa bibliográfica, com vistas a investigar os principais desafios na área da representação do conhecimento, avaliando como esses grafos podem impactar e melhorar a prática da Engenharia do Conhecimento por intermédio do uso de ferramentas. É indiscutível a importância dos fatores externos na ampliação do uso dos grafos de conhecimento.

Para tanto, na pesquisa bibliográfica, foram utilizados bancos de dados como SciELO, Springer Link, IEEE Xplore, ACM Digital Library, Science Direct, Neo4j, entre outros, com foco em publicações entre 2012 e 2023. A pesquisa incluiu artigos de periódicos, conferências, *workshops*, dissertações de mestrado e doutorado, localizados por meio do uso de palavras-chave como “grafos de conhecimento”, “banco de dados em grafos”, “representação do conhecimento”, “ferramentas de grafos” e “Engenharia do Conhecimento”.

Ainda, para demonstrar a aplicação prática dos grafos de conhecimento na Engenharia de Conhecimento, foram utilizados os dados do Congresso Internacional de Conhecimento e Inovação – CIKI/UFSC 2023 e do repositório institucional da UFSC para construir um grafo de conhecimento (GC) com o objetivo de demonstrar sua importância na organização e análise de informações de eventos e trabalhos acadêmicos, por meio da conversão de dados brutos em conhecimento acionável. Quanto aos dados do CIKI 2023, obtidos via *web scraping*, foram recortados especificamente os relacionados aos temas discutidos, sendo priorizados os títulos dos artigos, os autores e coautores, e as instituições participantes; do repositório, foram recortados os dados de teses e dissertações do PPGEGC do ano de 2023.

Para melhor compreensão dos procedimentos metodológicos adotados nesta pesquisa, no Quadro 3 a seguir, foram detalhadas e ilustradas de forma sequencial as etapas essenciais do estudo, desde a definição do tema e dos objetivos até considerações finais e recomendações. Cada etapa descrita busca garantir uma abordagem sistemática e estruturada, de modo a facilitar a replicação e o entendimento das metodologias empregadas, bem como a análise crítica dos resultados obtidos.

Quadro 3 – Descrição das etapas

<b>Descrição das etapas</b>	
Definição do tema e objetivos	Escolha do tema da pesquisa (grafos de conhecimento). Definição dos objetivos gerais e específicos da dissertação.
Revisão da literatura	Pesquisa em artigos, livros e outros materiais relevantes. Identificação de lacunas na literatura existente.
Metodologia	Escolha da abordagem metodológica (qualitativa). Definição das técnicas de coleta e análise de dados.
Coleta de dados	Utilização de base de dados como Springer Link, IEEE Xplore, ACM Digital Library, Science Direct. Coletar dados relevantes que suportem a pesquisa.
Análise de dados e desenvolvimento de solução	Aplicação de técnicas de análise de dados apropriadas. Utilização de ferramentas de grafos, bancos de dados de grafos etc.
Discussão dos resultados	Interpretação dos resultados obtidos.
Considerações finais e recomendações	Resultado dos principais achados da pesquisa. Recomendações para futuras pesquisas e aplicações práticas.

Fonte: Elaborado pelo autor (2024).

## 1.7 LIMITAÇÕES DA PESQUISA

Nesta pesquisa, objetiva-se desenvolver um grafo de conhecimento baseado nos dados dos artigos do CIKI, além dos dados de teses e dissertações do Repositório Institucional da UFSC relacionadas ao Programa de Engenharia e Gestão do Conhecimento para o ano de 2023, com o propósito de demonstrar as potencialidades dessa abordagem. Por meio da análise e organização desses dados, pretende-se estabelecer conexões e relações entre diferentes áreas de conhecimento, com vistas a proporcionar uma compreensão mais ampla e integrada das pesquisas realizadas. O grafo de conhecimento será um recurso valioso para estudantes, pesquisadores e profissionais que desejam ampliar seus conhecimentos em várias áreas acadêmicas; sua implementação poderá estimular pesquisas interdisciplinares, revelar lacunas e fortalecer a produção científica e acadêmica.

## 1.8 ESTRUTURA DO TRABALHO

A estrutura deste trabalho está dividida em cinco capítulos. O primeiro, designado “Introdução”, é composto pelas considerações iniciais sobre a temática abordada, a identificação do problema de pesquisa, os objetivos geral e específicos, a justificativa, a aderência ao programa de pós-graduação, o procedimento metodológico utilizado e respectivas limitações consideradas para a pesquisa. No capítulo 2, é apresentada uma revisão da literatura sobre Engenharia do Conhecimento, bem como são definidos os conceitos e a arquitetura da *web*, com destaque para a *web* semântica (SW), *Resource Description Framework* (RDF),

*Resource Description Framework Schema (RDFS)* e *Ontology Web Language (OWL)*. Já no Capítulo 3, aborda-se o histórico, os conceitos e os modelos de grafos de conhecimento, assim como bases de conhecimento como a DBpedia, Yago, Freebase, Wikidata e OpenCyc. No Capítulo 4, é delineado o desenvolvimento do grafo de conhecimento. Por fim, no Capítulo 5, são apresentadas as considerações finais.

## 2 ENGENHARIA DO CONHECIMENTO

Schreiber *et al.* (2000) observam que o conhecimento é um conjunto de dados e informações utilizado para realizar tarefas e gerar novas informações. Davenport e Prusak (1998) acrescentam que o conhecimento é oriundo da mente das pessoas e está intimamente ligado à ação. O conhecimento é uma competência individual e dinâmica, ligada à capacidade de ação e exige habilidades para entender, compreender, utilizar e integrar dados e informações disponíveis para obter melhores resultados.

Engenharia do Conhecimento refere-se ao processo de captura e aplicação de conhecimentos para diversas finalidades, incluindo o desenvolvimento de sistemas especialistas e inteligência artificial. Esse campo também está intimamente ligado à lógica matemática e participa amplamente das ciências cognitivas e da engenharia sociocognitiva. Nesse contexto, o conhecimento é gerado por agrupamentos sociocognitivos, principalmente humanos, e é organizado com base na nossa compreensão do funcionamento do raciocínio e da lógica humana (Ting, 2010).

Considerando um fluxo que envolve criação, organização, formalização, compartilhamento, aplicação e aprimoramento do conhecimento, Rautenberg, Todesco e Steial (2010) apontam que a ontologia auxilia o engenheiro do conhecimento a estruturar e codificar o conhecimento dentro da perspectiva da gestão do conhecimento, da mesma maneira que o gestor de conhecimento entende as soluções tecnológicas oferecidas pela Engenharia do Conhecimento.

Conforme destacado por Allen, Ilievski e Joshi (2023), a Engenharia do Conhecimento é um campo envolvente e dinâmico, vital para o *design* de sistemas inteligentes. Esse ramo busca criar sistemas capazes de capturar, codificar e aplicar o conhecimento humano na resolução de problemas complexos. A elaboração de modelos que refletem a perícia humana é essencial para o desenvolvimento de sistemas especialistas e inteligentes. Sendo inerentemente interdisciplinar, a Engenharia do Conhecimento reúne aspectos de inteligência artificial, ciência da computação e gestão do conhecimento para trazer soluções inovadoras. Para Schreiber *et al.* (2000), a colaboração entre especialistas de domínio e engenheiros do conhecimento é crucial para captar e formalizar o conhecimento, convertendo habilidades humanas em sistemas que aprendem e se adaptam. Segundo os autores, essa área é identificada pelos pesquisadores como o desenvolvimento de sistemas que incorporam, codificam e aplicam a sabedoria humana para solucionar problemas complexos, fazendo desses modelos elementos fundamentais para sistemas computacionais avançados. Segundo Pacheco (2006), a Engenharia do Conhecimento

evidencia a pesquisa e o desenvolvimento de técnicas e ferramentas para formalizar, codificar e gerenciar conhecimento, integrando processos conhecidos como “gestão do conhecimento”.

A Engenharia do Conhecimento também é considerada um método sistemático para adquirir e codificar conhecimentos especializados, abrangendo a natureza vaga e genérica do conceito. Geralmente apresentada como um conjunto estruturado de representações contextualizadas, visa compreender e manter informações detalhadas aplicadas a contextos ou organizações específicas (Boeres *et al.*, 2014).

De acordo com Studer, Benjamins e Fensel (1998), as ontologias são uma ferramenta valiosa e fundamental na representação de conhecimento. Elas possibilitam a definição e a organização dos conceitos e das suas relações dentro de um determinado domínio, proporcionando uma base sólida para a interoperabilidade entre sistemas e a reutilização do conhecimento. Segundo os autores, as técnicas baseadas em grafos, como os grafos de conhecimento, são destacadas por sua capacidade de representar relações complexas entre dados.

Grimm, Hitzler e Abecker (2007) destacam que, no campo da IA, a representação do conhecimento é um conceito essencial que abrange os métodos de formalização, armazenamento e manipulação do conhecimento em sistemas computacionais, que visa capacitá-los para tomarem decisões e realizarem raciocínios de forma semelhante à inteligência humana.

## 2.1 REPRESENTAÇÃO DO CONHECIMENTO

A representação do conhecimento envolve: definição e propósito, representação simbólica, domínios de interesse, base de conhecimento e raciocínio, formas de representação do conhecimento.

- a) Definição e propósito: a representação do conhecimento envolve criar sistemas computacionais que raciocinem sobre o mundo com base em um modelo de conhecimento compreensível pela máquina. Isso inclui desenvolver um modelo computacional que simbolize variados aspectos do mundo real ou sistemas hipotéticos, o que possibilita que computadores analisem essas informações para executar tarefas e tomar decisões.
- b) Representação simbólica: nos sistemas baseados em conhecimento, utilizam-se símbolos para representar entidades do mundo real, como objetos, eventos e relacionamentos. Esses símbolos são armazenados em uma base de conhecimento

e manipulados por intermédio de processos de raciocínio para gerar novas informações ou tomar decisões fundamentadas no conhecimento existente.

- c) Domínios de interesse: a amplitude da representação do conhecimento pode ser bastante diversa, estendendo-se tanto a partes específicas do mundo real quanto a sistemas conceituais abrangentes. Essa versatilidade possibilita a aplicação de sistemas baseados em conhecimento em uma vasta gama de campos e para multifacetados objetivos, desde a realização de pesquisas científicas até o desenvolvimento de aplicações empresariais diversas.
- d) Base de conhecimento e raciocínio: um elemento crucial da representação do conhecimento é a base de conhecimento, que contém os símbolos que representam o saber sobre o domínio. Os sistemas executam o raciocínio ao manipular esses símbolos, possibilitando que respondam a perguntas pertinentes ao domínio ou tomem decisões fundamentadas com base nas informações armazenadas.
- e) Formas de representação do conhecimento: o conhecimento pode ser representado de diversas formas, como redes semânticas, regras e lógica. Redes semânticas, usadas em representações de grafos RDF ou mapas de tópicos, organizam visualmente as relações entre conceitos. As regras, geralmente no formato “se, então”, formalizam operações comerciais ou lógicas. A lógica, por sua vez, oferece uma semântica formal sólida que facilita a dedução automatizada e a interpretação precisa (Grimm; Hitzler; Abecker, 2007),

Nos últimos anos, engenheiros do conhecimento criaram princípios, métodos e ferramentas para tornar a aquisição de conhecimento mais eficiente, conforme ressalta Ting (2010). Ele afirma que esses avanços são importantes para a gestão do conhecimento, destacando o crescente interesse na área por instituições acadêmicas e empresas. A inteligência artificial, com raízes teóricas e práticas amplas, beneficia-se de influências da filosofia, da psicologia e da linguística, que oferecem uma sólida base de ideias aplicadas.

Sistemas baseados em conhecimento, ou sistemas especializados, são essenciais na Engenharia do Conhecimento, mas desenvolvê-los exige altos custos e complexidade na aquisição de conhecimento. O CommonKADS foi proposto para simplificar a metodologia, tornando-a mais acessível e fácil de aprender tanto para usuários iniciantes como experientes, o que pode ser particularmente benéfico em ambientes organizacionais (Flórez; Carbó; Fernández, 2012).

Além do CommonKADS, que é uma metodologia amplamente utilizada na Engenharia do Conhecimento, existem várias outras ferramentas e metodologias que podem ser empregadas

em diferentes aspectos da construção e gestão do conhecimento. No Quadro 4, são apresentadas algumas ferramentas para a Engenharia do Conhecimento. Essas diversas ferramentas e metodologias aprimoram coletivamente a eficiência e a eficácia da Engenharia do Conhecimento em vários campos.

Quadro 4 – Ferramentas para Engenharia do Conhecimento

<b>Ferramenta</b>	<b>Características principais</b>	<b>Referências</b>
Protégé	Ambiente de desenvolvimento de ontologias. Suporta OWL RDF(S). Interface gráfica amigável.	(Noy; McGuinness, 2001)
CommonKADS	Metodologia abrangente para modelagem de conhecimento. Foco em análise de tarefas e agentes.	(Schreiber <i>et al.</i> , 2000)
OntoStudio	Ferramenta de modelagem de ontologias. Suporte para OWL. Integração com sistemas de bases de dados.	(Staab; Studer, 2009)
CmapTools	Ferramenta para construção de mapas conceituais. Facilita a visualização e organização do conhecimento.	(Novak; Cañas, 2007)
D3.js	Biblioteca JavaScript para visualização de dados. Suporte para gráficos interativos e dinâmicos.	(Bostock; Ogievetsky; Heer, 2011)
Apache Jena	<i>Framework</i> Java para construção de aplicações semânticas. Suporte para RDF, SPARQL.	(Mcbride, 2002)
ConceptNet	Rede semântica para representação de conhecimento comum. Suporte para integração com IA.	(Speer; Havasi, 2012)
RapidMiner	Plataforma de mineração de dados e aprendizado de máquina. Suporte para análise preditiva.	(Hofmann; Klinkenberg, 2014)
KNIME	Plataforma de integração de dados e análise. Suporte para <i>workflows</i> de ciência de dados.	(Berthold <i>et al.</i> , 2008)
IBM Watson	Plataforma de IA com capacidades de processamento de linguagem natural e aprendizado de máquina.	(Ferrucci <i>et al.</i> , 2011)

Fonte: Elaborado pelo autor (2024).

Recentemente, ferramentas criadas pela indústria para captar e reutilizar conhecimento foram eficazmente integradas em contextos educacionais para potencializar a aprendizagem. Essa integração proporcionou aos alunos acesso a tecnologias avançadas e habilidades técnicas relevantes para o mercado de trabalho, promovendo um ensino mais prático. (Sauza-Bedolla *et al.*, 2016).

## 2.2 CONCEITOS E ARQUITETURAS DA WEB

O objetivo da Web 3.0, destaca Ibrahim (2021), é organizar a maneira com que os usuários consultam e recebem conteúdo na internet. Com a introdução de linguagens como RDF e OWL, essa nova geração da Web se moldou para utilizar informações adquiridas de *sites* e dados disponibilizados pelos próprios usuários como base nessa estrutura. Para Ibrahim (2021), esse ambiente foi transformado em uma plataforma dedicada ao gerenciamento de dados,

fundamentada na análise da atividade dos usuários, em motores de busca e na inferência realizada. A Web 3.0 inovou ao atribuir significado aos dados por meio de linguagens como a OWL, baseando-se no princípio de redes semânticas, que remontam ao Extensible Markup Language (XML).

Blumauer e Nagy (2020) afirmam que, em 1999, o modelo RDF foi publicado como uma recomendação do World Wide Web Consortium (W3C) para estabelecer as bases para a *web* semântica. De acordo com Allemang, Gandon e Hendler (2020), RDF, RDFS e OWL são consideradas as linguagens fundamentais para a representação da *web* semântica.

As três linguagens de ontologia que foram propostas para descrever os recursos da *web*, conforme Abiteboul *et al.* (2011) são, inicialmente, a linguagem RDF, uma linguagem para expressar fatos com foco principal no banco de dados, e as linguagens RDFS e OWL, que possibilitam restringir fatos RDF em domínios de aplicativos específicos. Os autores destacam que o RDFS é bastante simples, enquanto o OWL é muito mais rico.

Segundo Milea, Frasinca e Kaymak (2012), as ferramentas e linguagens de última geração fornecidas sob esse “guarda-chuva”, como RDF, RDFS e OWL, vão além da tecnologia padrão da *web* e fornecem meios para compartilhamento e reutilização de dados fora dessa plataforma, como dados semânticos. Os autores explicam que a necessidade considerável e crescente de acesso ao grande volume de dados presentes na World Wide Web atualmente motiva uma migração de representações de dados em texto livre para representações de informações semanticamente ricas, sendo empreendidos sob um denominador comum como a *web* semântica, apresentada a seguir.

### 2.3 WEB SEMÂNTICA

Nas últimas décadas, a internet consolidou-se como o principal recurso para acessar informações, graças à sua presença em todos os locais e à vasta quantidade de dados disponíveis. Como sua expansão acelerada tornou a missão de recuperação da informação cada vez mais complexa, sua eficiência está associada à implementação de algoritmos de categorização específicos. Com isso, foi sugerida por Tim Berners-Lee uma versão renovada da internet, conhecida como *web* semântica (Machado; Souza; Simões, 2019).

A *web* semântica, segundo Castelo Branco Neto (2006), foi criada por causa da necessidade de organizar e detalhar o conteúdo das páginas da *web*, assim como representar entidades ausentes no ambiente digital. O autor pontua que, mesmo após encontrarem as

informações desejadas na *web*, os usuários muitas vezes enfrentam dificuldades em decorrência da estruturação dessas informações.

De acordo com Hogan (2020), o começo da *web* semântica abarcou as edições inaugurais dos padrões RDF e RDFS. O autor ressalta a visão de Berners-Lee sobre a importância do RDF para efetivar a *web* semântica, a qual visa organizar o conteúdo das páginas da internet de modo significativo, o que possibilita a navegação e a realização de diversas atividades pelos *softwares* de forma eficiente para o usuário.

A *web* semântica, caracterizada por Bukhari, Mir e Ahmad (2017), é uma evolução avançada da *web* tradicional, enriquecida com uma variedade de tecnologias padronizadas. Para os autores, essas tecnologias habilitam máquinas a compreender e processar informações, que são distribuídas e reaproveitadas para além de seus usos iniciais, criando uma ampla rede de dados. Ainda, de acordo com Bukhari, Mir e Ahmad (2017) e Hitzler (2021), as tecnologias fundamentais para a *web* semântica compreendem o RDF, o OWL e o Protocol and RDF Query Language (SPARQL), e estão intimamente ligadas aos grafos de conhecimento.

A finalidade principal da *web* semântica, destacam Berners-Lee, Hendler e Lassila (2001), é facilitar a busca, o manuseio e o processamento da informação, visando enriquecer a experiência dos usuários. Para os autores, a ideia é possibilitar que os sistemas de computador compreendam os dados de acordo com os domínios específicos em que estão inseridos, oferecendo aos usuários informações enriquecidas e de maior valor agregado.

Breitman (2010) defende que o propósito da *web* semântica é possibilitar que máquinas realizem de forma automatizada o processamento de dados que, anteriormente na *web* convencional, dependiam de interação humana. Para atingir essa meta, o autor aponta que é importante contar com linguagens de programação adequadas que proporcionem a publicação de ontologias de maneira que os computadores possam interpretar e manipular as informações sem a necessidade de ajuda humana.

A *web* semântica se baseia na criação de modelos abstratos que capturam as complexidades do mundo utilizando conceitos simples, incluindo tanto a modelagem conceitual quanto a integração do conhecimento computacional. Conforme Hitzler, Krötzsch e Rudolph (2009), o objetivo é representar informações de forma que máquinas possam realizar inferências e raciocínio lógico de maneira automatizada, impulsionando o avanço da inteligência artificial na área da computação. Ainda, segundo os autores, a *web* semântica refere-se à partilha de informações entre computadores, à elaboração de modelos conceituais e à aplicação de inteligência artificial. Isso, para eles, possibilita a transmissão de recursos complexos, a

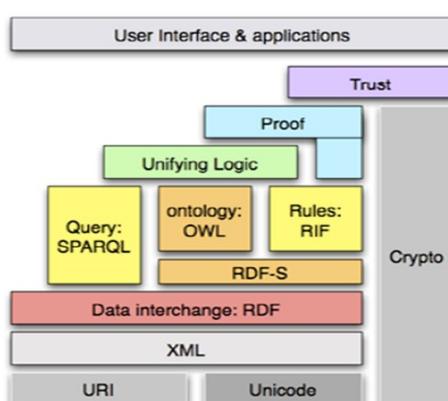
descrição abstrata de domínios de interesse e o desenvolvimento de máquinas capazes de construir raciocínio lógico para inferir conhecimentos codificados.

Blumauer e Nagy (2020) citam a publicação do artigo “A *web* semântica”, por Tim Berners-Lee, Jim Hendler e Ora Lassila, na revista *Scientific American*, em 2001, como uma concepção da internet em que tanto os humanos quanto as máquinas podem entender que as informações são compreensíveis, o que facilita, para os autores, uma conexão de dados mais inteligente e eficaz.

A *web* semântica, descrita por Berners-Lee, Hendler, Lassila (2001), é uma extensão da atual *web* que visa dar um sentido mais definido aos dados, de modo que máquinas e pessoas trabalhem em colaboração. Os autores afirmam que, para que isso se torne realidade, é crucial que os computadores processem conteúdo *web* com maior inteligência e eficiência, o que exige dados organizados e sistemas de inferência lógica. Ferreira (2014) pontua que a utilização da inferência enriquece a interpretação dos dados e aprimora a extração de significado e conhecimento. Para o pesquisador, a inferência amplia a capacidade dos sistemas de processar dados de forma mais inteligente e é crucial para descobrir novas relações em dados ao utilizar processos automatizados e regras lógicas pré-definidas, o que possibilita revelar padrões que aprofundam o entendimento dos dados ao transformá-los em um conjunto de interconexões.

A arquitetura em camadas da *web* semântica é descrita por El-Seoud, El-Sofany e Karam (2015) como uma organização hierárquica, que pode se manifestar em diferentes formas geométricas. Segundo os autores, cada camada suporta diretamente o nível superior e, por sua vez, se baseia no inferior, agindo como cliente. Eles também argumentam que mudanças em qualquer camada podem causar impactos que afetam até duas outras camadas adjacentes. Na Figura 1 a seguir, é exibido o modelo desenvolvido e apresentado por El-Seoud, El-Sofany e Karam (2015).

Figura 1 – Camadas da *web* semântica



Fonte: El-Seoud, El-Sofany e Karam (2015).

No contexto das tecnologias da *web* semântica (SWT), destacam Patel, Debnath e Bhushan (2023), é comum dar ênfase à descrição e representação das entidades nas diversas camadas existentes. Os autores apontam que a estrutura de camadas que compõe a *web* semântica sofreu adaptações ao longo do tempo para acompanhar as tendências atuais, particularmente em resposta ao surgimento de diversas inovações tecnológicas. Apesar de as tecnologias da *web* semântica terem recebido várias categorizações e descrições na literatura, essas tecnologias foram reunidas ao longo de um triplo de normas, métodos e ferramentas. Para os autores, foi considerado que as tecnologias da *web* semântica são coleções de normas, métodos e ferramentas que facilitam a visão de uma compreensão perfeita de dados ou informações por parte das máquinas atuando como substitutas de tarefas cognitivas. No Quadro 5, é apresentada a categorização das tecnologias da *web* semântica.

Quadro 5 – Categorização das tecnologias da *web* semântica

<b>Categoria</b>	<b>Descrição</b>	<b>Instâncias</b>
Normas	Formatos, linguagens ou vocabulários	UDI/IRI, XML, RDF/RDFS, SKOS, OWL, SPARQL, RIF.
Métodos	Abordagens, práticas ou técnicas	Análise contextual, mecanismo de raciocínio, compreensão de linguagem natural, grafo de conhecimento, dados vinculados e ontologia.
Ferramentas	Editores, ambientes de desenvolvimento ou <i>plug-ins web</i>	Ferramentas de anotação (ex.: Kmwotator), ferramenta de aquisição (ex.: OntoEdit), ferramentas de renderização (ex.: Protege) e raciocionadores (ex.: Pellet).

Fonte: Elaborado pelo autor (2024).

Paulheim (2017) reafirma a importância de estruturar conjuntos de tipos e relações em ontologias ou esquemas, estabelecendo suas interligações e limitações de aplicação. Para o autor, a ideia do Linked Data gerou a conexão entre diversos bancos de dados na *web* semântica, formando extensos grafos de conhecimento global, mesmo considerando sua diversidade de dados. O autor explica que, desde o início, a comunidade da *web* semântica defende o uso de grafos para a representação do conhecimento, com ênfase no formato padrão RDF, e que nessa representação do conhecimento, as entidades, que servem como nós dos grafos, são conectadas por relacionamentos, representados pelas arestas dos grafos.

Na seção seguinte, faz-se uma breve descrição do modelo RDF.

### 2.3.1 Resource Description Framework (RDF)

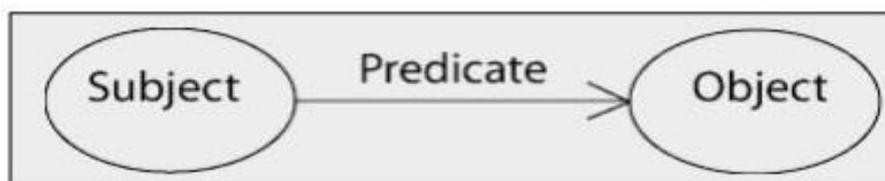
O modelo de descrição de recursos Resource Description Framework (RDF), de acordo com Papadaki, Tzitzikas e Mountantonakis (2023), é um modelo de dados baseado em

grafos que foi proposto para a implementação da ideia de *web* semântica e é o principal formato do método de publicação de Linked Data. Para Wang *et al.* (2017), o RDF é uma especificação do W3C, sendo muito utilizado na *web* semântica.

Segundo Chah (2022), o RDF é tratado como um modelo de representação de recursos, e, por vezes, é considerado apenas uma tecnologia da *web* semântica, pelo seu uso nas mais diversas ferramentas que estão dentro do escopo desse campo de estudo. Para o autor, o conceito que embasa a representação de recursos utilizando o modelo RDF tem extrapolado o próprio RDF, uma vez que há uma série de aplicações que tem em sua base os conceitos da *web* semântica e que utilizam o princípio do RDF sem utilizar o modelo em si.

Conforme Gelling, Fletcher e Schmidt (2023) e Chah (2018), um conjunto de dados RDF consiste basicamente em triplas de elementos com sujeito, predicado e objeto, representando fatos individuais. Essas triplas podem ser organizadas em estruturas conhecidas como “grafos nomeados”, desse modo, um conjunto de dados RDF é uma coleção desses grafos. Em consonância com os autores citados, Donkers, Yang e Baken (2020) e Klyne, Carroll e McBride (2004) descrevem o sujeito como um nó, o predicado como conexão e o objeto como outro nó ou propriedade. A direção do arco é significativa, pois sempre aponta para o objeto, como mostra a tripla RDF na Figura 2.

Figura 2 – Tripla RDF



Fonte: Klyne e Carroll (2004).

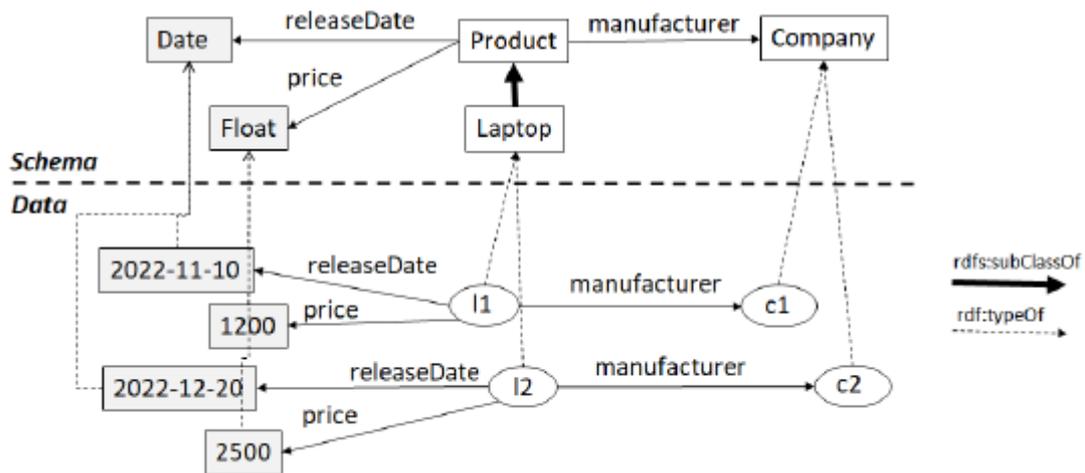
Além do RDF, de acordo com Tomaszuk, Angles e Thakkar (2020), o W3C desenvolveu e padronizou linguagens de alto nível que suportam modelagem e inferência RDFS e OWL, bem como uma linguagem de consulta declarativa, o SPARQL. Para os autores, embora não seja explicitamente definido usando conceitos de grafos, como vértices e arestas, há uma conexão estreita entre RDF e grafos: cada triplo pode ser entendido como uma aresta de um nó sujeito sob um determinado rótulo de predicado para um nó objeto. Na sequência, faz-se uma breve explicação do RDFS.

### 2.3.2 Resource Description Framework Schema (RDFS)

Conforme Papadaki, Tzitzikas e Mountantonakis (2023), o RDFS é um vocabulário que engloba diversas classes e propriedades, ampliando o modelo de dados RDF para a representação de conhecimento, com o objetivo de estruturar os elementos do RDF, uma vez que carece de expressividade semântica. Ele utiliza classes para categorizar recursos e emprega propriedades para estabelecer conexões entre entidades em uma classe e modelar restrições.

Na Figura 3 a seguir, é apresentado um grafo RDF com informações sobre produtos. A parte superior ilustra o *schema*, enquanto a parte inferior ilustra os dados.

Figura 3 – Grafo RDF sobre produtos



Fonte: Papadaki, Tzitzikas e Mountantonakis (2023).

De acordo com Wang *et al.* (2017), o RDF Schema (RDFS) é uma extensão do RDF que possibilita a definição de classes, propriedades e restrições. Os autores destacam que o RDFS é usado para criar modelos de dados mais complexos e é bastante utilizado na *web* semântica.

Segundo Patel, Debnath e Bhushan (2023), o Resource Description Framework Schema (RDFS), em vez de ser um modelo de dados, é uma linguagem que consiste em vocabulários para descrever as classes de recursos, bem como as propriedades que descrevem os recursos. De forma resumida, o RDFS pode ser visto como um vocabulário para modelagem de dados RDF.

Para Patel, Debnath e Bhushan (2023), o RDFS é a próxima camada que é a linguagem de vocabulário para RDF. O RDFS descreve classes e propriedades e registra a semântica associada aos recursos descritos pelo RDF, sendo as classes e propriedades organizadas na

forma de hierarquias de generalização e especialização. Ainda, acrescentam os autores, para propriedades, domínio e intervalo podem ser definidos usando RDFS.

O recurso básico para definir uma ontologia é suportado pelo RDFS, mas não é capaz de representar o relacionamento complexo entre as classes; para tanto, é adicionada a próxima camada, denominada “vocabulário de ontologia” (Patel; Debnath; Bhushan, 2023).

O vocabulário de ontologia representa ontologias em OWL, que é mais expressivo que o RDFS. Na seção a seguir, apresenta-se resumidamente a OWL.

### 2.3.3 Web Ontology Language (OWL)

Na Ciência da Computação, as ontologias são amplamente utilizadas para organizar e categorizar informações em diversos domínios. Segundo Rautenberg, Todesco e Gauthier (2009), ontologias definem conceitos, propriedades e relações, oferecendo uma estrutura semântica de conhecimento. Essas ferramentas são essenciais para lidar com a complexidade e heterogeneidade das informações, visto que aprimoram a organização e a compreensão do conhecimento em áreas como *web* semântica, recuperação de informações, integração de sistemas e inteligência artificial.

Na introdução sobre a OWL, Hogan (2020) destaca a importância de entender a origem filosófica do termo “ontologia” – que trata do estudo da existência das coisas e sua classificação. De acordo com Milea, Frasincar e Kaymak (2012), a OWL é a linguagem mais avançada para criar ontologias, essenciais para a *web* semântica.

Conforme Zhu *et al.* (2022), a linguagem OWL aproveita os recursos de declaração de fatos do padrão RDF e a arquitetura hierárquica de classes e atributos fornecidos pelo RDFS. Os autores destacam que a OWL amplia a modelagem de dados com numerosos termos novos, o que possibilita a representação de semânticas complexas e melhora a precisão de expressões, bem como contribui para aumentar a eficiência no processamento.

A finalidade da OWL, esclarece Hogan (2020), é estabelecer uma linguagem avançada para a criação de ontologias detalhadas que possam ser formalmente expressas e interpretadas de modo claro, e processadas por computadores com vistas a promover a interoperabilidade e aprimorar o compartilhamento e a integração tanto dos dados quanto das ontologias na *web*. Portanto, para o autor, a OWL estabelece também variações limitadas da linguagem, conhecidas como “sublinguagens”, as quais proporcionam vantagens concretas quanto à decisão e à complexidade computacional em variados contextos de aplicação.

A OWL, segundo Wang *et al.* (2017), é uma linguagem de ontologia formal que possibilita a criação de ontologias com alto teor semântico pelos desenvolvedores. De acordo com os autores, a utilização de OWL envolve a representação de conceitos, propriedades e axiomas, comum em *web* semântica e em diversos contextos da ciência da computação.

Frisendal (2016) enfatiza a vantagem da ontologia OWL na *web* semântica em razão de sua extensa variedade de funcionalidades. Seu alto desenvolvimento e eficiência a tornam um componente crucial em sistemas complexos que demandam precisão nos resultados de pesquisa.

De acordo com Paulheim (2017), é fundamental determinar a ontologia no começo da implementação dos grafos de conhecimento, pois ela direciona a estrutura, tornando uma representação clara das entidades e suas interconexões dentro de um domínio específico.

A melhor forma de representar grafos de conhecimento, segundo Zhu *et al.* (2022), consiste em expandir a base da ontologia e integrar informações mais detalhadas referentes à entidade. De acordo com os autores, a ontologia define o esquema de dados nos grafos de conhecimento; assim, os resultados da pesquisa de construção da ontologia podem ser extremamente benéficos na modelagem do grafo de conhecimento. Para os autores, a modelagem dos grafos de conhecimento utiliza a base lógica da OWL, que é uma linguagem de representação de ontologia baseada na teoria da lógica de descrição.

Considerando o que foi apresentado neste capítulo, conclui-se que, para assegurar que a informação seja representada de maneira formal e explícita, possibilitando inferências ao conteúdo representado, as ontologias são expressas em linguagens formais específicas no contexto da *web*. Os formatos mais utilizados para representação formal das ontologias são as linguagens Resource Description Framework (RDF) e Ontology Web Language (OWL), ambas criadas e padronizadas pelo W3C.

O modelo RDF oferece uma semântica simplificada, com boa representação para o tratamento de metadados, mas não fornece subsídios necessários para a expressividade exigida de uma ontologia (Mika, 2007). Já a linguagem OWL possibilita a descrição dos aspectos semânticos dos termos utilizados e seus respectivos relacionamentos, resultando em uma representação mais abrangente da RDF, o que contempla a interoperabilidade e proporciona a construção de ontologias. A OWL e o RDF incorporam as novas necessidades de representação de sistemas por ontologias (Ramalho; Vidotti; Fujita, 2007), portanto, apresentam semântica formal, assegurando que uma linguagem seja interpretada sem ambiguidades, o que é ponto crítico no compartilhamento de informações na *web*.

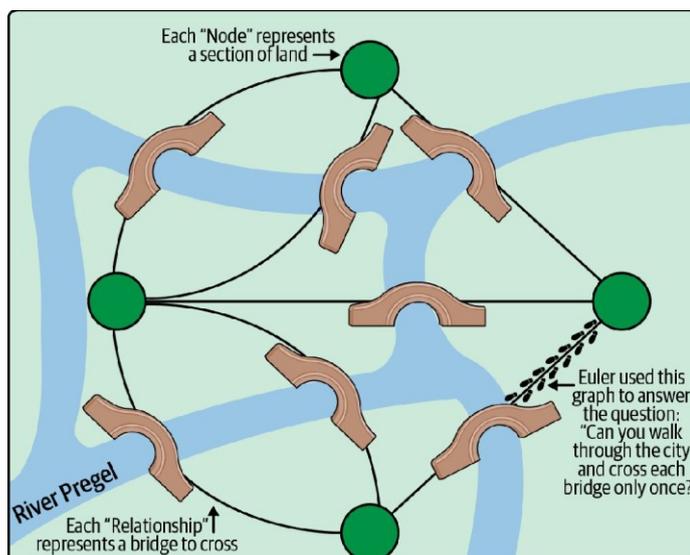
### 3 GRAFOS DE CONHECIMENTO

O presente capítulo tem por objetivo abordar, com base em uma revisão de literatura, o surgimento dos grafos, bem como suas definições. Para tanto, serão apresentados alguns modelos de grafos de conhecimento, o armazenamento de grafos de conhecimento, a construção do grafo de conhecimento, ferramentas para criar e gerenciar grafos de conhecimento, o modelo de grafo de propriedade, princípios de organização de grafo, exemplos de bases de conhecimento em grafos, e, por fim, conceitos e arquitetura da *web*.

#### 3.1 UM BREVE HISTÓRICO SOBRE O SURGIMENTO DOS GRAFOS

De acordo com Kejriwal (2022), a teoria dos grafos como é conhecida atualmente tem suas origens na proposta feita por Leonhard Euler para resolver o Problema da Ponte de Königsberg. Blumauer e Nagy (2020) sustentam que a teoria dos grafos se originou, em 1736, com base nas ideias de Euler. Conforme destacado por Barrasa, Hodler e Webber (2021) foi no século XVIII que o eminente matemático suíço concebeu essa teoria visando determinar a menor distância que o imperador da Prússia poderia percorrer para visitar a cidade de Königsberg (atual Kaliningrado), assegurando que atravessasse cada uma das sete pontes no trajeto apenas uma vez (Figura 4). Segundo os autores, Euler se utilizou desse grafo para solucionar o problema de percorrer a cidade e cruzar cada ponte apenas uma vez, sendo cada conexão correspondente a uma ponte a ser atravessada, e, cada ponto, uma parte da terra.

Figura 4 – Representação gráfica de Königsberg



Fonte: Barrasa, Hodler e Webber (2021).

Os autores mencionam que a interpretação de Euler sobre o problema delineado na Figura 4 poderia ser simplificada para uma forma lógica, eliminando todas as interferências do mundo real e concentrando-se apenas nas conexões entre as coisas. Para eles, Euler conseguiu demonstrar que o problema não necessariamente envolvia pontes, ilhas ou imperadores, pois comprovou-se que a geografia física de Königsberg era totalmente irrelevante e que o imperador não conseguiria percorrer toda a cidade cruzando cada ponte apenas uma vez. Era essencial haver no mínimo uma ilha (nó) com um número de pontes conectadas (relações) na qual o imperador poderia iniciar seu percurso. Para Barrasa, Hodler e Webber (2021), os matemáticos estudaram vários modelos de grafos e todas as variações sobre o tema dos nós conectados por relacionamentos com base no trabalho de Euler. Dessa forma, os grafos desempenharam um papel fundamental no desenvolvimento das ciências computacionais, da matemática e, sobretudo, da Inteligência Artificial.

Já em 1976, John Florian Sowa apresentou os inovadores Grafos Conceituais por intermédio de um artigo pioneiro, conforme relatado por Blumauer e Nagy (2020). Em 1982, a Holanda se tornou o berço dos grafos de conhecimento, graças aos esforços conjuntos do matemático Cornelis Hoede, da Universidade de Twente, e do sociólogo e matemático Frans N. Stokman, da Universidade de Groningen, que juntos lançaram as bases para a Teoria dos Grafos de Conhecimento. Cornelis Hoede ficou conhecido por suas contribuições para a teoria dos grafos e otimização combinatória. Suas pesquisas têm impacto significativo nesse campo e são amplamente reconhecidas na comunidade acadêmica internacional. O sociólogo holandês Frans N. Stokman é conhecido por suas contribuições no campo da análise das redes sociais para a compreensão das interações sociais e organizacionais.

A teoria dos grafos, conforme abordado por Nicoletti e Hruschka Júnior (2018), se concentra na investigação e no campo dos sistemas matemáticos. De maneira informal, um grafo é uma coleção de elementos conhecidos como vértices, interconectados por um conjunto de linhas denominadas “arestas”, que estabelecem ligações entre pares desses elementos.

### 3.2 GRAFOS DE CONHECIMENTO (GC): UMA BREVE DEFINIÇÃO

Nos últimos dez anos, os grafos de conhecimento (GC) tornaram-se significativamente relevantes no campo da inteligência artificial. Esses grafos, que têm raízes na história da IA, são essencialmente estruturas dirigidas e multi-relacionais, marcadas com tipos específicos de semântica (Kejriwal, 2022). Em parte, segundo o autor, isso foi alimentado pelo aumento da

publicação de conjuntos de dados estruturados na *web* e sucessos bem divulgados de projetos de grande escala, como o Google Knowledge Graph e o Amazon Product Graph.

Grafos de conhecimento, para Bellomarini, Sallinger e Vahdati (2020), são uma estrutura matemática composta por vértices que funcionam como unidades de conhecimento, interligados por arestas que simbolizam relações de pré-requisito. Embora essa definição tenha emergido do contexto da aprendizagem colaborativa entre estudantes e professores, os autores argumentam que ela se adapta adequadamente aos usos contemporâneos. No entanto, ao longo da seção, pode-se depreender que o conceito de grafos não tem um consenso, em vista de estar em disputa.

Para Paulheim, Ristoski e Portischet (2023), o conceito de grafos de conhecimento ganhou destaque em 2012, após a Google anunciar no *blog* da empresa a implementação dessa tecnologia com o objetivo de aprimorar a ferramenta de busca. A ideia era evoluir além da tradicional análise de palavras-chave e semelhança textual, adotando um sistema que compreende a informação de maneira estruturada. Com isso, os autores esclarecem que o objetivo do Google era melhorar a interpretação dos termos buscados pelos usuários para fornecer respostas mais exatas e adequadas às suas pesquisas.

Segundo Davis, Shrobe e Szolovits (1993), o conceito de grafos de conhecimento também é entendido como uma forma de representação do conhecimento baseada em relações, com princípios e regras importantes em lógica e IA. Essa abordagem tem sido destacada na comunidade da *web* semântica, especialmente por Berners-Lee, Hendler e Lassila (2001), com o intuito de desenvolver uma internet orientada por dados acessíveis e interpretáveis por sistemas computacionais.

Os grafos de conhecimento, portanto, abarcam uma estrutura em forma de rede com conexões direcionais, em que tanto os nós representados pelos pontos de junção quanto as arestas indicadas pelas linhas recebem rótulos identificadores (Kejriwal, 2022). Nesse sentido, as entidades representadas por nós em uma rede podem incluir uma gama de elementos como indivíduos, empresas e lugares, ou até serem de natureza mais específica. Essas conexões entre as entidades denominadas “arestas”, para Kejriwal (2022), podem igualmente ser referidas como propriedades ou predicados, pois elas explicam a natureza das relações que ligam as entidades.

Ehrlinger e Wöß (2016) apresentam uma visão abrangente das diversas definições de grafos de conhecimento. Segundo os autores, um GC adquire e integra informação a uma ontologia e aplica um raciocínio para derivar novo conhecimento. Os grafos de conhecimento, para Fensel *et al.* (2020), constituem-se em um conjunto coeso de dados que detalham os

componentes fundamentais de um grafo, os quais são representados por conceitos, objetos, entidades e indivíduos. Os autores descrevem “um grafo como uma estrutura que se correlaciona a um conjunto de objetos onde certos pares de objetos estão interconectados de alguma maneira específica” (Fensel *et al.* 2020, p. 2).

Para Silva, Ziviani e Porto (2019), a interpretação desses grafos é feita por meio da caracterização de entidades representadas pelos vértices, e as interações ou relações significativas denotadas pelas arestas que os unem, muitas vezes empregando uma abordagem ou descrição ontológica para esses componentes e suas interconexões.

Em sua pesquisa, Bellomarini, Sallinger e Vahdati (2020) ressaltam que a ideia mais evidenciada de grafos de conhecimento presente em várias definições é a de uma ferramenta para representar o conhecimento. Segundo os autores, muitas definições concordam em caracterizá-lo como um grafo estruturado por regras formais com conhecimento integrado.

Entretanto, não existe uma definição única normalmente aceita de grafos de conhecimento. Paulheim (2017) afirma que a maioria das definições entende o tema por intermédio de uma lente da *web* semântica e inclui esses recursos:

- Relações flexíveis entre o conhecimento em tópicos de um domínio: um grafo de conhecimento (i) define classes abstratas e relações de entidades em um esquema, (ii) descreve principalmente entidades do mundo real e suas inter-relações, organizadas em um grafo, (iii) permite potencialmente se inter-relacionar arbitrariamente entidades entre si, e (iv) cobre vários tópicos de domínios.
- Estrutura geral: uma rede de entidades, seus tipos semânticos, propriedades e relacionamentos.
- Suporte ao raciocínio sobre ontologias inferidas: um grafo de conhecimento adquire e integra informações em uma ontologia e aplica um raciocínio para derivar novos conhecimentos. (Paulheim, 2017, p. 489-508).

Conforme abordado por Fensel *et al.* (2020), nem todos os conceitos associados a grafos de conhecimento são relevantes, uma vez que a definição é abrangente, por se tratar de uma estrutura digital que representa o conhecimento como conceitos e as relações entre eles.

Os grafos de conhecimento, para Ehrlinger e Wöß (2016), podem incluir uma ontologia que possibilita a humanos e máquinas entender seu conteúdo e raciocinar sobre ele. Desse modo, os grafos de conhecimento representam formalmente a semântica, descrevendo entidades e seus relacionamentos. Ainda, segundo Yan *et al.* (2018), um grafo de conhecimento é uma estrutura semântica composta por vértices ou nós e arestas, de modo que vértices representam conceitos ou entidades, e arestas, relações semânticas entre conceitos.

Os grafos de conhecimento descritos por Barrasa, Hodler e Webber (2021) são um tipo específico de grafo no qual se destaca a compreensão contextual. Segundo eles, os GC

estruturam informações interligadas sobre entidades, eventos e objetos reais, facilitando a compreensão e o processamento por pessoas e máquinas.

### 3.3 MODELO DE GRAFOS DE CONHECIMENTO

Os grafos RDF e grafos de propriedades são modelos de dados que estão sendo usados para representar grafos de conhecimento (Tomaszuk; Angles; Thakkar, 2020). Apesar dos modelos de grafos de propriedades e grafos RDF compartilharem conceitos fundamentais como nós e arestas, eles se divergem quanto à metodologia de modelagem, à capacidade de expressão, ao formato de serialização e ao tipo de aplicações em que são utilizados (Tomaszuk; Angles; Thakkar, 2020). Conforme os autores, o grafo RDF é efetivamente o modelo padrão para grafos que interligam conhecimento na *web* semântica, apoiado por um ambiente abrangente facilitador de inferência e processamento. Já o modelo de grafo de propriedade, segundo Abuoda *et al.* (2022), oferece vantagens em tarefas analíticas de grafos escalonáveis, como combinação de grafos, análise de direções e entrelaçamento de grafos.

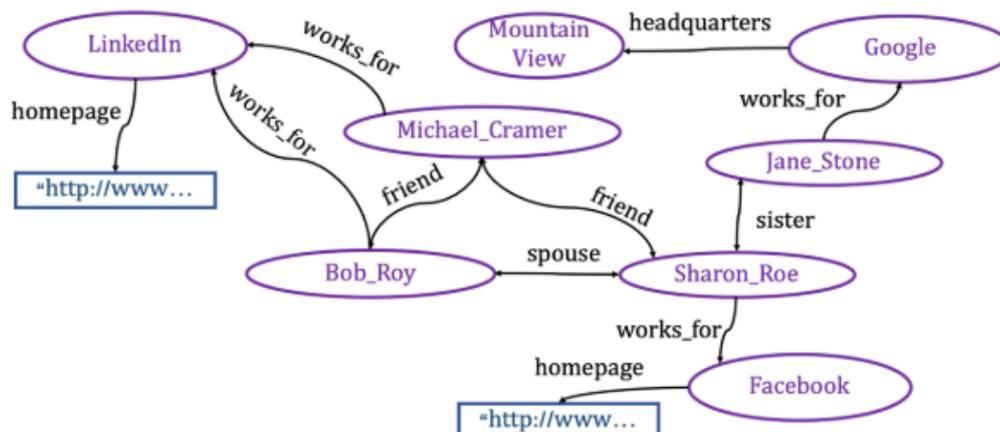
Gelling, Fletcher e Schmidt (2023) afirmam que o RDF, especificamente, é um padrão de modelo de dados proposto pelo W3C e se destina à troca de dados na *web* com uma semântica embutida; já a outra, conhecida como grafos de propriedades, é focada na gestão eficiente de grafos e tem bases sólidas nas comunidades de engenharia e desenvolvimento de *software*.

#### 3.3.1 Exemplo de grafo

Antes do advento dos grafos de conhecimento e algoritmos, grafos já eram empregados por físicos e cientistas para ilustrar redes sociais. O estudo das estruturas em sistemas complexos é o foco da ciência de redes, que tem se destacado na análise das interconexões entre proteínas, nas citações de estudos acadêmicos e nos modelos de redes sociais. A fácil obtenção de grandes volumes de dados tem resultado em progressos notáveis na compreensão dessas conexões. Na Figura 5, é ilustrado um exemplo de grafo de conhecimento de uma rede social (Kejriwal; Knoblock; Szekely, 2021).

Uma rede é um conjunto de itens nos quais certos pares estão interligados (Easley; Kleinberg, 2010). A natureza desses vínculos pode variar amplamente, o que a torna uma definição bastante abrangente. Essa adaptabilidade torna simples a presença de redes em vários campos, como nas redes sociais, por exemplo.

Figura 5 – Uma representação do grafo de conhecimento de uma rede social



Fonte: Kejriwal, Knoblock e Szekely (2021).

Para Kejriwal, Knoblock e Szekely (2021), a figura anterior exemplifica o domínio típico das redes sociais, e essa representação de uma rede social real demonstra como é possível retratar variadas relações complexas por meio de arestas direcionadas e rotuladas, além de relações mútuas entre várias entidades. Ainda, segundo os autores, os grafos de conhecimento de redes sociais podem ser incorporados pelo vocabulário *Friend Of A Friend* (FOAF) sem a necessidade de criar uma nova ontologia.

### 3.4 GRAFOS DE PROPRIEDADE

Gelling, Fletcher e Schmidt (2023) definem os grafos de propriedades como estruturas com base na teoria dos grafos, em que o modelo se baseia em conjuntos de nós interligados por arestas. Nesse modelo, nós e arestas são caracterizados por propriedades definidas em pares valores-chave.

O modelo de grafo de propriedades, também conhecido como “grafo de propriedades rotuladas”, facilita a criação de estruturas de grafos flexíveis com nós diversificados e arestas direcionadas, o que possibilita representar várias entidades e suas respectivas conexões (Angles, 2018).

De acordo com Barrasa, Hodler e Webber (2021), atualmente, no setor de TI, o modelo de grafos de propriedade é o predileto das empresas por se encaixar eficazmente em diversas questões de gerenciamento de dados, além de sua simplicidade de uso para especialistas em *software* e dados.

Os sistemas baseados em grafos de propriedades, conforme Tomaszuk, Angles e Thakkar (2020), foram desenvolvidos para o armazenamento e a consulta de grafos contendo

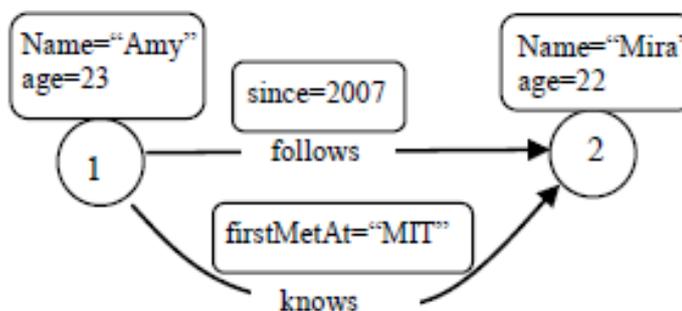
propriedades. Para os autores, um grafo de propriedade expande a concepção de um grafo direcionado e rotulado, possibilitando que nós e arestas contêm propriedades expressas como pares de valores-chave.

Barrasa, Hodler e Webber (2021) apontam que o modelo de grafo de propriedade é o modelo mais popular para bancos de dados de grafos mais modernos e, conseqüentemente, um método popular para a criação de grafos de conhecimento. Ele consiste em nós e relacionamentos:

- a) Nós que representam entidades no domínio:
  - Os nós podem conter zero ou mais propriedades, que são pares de valores-chave que representam dados da entidade, como preço ou data de nascimento;
  - Os nós podem ter zero ou mais rótulos, que declaram a finalidade do nó no grafo, como representar clientes ou produtos.
- b) Relacionamentos que representam como as entidades se inter-relacionam:
  - Os relacionamentos têm um tipo, como comprado ou curtido;
  - Os relacionamentos têm uma direção, indo de um nó para outro ou voltando para o mesmo nó;
  - Os relacionamentos podem conter zero ou mais propriedades, que são pares de valores-chave que representam alguma característica do link, como um registro de data e hora ou uma distância;
  - Os relacionamentos nunca ficam pendentes - sempre há um nó inicial e um final que pode ser o mesmo nó. (Barrasa; Hodler; Webber, 2021, p. 12).

Das *et al.* (2014) afirmam que, em um grafo de propriedades, cada vértice é definido com um identificador exclusivo. Cada aresta direcionada é identificada com um nome único e rotulada com uma *string*, conectando um vértice de origem a um vértice de destino. Um vértice ou uma aresta também pode ser associado a um conjunto de propriedades de chave/valor. Na Figura 6, é mostrado um exemplo de grafo de propriedade, que é baseado em chave/valor, direcionado e multi-relacional.

Figura 6 – Exemplo de grafo de propriedades



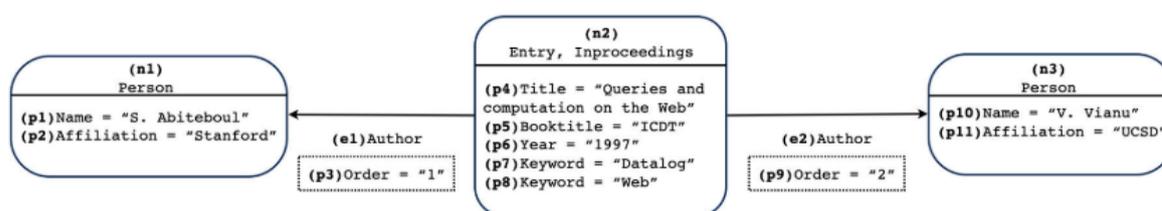
Fonte: Das *et al.* (2014).

De acordo com Donkers, Yang e Baken (2020), o modelo de grafos de propriedades é uma forma comum de representação de grafos, e se assemelha ao formato RDF, sendo empregado de forma nativa no sistema Neo4j. Ao contrário do RDF, neste modelo é possível

incluir propriedades de forma direta tanto nos vínculos quanto nos próprios elementos dos grafos. Para Donkers, Yang e Baken (2020), normalmente, os grafos de propriedade são centrados em nós, diferentemente das triplas RDF centradas em aresta.

Um grafo de propriedade, para Tomaszuk, Angles e Thakkar (2020), é essencialmente um multigrafo direcionado e com rótulos, o que se diferencia pelo fato de que tanto os nós quanto as arestas podem ter diversas propriedades. Uma propriedade tem um nome e um valor com um tipo de dado específico, enquanto rótulos podem ser atribuídos a nós, arestas e propriedades para detalhar suas classes, tipos ou funções dentro de um conjunto de dados. Na Figura 7, é apresentada uma representação do grafo de propriedade.

Figura 7 – Representação gráfica de um grafo de propriedade

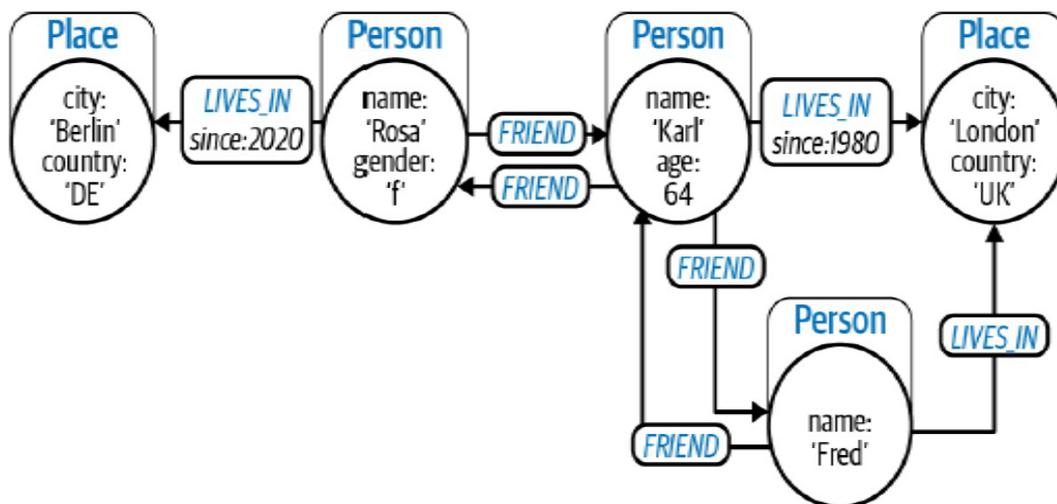


Fonte: Tomaszuk, Angles e Thakkar (2020).

No modelo de grafos de propriedade, segundo Anikin, Borisenko e Nedumov (2019), os vértices representam objetos distintos, e, as arestas, relacionamentos entre os vértices. Tanto os vértices quanto as arestas podem ter vários pares de chave/valor, chamados de “propriedades”. Conforme Das *et al.* (2014), o acesso aos dados de um grafo de propriedades geralmente é feito por intermédio da API padrão de Java, conhecida como “Blueprints”, ou mediante uma linguagem especializada de consulta própria.

Considerando as linguagens dedicadas à consulta de grafos, estas podem incluir propriedades, como o Cypher, e focam primordialmente na identificação de trajetos, uma vez que os nós de partida, ou conjuntos iniciais de nós, são escolhidos com base em chaves e valores específicos (Das *et al.*, 2014). Conforme Barrasa, Hodler e Webber (2021), as propriedades estão armazenadas dentro de nós. No exemplo da Figura 8 a seguir, cada nó tem um rótulo que representa sua função no grafo, sendo nomeados como Person e Place.

Figura 8 – Um grafo que representa pessoas, suas amigas e suas localizações



Fonte: Barrasa, Hodler e Webber (2021).

Ao se conectar os elementos “Person” e “Place”, tem-se respectivamente um relacionamento de saída tipo “Lives\_In”. Este segue do elemento “Person” – com as propriedades de nome “Rosa” e gênero “f” – em direção ao elemento “Place” – com a propriedade cidade de “Berlin”. Com isso, existe um atributo adicional, “since 2020”, declarando a data de início da residência; portanto, entende-se que Rosa reside em Berlim desde 2020 (Barrasa; Hodler; Webber, 2021).

No mesmo exemplo da Figura 8, é identificado também que Person Karl tem 64 anos, é amigo da Person Fred e ambos moram em London, sendo que Person Karl mora desde 1980. A Person Rosa é amiga de Person Karl, mas não é amiga da Person Fred. Para Barrasa, Hodler e Webber (2021), relacionamentos em grafos de propriedade não são simétricos e, na maioria dos domínios, os relacionamentos se aplicam em uma direção.

A falta de um modelo unificado para grafos de propriedade, afirmam Monteiro, Sá e Bernardino (2023), leva a uma grande variação de recursos entre diferentes implementações. A popularidade dos grafos de propriedade tem crescido tanto em estudos quanto em aplicações práticas, com seu suporte vindo de diversos sistemas de bancos de dados de grafos, como Neo4j, JanusGraph e TigerGraph.

### 3.5 GRAFOS RDF

O grafo RDF, de acordo com Gelling, Fletcher e Schmidt (2023), é utilizado como o padrão de troca de informações da *web* semântica e começou a ser usado primariamente para definir metadados de elementos na *web*. Segundo os autores, o W3C endossa uma variedade de

tecnologias relacionadas ao RDF que facilitam a formação e aplicação de grafos de conhecimento, seja no contexto da nuvem de dados interconectados, seja dentro de um sistema fechado.

Para Gelling, Fletcher e Schmidt (2023), mesmo sem uma definição que se utilize diretamente dos termos de grafos, tais como vértices e arestas, existe uma estreita relação entre o RDF e a estrutura de grafos. Para os autores, pode-se interpretar cada tripla do RDF como uma conexão entre um nó sujeito e um nó objeto, categorizada por um rótulo de predicado específico. Os autores acrescentam que o W3C classifica o RDF como um modelo baseado em grafos e utiliza essa abordagem visual para representar o RDF, embora não estabeleça um mapeamento formal entre o RDF e os elementos típicos de grafos, tais como vértices e arestas.

Os grafos de conhecimento baseados em RDF, de acordo com Le, Abel e Gouspillou (2023), têm a capacidade de representar entidades e seus relacionamentos usando ontologias. Ainda, segundo os autores, embora um grafo RDF possa ser considerado um tipo de ontologia, nem todos os grafos RDF são ontologias, bem como o foco principal de um grafo RDF é representar dados e relacionamentos entre entidades, em vez de definir um vocabulário formal para um domínio específico.

Dessa forma, um grafo RDF é formado por um conjunto de triplas (Le; Abel; Gouspillou, 2023; Pan *et al.*, 2017) e pode ser ilustrado por meio de um grafo direcional com nós e arestas em que cada tripla é caracterizada por uma conexão entre dois nós, unidos por uma aresta direcionada.

De acordo com Pan *et al.* (2017), um grafo RDF é composto por várias triplas e pode ser representado graficamente por meio de um diagrama orientado que exhibe cada tripla como uma conexão tipo nó-link-nó. Ainda, para os autores, em um grafo RDF identificamos três tipos de elementos característicos: IRIs, literais e nós em branco. Na constituição de uma tripla RDF, observa-se que o componente do sujeito é representável tanto por IRIs quanto por nós em branco; o componente do predicado é exclusivamente representado por IRIs; e o componente do objeto pode ser representado por IRIs, literais, ou ainda, por nós em branco.

Os sistemas de banco de dados RDF (RDB2RDF), segundo Tomaszuk, Angles e Thakkar (2020), também são conhecidos como “lojas de triplas”, e foram desenvolvidos especialmente para organizar e facilitar consultas em dados interconectados. Esses sistemas utilizam os padrões estabelecidos pela *web* semântica, com foco particular na estrutura destinada à descrição de recursos, enquanto que, para as consultas de linguagem, empregam SPARQL. Os autores ressaltam que o RDF é um modelo baseado em grafos utilizado para

representar informações sobre recursos na *web*, definindo relações e atributos por intermédio de multigrafos direcionados e rotulados.

Na perspectiva de Pan *et al.* (2017), os grafos RDF têm a capacidade de interligarem-se por meio de diversos tipos de mapeamentos, os quais abrangem tanto a conexão em nível de esquema, utilizando relações como `subClassOf`, quanto a conexão em nível de objeto, empregando termos como `sameAs`.

Na Figura 9, é apresentado um grafo RDF com uma tripla, em que o sujeito é `http://dbpedia.org/resource/Bolivia`, o predicado é `http://dbpedia.org/ontology/longName` e o objeto é o literal "Plurinational State of Bolivia". Essa tripla indica que a denominação oficial e completa da Bolívia é Estado Plurinacional da Bolívia.

Figura 9 – Exemplo de uma tripla



Fonte: Pan *et al.* (2017).

Karim, Vidal e Auer (2020) afirmam que, frequentemente nos grafos RDF, as entidades têm objetos comuns dentro de um grupo específico de propriedades, formando assim padrões que se assemelham a uma estrela composta por tais propriedades e objetos. Ainda, segundo os autores, caso haja muitos elementos ou características que sigam padrões estelares, tanto o tamanho do grafo RDF quanto a *performance* na execução de consultas pode ser prejudicada. Em um grafo RDF, combinar o mesmo padrão de estrela, apontam os autores, significa que as propriedades e os objetos são iguais, enquanto as entidades são diferentes.

A estrutura de triplas que caracteriza os grafos RDF é uma ferramenta poderosa para aprimorar a precisão de previsões em suas recuperações, defende Soliman (2020). Para o autor, é recomendada a incorporação de um classificador de *deep learning* para melhor antever resultados nas buscas desses grafos, visto que essa metodologia pode facilitar consultas de usuários na *web*.

Para Eddamiri, Benghabrit e Zemmouri (2020), conforme a complexidade dos dados RDF aumenta, torna-se cada vez mais essencial a realização de consultas eficientes para acessar e analisar essas grandes coleções de informações. De acordo com os autores, é possível extrair conhecimentos úteis de um grafo RDF ao utilizar o procedimento de descoberta de conhecimento em bancos de dados Knowledge Discovery in Databases (KDD). O KDD, para os autores, é um método iterativo e interativo composto por fases, como: 1) pré-processamento, que tem como objetivo eliminar dados com ruídos ou que estejam incompletos; 2) obtenção de

instâncias, que busca subgrafos relevantes; 3) transformação de grafos RDF em elementos utilizáveis na fase de extração de características; 4) fase de mineração de dados, responsável por identificar padrões de interesse nas representações escolhidas.

O desafio central reside no fato de que o aprendizado de máquina convencional exige dados organizados em tabelas para funcionar; em contrapartida, o formato RDF consiste em um vasto grafo interligado (Eddamiri; Benghabrit; Zemmouri, 2020).

Na seção seguinte, será apresentada uma breve revisão da literatura, que aborda o armazenamento e o gerenciamento de grafos de conhecimento.

### 3.6 ARMAZENAMENTO E GERENCIAMENTO DE GRAFOS DE CONHECIMENTO

Embora a pesquisa em bancos de dados focados em grafos tenha uma extensa história, que começou na década de 1980, de acordo com Angles e Gutiérrez (2018), foi somente em tempos mais recentes que desenvolvimentos tecnológicos significativos possibilitaram a criação de sistemas de bancos de dados de grafos verdadeiramente eficazes.

O avanço significativo na gestão de dados em grafos foi impulsionado por recursos de processamento e armazenamento de alta eficiência, sensores de última geração para captação de informações e dispositivos avançados para a análise e visualização de dados (Angles; Gutiérrez, 2018).

O campo de gerenciamento de dados de grafos, conforme destaca Sakr e Pardede (2012), é uma área de estudo e inovação tecnológica que se concentra na capacidade de lidar eficientemente com grandes conjuntos de dados estruturados como grafos. De acordo com os autores, esses dados são caracterizados por suas relações complexas e interconectadas, exigindo soluções especializadas para seu armazenamento, processamento e análise. Para os autores, a busca por métodos mais eficazes e escaláveis para gerir esses dados é um desafio constante para pesquisadores e profissionais da área de tecnologia da informação, sendo um campo em constante evolução e com impacto significativo no mundo digital contemporâneo.

Os sistemas de banco de dados de grafos, apontam Angles e Gutiérrez (2018), são definidos na literatura como sistemas projetados especificamente para gerenciar dados do tipo grafo seguindo os princípios básicos dos sistemas de banco de dados, ou seja, armazenamento contínuo de dados, independência física/lógica dos dados, integridade e consistência dos dados.

Angles e Gutiérrez (2018) esclarecem que existem duas principais categorias de sistemas destinadas ao gerenciamento de informações em grafo: bancos de dados de grafos e estruturas de processamento de grafos. Segundo os autores, a finalidade dos bancos de dados

de grafos é gerir dados em formato de grafo, mantendo ao mesmo tempo os princípios básicos de sistemas de gerenciamento de banco de dados; isso inclui o armazenamento persistente de dados, a distinção clara entre as disposições física e lógica dos dados, além da garantia de integridade e de constância das informações. Os autores relatam que esses últimos sistemas são projetados para gerenciar e examinar grafos de alta escala, utilizando várias máquinas para melhor desempenho. De acordo com os autores, esses sistemas têm duas estratégias diferenciadas para administrar e procurar dados em grafos com o objetivo de atender a propósitos diversos, respondendo prontamente às necessidades dos usuários, priorizando a eficiência, a facilidade de uso e a resolução efetiva de questões críticas associadas ao gerenciamento de dados.

Um modelo de banco de dados em grafo, para Angles e Gutiérrez (2018), é um modelo em que as estruturas de dados para o esquema e/ou instâncias são modeladas como um grafo direcionado, possivelmente rotulado, ou generalizações da estrutura de dados em grafo. A manipulação de dados é expressa por operações orientadas a grafo e construtores de tipos, e restrições de integridade apropriadas podem ser definidas sobre a estrutura do grafo.

Do *et al.* (2022) destacam que um banco de dados de grafos é um tipo de banco de dados *Not Only SQL* (NoSQL) que usa a estrutura de grafos para consultas semânticas por intermédio de nós, arestas e propriedades para representar e armazenar dados. Ainda, segundo os autores, a utilização de banco de dados de grafos expandiu-se por diversas áreas, incluindo educação, saúde, negócios e redes sociais, tendo como uma das principais vantagens o seu desempenho eficaz em consultas de dados.

Os bancos de dados de grafos, segundo Do *et al.* (2022), armazenam frequentemente dados de grafos de conhecimento e a lógica que descreve as interconexões e o contexto. Para os autores, sistemas de gestão de bancos de dados de grafos, entre os quais o Neo4j se destaca como o mais utilizado, proporcionam um armazenamento especializado e eficaz para estruturas de dados em grafos. Para Do *et al.* (2022), uma vantagem dos bancos de dados baseados em grafos é a eficiência em ações como efetuar consultas e análises que envolvem as conexões entre elementos, um contraste com o sistema *Structured Query Language* (SQL) que opera por meio de relações entre tabelas.

A escolha do tipo certo de banco de dados para armazenar um grafo de conhecimento, de acordo com Fensel *et al.* (2020), depende de vários fatores relacionados aos requisitos específicos do projeto. No Quadro 6, são apresentados os fatores imprescindíveis para armazenamento de grafos de conhecimento.

Quadro 6 – Requisitos fundamentais do projeto para armazenar grafos de conhecimento

Objetivos	Requisitos específicos do projeto
Entender os dados e consultas	<p><b>Complexidade dos relacionamentos:</b> os bancos de dados de grafos ou os bancos de dados de grafos de propriedades geralmente são os mais adequados, se os dados envolverem relacionamentos complexos e a realização de análises profundas.</p> <p><b>Dados semânticos e interoperabilidade:</b> os armazenamentos RDF (<i>triplestores</i>) podem ser os mais apropriados caso se trate de dados semânticos ou se a interoperabilidade e a adesão a padrões como RDF forem essenciais.</p> <p><b>Variedade de modelos de dados:</b> se o projeto exigir o armazenamento de dados em vários formatos (por exemplo, grafos, documentos, pares de valores-chave), pode-se considerar bancos de dados de vários modelos.</p>
Avaliar as necessidades de desempenho e escalabilidade	<p><b>Desempenho da consulta:</b> os bancos de dados de grafos são otimizados para a rápida passagem de relacionamentos complexos. Para projetos em que o desempenho da consulta em dados interconectados é fundamental, eles geralmente são a melhor opção.</p> <p><b>Escalabilidade:</b> considerar a necessidade de escalabilidade horizontal (a capacidade de distribuir dados em várias máquinas) ou escalabilidade vertical (atualização dos recursos de uma única máquina) é suficiente. Os bancos de dados NoSQL e de vários modelos geralmente oferecem melhor escalabilidade horizontal.</p>
Considerar a integridade dos dados e as transações	<p><b>Transações ACID:</b> se o aplicativo exigir integridade transacional forte (atomicidade, consistência, isolamento, durabilidade – ACID), os bancos de dados relacionais são tradicionalmente fortes nessa área. Entretanto, alguns bancos de dados grafos modernos também oferecem suporte robusto a transações ACID.</p> <p><b>Consistência de dados:</b> avaliar a necessidade de consistência imediata <i>versus</i> consistência eventual, o que pode influenciar a escolha entre bancos de dados relacionais mais tradicionais e determinadas opções de NoSQL.</p>
Avaliar a complexidade operacional e de desenvolvimento	<p><b>Familiaridade do desenvolvedor:</b> a curva de aprendizado associada à tecnologia de banco de dados deve ser considerada. Se a equipe já estiver familiarizada com SQL, por exemplo, utilizar um banco de dados relacional ou um banco de dados multimodelo compatível com SQL pode acelerar o desenvolvimento.</p> <p><b>Ferramentas operacionais e suporte:</b> analisar o ecossistema em torno do banco de dados, inclusive as ferramentas de gerenciamento, o suporte da comunidade e a disponibilidade de desenvolvedores ou administradores qualificados.</p>
Verificar as integrações e o ecossistema	<p><b>Integração com sistemas existentes:</b> considerar a facilidade com que o banco de dados pode se integrar à infraestrutura existente e a outras ferramentas que se planeja utilizar.</p> <p><b>Ecossistema e comunidade:</b> um ecossistema e uma comunidade forte podem fornecer recursos valiosos para a solução de problemas, otimização e desenvolvimento de práticas recomendadas.</p>
Analisar o licenciamento e os custos	<p><b>Código aberto <i>versus</i> comercial:</b> alguns bancos de dados são de código aberto e de uso gratuito, enquanto outros exigem licenças comerciais. Considerar o custo total de propriedade, incluindo custos de hospedagem, manutenção e desenvolvimento.</p>

Fonte: Adaptado de Fensel *et al.* (2020).

Como se pode perceber, não existe uma resposta única para a escolha de um banco de dados para um grafo de conhecimento. Do *et al.* (2022) concluem que é fundamental avaliar as necessidades específicas do projeto em relação aos pontos fortes e às limitações de cada tipo de banco de dados. Os autores sugerem que o teste piloto com um pequeno conjunto de dados também pode ser uma abordagem prática para avaliar se um banco de dados atende às suas necessidades antes de se comprometer totalmente com ele.

Para Barrasa e Hodler e Webber (2021), um bom grafo de conhecimento armazena e unifica os dados implícitos para que possam ser fundamentados. Segundo os autores, ele não procura necessariamente alterar os dados subjacentes, mas fornece orientação sobre como esses dados podem ser entendidos, independentemente de sua tecnologia de origem. Para os autores, um grafo de conhecimento rápido e atualizado capacita os usuários. Uma breve explicação a respeito de ferramentas para criar grafos é apresentada na seção seguinte.

### 3.7 FERRAMENTAS PARA CRIAR GRAFOS

Ao tratar das ferramentas para criação de grafos, Fensel *et al.* (2020) argumentam que é essencial levar em conta uma variedade de ferramentas quando se trata do desenvolvimento e da administração de grafos de conhecimento. Eles salientam que a gama de *softwares* inclui desde sistemas de banco de dados que são especificamente desenhados para conter grafos de dados até conjuntos de ferramentas e bibliotecas projetadas para facilitar a manipulação, a busca e a avaliação desses grafos. Para os autores, geralmente a escolha da ferramenta é ditada pelas necessidades do projeto, como a dimensão do grafo, a complexidade das consultas e a integração com sistemas ou bancos de dados adicionais.

Com o rápido crescimento das entidades acadêmicas, segundo Liu *et al.* (2020), torna-se cada vez mais difícil obter informações com rapidez e precisão em redes de grande escala. Para os autores, alguns sistemas já foram desenvolvidos para fornecer serviços de pesquisa e mineração, como *sites* chineses que fornecem serviços de gerenciamento de informações acadêmicas e recuperação de literatura.

Liu *et al.* (2020) apresentaram o Web of Scholars como um novo sistema que adapta os métodos genéricos para pesquisar, classificar e explorar com eficiência estudantes e seus relacionamentos em grandes redes acadêmicas heterogêneas, como também exibe ferramentas de visualização para apresentá-los. Os autores, em sua pesquisa, também fornecem vários aplicativos acadêmicos em um grande grafo de conhecimento de relacionamento acadêmico.

A classificação DB-Engines<sup>1</sup>, segundo Monteiro, Sá e Bernardino (2023) é uma iniciativa independente de análise de dados que fornece informações sobre sistemas de gerenciamento de bancos de dados. Seu principal produto é uma classificação mensal de popularidade de bancos de dados com base em vários fatores, incluindo adoção por empresas e desenvolvedores, popularidade *on-line*, recursos oferecidos, desempenho, escalabilidade,

---

<sup>1</sup> Disponível em: <https://db-engines.com/en/ranking/graph+dbms>. Acesso em: 18 set. 2024.

suporte da comunidade e *feedback* de usuários especializados em bancos de dados. Os autores apontam que uma combinação ponderada desses critérios é usada para determinar a classificação geral da DB-Engines. O Neo4j vem se destacando nos últimos anos como um dos melhores na classificação.

No desenvolvimento da tecnologia de banco de dados, afirmam Monteiro, Sá e Bernardino (2023), dos muitos modelos de dados já existentes, há três tipos normalmente usados, que são o modelo hierárquico, o modelo de grafo e o modelo relacional. Segundo os autores, o modelo relacional é apoiado por uma base matemática rigorosa, com alta independência e segurança de dados e operação simples. No entanto, os autores pontuam que, com a inovação contínua da tecnologia da internet, a escala de dados na rede continua a aumentar, logo, a complexidade dos dados também cresce.

Monteiro, Sá e Bernardino (2023) concluem que o modelo de relacionamento tem sido lento, portanto, incapaz de atender às necessidades dos campos relacionados. Além disso, cada vez mais problemas aparecem nos bancos de dados relacionais. Sendo assim, os autores afirmam que o surgimento do banco de dados de grafos atenuou efetivamente esse fenômeno.

O banco de dados de grafos, bem como sua estrutura de armazenamento de dados e seus métodos de consulta, de acordo com Monteiro, Sá e Bernardino (2023) é um novo banco de dados NoSQL baseado na teoria dos grafos. Para os autores, no cálculo de grafos, a expressão básica da estrutura de dados é:  $G = (V, E)$ ,  $V =$  vértice,  $E =$  aresta. No banco de dados de grafos, o modelo de dados é representado principalmente por nós e arestas e os pares de valores-chave também podem ser processados.

Para Monteiro, Sá e Bernardino (2023), o banco de dados Neo4j é um banco de dados de grafos NoSQL de alto desempenho escrito em Java e Scala, dedicado ao armazenamento de grafos de rede. Como um banco de dados de grafos, o Neo4j tem as seguintes vantagens:

- Operações de banco de dados mais rápidas;
- Dados mais intuitivos;
- Mais flexível;
- A velocidade das operações do banco de dados aumenta significativamente à medida que o banco de dados cresce;
- Linguagem de consulta autônoma (chamada Cypher);
- A estrutura do relacionamento entre entidades é muito natural e se ajusta ao senso de intuição dos indivíduos. (Monteiro; Sá; Bernardino, 2023).

O Neo4j é um sistema de gerenciamento de banco de dados de grafos de código aberto implementado em Java, considerado o sistema de gerenciamento de banco de dados de grafos mais abrangente e promissor (Monteiro; Sá; Bernardino, 2023). Em comparação com a modelagem de um banco de dados de grafos usando um sistema de gerenciamento de banco de

dados relacional, essa abordagem possibilita aplicar uma otimização no caso de dados com uma estrutura mais complexa e carregada. Na seção seguinte, será apresentado o Neo4j, que é particularmente adequado para lidar com dados complexos e interconectados como redes sociais, mapas e sistemas de recomendação.

### 3.7.1 Neo4j

Conforme Monteiro, Sá e Bernardino (2023), o Neo4j destaca-se no universo dos bancos de dados por ser uma solução inovadora baseada em grafos. Desenvolvido e disponibilizado inicialmente em 2007, esse banco de dados de código aberto é construído sobre a linguagem Java, o que lhe confere robustez e confiabilidade. Para os autores, a principal característica do Neo4j é sua capacidade de armazenar dados não como tabelas, que são típicas em bancos de dados relacionais, mas sim como estruturas de grafos, o que resulta em representações mais naturais e eficientes das relações entre os dados. Essa particularidade, segundo os autores, torna o Neo4j extremamente versátil e aplicável a uma vasta gama de setores, incluindo saúde, educação, setor governamental, indústria de manufatura automatizada e até mesmo aplicações militares. Os autores afirmam que sua popularidade enquanto banco de dados de grafo é notável, sendo frequentemente escolhido por desenvolvedores e empresas que buscam uma solução moderna e eficaz para o gerenciamento de dados complexos. O Neo4j utiliza a linguagem de consulta Cypher, que é inspirada em SQL.

Para Vukotic e Watt (2015), percorrer um grafo significa visitar diversos nós e deslocar-se entre nós conectados com relacionamentos; essa atividade é essencial na obtenção de informações armazenadas em um grafo e é uma característica distintiva desse modelo específico. O Neo4j é notabilizado por esses autores por apresentar uma avançada *Application Programming Interface* (API) Traversal, que tanto facilita a navegação pelo grafo como emprega a API REST ou as linguagens de consulta do próprio Neo4j para efetuar a exploração desses dados. Segundo os autores, a capacidade de atravessar grafos, uma das principais ferramentas do Neo4j, evidencia sua grande eficácia no gerenciamento de dados em grafos de ampla dimensão.

Monteiro, Sá e Bernardino (2023) apontam que, no *ranking* da DB-Engines de 2022, o Neo4j é o banco de dados que se destaca em primeiro lugar por causa de suas características notáveis. Os autores destacam que o Neo4j é um banco de dados escalável, otimizado para armazenar e consultar grandes grafos distribuídos em um *cluster* de múltiplas máquinas. Além disso, para os autores, é flexível e adequado para o tratamento de dados com formatos não

estruturados, o que o torna uma excelente escolha para diversas aplicações. Outro ponto forte é sua linguagem de consulta, que é fácil de entender e usar, o que facilita o trabalho dos desenvolvedores e analistas de dados.

As principais vantagens do Neo4j, destacam Monteiro, Sá e Bernardino (2023), são a arquitetura de *cluster* distribuída de alto desempenho, a minimização da latência da consulta em decorrência da fragmentação do grafo e o serviço de nuvem – AuraDB – totalmente gerenciado, com atualizações e *backups* automáticos. Contudo, segundo os autores, o Neo4j também apresenta algumas limitações: não aceita diretamente dados formatados em RDF e consome uma grande quantidade de memória.

De acordo com Vukotic e Watt (2015), a modelagem de dados para Neo4j é flexível e pode haver várias representações possíveis do mesmo domínio, as quais mostram a diversidade de aplicações para as quais a modelagem de dados grafos é adequada.

### 3.8 CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO

A construção de um grafo de conhecimento, conforme Madnick *et al.* (2009) e Zaveri *et al.* (2012), deve produzir uma estrutura de alta qualidade, com base em um modelo de dados robusto e expressivo para o grafo. Ainda, segundo os autores, a qualidade de um grafo de conhecimento e das fontes de dados pode ser medida ao longo de várias dimensões, como correção, atualização, abrangência e concisão.

Para Melnyk, Dognin e Das (2022), a construção de grafos de conhecimento é um campo de estudo em destaque que busca representar de forma mais organizada as informações presentes em grandes conjuntos de dados textuais, em uma configuração mais estruturada e compacta.

O processo de criar e manter grafos de conhecimento envolve múltiplas fases, com o objetivo de compilar dados relevantes oriundos de diversas fontes. As etapas exatas necessárias podem variar conforme os dados que precisam ser incorporados e os objetivos específicos do uso do grafo. No entanto, é geralmente preferível que essas fases sejam realizadas por intermédio de fluxos automatizados, minimizando a necessidade de intervenção e supervisão humana (Cudré-Mauroux, 2020).

Construídos com base em conteúdo, dados e anotações de serviço, os grafos de conhecimento, de acordo com Fensel *et al.* (2020, p. 69), podem melhorar os resultados do mecanismo de pesquisa e os sistemas de conversação, além de desenvolver os agentes conversacionais em diferentes pontos no *pipeline* do sistema de diálogo e as anotações de

serviço, cujo futuro é promissor, ajudando a construir sistemas flexíveis desacoplados dos serviços da *web* com os quais eles se comunicam.

Segundo Pan *et al.* (2017), as tecnologias relacionadas à construção dos grafos de conhecimento podem ser compreendidas por meio de diferentes componentes, como:

- Representação e raciocínio do conhecimento (linguagens, esquemas e vocabulários padrão).
- Armazenamento de conhecimento (bancos de dados e repositórios de grafos).
- Engenharia do Conhecimento (metodologias, editores e padrões de *design*).
- Aprendizado de conhecimento, incluindo aprendizado de esquema e população (Pan *et al.*, 2017).

De acordo com Fensel *et al.* (2020), essa abordagem é relevante em sistemas baseados em conhecimento e aponta que as estratégias e técnicas relacionadas com grafos de conhecimento consideram cuidadosamente a extensa quantidade de dados, além das bases convencionais de conhecimento.

Barrasa, Hodler e Webber (2021) também pontuam que, iniciando a construção com um grafo básico, porém útil, podem ser adicionadas sucessivas camadas de organização, o que demonstra como os grafos de conhecimento podem ser usados para resolver problemas cada vez mais sofisticados. Nesse sentido, a metodologia para a construção de grafos de conhecimento, segundo Simsek *et al.* (2022), consiste nas seguintes etapas.

- a) Criação de conhecimento, que se concentra na criação eficiente de grandes ABoxes e na engenharia de ontologias leves baseada em Schema.org para TBoxes.
- b) Hospedagem de conhecimento, que considera a proveniência do conhecimento criado com base em fontes heterogêneas e a contextualização do conhecimento.
- c) Curadoria de conhecimento, que visa avaliar, corrigir e enriquecer um grafo de conhecimento. Problemas de qualidade em grafos de conhecimento são inevitáveis em razão do grande tamanho e da heterogeneidade. A tarefa de avaliação afere a qualidade em diversas dimensões. Em seguida, a qualidade é melhorada principalmente nas dimensões de correção e completude, limpando e enriquecendo os grafos de conhecimento, respectivamente.
- d) Implantação de conhecimento, que coloca os grafos de conhecimento em uso (Simsek *et al.*, 2022).

A construção de grafos de conhecimento, conforme Hofer *et al.* (2023), requer uma abordagem multidisciplinar, que necessita de *expertise* em campos da pesquisa, tais como processamento de linguagem natural, integração de dados, representação do conhecimento e gestão do conhecimento.

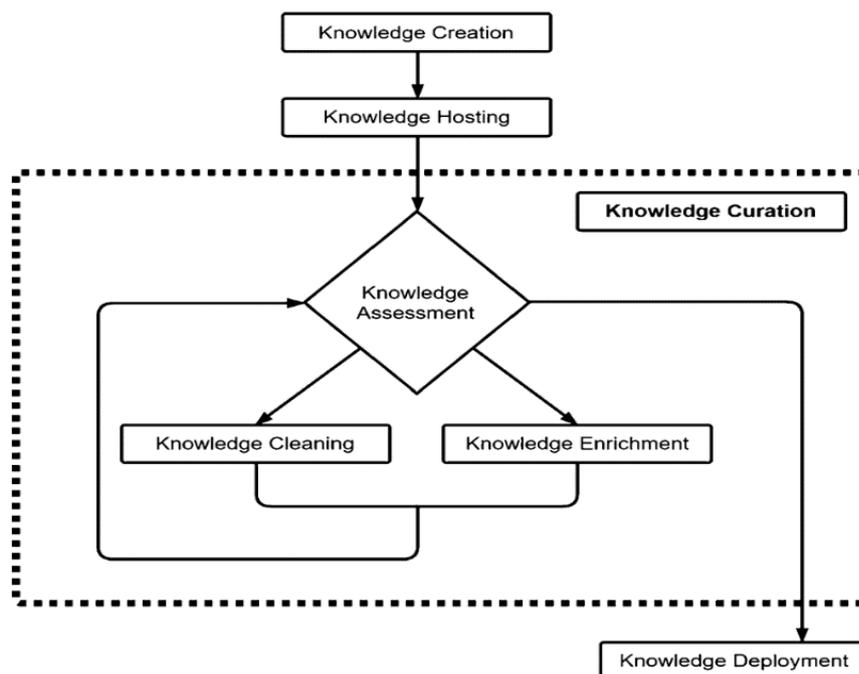
Na perspectiva de Abu-Salih (2021), a construção dos grafos de conhecimento é baseada nas soluções técnicas e em métodos empregados na navegação da *web* e em outros repositórios de informações de dados para inferir entidades e relações. O autor explica que a construção dos grafos de conhecimento pode ser feita por meio de um modelo ontológico existente com base em esquema, de um modelo livre de esquema ou por meio de uma estratégia combinada, que é independente de modelos ou esquemas.

Os grafos de conhecimento combinam diversas formas de informações oriundas de múltiplas fontes, incluindo tanto dados estruturados quanto semiestruturados, e abrangem tipos variados, tais como texto, imagens e áudio, assim como informações provenientes de bases de dados estruturadas e outros conjuntos de grafos de conhecimento (Hofer *et al.*, 2023).

Liu *et al.* (2020) ressaltam um ponto importante no processo de construção de grafos de conhecimento ao reconhecer que diferentes fontes de dados requerem abordagens específicas, sendo essenciais ao se criar um sistema eficiente e acurado. Para os autores, na construção de um grafo de conhecimento, é imprescindível não só classificar as fontes, mas também personalizar o processo de construção conforme as suas particularidades.

De modo geral, para Fensel *et al.* (2020), o processo de construção de grafos de conhecimento envolve as seguintes etapas: criação do conhecimento, hospedagem do conhecimento, curadoria do conhecimento e implantação do conhecimento. Na Figura 10, são identificadas as principais etapas em um padrão de processo global.

Figura 10 – Principais etapas do processo de construção de grafos de conhecimento



Fonte: Fensel *et al.* (2020).

O processo de criação do conhecimento envolve a coleta dos dados de diversas fontes, organizando-os e transformando-os em informações pertinentes. A forma como o conhecimento é representado muda conforme o sistema, podendo ser encontrado em documentos textuais, páginas da *web*, bancos de dados relacionais, entre outros repositórios de dados. Uma maneira de resolver essa discrepância na apresentação dos dados, segundo os autores, é por meio do uso de dados estruturados semanticamente (Fensel *et al.*, 2020).

Na fase de hospedagem de conhecimento, Fensel *et al.* (2020) explicam que o conhecimento é codificado como um conjunto de dados enriquecidos semanticamente, o que resulta em metadados incorporados aos dados. Esse processo é realizado por meio do emprego de um vocabulário padrão que é amplamente aceito seguindo as diretrizes do RDF.

Para Fensel *et al.* (2020), a finalidade principal de curadoria de conhecimento é apresentar métodos sensíveis ao custo para melhorar o padrão dos extensos grafos de conhecimento, assegurando que sejam eficazes para as utilizações pretendidas. O objetivo da curadoria de conhecimento é fornecer métodos sensíveis e eficientes do ponto de vista econômico para melhorar a qualidade e o padrão dos extensos grafos de conhecimento, de modo a assegurar sua funcionalidade para suas aplicações.

O processo de curadoria de conhecimento defendido por Paulheim (2017) estabelece grandes grafos de conhecimento de cobertura e qualidade significativas. Fensel *et al.* (2020) afirmam que, nas subetapas desse processo, são identificadas as seguintes atividades: avaliação do conhecimento, limpeza e enriquecimento. Ainda, de acordo com Paulheim (2017), existem dois focos principais no aperfeiçoamento de grafos de conhecimento: a inclusão de dados faltantes, chamada de “conclusão”, e a correção de dados incorretos no grafo, conhecida como “detecção de erros”.

O processo de avaliar dados e conhecimento, de acordo com Färber *et al.* (2018), consiste em examinar a qualidade de um grafo de conhecimento para, assim, qualificar a sua eficiência. O objetivo é medir a utilidade acentuando-a em duas dimensões cruciais, que são a correção e completude desse grafo.

No Quadro 7, são apresentadas as dimensões de correção e completude no processo de avaliação dos dados e conhecimento do grafo.

Quadro 7 – Processo de avaliação de dados e conhecimento do grafo de conhecimento

(continua)

<b>Qualidade</b>	<b>Correção e completude</b>	<b>Autor</b>
Acessibilidade	Requer que as informações sejam acessíveis e possam ser facilmente consultadas; incluam o devido licenciamento e garantam conectividade, segurança e eficiência; permaneçam atualizadas para atender às necessidades específicas de utilização ou aos seus usuários.	Färber <i>et al.</i> (2018); Zaveri <i>et al.</i> (2012); Wang e Strong (1996).
Precisão	Reflete a exatidão e fidedignidade dos dados quanto à sua conformidade sintática e semântica e à sua verdade atribuída a referências externas.	Wang e Strong (1996); Färber <i>et al.</i> (2018); Fürber e Hepp (2010).
Quantidade	Define a quantidade de dados que são apropriados para uma determinada tarefa.	Zaveri <i>et al.</i> (2012); Flemming (2011).
Credibilidade	Também parte da confiabilidade, define o grau em que os dados são considerados verdadeiros, credíveis, corretos, reais, confiáveis, imparciais e verificáveis.	Färber <i>et al.</i> (2018); Zaveri <i>et al.</i> (2012); Wang e Strong (1996); Pipino <i>et al.</i> (2002).
Completude dos dados	Refere-se ao fato de os dados de um grafo de conhecimento estarem completos no nível de esquema e de instância para um caso de utilização específica.	Fürber e Hepp (2010).
Representação concisa	Refere-se à maneira eficiente com que os dados estão estruturados, tanto no nível de esquema quanto de instância.	Mendes, Mühleisen e Bizer (2012); Zaveri <i>et al.</i> (2012); Wang e Strong (1996).
Representação consistente	Refere-se à dimensão da coerência em termos de formato, descrições formais e outros dados armazenados no grafo de conhecimento.	Wang e Strong (1996); Zaveri <i>et al.</i> (2012); Mendes, Mühleisen e Bizer (2012).
Relação custo-eficácia	Mede o custo total do recolhimento de uma quantidade adequada de dados precisos para apoiar a utilidade desejada.	Wang e Strong (1996).
Facilidade de manipulação	Refere-se a uma subdimensão de facilidade de operação, a facilidade de manipulação e aplicação dos dados em diferentes tarefas.	Wang e Strong (1996); Pipino <i>et al.</i> (2002).
Facilidade de operação	Trata da simplicidade com que os dados podem ser combinados, modificados, renovados, descarregados, enviados, replicados, implementados e adaptados para um objetivo ou necessidade particular. Essa característica também pode ser contemplada pelo conceito de facilidade de manuseio dos dados.	Wang e Strong (1996).
Facilidade de compreensão	Refere-se à facilidade com que os usuários compreendem os dados claros e sem ambiguidade.	Färber <i>et al.</i> (2018); Wang e Strong (1996); Zaveri <i>et al.</i> (2012).
Flexibilidade	Diz respeito à capacidade de adaptação, extensão e aplicabilidade a várias tarefas, estando intimamente ligada à facilidade de uso.	Wang e Strong (1996).

(conclusão)

<b>Qualidade</b>	<b>Correção e completude</b>	<b>Autor</b>
Livre de erros	Trata também da exatidão dos dados, avaliando a acurácia ao identificar o número de erros – que incluem afirmações de ocorrência inadequadas ou não realizadas, assim como as afirmações imprecisas de valor de propriedade em relação ao conjunto total de declarações numa base de conhecimento.	Wang e Strong (1996); Pipino <i>et al.</i> (2002).
Interoperabilidade	Refere-se ao nível de reutilização de padrões, ontologias ou vocabulários amplamente reconhecidos na configuração de um grafo de conhecimento, o que facilita o entendimento por sistemas computacionais.	Zaveri <i>et al.</i> (2012); Färber <i>et al.</i> (2018).
Objetividade	É um componente crucial da confiabilidade, atuando como um critério para avaliar a independência, a neutralidade e a imparcialidade dos dados. O grau de objetividade de um grafo de conhecimento, por exemplo, é determinado por especialistas e varia conforme o tipo de conteúdo e a possibilidade de verificação por meio de fontes externas e confiáveis.	Zaveri <i>et al.</i> (2012); Färber <i>et al.</i> (2018); Wang e Strong (1996); Pipino <i>et al.</i> (2002).
Relevância	Diz respeito a quão aplicáveis e pertinentes são os dados em uma tarefa específica.	Pipino <i>et al.</i> (2002); Zaveri <i>et al.</i> (2012).
Reputação	Está associada à percepção de confiança dos dados, sendo influenciada pela credibilidade da sua origem ou pela qualidade do seu conteúdo.	Färber <i>et al.</i> (2018); Wang e Strong (1996).
Segurança	Um aspecto-chave da acessibilidade diz respeito à limitação de acesso a dados para preservar a sua integridade e prevenir o uso inadequado. Isso pode ser avaliado pelo uso de procedimentos convencionais de segurança e proteção de dados pessoais.	Zaveri <i>et al.</i> (2012); Wang e Strong (1996); Färber <i>et al.</i> (2018).
Oportunidade	A velocidade na atualização de informações é essencial em atividades nas quais os dados necessitam ser frequentemente renovados. Isso envolve considerar tanto a frequência da renovação dos dados quanto o intervalo até que eles percam sua validade. O princípio da oportunidade se aplica especialmente em segmentos como o de hospedagem e viagens aéreas, nos quais os preços flutuam continuamente.	Pipino <i>et al.</i> (2002); Wang e Strong (1996); Zaveri <i>et al.</i> (2012).
Rastreabilidade	Envolve a documentação detalhada, a verificabilidade e a fácil localização dos dados de origem. É reconhecida não somente como uma verificação, mas também como um componente crucial da qualidade e confiabilidade.	Wang e Strong (1996); Färber <i>et al.</i> (2018).
Valor acrescentado	Descreve o nível de benefício e vantagens que os dados trazem quando usados, aumentando assim a eficácia das operações. Esse conceito também é visto como um aspecto da qualidade no que se refere à exaustividade dos dados.	Wang e Strong (1996); Färber <i>et al.</i> (2018).
Compreensibilidade	Indica a clareza com que uma pessoa consegue entender informações de forma inequívoca.	Färber <i>et al.</i> (2018); Zaveri <i>et al.</i> (2012).
Variedade	Refere-se a um número de fontes diferentes por meio das quais os dados são obtidos.	Fensel <i>et al.</i> (2020).

Fonte: Adaptado de Färber *et al.* (2018).

No enriquecimento de um grafo de conhecimento, segundo Fensel *et al.* (2020), o objetivo é aprimorá-lo com a adição de novas informações, o processo se inicia com a identificação de fontes de conhecimento pertinentes, reconhecendo a complexidade e abrangência da área de qualidade de dados, evidenciada pela diversidade de categorias e aspectos utilizados para avaliar e aprimorar a confiabilidade dessas informações.

Para Bizer *et al.* (2008), um meio de publicar dados abertamente de acordo com alguns princípios é por intermédio do Linked Open Data (LOD), que tem como base as tecnologias semânticas, as quais possibilitam que os dados sejam facilmente reutilizados em razão da legibilidade e da interpretabilidade implícitas da máquina. Dietrich, Jones e Wright (2008) explicam que dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos apenas, no máximo, à exigência de atribuição e compartilhamento iguais.

A implementação do LOD é apoiada por uma série de tecnologias e padrões, como o SPARQL, uma linguagem de consulta para bancos de dados RDF, o que possibilita que usuários realizem consultas complexas sobre os dados interligados (Bizer *et al.*, 2008). Além disso, ontologias como OWL são usadas para definir vocabulários comuns e estruturas de dados, o que ajuda na interpretação e no processamento automático das informações (Bizer *et al.*, 2008).

Os dados vinculados são apresentados de maneira estruturada para simplificar interconexões e consultas semânticas, podendo ser disponibilizados ou usados como dados abertos. A qualidade dos dados abertos vinculados é avaliada por meio de cinco critérios: se os dados forem fornecidos sob uma licença aberta; se os dados estiverem disponíveis como dados estruturados; se os dados também estiverem disponíveis em um formato não proprietário; se os Uniform Resource Identifier (URIs) forem usados para que os dados possam ser referenciados e se o conjunto de dados estiver vinculado a outros conjuntos de dados para fornecer contexto (Bizer *et al.*, 2008).

### 3.9 ALGUNS GRAFOS DE CONHECIMENTO

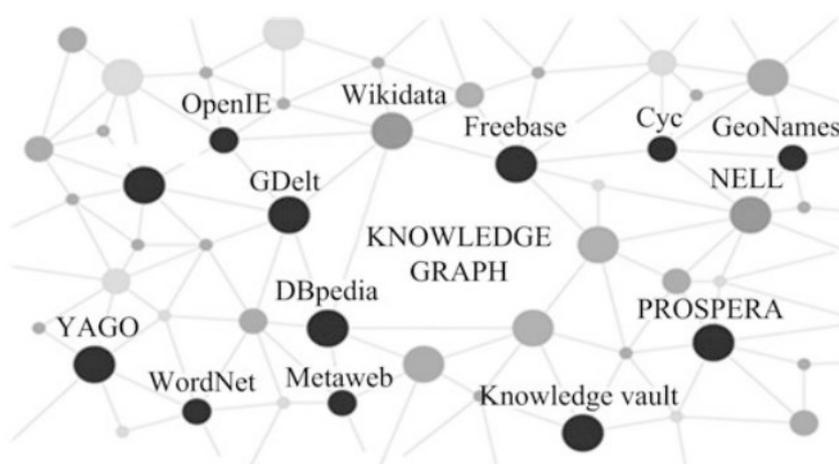
Os importantes avanços na extração de informação e o maior sucesso e qualidade das comunidades de compartilhamento de conhecimento, como a Wikipédia, possibilitaram a construção automática de grandes bases de conhecimento (Paulheim, 2017).

De acordo com Nakashole, Theobald e Weikum (2011), esforços significativos foram realizados nessa área por intermédio de projetos científicos inovadores, tais como opencyc.org, dbpedia.org, knowitall, stat-snowball, readtheweb e Yago. Além disso, iniciativas do setor

privado, como wolframalpha.com, freebase.com e trueknowledge.com, também são consideráveis. Com isso, esses bancos de dados reúnem milhões de entidades e relacionam centenas de milhões de informações sobre essas entidades.

Nos últimos anos, de acordo com Färber *et al.* (2018), tem ocorrido o desenvolvimento de diversos grafos de conhecimento notáveis, abrangentes e acessíveis em múltiplos domínios. Estes incluem o DBpedia, o Freebase, o OpenCyc, o Wikidata e o Yago, os quais, apesar de serem amplamente empregados, ainda carecem de uma análise comparativa minuciosa.

Figura 11 – Alguns grafos de conhecimento em grande escala bem conhecidos



Fonte: Feng (2023).

Paulheim (2017) aponta que existem diversas estratégias para se desenvolver grafos de conhecimento. Enquanto alguns projetos, tais como o Cyc, são cuidadosamente elaborados por especialistas, outros projetos, como o Freebase e o Wikidata, são fruto do trabalho colaborativo. O autor menciona também grafos como DBpedia e Yago, que são originados de extensas fontes semiestruturadas, a exemplo da Wikipédia. Além disso, existem técnicas para obter informações de dados semiestruturados ou não estruturados que criam grafos de conhecimento, como NELL, PROSPERA e KnowledgeVault.

Um dos projetos pioneiros na área de processamento de linguagem em grande escala foi o projeto conhecido como Never Ending Language Learning (NELL). Para Paulheim (2017), esse projeto está em constante desenvolvimento e funciona por meio da análise de um amplo espectro de páginas *web*. Ele utiliza um método integrado para reconhecer e assimilar padrões de texto que sinalizam afirmações sobre categorias e conexões, o que é usado para distinguir e retirar novas entidades e associações.

A estrutura do PROspering knOwledge with Scalability, PrEcision, and RecAll (PROSPERA), conforme descrita por Nakashole, Theobald e Weikum (2011), é inovadora no processamento de relações binárias, visto que se baseia em um *corpus* textual para extrair informações relevantes, atuando em três fases distintas: a primeira, na coleta de padrões, em que identifica estruturas de dados pertinentes; a segunda, na análise de padrões, em que interpreta e avalia as informações coletadas; a última, que conclui com o “raciocínio”, etapa em que aplica lógica para deduzir novos fatos.

Nakashole, Theobald e Weikum (2011) afirmam que o desenho do sistema visa à eficiência em ambientes distribuídos, o que favorece o processamento simultâneo de etapas e acelera o processo. Um elemento central é a natureza iterativa do sistema, que reutiliza as descobertas de uma iteração como base para as próximas análises, criando, assim, um processo contínuo de melhorias em que as pesquisas ficam cada vez mais refinadas e precisas.

O Knowledge Vault é um projeto do Google que se dedica a extrair informações de uma variedade de fontes, incluindo textos, tabelas em HTML e dados da *web* estruturados em formatos como microdados e microformatos. Essas informações são processadas e a confiabilidade de cada fato é determinada pela combinação da precisão das ferramentas de extração e das probabilidades anteriores das afirmações calculadas utilizando os grafos de conhecimento. Posteriormente, somente os dados considerados confiáveis conforme as medidas de confiança estabelecidas são incorporados ao acervo do Knowledge Vault (Paulheim, 2017).

Em concordância com Färber *et al.* (2018), foram selecionadas para a aplicação desta pesquisa as bases DBpedia, Yago, Freebase Wikidata e Cyc porque são grandes grafos de conhecimento, de vários domínios, disponíveis gratuitamente e que abrangem o conhecimento geral, não sendo apenas domínios específicos, o que os torna amplamente aplicáveis a várias necessidades de pesquisa. Além disso, segundo Färber *et al.* (2018), fazem parte da nuvem LOD, o que significa que são acessíveis e utilizáveis de forma gratuita, fornecendo uma rica fonte de informações para pesquisadores e desenvolvedores que buscam dados abrangentes em diferentes campos.

No Quadro 8, é apresentado um comparativo entre cinco bases de conhecimento de grafos. Esses grafos de conhecimento foram comparados entre si com o objetivo de ajudar pesquisadores e desenvolvedores a escolher o melhor grafo de conhecimento para suas necessidades individuais, aplicando critérios de qualidade de dados.

Quadro 8 – Comparativo de bases de conhecimento

Característica	DBpedia	Yago	Freebase	Wikidata	Cyc
Fonte de dados	Wikipédia.	Wikipédia, WordNET, GeoNames.	Diversas, incluindo contribuições de usuários.	Contribuições colaborativas e referências verificáveis.	Conhecimento curado por especialistas.
Estrutura	Extração de infobox da Wikipédia.	Combina fatos da Wikipédia com taxonomia do WordNET.	Modelo de dados baseado em grafos com curadoria da comunidade.	Modelo de dados entidade relacionamento com suporte multilíngue.	Base de conhecimento com lógica formal e ontologias complexas.
Atualizações	Sincronizado com atualizações da Wikipédia.	Atualizações regulares com novas versões.	Descontinuado (dados migrados para o Wikidata)	Contínuas pela comunidade global.	Constante, com adição de conhecimento especializado.
Tipos de dados	Enciclopédicos e estruturados.	Enciclopédicos, lexicais e geográficos.	Enciclopédicos, culturais e científicos.	Enciclopédicos com ênfase na verificabilidade.	Conhecimento geral e específico de domínio com lógica formal.
Exatidão	Dependente da qualidade dos <i>infoboxes</i> .	Alta, com validação cruzada de informações.	Alta, antes da migração para o Wikidata.	Alta, com revisões e referências verificáveis.	Muito alta, com foco em conhecimento validado por especialistas.
Interligação	Ligações para a <i>web</i> semântica.	Ligações semânticas internas e externas.	Interligação entre entidades e conceitos.	Interconectado com outras bases de dados semânticos.	Isolado, com poucas conexões externas.
Linguagem semântica	RDF e SPARQL.	RDF, SPARQL e RDFS/OWL.	RDF e API, própria para acesso aos dados.	RDF, SPARQL e outras interfaces de programação.	Cycl (linguagem própria de programação).
Licença	Creative Commons Attribution-ShareAlike 3.0 Unported License, GNU Free Documentation License.	Creative Commons Attribution License, Creative Commons Attribution 3.0 License.	Licenças Creative Commons, Contribuições dos Usuários, Licença de Dados Abertos.	Creative Commons CC0 1.0 Universal Public Domain Dedication (CC0 1.0).	OpenCyc, ResearchCyc, Cyc comercial.

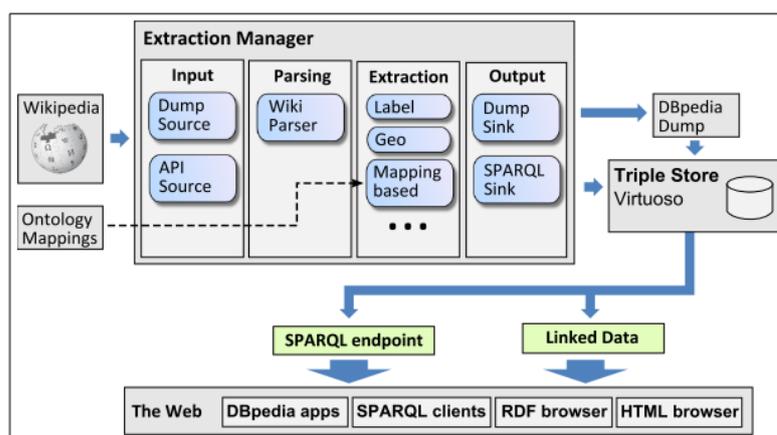
Fonte: Elaborado pelo autor (2024).

### 3.9.1 DBpedia

O DBpedia destaca-se como uma das principais bases de conhecimento no contexto dos grafos de conhecimento, atuando como um ponto de conexão entre as informações estruturadas da Wikipédia e a *web* semântica. O DBpedia é um projeto comunitário que tem como objetivo extrair informações organizadas e em diversos idiomas da Wikipédia. Esses dados são, em seguida, disponibilizados de forma gratuita na internet, fazendo uso das tecnologias associadas à *web* semântica e Linked Data. A estruturação da informação é realizada por meio da extração de dados dos *infoboxes* dos artigos da Wikipédia, convertendo-os em um formato estruturado e de fácil consulta. Os mapeamentos são concebidos por meio de um esforço global de *crowdsourcing*<sup>2</sup>, que possibilita a combinação de conhecimento das diferentes edições da Wikipédia (Lehmann *et al.*, 2012).

Para Lehmann *et al.* 2012, o projeto publica regularmente lançamentos de todas as bases de conhecimento da DBpedia para *download* e oferece acesso para consultas SPARQL, o que possibilita a realização eficiente de consultas complexas, recuperação e integração de informações. Os dados na DBpedia são representados no formato RDF, o que facilita a interoperabilidade com outras bases de dados semânticos. Na Figura 12, é apresentado um diagrama que detalha quatro etapas técnicas da extração da DBpedia, a saber, entrada, análise, extração e saída.

Figura 12 – Visão Geral da estrutura de extração DBpedia



Fonte: Lehmann *et al.* (2012).

Färber *et al.* (2018) apontam que o DBpedia é o grafo de conhecimento mais proeminente na nuvem Linked Open Data (LOD). Os pesquisadores destacam que a DBpedia

<sup>2</sup> *Crowdsourcing* é um termo em inglês que se refere ao processo de obter serviços, ideias ou conteúdo por meio da contribuição de um grande grupo de pessoas, geralmente *on-line*.

tem sido amplamente utilizada na comunidade de pesquisa da *web* semântica e também se mostrado relevante em contextos comerciais. Centenas de conjuntos de dados na *web* publicam *links* RDF apontando para a DBpedia, estabelecendo-a como um dos *hubs* centrais de interligação na nuvem LOD.

Esse projeto, conforme Färber *et al.* (2018), tem impacto significativo na disseminação e integração de conhecimento estruturado, o que promove a evolução da *web* semântica e fomenta a interconexão eficiente entre diferentes fontes de informação. Os autores concluem que o DBpedia desempenha um papel crucial ao disponibilizar e compartilhar dados estruturados, impulsionando avanços em áreas como pesquisa acadêmica, desenvolvimento de aplicações inteligentes e enriquecimento do conhecimento humano.

### 3.9.2 Yago

O Yet Another Great Ontology (Yago) é um grande grafo de conhecimento semântico que desempenha um papel fundamental na disponibilização de uma base de dados estruturada e precisa. Ele surgiu no Instituto Max Planck de Ciência da Computação, em Saarbrücken, a partir de 2007. Färber *et al.* (2018) relatam que essa iniciativa envolve o desenvolvimento de uma ontologia que reúne informações provenientes de várias fontes, como a Wikipédia, o WordNet e o GeoNames. Uma das principais características do Yago é a fusão de dados da Wikipédia e do WordNet, o que possibilita a união de conhecimento enciclopédico com a compreensão lexical, resultando em uma ontologia rica que oferece uma representação detalhada do conhecimento.

Färber *et al.* (2018) destacam que o Yago é capaz de integrar declarações de diferentes versões linguísticas da Wikipédia em um único grafo de conhecimento, bem como apresenta informações contextuais e detalhes sobre a fonte das informações. Os autores relatam que um exemplo de informação contextual são os textos âncora de todos os *links* da Wikipédia, e que esse tipo de integração e contextualização das informações torna o Yago uma ferramenta valiosa para a análise e extração de conhecimento de dados da Wikipédia.

De acordo com Suchanek, Kasneci e Weikum (2007), o Yago se fundamenta em um modelo de dados que compreende entidades e relações binárias, e que, por meio da introdução de identificadores para instâncias de relação, também é possível representar vínculos entre instâncias de relação e características gerais das relações. Um exemplo, segundo os autores, seria a avaliação da popularidade de pares de jogadores de futebol e seus times, além de características gerais das relações, como transitividade ou cumplicidade.

Para Suchanek, Kasneci e Weikum (2007), essa integração resulta em uma ontologia que oferece uma ampla variedade de classes e relações, proporcionando uma estrutura ontológica robusta. O Yago é reconhecido por sua alta precisão e confiabilidade, resultado de um processo cuidadoso de extração, filtragem e consolidação de dados. Segundo os autores, essa ferramenta é disponibilizada gratuitamente à comunidade e pode ser empregada em diversas atividades, inclusive para enriquecer dados, busca semântica e investigações acadêmicas. Ademais, os autores afirmam que a DBpedia e o Yago assemelham-se, o que possibilita a realização de consultas por meio da linguagem SPARQL e a recuperação de informações específicas pelos usuários de forma eficiente. O Yago é mantido por uma comunidade acadêmica ativa, que continua a desenvolver e expandir os grafos de conhecimento (Suchanek; Kasneci; Weikum, 2007).

### 3.9.3 Freebase

Färber *et al.* (2018) relatam que o Freebase foi um conhecido grafo de conhecimento introduzido pela Metaweb Technologies, Inc., em 2007, que posteriormente foi adquirido pela Google Inc., no ano de 2010. Diferentemente da DBpedia, o Freebase possibilitava que os usuários finais contribuíssem para os grafos editando dados estruturados. Além das contribuições dos usuários, o Freebase também integrava dados de fontes como Wikipédia, NNDB, FMD e MusicBrainz. Ele se utilizava de um modelo de grafos próprio para armazenar declarações complexas. Porém, em agosto de 2016, o Freebase encerrou completamente seus serviços, restando apenas o conjunto de dados mais recente, que ainda está disponível. Atualmente, a Wikimedia Deutschland e a Google integram os dados do Freebase no Wikidata por meio da ferramenta Primary Sources Tool.

A extensiva abrangência nesses grafos do Freebase decorre da incorporação de diversos conjuntos de dados, por exemplo, o MusicBrainz. Um dos principais aspectos do Freebase foi o seu modelo de dados baseado em grafos, que representava entidades como nós e relações como arestas, o que possibilitava uma representação eficiente e flexível do conhecimento. A disponibilização de APIs públicas propiciou que desenvolvedores acessassem e integrassem os dados do Freebase em seus próprios aplicativos e serviços, ampliando o alcance e a utilidade da plataforma (Färber *et al.*, 2018).

Bollacker *et al.* (2008) menciona que o Freebase teve uma influência duradoura no desenvolvimento da *web* semântica, visto que forneceu acesso aberto a uma base de conhecimento tão vasta e estruturada, o que influenciou positivamente o desenvolvimento de

outros projetos na área. Para o autor, embora o Freebase como plataforma independente não exista mais, seu legado continua vivo por meio dos dados que foram migrados para o Wikidata e sua influência em projetos semânticos subsequentes. O impacto do Freebase no compartilhamento e na estruturação de conhecimento colaborativo é inegável, e seu papel na evolução da *web* semântica continua a ser reconhecido e valorizado na comunidade de tecnologia e informação.

O Google Knowledge Graph é alimentado em parte pelo Freebase e tem mais de 22 milhões de entidades e 350 milhões de relações em cerca de 100 domínios. A aquisição pelo Google, em 2010, marcou um ponto de virada para o Freebase, que eventualmente integrou seus dados ao Projeto Knowledge Graph do Google, ampliando ainda mais o impacto e o alcance do conhecimento coletado. A migração para o Wikidata, em 2014, representou uma transição significativa para os dados do Freebase, garantindo a continuidade e acessibilidade do conhecimento estruturado para futuras aplicações (Bollacker *et al.*, 2008).

### 3.9.4 Wikidata

Färber *et al.* (2018) esclarece que o projeto Wikidata, iniciado em outubro de 2012 pela Wikimedia Deutschland (WMDE), tem como propósito disponibilizar dados úteis para qualquer projeto da Wikimedia, incluindo a Wikipédia. Além de armazenar fatos, o Wikidata também inclui as respectivas fontes para garantir a validade dos dados. As informações sobre entidades no Wikidata estão disponíveis em quase 400 idiomas por meio de rótulos, aliases e descrições. Para os autores, esse é um esforço colaborativo, no qual os usuários contribuem e editam informações, enquanto o esquema é atualizado e expandido com base em acordos comunitários. O Wikidata está em processo de crescimento exponencial em decorrência da integração dos dados do Freebase.

O Wikidata é projetado para ser facilmente integrado a outros grafos de conhecimento e sistemas de informação, por meio do uso de padrões da *web* semântica como RDF (Färber *et al.*, 2018). O serviço de consulta SPARQL do Wikidata possibilita que usuários façam perguntas complexas e recuperem dados de formas sofisticadas, aproveitando a natureza interconectada dos grafos. Considerado um recurso valioso para pesquisadores, desenvolvedores e qualquer pessoa interessada tanto em acessar conhecimento estruturado em escala global como contribuir com ele, para Färber *et al.* (2018), o Wikidata é uma ferramenta poderosa para a *web* semântica, que oportuniza a integração e a análise de grandes conjuntos de dados estruturados.

O Wikidata é reconhecido como uma das principais fontes de dados estruturados disponíveis na *web*. Sendo uma fonte secundária, é crucial que o seu conteúdo seja respaldado por referências confiáveis. Isso se torna especialmente relevante, pois o Wikidata encoraja os seus editores a incluir afirmações que possam não ser amplamente aceitas, contanto que sejam suportadas por referências. No entanto, apesar da relação fundamental entre conteúdo e referências, a capacidade do Wikidata para sistematicamente avaliar e garantir a qualidade das referências permanece limitada (Amaral *et al.*, 2021).

### 3.9.5 Cyc

Färber *et al.* (2018) relatam que o projeto Cyc foi iniciado, em 1984, pelo consórcio de pesquisa e desenvolvimento da indústria Microelectronics and Computer Technology Corporation. O objetivo do Cyc é armazenar, de forma processável por máquina, milhões de fatos de senso comum, como “Toda árvore é uma planta”. O principal objetivo do Cyc tem sido a inferência e o raciocínio. Uma vez que o Cyc é proprietário, foi lançada uma versão menor do grafo, denominada OpenCyc, por meio da licença de código aberto Apache Versão 2.

De acordo com Färber *et al.* (2018), em julho de 2006, foi publicado o ResearchCyc para a comunidade científica, contendo mais dados que o OpenCyc. O Cyc e o ResearchCyc são grafos que não são disponibilizados para serem usados livremente, sendo empregados em um contexto específico. Os autores afirmam que o Cyc é um projeto de inteligência artificial de longo prazo que tem como objetivo codificar conhecimento e raciocínio comum de forma que possam ser utilizados por computadores. Ele funciona como uma base de conhecimento e um motor de inferência, que possibilita o raciocínio sobre os dados armazenados para gerar novas conclusões.

Os autores explicam que o conhecimento armazenado no Cyc pode ser aplicado a uma variedade de tarefas complexas, desde a melhoria da busca por informações até o desenvolvimento de sistemas de inteligência artificial mais avançados.

Uma das principais características do Cyc é o seu conhecimento profundo, projetado para armazenar uma vasta quantidade de conhecimento do mundo real, incluindo senso comum e fatos específicos de domínio (Färber *et al.*, 2018). Além disso, o Cyc utiliza uma lógica formal para representar seu conhecimento, o que promove um raciocínio sofisticado e inferências complexas. Segundo Färber *et al.* (2018), o Cyc contém uma das maiores ontologias existentes, com centenas de milhares de conceitos e milhões de fatos que descrevem como esses conceitos se relacionam entre si.

No entanto, sua interoperabilidade é limitada em comparação com outros grafos de conhecimento, como DBpedia e Wikidata, embora esforços tenham sido feitos para conectá-lo a padrões da *web* semântica.

Färber *et al.* (2018) concluem que o Cyc é um dos primeiros e mais ambiciosos projetos no campo da inteligência artificial e grafos de conhecimento, com foco particular em capturar e utilizar conhecimento de senso comum que é frequentemente implícito para humanos, mas difícil de formalizar em sistemas computacionais. Sua capacidade de armazenar uma vasta quantidade de conhecimento e realizar raciocínio sofisticado o torna uma ferramenta poderosa para aplicações em diversas áreas, contribuindo para avanços significativos no campo da inteligência artificial.

## 4 DESENVOLVIMENTO DE UM GRAFO DE CONHECIMENTO

### 4.1 CONSIDERAÇÕES INICIAIS

Para assegurar a transparência e a reprodutibilidade das metodologias e tecnologias empregadas no desenvolvimento de um grafo de conhecimento, recomenda-se a documentação de todo o processo. Essa ação, ainda, facilita a aplicação do conhecimento adquirido em projetos futuros, bem como colabora com aqueles que possam enfrentar problemas semelhantes ao oferecer-lhes uma descrição minuciosa dos desafios encontrados e das soluções implementadas.

Nesta pesquisa, propõe-se a construção de um grafo de conhecimento com base nos dados obtidos do Congresso Internacional de Conhecimento e Inovação (CIKI) (Figuras 13 e 14) e do Repositório Institucional da UFSC associados ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC) (Figura 15), ambos de 2023, a serem aplicados no sistema de gestão de banco de dados Neo4j. O processo compreendeu as etapas de extração, modelagem e inserção dos dados no banco de dados de grafo, com a finalidade de representar e analisar as relações entre os diversos elementos de conhecimento, facilitando assim a compreensão e a exploração das interconexões existentes.

Figura 13 – Tela da página inicial do *site* CIKI 2023

**Edição Atual**

2023: Anais do XIII Congresso Internacional de Conhecimento e Inovação (ciKi) - Versão Preliminar

O Congresso Internacional de Conhecimento e Inovação (CIKI) é um evento interamericano que promove avanços conceituais, metodológicos e práticos em gestão do conhecimento, capital intelectual e gestão da inovação.

A Conferência Internacional sobre Gestão do Conhecimento (ICKM) tem como objetivo proporcionar discussões mundiais sobre os modelos mais contemporâneos de gestão do conhecimento.

Em 2023, foi realizada uma edição conjunta da 18ª Conferência Internacional sobre Gestão do Conhecimento (ICKM2023) e do 13º Congresso Internacional sobre Conhecimento e Inovação (CIKI2023), na cidade de Florianópolis, Brasil, organizada pelo Programa de Pós-Graduação em Engenharia, Gestão e Mídia do Conhecimento (EGC) da Universidade Federal de Santa Catarina – UFSC (Brasil). Diferentes universidades e organizações se reuniram compartilhando suas experiências e pesquisas em todos os aspectos da criação, inovação, coprodução, distribuição e utilização do conhecimento na sociedade.

Agradecemos aos autores e avaliadores e apresentamos a comunidade os artigos selecionados. Boa leitura.

Prof. Fernando Ostuni Gauthier  
Editor dos Anais

Idioma  
English  
Español (España)  
Português (Brasil)

Fonte: Extraído do *site* CIKI (2024).

Figura 14 – Tela de visualização da lista de artigos do CIKI

Artigo completo

**AVALIAÇÃO DA MATURIDADE DA GESTÃO DO CONHECIMENTO DA REITORIA DE UMA INSTITUIÇÃO DE ENSINO**  
UM ESTUDO DE CASO  
Joelias Silva Pinto Júnior, Léa Paula Vanessa Xavier Correa de Morais, Gertrudes Aparecida Dandolini  
[PDF](#)

**ABORDAGEM DE PESQUISA COLABORATIVA**  
UMA ANÁLISE BIBLIOMÉTRICA DA LITERATURA  
Gertrudes Dandolini, ALINE WREGE VASCONCELOS, Cecília Kotzias, Laura Pertile, Marily Dilamar da Silva  
[PDF](#)

**AMBIENTES VIRTUAIS DE APRENDIZAGEM SOB A PERSPECTIVA DA GESTÃO DO CONHECIMENTO: UMA REVISÃO SISTEMÁTICA DE LITERATURA**  
Roberta Begrow, Marcelo Macedo  
[PDF](#)

**ANÁLISE DE PATENTES DEPOSITADAS PELA UNIVERSIDADE FEDERAL DE SANTA CATARINA**  
Gabriel Rogedo Chiarelli, Tainara Volan, Caroline Rodrigues Vaz, Mauricio Uriona-Maldonado  
[PDF](#)

**Aplicação da Knowledge Management Assessment Tool – KMAT para sistematização da Gestão do Conhecimento**  
Emerson Cleister Lima Muniz, Heloisa Fernandes do Nascimento

Fonte: Extraído do *site* do CIKI (2024).

Figura 15 – Tela página inicial Repositório Institucional Teses Dissertações/UFSC

Entrar A- A A+

**RI UFSC** REPOSITÓRIO INSTITUCIONAL | UFSC

Repositório Institucional da UFSC ⇨ Teses e Dissertações ⇨ Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento ⇨ Navegação Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento por data do documento

**Buscar DSpace**  
   
 Buscar DSpace  
 Esta coleção

**Navegar**  
 Todo o repositório  
[Comunidades e Coleções](#)  
[Por data do documento](#)  
[Autores](#)  
[Títulos](#)  
[Assuntos](#)  
 Esta coleção  
[Por data do documento](#)  
[Autores](#)  
[Títulos](#)  
[Assuntos](#)

**Navegação Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento por data do documento**

Ir para um ponto do índice: (Escolha o mês) 2023  
 Ou digite um ano:

Classificar por: data do documento Ordenar: Ascendente Resultados: 20

[Página anterior](#) Itens para a visualização no momento 738-757 of 796 [Próxima página](#)

**Marketing do conhecimento: o uso da plataforma digital do Youtube na divulgação online do conhecimento científico**  
Paz, Marisol da Silva (2023)

**Lean leadership: atributos implícitos da liderança em um ambiente lean**  
Rosa, Kellyn Clyciane Mendes (2023)

**Modelo operacional de educação continuada para profissionais de engenharia baseado em competências digitais para atender a desafios e oportunidades da transformação digital**  
Gargioni, Sérgio Luiz (2023)

**Fatores críticos de sucesso para internacionalização de empresas de base tecnológica**  
Silva, Yuri Caldeira da (2023)

Fonte: Extraído do Repositório Institucional da UFSC (2024).

Em 2023, a 18.<sup>a</sup> Conferência Internacional sobre Gestão do Conhecimento (ICKM 2023) e o 13.<sup>o</sup> Congresso Internacional sobre Conhecimento e Inovação (CIKI 2023) foram realizados conjuntamente na cidade de Florianópolis, Santa Catarina, Brasil, sendo promovidos pelo PPGEGC/UFSC. Diversas universidades e organizações se reuniram nesse evento para compartilhar experiências e pesquisas sobre criação, inovação, coprodução, disseminação e aplicação do conhecimento na sociedade.

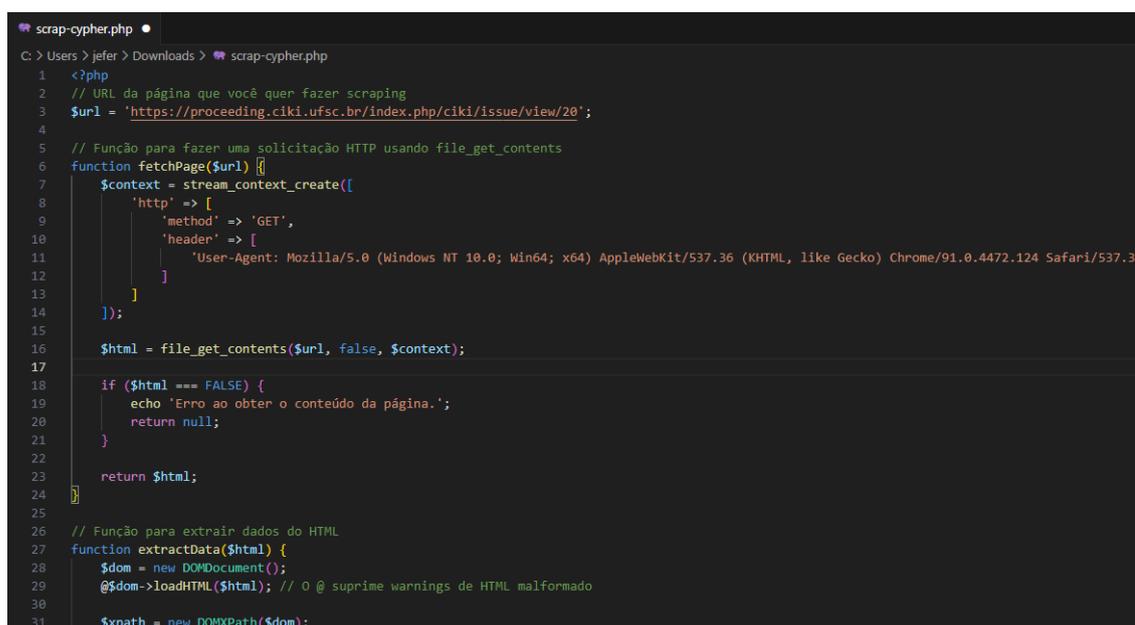
Para construir o grafo de conhecimento com base nos dados extraídos do Congresso CIKI 2023 e do Repositório PPGEGC/UFSC 2023, foi crucial modelar várias informações, capturando elementos individuais e suas relações.

#### 4.2 GRAFO DE CONHECIMENTO SOBRE ARTIGOS DO CIKI 2023

Para dar início à construção do grafo de conhecimento baseado nos artigos do CIKI 2023, inicialmente se modelou informações de artigos científicos, como título, data de criação, submissão, emissão, descrições e palavras-chave. Autores e coautores foram identificados pelos nomes e afiliações, o que possibilitou mapear colaborações e redes de pesquisa.

O processamento dos dados do CIKI 2023 no Neo4j foi realizado por meio da técnica de *web scraping* (Figura 16) para extrair informações diretamente do *site* do Congresso. Segundo Mitchell (2019), essa técnica é uma abordagem eficiente para coletar dados estruturados de *sites*, pois é possível criar arquivos que podem ser facilmente migrados para bancos de dados como o Neo4j.

Figura 16 – Tela do *web scraping*



```

scrap-cypher.php
C:\Users\jefer\Downloads\scrap-cypher.php
1 <?php
2 // URL da página que você quer fazer scraping
3 $url = 'https://proceeding.ciki.ufsc.br/index.php/ciki/issue/view/20';
4
5 // Função para fazer uma solicitação HTTP usando file_get_contents
6 function fetchPage($url) {
7     $context = stream_context_create([
8         'http' => [
9             'method' => 'GET',
10            'header' => [
11                'User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36'
12            ]
13        ]
14    );
15
16    $html = file_get_contents($url, false, $context);
17
18    if ($html === FALSE) {
19        echo 'Erro ao obter o conteúdo da página.';
20        return null;
21    }
22
23    return $html;
24 }
25
26 // Função para extrair dados do HTML
27 function extractData($html) {
28     $dom = new DOMDocument();
29     @$dom->loadHTML($html); // 0 @ suprime warnings de HTML malformatado
30
31     $xpath = new DOMXPath($dom);

```

Fonte: Elaborado pelo autor (2024).

Acessar e analisar grandes volumes de dados é essencial em qualquer pesquisa. Com o uso de *web scraping*, a extração automática de informações do *site* do CIKI 2023 ocorreu de forma rápida e eficiente, em vista de valer-se de serviços de *software* ou programas específicos que percorrem diversos níveis de páginas. Essa técnica é amplamente utilizada por indivíduos e empresas para obter vantagens da imensa quantidade de dados disponíveis *on-line*, aprimorando suas decisões. Com o uso de *web scraping*, os dados são extraídos no formato mais relevante ao contexto de utilização e, posteriormente, podem ser convertidos para qualquer formato desejado (Abodayeh *et al.*, 2023). Na Figura 17, é apresentado o código fonte HTML contendo informações essenciais do processo de *web scraping* para extração de dados.

Figura 17 – Tela do código-fonte HTML

```

223 </nav>
224
225 <article class="obj_article_details">
226
227
228 <h1 class="page_title">
229 GESTÃO DO CONHECIMENTO COMO MEIO PARA O DESENVOLVIMENTO NO SETOR AGRÍCOLA
230 </h1>
231
232
233 <div class="row">
234 <div class="main_entry">
235
236 <section class="item authors">
237 <h2 class="pkp_screen_reader">Autores</h2>
238 <ul class="authors">
239 <li>
240 <span class="name">
241 Braz Ferraz Filho
242 </span>
243 <span class="affiliation">
244 Universidade Federal de Santa Catarina (UFSC) - Instituto Federal Catarinense (IFC) - Campus Santa Rosa do Sul
245 </span>
246 <span class="orcid">
247
248 <a href="https://orcid.org/0000-0001-8791-5569" target="_blank">
249 https://orcid.org/0000-0001-8791-5569
250 </a>
251 </span>
252 </li>
253 <li>
254 <span class="name">
255 Thalita Bez Batti
256 </span>
257 <span class="affiliation">
258 Universidade Federal de Santa Catarina (UFSC)
259 </span>
260 </li>
261 <li>
262 <span class="name">
263 Thais Regina Umbelino

```

Fonte: Elaborado pelo autor (2024).

Os metadados de uma página HTML, como `<meta>`, `<title>` e `<link>`, fornecem informações essenciais sobre o conteúdo e a estrutura da página, desempenhando um papel crucial no processo de *web scraping*. Eles ajudam a identificar e classificar o conteúdo de forma precisa, o que facilita a extração e organização dos dados. Metadados podem incluir descrições, palavras-chave e informações de autor, que são úteis para entender o contexto e a relevância das informações extraídas, bem como melhorar a eficiência e a precisão do *scraping*. Na Figura 18, é mostrada a tela de metadados em HTML de um artigo apresentado no congresso CIKI 2023, contendo informações utilizadas na construção do grafo.

Figura 18 – Tela metadado da página HTML

```

1 <!DOCTYPE html>
2 <html lang="pt-BR" xml:lang="pt-BR">
3 <head>
4   <meta charset="utf-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0">
6   <title>
7     GESTÃO DO CONHECIMENTO COMO MEIO PARA O DESENVOLVIMENTO NO SETOR AGRÍCOLA
8     | Anais do Congresso Internacional de Conhecimento e Inovação - ciki
9   </title>
10
11
12 <link rel="icon" href="https://proceeding.ciki.ufsc.br/public/journals/1/favicon_pt_BR.png">
13 <meta name="generator" content="Open Journal Systems 3.2.1.1">
14 <meta name="gs_meta_revison" content="1.1"/>
15 <meta name="citation_journal_title" content="Anais do Congresso Internacional de Conhecimento e Inovação - ciki"/>
16 <meta name="citation_journal_abbrev" content="ciki"/>
17 <meta name="citation_issn" content="2318-5376"/>
18 <meta name="citation_author" content="Braz Ferraz Filho"/>
19 <meta name="citation_author_institution" content="Universidade Federal de Santa Catarina (UFSC) - Instituto Federal Catarinense (IFC) - Campus Santa Rosa do Sul"/>
20 <meta name="citation_author" content="Thalita Bez Batti"/>
21 <meta name="citation_author_institution" content="Universidade Federal de Santa Catarina (UFSC)"/>
22 <meta name="citation_author" content="Thais Regina Umbelino"/>
23 <meta name="citation_author_institution" content="Universidade Federal de Santa Catarina (UFSC)"/>
24 <meta name="citation_author" content="Alexandre Leopoldo Gonçalves"/>
25 <meta name="citation_author_institution" content="Universidade Federal de Santa Catarina (UFSC)"/>
26 <meta name="citation_title" content="GESTÃO DO CONHECIMENTO COMO MEIO PARA O DESENVOLVIMENTO NO SETOR AGRÍCOLA"/>
27 <meta name="citation_language" content="pt"/>
28 <meta name="citation_date" content="2023/11/21"/>
29 <meta name="citation_abstract_html_url" content="https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/1472"/>
30 <meta name="citation_keywords" xml:lang="pt" content="Agricultura"/>
31 <meta name="citation_keywords" xml:lang="pt" content="Produção"/>
32 <meta name="citation_keywords" xml:lang="pt" content="Gestão do Conhecimento"/>
33 <meta name="citation_pdf_url" content="https://proceeding.ciki.ufsc.br/index.php/ciki/article/download/1472/794"/>
34 <link rel="schema:DC" href="http://purl.org/dc/elements/1.1/" />
35 <meta name="DC.Creator.PersonalName" content="Braz Ferraz Filho"/>
36 <meta name="DC.Creator.PersonalName" content="Thalita Bez Batti"/>
37 <meta name="DC.Creator.PersonalName" content="Thais Regina Umbelino"/>
38 <meta name="DC.Creator.PersonalName" content="Alexandre Leopoldo Gonçalves"/>
39 <meta name="DC.Date.created" scheme="ISO8601" content="2023-11-21"/>
40 <meta name="DC.Date.dateSubmitted" scheme="ISO8601" content="2023-10-16"/>

```

Fonte: Elaborado pelo autor (2024).

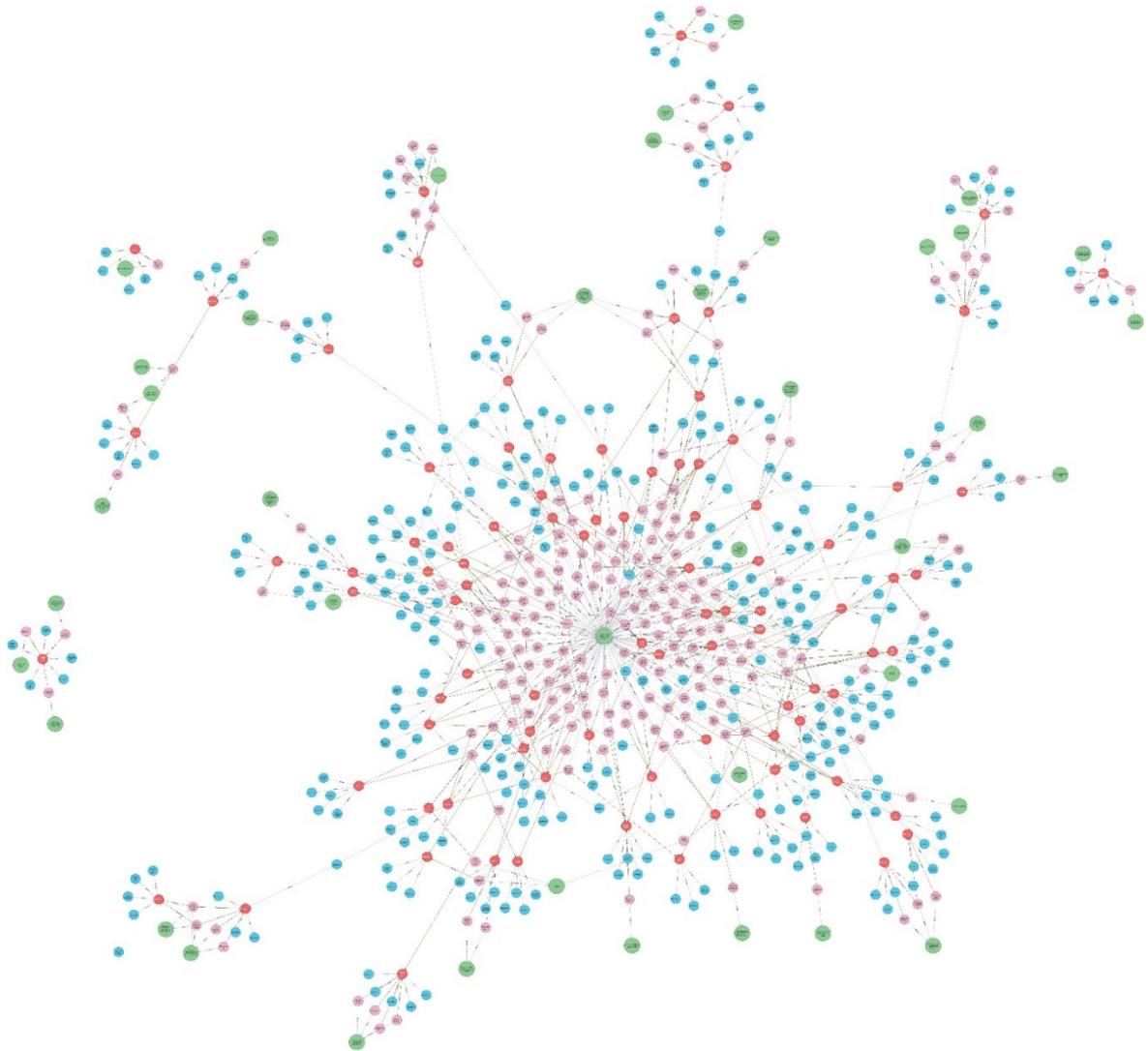
Para garantir que os dados estejam alinhados e consistentes no grafo, foi utilizada a técnica de Entity Resolution (ER) de forma manual. Trata-se da identificação e fusão de registros que representam a mesma entidade, mas que podem estar descritos de maneiras diferentes. Esse processo envolve normalização, comparação e fusão de registros, usando desde métodos baseados em regras até algoritmos de aprendizado de máquina. Com isso, evita-se redundâncias e assegura-se que cada entidade seja representada de forma única no grafo, de modo a obter análises e consultas mais precisas. Essa técnica ainda garante a consistência e a integridade dos dados, essencial em grandes volumes de informações variadas. Logo, com a aplicação de ER, é possível criar uma base de dados precisa e consolidada, fundamental para análises e decisões informadas (Elmagarmid; Ipeirotis; Verykios, 2007).

O gerenciamento e a visualização de dados complexos no Neo4j após a implementação dos dados do CIKI 2023 demonstraram-se eficaz e interativo. De acordo com Robinson, Webber e Eifrem (2015), bancos de dados orientados a grafos como o Neo4j proporcionam uma modelagem mais natural e intuitiva das relações intrincadas entre dados, sendo particularmente vantajosos em ambientes de pesquisa científica.

No CIKI 2023, destacou-se como a gestão do conhecimento pode ser fortalecida através da visualização das conexões e relações entre diversos artigos e conceitos. Essa metodologia facilita a identificação de novas informações e insights, promovendo um maior entendimento e utilização do conhecimento existente. Além disso, a aplicação da tecnologia no Neo4j pode funcionar como uma vitrine para as capacidades avançadas de análise de dados, promovendo a adoção de tecnologias de grafos em outros setores e áreas de pesquisa.

Na Figura 19, apresenta-se uma visão ampla das informações do CIKI 2023 no Neo4j. No grafo de conhecimento, as cores dos nós têm a função de diferenciar os tipos de entidades ali representadas. Os nós verdes correspondem às instituições, enquanto os nós rosas representam qualquer pessoa que tenha escrito um artigo ou participado dele. Em contrapartida, os nós vermelhos simbolizam os próprios artigos e os nós azuis são alusivos às palavras-chave associadas a estes. Essa legenda de cores facilita a visualização e compreensão das interrelações e conexões entre as diversas entidades no grafo.

Figura 19 – Grafo de conhecimento do CIKI 2023



Fonte: Elaborado pelo autor (2024).

### 4.3 GRAFO DE CONHECIMENTO SOBRE TESES E DISSERTAÇÕES DO PPGEGC/UFSC EM 2023

O Repositório Institucional de Teses e Dissertações da UFSC reúne informações sobre teses e dissertações de todos os Programas de Pós-graduação da universidade, inclusive do EGC. Neste estudo, foram utilizados os dados do PPGEGC/UFSC referentes ao ano de 2023.

As informações de cada tese ou dissertação são formadas, entre outros, por dados como o título do trabalho (dc.title), o nome do autor (dc.contributor.author) e do orientador (dc.contributor.advisor), o ano de defesa (dc.date.issued), a URI (dc.identifier.uri) e o resumo (dc.description.abstract). Adicionalmente, é mencionada a tipologia do documento (dc.type).

Na Figura 20, é ilustrado o arquivo detalhado de uma dissertação de mestrado do EGC/UFSC que integra os dados coletados, e que pode ser acessada no Repositório Institucional da UFSC.

Figura 20 – Informações completas de uma das dissertações dispostas no Repositório Institucional de Tese e Dissertações da UFSC de 2023

dc.contributor	Universidade Federal de Santa Catarina
dc.contributor.advisor	Macedo, Marcelo
dc.contributor.author	Almeida, Natália Aline Jesus
dc.date.accessioned	2024-03-21T23:24:27Z
dc.date.available	2024-03-21T23:24:27Z
dc.date.issued	2023
dc.identifier.other	386625
dc.identifier.uri	<a href="https://repositorio.ufsc.br/handle/123456789/254819">https://repositorio.ufsc.br/handle/123456789/254819</a>
dc.description	Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2023.
dc.description.abstract	A pesquisa em questão tem como objetivo primordial a compreensão da relação entre os espaços de aprendizagem comuns e a gestão do conhecimento no contexto do ensino profissional e tecnológico na Biblioteca da Faculdade Senac Palhoça. Para tanto, a investigação contextualiza a evolução da Educação Profissional no Brasil e examina a Gestão do Conhecimento como um processo intrínseco de criação, difusão e incorporação de conhecimento nas estruturas organizacionais. É ressaltada a relevância da Gestão do Conhecimento na interseção entre a academia e o mundo do trabalho, pois desempenha um papel crucial na formação do conhecimento e no desenvolvimento de habilidades, especialmente ao explorar os espaços de Aprendizagem Colaborativa presentes nas instituições de ensino, os quais têm foram reconhecidas como essenciais para aprimorar o ensino e a aprendizagem. O conceito de learning commons é abordado como um conceito em constante evolução e de natureza multidisciplinar, com sua aplicação frequentemente observada no âmbito das bibliotecas. Esses espaços têm como objetivo facilitar a produção de conhecimento e promover uma aprendizagem interativa, sendo percebidos como potenciais para a gestão do conhecimento. O estudo de caso se concentra na Biblioteca da Faculdade Senac Palhoça, fornecendo uma compreensão aprofundada do tema em questão e contribuindo para a realização dos objetivos específicos do trabalho. Os resultados obtidos descrevem o nível de implementação do processo de gestão do conhecimento em relação aos espaços disponíveis na biblioteca, utilizando como referência o modelo da APO. Observe-se que a biblioteca adota uma abordagem focada no cliente, priorizando o atendimento às demandas dos usuários em todos os espaços disponíveis.
dc.description.abstract	Abstract: The research in question has as its primordial objective the understanding of the relationship between learning commons spaces and knowledge management in the context of professional and technological education at the Senac Palhoça College Library. To this end, the investigation contextualizes the evolution of Professional Education in Brazil and examines Knowledge Management as an intrinsic process of creation, dissemination and incorporation of knowledge into organizational structures. The relevance of Knowledge Management at the intersection between academia and the world of work is highlighted, as it plays a crucial role in the formation of knowledge and the development of skills, especially when exploring the Collaborative Learning spaces present in educational institutions, which have been recognized as essential for improving teaching and learning. The concept of learning commons is approached as a concept in constant evolution and of a multidisciplinary nature, with its application frequently observed in the context of libraries. These spaces aim to facilitate the production of knowledge and promote interactive learning, being perceived as potential for knowledge management. The case study focuses on the Senac Palhoça Faculty Library, providing an in-depth understanding of the topic in question and contributing to the achievement of the specific objectives of the work. The results obtained describe the level of implementation of the knowledge management process in relation to the spaces available in the library, using the APO model as a reference. It is observed that the library adopts a customer-centric approach, prioritizing meeting users' demands in all available spaces.
dc.format.extent	120 p.   tabs.
dc.language.iso	por
dc.subject.classification	Gestão do conhecimento
dc.subject.classification	Engenharia e gestão do conhecimento
dc.subject.classification	Bem comum
dc.subject.classification	Ensino profissional
dc.title	A relação dos espaços de learning commons com a gestão do conhecimento no ensino profissional: um estudo de caso
dc.type	Dissertação (Mestrado)
dc.contributor.advisor-co	Santos, Geneia Lucas dos

Fonte: Extraído do Repositório Institucional da UFSC (2024).

As informações do Repositório Institucional de Teses e Dissertações da UFSC referentes a teses e dissertações do EGC/UFSC produzidas em 2023 foram coletadas manualmente e inseridas em uma planilha, incluindo dados como autor, orientador, data de publicação, URI, título, tipo e descrição. Esse processo facilitou o acesso ao conteúdo acadêmico, garantindo informações precisas e confiáveis, bem como contribuiu para a construção do *script* Cypher. Na Figura 21, é apresentada a planilha eletrônica com informações extraídas do Repositório.

Figura 21 – Planilha eletrônica com informações de Teses e Dissertações

A	B	C	D	E	F
de.contributor.advisor	de.contributor.author	dc.date.issu	dc.identifier.uri	dc.title	dc.type
Francisco Antonio Pereira Fialho	Edgêthon de Souza	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251267">https://repositorio.ufsc.br/handle/123456789/251267</a>	A complexidade como suporte à modelagem de competências docentes, com Tese (Doutorado)	A implementação das novas metodologias de
Márcelo Mascado	Nathalia Alaine Jena Almeida	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251268">https://repositorio.ufsc.br/handle/123456789/251268</a>	A relação dos espaços de learning commons com a gestão do conhecimento Dissertação (Mestrado)	A pesquisa em questão tem como objetivo pr
Gertrudes Aparecida Dandolini	Rosana Malvestri	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251269">https://repositorio.ufsc.br/handle/123456789/251269</a>	A laboratório prático e os desafios de doutorado interdisciplinar: um quadro de Tese (Doutorado)	A filosofia ou subteoria prático é uma virt
Araçá Hack Otávan	Claudia Regina Damasceno Luciani	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251270">https://repositorio.ufsc.br/handle/123456789/251270</a>	Aprendizagem significativa e cidadania: um estudo de caso no TRE-SC Dissertação (Mestrado)	A presente pesquisa analisa o curso de forma
Fernando Alvaro Ostuni Gauthier	Lodacir Rodrigo Silva da Rosa	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251284">https://repositorio.ufsc.br/handle/123456789/251284</a>	Assistente virtual baseado em conhecimento no apoio a estudantes de curso de pós-graduação	Dissertação (Mestrado) O trabalho tem como objetivo o desenvolvi
Lia Castano Bastos	Pastry Geraldine Balconi Mandelli	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251288">https://repositorio.ufsc.br/handle/123456789/251288</a>	Associação de práticas de Gestão do Conhecimento e a evolução de organiz Tese (Doutorado)	O indivíduo é o principal ativo estratégico p
Gertrudes Aparecida Dandolini	Karyne Malischewski	2023	<a href="https://repositorio.ufsc.br/handle/123456789/253749">https://repositorio.ufsc.br/handle/123456789/253749</a>	Colaboração e inovação no setor público: diretrizes aos i-Labs Dissertação (Mestrado)	A inovação no setor público, antes consid
Édis Mafra Lapoli	Guilherme Agnolin	2023	<a href="https://repositorio.ufsc.br/handle/123456789/249523">https://repositorio.ufsc.br/handle/123456789/249523</a>	Compartilhamento do conhecimento e inclusão social: estudo de caso de um Dissertação (Mestrado)	As pessoas com deficiência têm direito em p
Édis Mafra Lapoli	Patricia Gesser da Costa	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251286">https://repositorio.ufsc.br/handle/123456789/251286</a>	Competências empreendedoras emergentes: o caso de profissionais da saúde Dissertação (Mestrado)	Empreender se torna cada vez mais desafiad
Édis Mafra Lapoli	Rosani Cesário Pereira	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251074">https://repositorio.ufsc.br/handle/123456789/251074</a>	Competências essenciais dos propositores: um estudo com base na percepção Dissertação (Mestrado)	O modelo de gestão que vem se consolidand
Richard Perassi Luiz de Sousa	William Cordero Costa	2023	<a href="https://repositorio.ufsc.br/handle/123456789/252483">https://repositorio.ufsc.br/handle/123456789/252483</a>	Conhecimento de artes visuais e mídia digital no ensino escolar do território Dissertação (Mestrado)	Esta dissertação é apresentado um estudo si
Luciana Maria Fadel	Andréas Carlos Fialho	2023	<a href="https://repositorio.ufsc.br/handle/123456789/247255">https://repositorio.ufsc.br/handle/123456789/247255</a>	Das Bibliotecas aos buscadores: rotando técnicas avançadas para a recuperação Dissertação (Mestrado)	Pesquisar on-line é uma atividade presente n
Lia Castano Bastos	Rafael Maia Pinto	2023	<a href="https://repositorio.ufsc.br/handle/123456789/249503">https://repositorio.ufsc.br/handle/123456789/249503</a>	Deteção de erros e fraudes em gastos públicos nas áreas de saúde e educação Dissertação (Mestrado)	A Constituição Federal estabelece limites m
Fernando José Spanhol	Rangel Machado Simoes	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251478">https://repositorio.ufsc.br/handle/123456789/251478</a>	Educação digital superior: desenvolvendo as competências digitais no curso Tese (Doutorado)	Este trabalho propõe um modelo conceitual
Édis Mafra Lapoli	Aline Débora de Fraga	2023	<a href="https://repositorio.ufsc.br/handle/123456789/248795">https://repositorio.ufsc.br/handle/123456789/248795</a>	Empreses de economia solidária: experiências vividas e sentidas de pessoas que se Dissertação (Mestrado)	As mudanças nos perfis demográficos globais
Neri dos Santos	Yuri Caldeira da Silva	2023	<a href="https://repositorio.ufsc.br/handle/123456789/253665">https://repositorio.ufsc.br/handle/123456789/253665</a>	Fatores críticos de sucesso para internacionalização de empresas de base te Dissertação (Mestrado)	A internacionalização para as empresas é ur
Tarcísio Vazini	Thiago da Silva Mota	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251261">https://repositorio.ufsc.br/handle/123456789/251261</a>	Formação de conhecimento para usuários idosos baseada em personagen r Dissertação (Mestrado)	As Histórias em Quadrinhos são ferramentas
Édis Mafra Lapoli	Juliano Keller Alves	2023	<a href="https://repositorio.ufsc.br/handle/123456789/247773">https://repositorio.ufsc.br/handle/123456789/247773</a>	Framework adaptativo de gestão do conhecimento para a aplicação da ISO Tese (Doutorado)	A ISO 30401 é a norma internacional para s
Alexandre Augusto Biz	Leonardo Lincoln Leite de Lacerda	2023	<a href="https://repositorio.ufsc.br/handle/123456789/247241">https://repositorio.ufsc.br/handle/123456789/247241</a>	Framework conceitual de knowledge commons para cognição de destino tur Tese (Doutorado)	A gestão do conhecimento é responsável po
Denilson Seli	Rudger Nowakki do Nascimento Ti	2023	<a href="https://repositorio.ufsc.br/handle/123456789/249926">https://repositorio.ufsc.br/handle/123456789/249926</a>	Framework de ciência de dados e engenharia do conhecimento para apoio a Tese (Doutorado)	A ciência brasileira está fortemente ligada à
Clárisia Stefani Teixeira	Carlos Marcelo Fastino da Silva	2023	<a href="https://repositorio.ufsc.br/handle/123456789/249568">https://repositorio.ufsc.br/handle/123456789/249568</a>	Framework de mecanismos de inovação aberta para a promoção da ambidir Dissertação (Mestrado)	No ambiente de negócios contemporâneo a i
Patrícia de Sá Freire	Gisely Jussily Tonello Martins	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251740">https://repositorio.ufsc.br/handle/123456789/251740</a>	Framework para a comunicação como ponte entre stakeholders da governa Tese (Doutorado)	O crescimento da urbanização impõe desafio
Márcelo Mascado	Fernando Ferrero Aguiar	2023	<a href="https://repositorio.ufsc.br/handle/123456789/249573">https://repositorio.ufsc.br/handle/123456789/249573</a>	Framework para criação e manutenção de bases de conhecimento para auto Tese (Doutorado)	A presente tese aborda a problemática da bu
Fernando José Spanhol	Rafael Santos da Silva	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251292">https://repositorio.ufsc.br/handle/123456789/251292</a>	Framework para desenvolvimento de competências digitais de enfermeiros Tese (Doutorado)	A prestação de serviços de Atenção Primária
Alexandre Augusto Biz	Luana Emmendorfer	2023	<a href="https://repositorio.ufsc.br/handle/123456789/254475">https://repositorio.ufsc.br/handle/123456789/254475</a>	Framework para sistema de inteligência turística suportado pela governanç Tese (Doutorado)	Os destinos turísticos estão cada vez mais de
Eduardo Juan Soriano-Sierra	Fabiola Provenzi	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251146">https://repositorio.ufsc.br/handle/123456789/251146</a>	Gestão do conhecimento e cultura de segurança: estudo de caso em uma inst Dissertação (Mestrado)	No âmbito deste estudo, no que tang
Édis Mafra Lapoli	Maria Carolina Milani Caldas Sart	2023	<a href="https://repositorio.ufsc.br/handle/123456789/251187">https://repositorio.ufsc.br/handle/123456789/251187</a>	Gestão feminina: a experiência de mulheres que se destacam na polícia civil Tese (Doutorado)	A entrada da mulher no mercado de trabalho

Fonte: Elaborado pelo autor (2024).

O *script* em Cypher para o Neo4j, destinado à inserção e interconexão de informações sobre dissertações e teses, encontra-se ilustrado na Figura 22. O propósito desse *script* é estruturar e criar vínculos entre as informações dos trabalhos acadêmicos e as informações do Congresso CIKI 2023.

Figura 22 – *Script* na linguagem Cypher com as informações de Teses e Dissertações

```

783 // Relaciona o Autor do trabalho com a Instituição
784 MATCH (a:Pessoa (name: 'Luciano Zamperetti Wolski'))
785 MATCH (i:Institution (name: 'Universidade Federal de Santa Catarina'))
786 MERGE (a)-[:Filiado]->(i);
787
788 // Relaciona a Pessoa orientador com a Instituição
789 MATCH (o:Pessoa (name: 'Alexandre Leopoldo Gonçalves'))
790 MATCH (i:Institution (name: 'Universidade Federal de Santa Catarina'))
791 MERGE (o)-[:Filiado]->(i);
792
793 // Insete o Trabalho
794 MERGE (:Trabalho {title:'Assistente virtual baseado em conhecimento no apoio a estudantes de curso de pós-graduação',
795           tipo: 'Dissertação',
796           DC_Date_issued: '2023',
797           DC_Identifier_URI: 'https://repositorio.ufsc.br/handle/123456789/251684',
798           DC_Description: 'O trabalho tem como objetivo o desenvolvimento de um Assistente Virtual baseado em conhecimento, aplicado ao atendimento de
demandas administrativas providas de alunos de curso de pós-graduação. A metodologia adotada para Design Science Research (DSR). Partindo
das revisões de literatura, foi possível identificar as principais ferramentas de chatbot utilizadas na solução de problemas no meio
acadêmico existente. Para a modelagem do conhecimento, foram realizadas entrevistas agregadas a outras técnicas de modelagem do conhecimento
e coleta de dados, tornando possível o desenvolvimento de uma ontologia, que foi incorporada a um sistema de chatbot baseado em interesse,
viabilizando assim, o Assistente Virtual. com o propósito de avaliação, o artifício foi disponibilizado para o uso pela comunidade
acadêmica, onde foram apresentados recursos, podendo ser aplicado a diversos setores como acadêmico, administrativo, marketing, produção,
entre outros, permitindo também a atualização e a inclusão de novos conhecimentos de forma constante. Os testes realizados apontaram um bom
desempenho na interação do usuário para com o Assistente Virtual, e na assertividade das respostas fornecidas pelo Assistente Virtual para o
usuário.'});
799
800 // Insete o Autor do trabalho
801 MERGE (:Pessoa (name: 'Lodacir Rodrigo Silva da Rosa'));
802
803 // Relaciona o Trabalho com o Autor do trabalho
804 MATCH (a:Pessoa (name: 'Lodacir Rodrigo Silva da Rosa'))
805 MATCH (t:Trabalho {title: 'Assistente virtual baseado em conhecimento no apoio a estudantes de curso de pós-graduação', tipo: 'Dissertação'})
806 MERGE (a)-[:Defendeu]->(t);
807
808 // Insete a Pessoa orientador
809 MERGE (:Pessoa (name: 'Fernando Alvaro Ostuni Gauthier'));
810

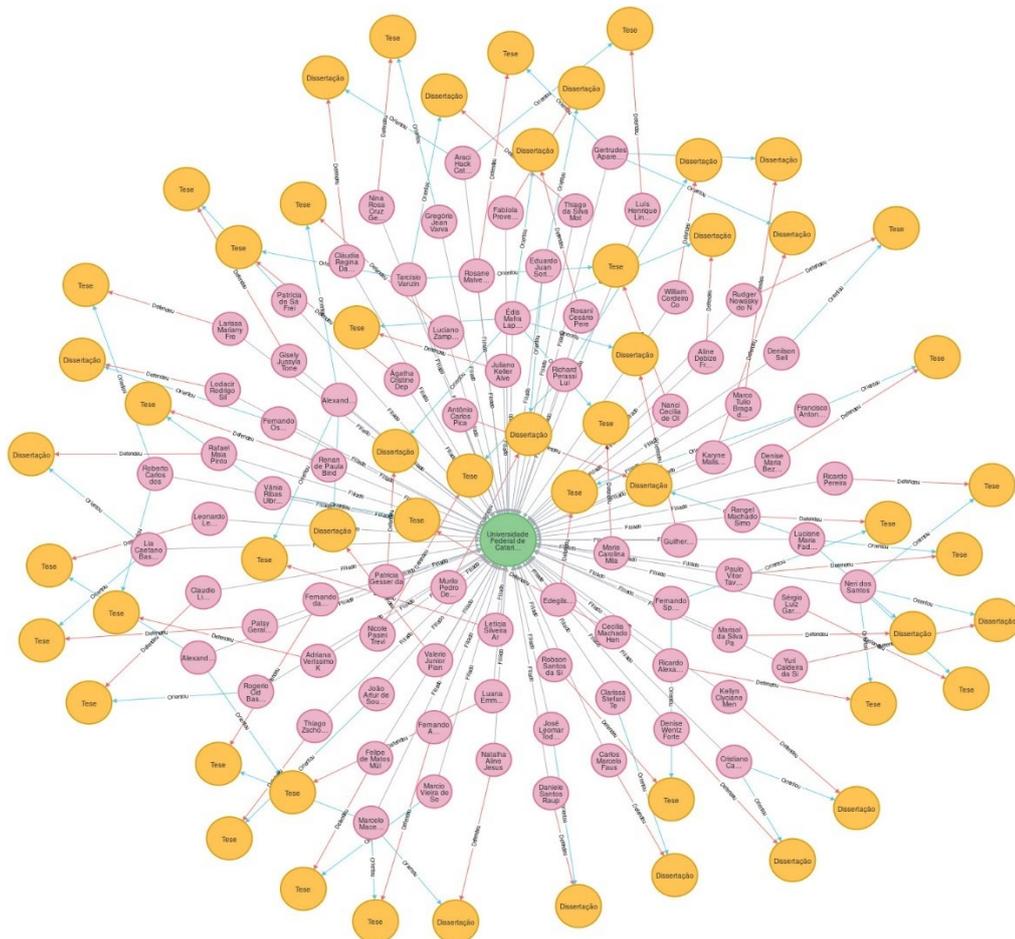
```

Fonte: Elaborado pelo autor (2024).

O grafo que exhibe os dados dos trabalhos de teses e dissertações coletados, acompanhados dos pesquisadores que os elaboraram e seus respectivos orientadores, é apresentado na Figura 23. No grafo, as cores dos nós também servem para diferenciar tipos de entidades e suas interações. A instituição UFSC é partilhada por todos os indivíduos no grafo, sendo representada na cor verde. Os nós em rosa indicam pessoas, que podem ser tanto alunos (Defendeu) quanto professores (Orientador). Os nós amarelos correspondem aos trabalhos apresentados, sejam eles, teses ou dissertações. Essa diferenciação por cores facilita a

compreensão das conexões e interações entre a instituição, os indivíduos envolvidos e os trabalhos acadêmicos realizados.

Figura 23 – Grafo de conhecimento de Teses e Dissertações do PPGEGC/UFSC



Fonte: Elaborado pelo autor (2024).

#### 4.4 GRAFO DE CONHECIMENTO AGREGADO

O processo de criação do grafo de conhecimento dos artigos do CIKI 2023 e do grafo de conhecimento de teses e dissertações do EGC/UFSC em 2023 com o uso do Ne4j resultou em um único grafo para armazenamento, que agrega as informações das duas fontes e possibilita analisar detalhadamente as relações entre as pessoas que escreveram artigos e aquelas que defenderam teses ou dissertações (Figura 24).

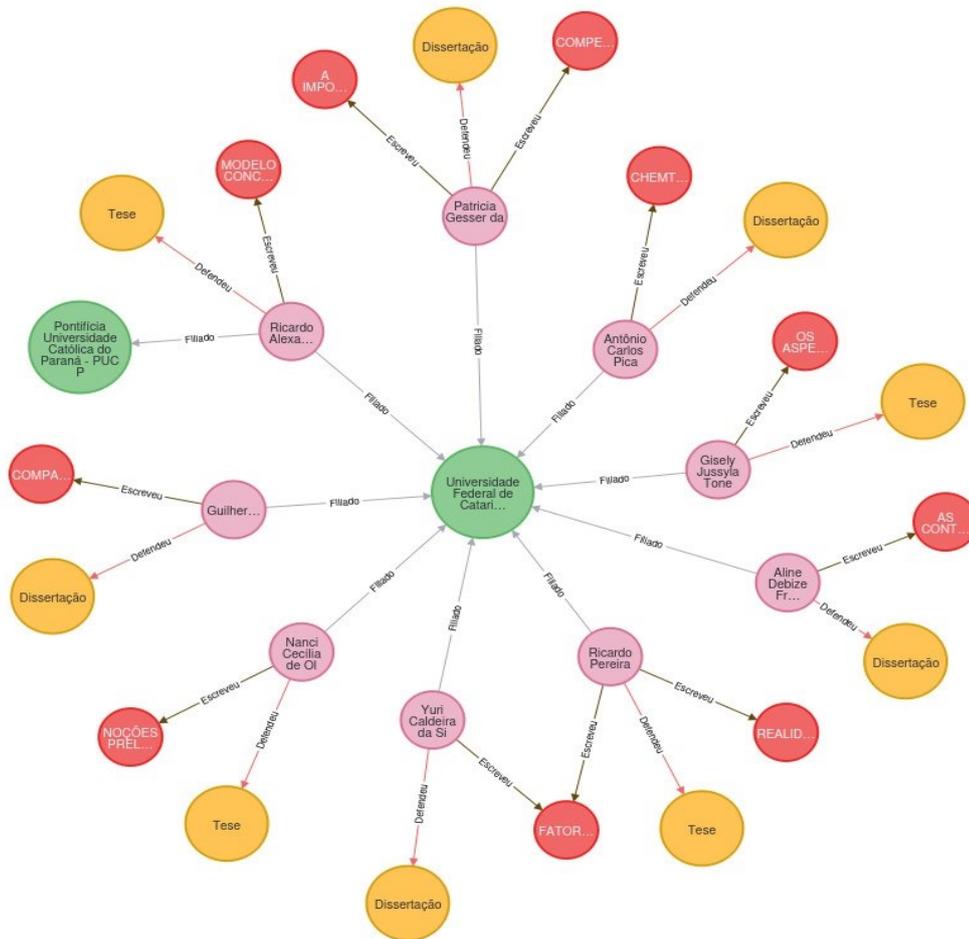
Figura 24 – Visão do grafo gerado pelo Neo4j com base nos dados de artigos CIKI 2023 e de Teses e Dissertações PPGEGC/UFSC



Fonte: Elaborado pelo autor (2024).

Ao agregar as informações, obtém-se uma visão unificada e coerente das contribuições acadêmicas, essencial para a extração de conhecimentos. Essa abordagem possibilita que pesquisadores e instituições visualizem redes de coautoria e compreendam melhor as dinâmicas de colaboração científica. Ao integrar essas informações em um grafo, facilita-se a identificação de padrões de colaboração, o que promove uma análise mais profunda das interações entre pesquisadores e instituições, bem como revela novas oportunidades para colaborações futuras e avanços científicos. Na Figura 25, são apresentadas as pessoas que publicaram artigos no CIKI 2023 e que defenderam teses ou dissertações na UFSC no mesmo ano.

Figura 25 – Visão do Neo4j: pessoas que escreveram artigos no CIKI 2023 e defenderam Teses/Dissertações em 2023



Fonte: Elaborado pelo autor (2024).

Ainda que o Neo4j seja um banco de dados orientado a grafos, a linguagem de consulta possibilita retornar resultados em formato de tabela ou grafo. Na tabela, cada linha representa um conjunto de propriedades de nós e relações. No grafo, as conexões entre nós são visualmente representadas. Essa flexibilidade é valiosa, visto que os usuários podem escolher o melhor formato de visualização baseado no contexto da consulta. Ao buscar dados sobre autores e suas contribuições acadêmicas, uma tabela facilita uma visão ordenada e rápida, enquanto um grafo visual ajuda a identificar padrões e conexões complexas de forma intuitiva.

No Quadro 9, é apresentado o resultado da consulta Cypher sobre pessoas que escreveram artigos no CIKI 2023 e defenderam teses ou dissertações na UFSC em 2023.

Quadro 9 – Pessoas: Artigo(s) e Tese/Dissertação

Pessoa	Artigo	Trabalho	Tipo
Aline Debize de Fraga	As contribuições da maturidade e experiência de carreira para o empreendedorismo sênior	Empreendedorismo sênior: experiências vividas e sentidas de pessoas que empreendem depois dos 50 anos de idade	Dissertação
Antônio Carlos Picalho	Chemtrails Over The Country Club de Lana Del Rey sob a ótica do Cinema Of Attraction de Tom Gunning	Das bibliotecas aos buscadores: testando técnicas avançadas para a recuperação da informação em pesquisas por documentos na <i>web</i>	Dissertação
Gisely Jussyla Tonello Martins	Os aspectos tecnológicos e de governança no contexto das cidades inteligentes	<i>Framework</i> para a comunicação como ponte entre <i>stakeholders</i> da governança multinível de cidades inteligentes	Tese
Guilherme Agnolin	Compartilhamento do conhecimento e inclusão social: estudo de caso de uma organização voluntária para pessoas com deficiência	Compartilhamento do conhecimento e inclusão social: estudo de caso de uma organização voluntária para pessoas com deficiência	Dissertação
Nanci Cecília de Oliveira Veras	Noções preliminares sobre o compartilhamento de conhecimento no campo semântico em Língua Brasileira de Sinais	Os atos da fala e a comunicação da intencionalidade no uso de metáforas na interpretação da Língua de Sinais – Libras	Tese
Patricia Gesser da Costa	A importância da inteligência competitiva aplicada às organizações de saúde: uma revisão integrativa da literatura Competências empreendedoras emergentes como fomento ao empreendedorismo em saúde estética: uma revisão integrativa	Competências empreendedoras emergentes: o caso de profissionais da saúde estética	Dissertação
Ricardo Alexandre Diogo	Modelo conceitual para formulação de diretrizes estratégicas para a educação em Engenharia	Modelo conceitual para formulação de diretrizes estratégicas na concepção e atualização de cursos de Engenharia no contexto da transformação digital	Tese
Ricardo Pereira	Fatores críticos de sucesso para internacionalização de empresas de base tecnológica: uma revisão integrativa Realidade virtual e metaverso no ensino de educação física: evidências e desafios para uma aprendizagem engajadora e inclusiva	PlatforMass: modelo conceitual para personalização em massa na Indústria 5.0, baseado em plataforma	Tese
Yuri Caldeira da Silva	Fatores críticos de sucesso para internacionalização de empresas de base tecnológica: uma revisão integrativa	Fatores críticos de sucesso para internacionalização de empresas de base tecnológica	Dissertação

Fonte: Elaborado pelo autor (2024).

A combinação dos grafos do CIKI 2023 e das Teses e Dissertações do PPGE/GC/UFSC de 2023 gerou importantes *insights* ao integrar diferentes conjuntos de dados acadêmicos. Com isso, foi possível identificar conexões antes invisíveis entre estudos e pesquisadores, o que

proporcionou uma visão mais completa do contexto acadêmico. A habilidade de mapear dados de diversas fontes e formatos é essencial para enriquecer as análises e facilitar a identificação de padrões, tendências e colaborações potenciais. Essa fusão não somente aprimora a qualidade dos dados, mas também expande as possibilidades de pesquisa e inovação.

Em sua análise sobre bancos de dados de grafos, Angles e Gutierrez (2008) destacam a importância da linguagem Cypher utilizada no Neo4j pela sua flexibilidade e capacidade expressiva na modelagem de dados complexos, bem como na execução de análises detalhadas. Por intermédio do Cypher, é possível realizar operações como correspondência de padrões, agregações e manipulação de caminhos, aspectos essenciais para se obter *insights* significativos de grandes volumes de dados inter-relacionados.

#### 4.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Por meio do desenvolvimento de um grafo de conhecimento baseado em dados extraídos do Congresso Internacional de Conhecimento e Inovação (CIKI 2023) e do Repositório Institucional da UFSC, demonstrou-se a eficácia do Neo4j na modelagem e análise de relações complexas entre elementos de conhecimento. Com a utilização das técnicas de *web scraping* e *Entity Resolution*, foi possível construir um grafo robusto e coerente, que integrou informações sobre artigos, teses e dissertações, facilitando a visualização das interconexões entre pesquisadores, instituições e suas produções acadêmicas.

Com a fusão dos grafos, obteve-se uma análise detalhada das contribuições acadêmicas, que revelou padrões de colaboração e possibilitou a identificação de novas oportunidades para parcerias futuras. A capacidade do Neo4j de gerenciar e visualizar dados complexos de forma interativa mostrou-se vantajosa para a compreensão das dinâmicas de criação e disseminação do conhecimento. A utilização da linguagem Cypher foi fundamental para a execução de consultas precisas e a obtenção de *insights* significativos com base nos dados inter-relacionados.

Em síntese, a integração dos dados do CIKI 2023 e do PPGEGC/UFSC 2023 em um único grafo de conhecimento não somente aprimorou a qualidade das análises, mas também expandiu as possibilidades de pesquisa e inovação. Essa abordagem multidimensional contribui para a identificação de lacunas no conhecimento, a exploração de novas áreas de estudo e o fomento de colaborações interdisciplinares, estabelecendo uma base sólida para futuros avanços científicos e tecnológicos. Esta pesquisa ilustra o potencial dos bancos de dados orientados a

grafos de enriquecer a infraestrutura de pesquisa e apoiar a tomada de decisões estratégicas no contexto acadêmico.

Os dados utilizados neste estudo estão disponíveis para acesso público na plataforma de compartilhamento de código da Universidade Federal de Santa Catarina (UFSC). Essa disponibilização visa promover a transparência e a reprodutibilidade da pesquisa, possibilitando que outros pesquisadores explorem e utilizem o grafo de conhecimento acadêmico em suas próprias investigações. Eles podem ser encontrados na seguinte URL: <https://codigos.ufsc.br/jeferson.mello/grafodeconhecimentoacademico>.

## CONSIDERAÇÕES FINAIS

Nos últimos anos, a Engenharia do Conhecimento passou por mudanças profundas. Inicialmente, os gestores das organizações focavam a coleta de informações de especialistas altamente qualificados e a transcrição em formatos digitais para serem armazenadas em sistemas de base de conhecimento. Este era um processo majoritariamente manual, que se concentrava na estruturação e organização do conhecimento por meio de esquemas ou ontologias, muitas vezes desconsiderando a integração de dados preexistentes.

Esse método exigia sistemas especializados e um alto investimento inicial. Atualmente, com uma maior disponibilidade de diferentes tipos de dados e avanços em gerenciamento de dados, aprendizado de máquina e colaboração coletiva, as bases de conhecimento agora agregam uma ampla variedade de informações. O uso dessas bases, principalmente sob a forma de grafos de conhecimento, tornou-se essencial em diversas aplicações e desempenha um papel crucial na difusão do conhecimento e na promoção da inovação acadêmica. Essa evolução fez das bases de conhecimento uma ferramenta indispensável na Engenharia do Conhecimento.

A complexidade crescente do conhecimento e dos sistemas baseados nele resultou no desenvolvimento de metodologias de Engenharia do Conhecimento, como os grafos de conhecimento. Diferentemente das bases de conhecimento tradicionais, os grafos de conhecimento representam informações de forma mais abrangente, utilizando um grande conjunto de declarações explícitas e ontologias mais simples e menores. A criação desses grafos é fruto da fusão de sistemas de representação de conhecimento com estratégias de gerenciamento de dados e métodos de busca aprimorados, o que promove uma melhor compreensão das informações em decorrência de sua natureza interconectada. Além disso, a aplicação de processos comerciais eficientes e inovadores proporciona soluções de alto nível.

Um dos grandes desafios da Engenharia do Conhecimento envolve a aquisição e a melhoria do conhecimento. É evidente que a complexidade da captura do conhecimento, a variedade das fontes, a confiabilidade das informações e a quantidade de conhecimento capturado para processar e organizar as informações de forma eficiente torna esse desafio incitador.

Para garantir a atualização permanente das informações, que mudam rapidamente, é preciso manter consistência dentro da base de conhecimento, proporcionando aos usuários acesso fácil e fidedigno. Nessa direção, o grafo de conhecimento exerce uma função importante na representação visual e estrutural das relações e colaborações no universo acadêmico.

Este estudo objetivou analisar os grafos de conhecimento quanto aos seus modelos, conceitos, ferramentas e contribuições para a Engenharia do Conhecimento. Para tanto, inicialmente, buscou-se expandir as informações sobre os principais conceitos e arquiteturas da *web* e fazer uma breve apresentação sobre as bases de dados de armazenamento do conhecimento, temas considerados importantes pilares para o entendimento dos grafos. Ainda, tratou de apresentar definições, modelos e propriedades dos grafos de conhecimento, características de representação e uso do conhecimento, bem como ferramentas, armazenamento, recuperação visual e operação de grafos de conhecimento.

Em seguida, propôs-se a elaboração passo a passo de grafos de conhecimento utilizando tanto dados do Congresso Internacional de Conhecimento e Inovação (CIKI) de 2023, organizado pelo Programa de Pós-Graduação em Engenharia, Gestão e Mídia do Conhecimento da Universidade Federal de Santa Catarina (PPGEGC/UFSC), como as teses e dissertações produzidas no EGC no mesmo ano. O objetivo era verificar a conexão entre os participantes do Congresso que escreveram algum artigo e aqueles que defenderam seus trabalhos de Pós-graduação no mesmo período.

A união de diferentes conjuntos de dados acadêmicos por meio da integração dos grafos do CIKI 2023 e das Teses e Dissertações do PPGEGC/UFSC resultou em *insights* significativos, por exemplo, ao identificar áreas de pesquisa emergentes no CIKI 2023 no processo de mapeamento e análise de interconexões entre artigos, autores, palavras-chave e temas discutidos. Quanto ao resultado obtido, no total, foram identificadas nove pessoas que participaram do Congresso, das quais cinco defenderam dissertações e quatro defenderam teses, sendo que duas dessas pessoas apresentaram dois artigos e as outras sete apenas um.

Com base nos resultados, pode-se apontar que, ao examinar a frequência e a coocorrência das palavras-chave, é possível rapidamente reconhecer os tópicos de maior destaque. Adicionalmente, o grafo pode monitorar as redes de colaboração entre pesquisadores e instituições, evidenciando a formação de novos grupos de pesquisa focados em temas inovadores. A análise de citações e referências sinaliza trabalhos influentes, que contribuem significativamente para o desenvolvimento de novas áreas de estudo. Com a inclusão de dados temporais, observa-se a evolução do interesse por certos tópicos ao longo do tempo. Ferramentas de visualização interativas, aliadas a algoritmos de análise de grafos, como detecção de comunidades e análise de centralidade, facilitam a identificação de padrões, destacando áreas emergentes com maior potencial.

Por fim, é crucial direcionar esforços para as futuras pesquisas, ampliando a amostra e diversificando os contextos estudados, a fim de garantir resultados mais representativos e

robustos. Além disso, a integração de novas tecnologias e metodologias inovadoras será fundamental para aprofundar a compreensão dos fenômenos observados. Os desafios mais importantes de serem enfrentados são tanto a obtenção de financiamento adequado quanto a superação de barreiras éticas, além da necessidade de manter a relevância científica em um cenário de mudanças rápidas. Nesse contexto, a colaboração interdisciplinar e internacional revela-se imprescindível para superar essas dificuldades e promover avanços significativos no campo de estudo. Buscar soluções para essas questões é fundamental para o progresso e a inovação na área, assegurando que as pesquisas continuem contribuindo de maneira significativa para o conhecimento científico.

## REFERÊNCIAS

- ABITEBOUL, S. *et al.* Ontologies, RDF, and OWL. *In: ABITEBOUL, S. et al. Web Data Management*. Cambridge: Cambridge University Press, 2011. p. 143-170.
- ABODAYEH, A. *et al.* Web Scraping for Data Analytics: A BeautifulSoup Implementation. *In: INTERNATIONAL CONFERENCE OF WOMEN IN DATA SCIENCE AT PRINCE SULTAN UNIVERSITY*, 6., 2023, Riyadh. **Proceedings** [...]. Riyadh: IEEE, 2023. p. 65-69. DOI: 10.1109/WiDS-PSU57071.2023.00025. Disponível em: <https://ieeexplore.ieee.org/document/10145369>. Acesso em: 30 ago. 2024.
- ABU-SALIH, B. Domain-specific knowledge graphs: A survey. **Journal of Network and Computer Applications**, v. 185, p. 103076, April 2021.
- ALLEMANG, D.; GANDON, F.; HENDLER, J. A. **Semantic web for the working ontologist effective modeling for linked data, RDFS, and OWL**. 3rd. ed. New York: ACM, 2020.
- ALLEN, B. P.; ILIEVSKI, F.; JOSHI, S. Identifying and Consolidating Knowledge Engineering Requirements. **arXiv preprint arXiv:2306.15124**, 2023.
- AMARAL, G. *et al.* Assessing The Quality of Sources in Wikidata Across Languages: A hybrid Approach. **Journal of Data and Information Quality**, v. 13, n. 4, p. 1-35, 2021.
- ANGLES, R.; GUTIERREZ, C. Survey of graph database models. **ACM Computing Surveys**, v. 40, n. 1, p. 1-39, 2008.
- ANGLES, R. The property graph database model. *In: ALBERTO MENDELZON INTERNATIONAL WORKSHOP ON FOUNDATIONS OF DATA MANAGEMENT*, 12., 2018, Cali. **Proceedings** [...]. Aachen: CEUR-WS, 2018.
- ANIKIN, D.; BORISENKO, O.; NEDUMOV, Y. Labeled Property Graphs: SQL or NoSQL? *In: IVANNIKOV MEMORIAL WORKSHOP*, 2., 2019, Veliky Novgorod. **Proceedings** [...]. New York: IEEE, 2019. p. 7-13.
- AUER, S. *et al.* **DBpedia: A Nucleus for a Web of Open Data**. 6. ed. Berlin: Springer, 2007. v. 4.825. p. 722-735.
- BARRASA, J.; HODLER, A. E.; WEBBER, J. **Knowledge Graphs: Data in Context for Responsive Businesses**. Sebastopol, CA: O'Reilly Media, 2021.
- BELLOMARINI, L.; SALLINGER, E.; VAHDATI, S. Knowledge Graphs: The Layered Perspective. *In: JANEV, V. et al. (ed.). Knowledge Graphs and Big Data Processing*. Cham: Springer, 2020. p. 20-34.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, p. 34-43, 2001.

BERTHOLD, M. R. *et al.* KNIME: The Konstanz information miner. *In: PREISACH, C. et al. (ed.). Studies in Classification, Data Analysis, and Knowledge Organization.* Heidelberg: Springer, 2008. p. 319-326.

BIZER, C. *et al.* Linked data on the web (LDOW2008). *In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 17., 2008, Beijing. Proceedings [...].* New York: ACM, 2008.

BLUMAUER, A.; NAGY, H. **The Knowledge Graph Cookbook: Recipes that Work-**mono\_monochrom.l. Wien: edition mono/monochrom, 2020.

BOERES, S. A. de A. *et al.* A Engenharia do Conhecimento e a Ciência da Informação. **Biblios**, Lima, n. 57, p. 59-66, 2014.

BOLLACKER, K. *et al.* Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge. *In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 8., 2008, Vancouver. Proceedings [...].* New York: ACM, 2008. p. 1.247-1.249.

BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D3 data-driven documents. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2.301-2.309, 2011.

BREITMAN, K. **Web Semântica: a internet do futuro.** Rio de Janeiro: LTC, 2010.

BUKHARI, S. N.; MIR, J. A.; AHMAD, U. Study and review of recent trends in semantic web. **International Journals of Advanced Research in Computer Science and Software Engineering**, v. 7, n. 6, p. 304-310, 2017.

CASTELLO BRANCO NETO, W. **Web Semântica na construção de sistemas de aprendizagem adaptativos.** 2006. 237 f. Tese (Doutorado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, SC, 2006.

CHAH, N. **Data Profiling, Machine Learning, and Data Visualizations for a Multilingual Crowdsourced Knowledge Graph: The Case of Wikidata.** 2022. 144 f. Tese (Doutorado em Filosofia) – Graduate Department of Faculty of Information, Universidade de Toronto, Toronto, 2022.

CHAH, N. OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google's Knowledge Graph. **arXiv preprint arXiv:1805.03885**, 2018.

CUDRÉ-MAUROUX, P. Leveraging Knowledge Graphs for Big Data Integration: The XI Pipeline. **Semantic Web**, v. 11, n. 1, p. 13-17, 2020.

DAS, S. *et al.* A tale of two graphs: Property graphs as RDF in Oracle. *In: INTERNATIONAL CONFERENCE ON EXTENDING DATABASE TECHNOLOGY, 17., 2014, Athens. Proceedings [...].* Konstanz: Open Proceedings, 2014. p. 762-773.

DAVENPORT, T. H.; PRUSAK, L. **Working knowledge: how organizations manage what they know.** Boston: Harvard Business School, 1998.

DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What Is Knowledge Representation? **AI MAGAZINE**, v. 14, n. 1, p. 17-33, 1993.

DIETRICH, J.; JONES, N.; WRIGHT, J. Using social networking and semantic web technology in software engineering: Use cases, patterns, and a case study. **The Journal of Systems and Software**, v. 81, n. 12, p. 2.183-2.193, 2008.

DO, T.-T.-T. *et al.* **Query-based Performance Comparison of Graph Database and Relational Database**. New York: ACM, 2022.

DONKERS, A. J. A.; YANG, D.; BAKEN, N. Linked data for smart homes: Comparing RDF and labeled property graphs. *In*: LINKED DATA IN ARCHITECTURE AND CONSTRUCTION WORKSHOP, 8., 2020, Dublin. **Proceedings** [...]. Aachen: CEUR-WS, 2020.

DRESCH, A.; LACERDA, D. P.; ANTUNES JÚNIOR, J. A.V. **Design Science Research: Método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2015.

EASLEY, D.; KLEINBERG, J. **Networks, crowds, and markets**. Cambridge: Cambridge University Press, 2010.

EDDAMIRI, S.; BENGHABRIT, A.; ZEMMOURI, E. RDF graph mining for cluster-based theme identification. **International Journal of Web Information Systems**, v. 16, n. 2, p. 223-247, 2020.

EHRLINGER, L.; WÖB, W. Towards a Definition of Knowledge Graphs. *In*: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS, 12.; INTERNATIONAL WORKSHOP ON SEMANTIC CHANGE & EVOLVING SEMANTICS, 1., Leipzig. **Joint Proceedings** [...]. Leipzig, 2016. v. 1.695.

EL-SEOUD, S. A.; EL-SOFANY, H. F.; KARAM, O. H. The semantic web architecture and its impact on E-learning systems development. **International Journal of Emerging Technologies in Learning**, v. 10, n. 5, p. 29-34, 2015.

ELMAGARMID, A. K.; IPEIROTIS, P. G.; VERYKIOS, V. S. Duplicate record detection: A survey. **IEEE Transactions on Knowledge and Data Engineering**, v. 19, n. 1, p. 1-16, 2007.

FÄRBER, M. *et al.* Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. **Semantic Web**, v. 9, n. 1, p. 77-129, 2018.

FENG, Z. **Formal Analysis for Natural Language Processing: A Handbook**. China: University of Science and Technology of China Press, 2023.

FENSEL, D. *et al.* **Knowledge Graphs: methodology, tools and selected use cases**. Austria: Springer International Publishing, 2020.

FERREIRA, J. A. **Wikis semânticos: da web para a web semântica**. São Paulo: Unesp, 2014.

FERRUCCI, D. *et al.* Building watson: An overview of the deepQA project. **AI Magazine**, v. 31, n. 3, p. 59-79, 2011.

FLEMMING, A. **Quality characteristics of data sources publishing linked data**. Thesis – Humboldt University of Berlin, Berlin, 2010.

FLÓREZ, J. E.; CARBÓ, J.; FERNÁNDEZ, F. A meta-tool to support the development of knowledge engineering methodologies and projects. **International Journal of Software Engineering and Knowledge Engineering**, v. 22, n. 8, p. 1.055-1.083, 2012.

FRISENDAL, T. **Graph Data Modeling or Nosql and Sql: Visualize Structure and Meaning**. New Jersey: TechnicsPub, 2016.

FÜRBER, C.; HEPP, M. Using SPARQL and SPIN for data quality management on the Semantic Web. *In: INTERNATIONAL CONFERENCE ON BUSINESS INFORMATION SYSTEMS*, 13., 2010, Berlin. **Proceedings** [...]. Berlin: Springer, 2010. p 35-46.

GELLING, E.; FLETCHER, G.; SCHMIDT, M. Bridging graph data models: RDF, RDF-Star, and property graphs as directed acyclic. **arXiv:2304.13097v1**, 2023. DOI <https://doi.org/10.48550/arXiv.2304.13097>.

GRIMM, S.; HITZLER, P.; ABECKER, A.: Knowledge Representation and Ontologies, Logic, Ontologies and Semantic Web Languages. *In: STUDER, R.; GRIMM, S.; ABECKER, A. (ed.). Semantic Web Services*, New York: Springer, 2007. p. 51-105.

HITZLER, P. A review of the semantic web field. **Communications of the ACM.**, v. 64, n. 2, p. 76-83, 2021. DOI <https://doi.org/10.1145/3397512>. Acesso em: 30 ago. 2024.

HITZLER, P.; KRÖTZSCH, M.; RUDOLPH, S. **Foundations of Semantic Web Technologies**. Florida: CRC Press, 2009.

HOFER, M. *et al.* Construction of Knowledge Graphs: Current State and Challenges. **Information**, v. 15, n. 8, p. 1-61, 2024. DOI <https://doi.org/10.3390/info15080509>.

HOFMANN, M.; KLINKENBERG, R. **RapidMiner: Data Mining Use Cases and Business Analytics Applications**. Boca Raton, FL: CRC Press, 2014.

HOGAN, A. The Semantic Web: Two decades on. **Semantic Web**, v. 11, n. 1, p. 169-185, 2020.

HOGAN, A. *et al.* Knowledge graphs. **ACM Computing Surveys**, v. 54, n. 4, 2021.

HUR, A.; JANJUA, N.; AHMED, M. A Survey on State-of-the-art Techniques for Knowledge Graphs Construction and Challenges ahead. *In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND KNOWLEDGE ENGINEERING*, 4., 2021, Laguna Hills. **Proceedings** [...]. New York: IEEE, 2021. p. 99-103.

IBRAHIM, A. K. Evolution of the Web: from Web 1.0 to 4.0. **Qubahan Academic Journal**, v. 1, n. 3, p. 20-28, 2021. Disponível em: <https://journal.qubahan.com/index.php/qaj/article/view/75>. Acesso em: 30 ago. 2024.

Ji, S. *et al.* A Survey on Knowledge Graphs: Representation, Acquisition and Applications. **IEEE Transactions On Neural Networks And Learning Systems**, v. 4, p. 1-27, 2021.

KARIM, F.; VIDAL, M. E.; AUER, S. Compacting frequent star patterns in RDF graphs. **Journal of Intelligent Information Systems**, v. 55, n. 3, p. 561-585, 2020.

KEJRIWAL, M. Knowledge Graphs: A Practical Review of the Research Landscape. **Information**, v. 13, n. 4, 2022. DOI <https://doi.org/10.3390/info13040161>.

KEJRIWAL, M.; KNOBLOCK, C. A.; SZEKELY, P. **Knowledge Graphs: Fundamentals, Techniques, and Applications**. Cambridge, MA: MIT Press, 2021.

KLYNE, G.; CARROLL, J. J.; MCBRIDE, B. Resource Description Framework (RDF): Concepts and Abstract Syntax. **W3C Recommendation**, Feb. 25, 2004. Disponível em: <https://www.w3.org/TR/rdf-concepts/>. Acesso em: 30 ago. 2024.

LE, N. L.; ABEL, M. H.; GOUSPILLOU, P. A Constraint-based Recommender System via RDF Knowledge Graphs. *In: INTERNATIONAL CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK IN DESIGN*, 26., 2023, Rio de Janeiro. **Proceedings** [...]. Compiègne: HUT, 2023. p. 849-854.

LEHMANN, J. *et al.* DBpedia: a large-scale, multilingual knowledge base extracted from Wikipedia. **Semantic Web**, v. 6, n. 2, p. 167-195, 2012.

LIU *et al.* Web of Scholars: A Scholar Knowledge Graph. *In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL*, 43., Xi'an. **Proceedings** [...]. New York: ACM, 2020.

MACHADO, L. M. O.; SOUZA, R. R.; SIMÕES, M. da G. Semantic web or web of data? A diachronic study (1999 to 2017) of the publications of Tim Berners-Lee and the World Wide Web Consortium. **Journal of the Association for Information Science and Technology**, v. 70, n. 7, p. 701-714, 2019.

MADNICK, S. E. *et al.* Overview and framework for Data and information quality research. **Journal of Data and Information Quality**, v. 1, n. 1, p. 1-22, 2009.

MCBRIDE, B. Jena: A semantic web toolkit. **IEEE Internet Computing**, v. 6, n. 6, p. 55-58, 2002.

MELNYK, I.; DOGNIN, P.; DAS, P. Knowledge Graph Generation From Text. *In: EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*, 2022, Abu Dhabi. **Findings**. Stroudsburg: ACL, 2022. p. 1.610-1.622.

MENDES P. N., MÜHLEISEN, H., BIZER, C. Sieve: linked data quality assessment and fusion. *In: JOINT EDBT/ICDT WORKSHOPS*, 2012, Berlin. **Proceedings** [...]. New York: ACM, 2012. p. 116-123. DOI <https://doi.org/10.1145/2320765.2320803>.

- MIKA, P. **Social networks and the semantic web**. New York: Springer, 2007.
- MILEA, V.; FRASINCAR, F.; KAYMAK, U. TOWL: A temporal web ontology language. **IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics**, v. 42, n. 1, p. 268-281, 2012.
- MITCHELL, R. **Web Scraping com Python: coletando mais dados na web moderna**. 2. ed. São Paulo: Novatec, 2019.
- MONTEIRO, J.; SÁ, F.; BERNARDINO, J. Experimental Evaluation of Graph Databases: JanusGraph, Nebula Graph, Neo4j, and TigerGraph. **Applied Sciences**, Basel, v. 13, n. 9, 2023.
- NAKASHOLE, N.; THEOBALD, M.; WEIKUM, G. Scalable knowledge harvesting with high precision and high recall. In: INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 4., 2011, Hong Kong. **Proceedings [...]**. New York: ACM, 2011. p. 227-236.
- NICOLETTI, M. do C.; HRUSCHKA JÚNIOR, E. R. **Fundamentos da Teoria dos Grafos para Computação**. 3. ed. Rio de Janeiro: LTC, 2018.
- NOVAK, J. D.; CAÑAS, A. J. Theoretical Origins of Concept Maps, How to Construct Them, and Uses in Education. **Reflecting Education**, v. 3, n. 1, p. 29-42, 2007.
- NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory, 2001.
- PACHECO, R. C. S. Coprodução em ciência, tecnologia e inovação: fundamentos e visões. In: PEDRO, J. M.; FREIRE, P. de S. (org.). **Interdisciplinaridade: Universidade e Inovação Social e Tecnológica**, Curitiba, v. 1, p. 21-62, 2016.
- PAN, J. Z. *et al.* **Exploiting Linked Data and Knowledge Graphs in Large Organizations**. Cham: Springer, 2017.
- PAPADAKI, M.-E.; TZITZIKAS, Y.; MOUNTANTONAKIS, M. A Brief Survey of Methods for Analytics over RDF Knowledge Graphs. **Analytics**, v. 2, n. 1, p. 55-74, 2023.
- PATEL, A.; DEBNATH, N. C.; BHUSHAN, B. **Semantic Web: Technologies Research and applications**. New York: Routledge, 2023.
- PAULHEIM, H. Knowledge Graph refinement: A survey of approaches and evaluation methods. **Semantic Web**, v. 8., p. 489-508, 2017.
- PAULHEIM, H.; RISTOSKI, P.; PORTISCH, J. **Embedding Knowledge Graphs with RDF2vec**. Cham: Springer Nature, 2023.
- PENG, C. *et al.* Knowledge Graphs: Opportunities and Challenges. **Artificial Intelligence Review**, v. 56, n. 11, p. 13.071-13.102, 2023.

PÉREZ, J.; ARENAS, M.; GUTIERREZ, C. Semantics and complexity of SPARQL. **ACM Transactions on Database Systems**, v. 34, n. 3, p. 1-45, 2009.

PIPINO, L. L. *et al.* Data quality assessment. **Communications of the ACM**, v. 45, n. 4, p. 211-218, April 2002. DOI <https://doi.org/10.1145/505248.506010>.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 8, n. 6, dez. 2007.

RAUTENBERG, S.; TODESCO, J. L.; GAUTHIER, F. A. O. Processo de desenvolvimento de ontologias: uma proposta e uma ferramenta. **Rev. Tecnol.**, Fortaleza, v. 30, n. 1, p. 133-144, jun. 2009.

RAUTENBERG, S.; TODESCO, J. L.; STEIAL, V. S. Ontologias de domínio no mapeamento de instrumentos da gestão do conhecimento e de agentes computacionais da Engenharia do Conhecimento: o estado da arte. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 163-182, maio/ago. 2010.

ROBINSON, I.; WEBBER, J.; EIFREM, E. **Graph Databases: New Opportunities for Connected Data**. 2nd. ed. Sebastopol, CA: O'Reilly Media, 2015.

ROCHADEL, W.; SOUZA, J. A.; DANDOLINI, G. A. Ontology And E-Government In The Context Of Knowledge Engineering. *In: INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS & TECHNOLOGY MANAGEMENT*, 13., 2016, São Paulo. **Proceedings** [...]. São Paulo: Tecsi USP, 2016. DOI 10.5748/9788599693124-13. Disponível em: <https://www.tecsi.org/contecsi/index.php/contecsi/13CONTECSI/paper/view/3913>. Acesso em: 16 dez. 2023.

SAKR, S.; PARDEDE, E. **Graph data management: techniques and applications**. Hershey: IGI Global, 2012.

SAUZA-BEDOLLA, J. *et al.* Industrial knowledge management tools applied to engineering education. **IFIP Advances in Information and Communication Technology**, v. 492, p. 3-12, 2016.

SCHREIBER, G. *et al.* **Knowledge Engineering and Management: The CommonKADS Methodology**. London: MIT Press, 2000.

SILVA, D. da; ZIVIANI, A.; PORTO, F. Aprendizado de máquina e inferência em Grafos de Conhecimento. *In: CAVALCANTI, M. C.; TRAINA, A. (org.). Simpósio Brasileiro de Banco de Dados, 7 a 10 de outubro de 2019*. Fortaleza: Sociedade Brasileira de Computação, 2019.

SIMSEK, U. *et al.* A Knowledge Graph Perspective on Knowledge Engineering. **SN Computer Science**, v. 4, n. 1, p. 1-16, 2022.

SIVARAJAH, U. *et al.* Critical analysis of Big Data challenges and analytical methods. **Journal of Business Research**, v. 70, p. 263-286, 2017.

- SOLIMAN, H. Deep learning based searching approach for RDF graphs. **PLoS ONE**, v. 15, n. 3, p. 1-16, 2020. DOI <https://doi.org/10.1371/journal.pone.0230500>.
- SPEER, R.; HAVASI, C. Representing general relational knowledge in concept net 5. *In*: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 8., 2012, Istanbul. **Proceedings** [...]. Paris: Elra, 2012. p. 3.679-3.686.
- STAAB, S.; STUDER, R. **Handbook on Ontologies**. 2nd. ed. New York: Springer, 2009.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge Engineering: Principles and methods. **Data and Knowledge Engineering**, v. 25, n. 1-2, p. 161-197, 1998.
- SUCHANEK, F. M. *et al.* SOFIE: A self-organizing framework for information extraction. *In*: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 18., 2009, Madrid. **Proceedings** [...]. New York: ACM, 2009. p. 631-640.
- SUCHANEK, F. M.; KASNECI, G.; WEIKUM, G. Yago: A core of semantic knowledge Unifying WordNet and Wikipedia. *In*: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 16., 2007, Banff. **Proceedings** [...]. New York: ACM, 2007. p. 697-706.
- TING, Z. The application of basic principles of knowledge engineering in knowledge management. **Trans Tech Publications**, v. 108-111, n. 1, p. 123-128, 2010.
- TOMASZUK, D.; ANGLES, R.; THAKKAR, H. PGO: Describing Property Graphs in RDF. **IEEE Access**, v. 8, p. 118355-118369, 2020.
- VUKOTIC, A.; WATT, N. **Neo4j in Action**. Shelter Island, NY: Manning Publications Co, 2015.
- W3C. **Semantic web layer cake**. 2007. Disponível em: <https://www.w3.org/2007/03/layerCake.svg>. Acesso em: 1 ago. 2019.
- WANG, Q. *et al.* Knowledge Graph Embedding: A Survey of Approaches and Applications. **IEEE Trans. Knowl. Data Eng.**, v. 29, p. 2.724-2.743, 2017.
- WANG, R. Y.; STRONG, D. M. Beyond Accuracy: What Data Quality Means to Data Consumers. **Journal of Management Information System**, v. 12, n. 4, p. 5-34, 1996. Disponível em: <https://www.jstor.org/stable/40398176?origin=JSTOR-pdf>. Acesso em: 30 ago. 2024.
- YAN, J. *et al.* A retrospective of Knowledge Graphs. **Front. Comput. Sci.**, Berlin, v. 12, n. 1, p. 55-74, fev. 2018. DOI <https://doi.org/10.1007/s11704-016-5228-9>.
- ZAVERI, A. *et al.* Quality Assessment for Linked Data: A Survey A Systematic Literature Review and Conceptual Framework. **Semantic Web**, v. 1, p. 1-5, 2012.
- ZHU, L. *et al.* A General Characterization of Representing Spatiotemporal Knowledge Graph based on OWL. **Earth Science Informatics**, v. 15, n. 1, p. 1-26, 2022.