



RESTAURAÇÃO DA BASE DE DADOS DE DISSERTAÇÕES E TESES DO PPGEKC UTILIZANDO GRAFOS DO CONHECIMENTO

Márcio Clemes¹;

Rogério Cid Bastos².

Abstract: *This article presents an innovative approach to the restoration and enhancement of the dissertation and thesis database of the Graduate Program in Knowledge Engineering and Management (PPGEGC) at the Federal University of Santa Catarina (UFSC), using knowledge graphs. The application of standard ontologies, such as FOAF and Dublin Core, together with a custom ontology, ensures efficient organization and retrieval of academic information. The paper discusses the knowledge graph structure, practical implementation challenges, and the expected benefits for academic data management, focusing on interoperability, scalability, and long-term sustainability.*

Keywords: *knowledge graphs; ontologies; academic databases; FOAF*

Resumo: Este artigo apresenta uma abordagem inovadora para a restauração e aprimoramento da base de dados de dissertações e teses do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC) da Universidade Federal de Santa Catarina (UFSC), utilizando grafos de conhecimento. A aplicação de ontologias padrão, como FOAF e Dublin Core, juntamente com uma ontologia customizada, garante a organização eficiente e a recuperação de informações acadêmicas. O artigo discute a estruturação do grafo de conhecimento, os desafios da implementação prática, e os benefícios esperados para a gestão de dados acadêmicos, com foco na interoperabilidade, escalabilidade e sustentabilidade a longo prazo.

Palavras-chave: grafos de conhecimento; ontologias; bases acadêmicas; FOAF.

1 INTRODUÇÃO

O gerenciamento de informações acadêmicas é uma tarefa crítica para instituições de ensino superior (Silva & Bax, 2023; Bittencourt et al., 2024; Silva et al., 2024). A eficiência na organização, preservação e recuperação desses documentos é essencial para o avanço do conhecimento e para o apoio à pesquisa.

As bases de dados acadêmicas tradicionais muitas vezes enfrentam desafios relacionados à interoperabilidade, atualização e recuperação de informações, o que pode limitar o acesso a dados relevantes e dificultar o uso eficiente desses recursos. A restauração e melhoria

¹ Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento - Universidade Federal de Santa Catarina (UFSC) Florianópolis - Brasil. ORCID: <https://orcid.org/0000-0003-1260-3901>. e-mail: marcio.clemes@ufsc.br

² Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento - Universidade Federal de Santa Catarina (UFSC) Florianópolis - Brasil. ORCID: <https://orcid.org/0000-0002-3775-1516>. e-mail: rogerio.bastos@ufsc.br



dessas bases de dados são, portanto, uma necessidade premente, e o uso de tecnologias avançadas, como grafos de conhecimento, surge como uma solução promissora.

Grafos de conhecimento são estruturas que permitem modelar dados complexos e suas inter-relações de forma intuitiva e flexível. Eles são especialmente úteis em contextos onde a integração e a interoperabilidade de dados são cruciais, como é o caso das bases de dados acadêmicas. Ao utilizar grafos de conhecimento, é possível criar uma rede de informações interconectadas que melhora significativamente a recuperação e o uso dos dados.

A aplicação de grafos de conhecimento na administração universitária oferece um potencial significativo para melhorar a eficiência, a tomada de decisões e a qualidade geral dos serviços educacionais. Por exemplo, grafos de conhecimento oferecem uma maneira eficiente de organizar e representar o vasto conhecimento presente em ambientes universitários. Eles permitem a criação de uma estrutura interconectada que representa entidades (como cursos, professores, alunos, departamentos) e suas relações de forma intuitiva e flexível. Isso facilita a compreensão holística da estrutura universitária e seus processos (Silva & Bax, 2023).

Na descoberta de relações, grafos de conhecimento tem possibilitado descobrir relações não óbvias entre diferentes elementos da universidade (Goulart et al., 2023). Por exemplo, podem ajudar a identificar conexões entre áreas de pesquisa, colaborações interdepartamentais potenciais, ou padrões de desempenho acadêmico. Essas descobertas podem informar decisões estratégicas sobre alocação de recursos, desenvolvimento de novos programas ou iniciativas de pesquisa.

Com relação ao processo de tomada de decisão (Filho, 2023), o uso de grafos de conhecimento permite uma representação rica e interconectada do conhecimento institucional, e podem auxiliar administradores universitários na tomada de decisões complexas. Por exemplo, podem ajudar na análise de impacto de mudanças curriculares, previsão de tendências de matrícula, abandono, formação etc ou até mesmo otimizar a alocação de recursos. Outras aplicações dizem respeito a

- ***Gestão de Currículos*** - grafos de conhecimento podem ser aplicados para mapear e analisar estruturas curriculares. Isso permite uma visualização clara das relações entre disciplinas, pré-requisitos e sequências de aprendizagem, facilitando a revisão e atualização de programas acadêmicos;
- ***Personalização da Experiência do Aluno*** - utilizando grafos de conhecimento, as universidades podem criar sistemas de recomendação mais sofisticados para os alunos. Isso pode incluir sugestões personalizadas de cursos, oportunidades



de pesquisa ou atividades extracurriculares baseadas no perfil acadêmico e interesses do aluno;

- **Gestão de Pesquisa e Colaborações** - grafos de conhecimento podem mapear as competências de pesquisa dentro da universidade, facilitando a formação de equipes interdisciplinares e a identificação de oportunidades de colaboração interna e externa. Isso pode melhorar significativamente a produtividade e o impacto da pesquisa universitária;
- **Eficiência Administrativa** - ao conectar diferentes aspectos da administração universitária em um grafo de conhecimento, é possível identificar ineficiências, redundâncias ou gargalos nos processos administrativos. Isso pode levar a melhorias na eficiência operacional e na qualidade dos serviços prestados;
- **Análise de Desempenho e Qualidade** - grafos de conhecimento podem ser usados para criar métricas mais abrangentes e contextualizadas para avaliar o desempenho institucional, considerando múltiplas dimensões e relações complexas entre diferentes aspectos da vida universitária;
- **Adaptação a mudanças** - em um ambiente educacional em constante evolução, os grafos de conhecimento oferecem a flexibilidade necessária para adaptar rapidamente os sistemas de informação às mudanças. Isso é particularmente valioso em tempos de transformação digital e inovações pedagógicas.

Para ilustrar o potencial do tema, é proposto, neste trabalho, a utilização de grafos de conhecimento para a restauração e melhoria da base de dados de dissertações e teses do PPGEGC. Para isso, serão empregadas ontologias amplamente reconhecidas, como FOAF (*Friend of a Friend*) e *Dublin Core*, bem como uma ontologia customizada desenvolvida especificamente para atender às necessidades do PPGEGC. Essa abordagem visa não apenas restaurar a base de dados existente, mas também aprimorá-la, garantindo maior acessibilidade e eficiência na recuperação de informações.

O trabalho está organizado em seis seções: esta introdução que mostra o potencial de uso de grafos de conhecimento no ambiente acadêmico. Em seguida é apresentada uma revisão sobre o tema Grafos de Conhecimento e Ontologias. Por fim, é apresentada uma aplicação na base de dados de dissertações e teses do Programa de pós-graduação em Engenharia e Gestão do Conhecimento da UFSC – Florianópolis – SC – Brasil. Ao final os principais resultados são discutidos.



2 REVISÃO DE CONCEITOS

Nesta seção são apresentados os conceitos de Ontologias e de Grafos de Conhecimento e destacado, como essas ferramentas podem ser úteis no processo de gestão universitária.

2.1 ONTOLOGIAS EM BASES DE DADOS ACADÊMICAS

Ontologias são essenciais na estruturação de grafos de conhecimento, pois definem os conceitos, as propriedades e as relações que serão representados no grafo. Elas fornecem uma base semântica que permite a interoperabilidade entre diferentes sistemas e facilita a organização de informações complexas (Gruber, 1993).

Em bases de dados acadêmicas, ontologias como FOAF e *Dublin Core* são amplamente utilizadas para padronizar a descrição de recursos, como teses e dissertações, garantindo que as informações sejam consistentes e facilmente recuperáveis (Miller, 2001; Wolf et al., 1998). A utilização dessas ontologias em conjunto com grafos de conhecimento permite uma organização mais eficiente e uma melhor exploração dos dados acadêmicos, promovendo maior visibilidade e acessibilidade para a comunidade científica.

2.2 FOAF E DUBLIN CORE

FOAF (*Friend of a Friend*) é uma ontologia que facilita a descrição de pessoas, suas atividades e suas conexões com outros indivíduos e objetos. Originalmente desenvolvida para modelar redes sociais na *web* semântica, FOAF pode ser adaptada para descrever autores e suas relações com as obras acadêmicas, fornecendo uma estrutura robusta para conectar informações sobre pesquisadores, seus trabalhos e suas colaborações (Brickley & Miller, 2010).

Dublin Core, por outro lado, é uma das ontologias mais amplamente utilizadas para a descrição de recursos de informação. Composta por um conjunto de 15 elementos essenciais, como título, autor, data e tipo de recurso, o *Dublin Core* é uma escolha natural para a catalogação de dissertações e teses em bases de dados acadêmicas. Sua ampla adoção e padronização garantem que os metadados descritos sejam compreensíveis e utilizáveis em diferentes contextos e sistemas (Wolf et al., 1998).

2.3 GRAFOS DE CONHECIMENTO

Os grafos de conhecimento têm se consolidado como uma ferramenta poderosa para a modelagem e integração de informações em diversas áreas, incluindo a *web* semântica, a



inteligência artificial e a gestão de dados acadêmicos (Hogan et al., 2021). Sua capacidade de representar entidades e suas relações de forma explícita e estruturada os torna ideais para ambientes onde a complexidade e a interconexão dos dados são características fundamentais.

No contexto acadêmico, os grafos de conhecimento oferecem várias vantagens sobre os sistemas tradicionais de banco de dados. Eles permitem a criação de redes semânticas que conectam documentos, autores, temas e outras entidades relevantes, facilitando a descoberta de novas relações e insights a partir dos dados disponíveis (Nickel et al., 2016). Além disso, os grafos de conhecimento suportam consultas complexas e flexíveis, o que é essencial em ambientes de pesquisa onde a necessidade de cruzamento de informações é constante.

2.4 APLICAÇÕES DE GRAFOS DE CONHECIMENTO EM EDUCAÇÃO

Grafos de conhecimento têm encontrado diversas aplicações no campo da educação, particularmente na organização e gestão de recursos educacionais, na análise de desempenho acadêmico e na personalização do aprendizado. Por exemplo, Sun et al., (2016) demonstram como grafos de conhecimento podem ser utilizados para organizar e recomendar recursos educacionais de forma eficiente, enquanto Alyahyan & Düşteğör (2020) discutem o uso de grafos para analisar e prever o desempenho acadêmico dos estudantes, identificando padrões que poderiam passar despercebidos em análises tradicionais.

Albreiki et al. (2024) exploram o potencial dos grafos de conhecimento para a detecção precoce de evasão escolar, utilizando dados acadêmicos e comportamentais para identificar alunos em risco. Essas aplicações sublinham o valor dos grafos de conhecimento como ferramentas para melhorar tanto a gestão de dados educacionais quanto o suporte à tomada de decisões em ambientes acadêmicos.

2.5 DISCUSSÃO TEÓRICA: GRAFOS DE CONHECIMENTO E ONTOLOGIAS

Grafos de conhecimento representam uma abordagem poderosa para modelar, integrar e interligar informações complexas. Eles têm sido amplamente utilizados em diversas áreas, desde a web semântica até a inteligência artificial, devido à sua capacidade de representar relações semânticas entre entidades de maneira estruturada e navegável (Hogan et al., 2021).

A força dos grafos de conhecimento reside na sua flexibilidade e na capacidade de lidar com dados heterogêneos. Ao contrário de bancos de dados relacionais tradicionais, que dependem de esquemas rígidos, os grafos de conhecimento permitem a modelagem de relações dinâmicas entre entidades, o que é crucial em contextos onde a estrutura dos dados pode evoluir



com o tempo (Nickel et al., 2016). Isso é particularmente relevante para bases de dados acadêmicas, onde novas pesquisas e descobertas frequentemente introduzem novas relações e conceitos.

Ontologias desempenham um papel central na construção de grafos de conhecimento, fornecendo um vocabulário controlado que define os tipos de entidades e as relações que podem existir entre elas. Gruber (1993) definiu uma ontologia como uma especificação explícita de uma conceitualização compartilhada, o que significa que uma ontologia formaliza a maneira como um domínio é estruturado e compreendido coletivamente. No contexto de grafos de conhecimento, ontologias como FOAF e *Dublin Core* garantem que os dados sejam consistentes e interoperáveis, facilitando a integração e o compartilhamento de informações entre diferentes sistemas (Miller, 2001; Weibel et al., 1998).

FOAF, por exemplo, é amplamente utilizado para modelar redes sociais e pode ser adaptado para descrever as relações entre autores e suas publicações. *Dublin Core*, por outro lado, é um padrão amplamente adotado para a descrição de metadados, especialmente em bibliotecas e repositórios digitais (Weibel et al., 1998). A utilização dessas ontologias no contexto do PPGEGC permite que as teses e dissertações sejam descritas de maneira rica e padronizada, facilitando a recuperação de informações e a descoberta de novos conhecimentos.

Além disso, a capacidade dos grafos de conhecimento de suportar consultas semânticas avançadas, como as oferecidas por SPARQL, permite que os usuários explorem os dados de maneiras complexas e não lineares. Isso é especialmente útil em ambientes acadêmicos, onde a descoberta de conhecimento muitas vezes depende da capacidade de cruzar informações de diferentes fontes e identificar padrões ocultos (Duda et al., 2022).

2.6 ESTRUTURA DO GRAFO DE CONHECIMENTO

A estrutura do grafo de conhecimento desenvolvido para a base de dados do PPGEGC segue um modelo de nós e arestas, onde cada nó representa uma entidade (como uma tese, um autor, ou uma instituição), e cada aresta representa uma relação entre essas entidades. Esse modelo permite a representação de dados complexos de maneira intuitiva e navegável, facilitando a exploração e a recuperação de informações.

2.6.1 Nós e Relações

- **Nós:** No grafo de conhecimento, cada dissertação ou tese é representada como um nó central. Este nó é conectado a outros nós que representam os metadados associados,



como o título, tipo, autor(es), local de publicação, data de publicação, URL e resumo. Além disso, autores são representados como nós separados, permitindo a criação de uma rede de colaboração que pode ser explorada para identificar conexões entre diferentes trabalhos.

- **Relações:** As arestas que conectam os nós representam as relações definidas pela ontologia. Por exemplo, uma aresta pode conectar um nó de autor a um nó de dissertação, indicando que o autor escreveu aquela dissertação. Da mesma forma, uma aresta pode conectar uma dissertação a uma instituição, indicando o local de publicação. Essas relações são fundamentais para a estrutura do grafo, pois permitem a navegação e a realização de consultas complexas, como "todos os trabalhos publicados por um determinado autor em um ano específico".

2.7 CONSULTAS E NAVEGAÇÃO

Uma das principais vantagens dos grafos de conhecimento é a capacidade de realizar consultas complexas de maneira eficiente. Por exemplo, o sistema pode ser consultado para identificar todas as dissertações defendidas em um determinado ano, ou para encontrar trabalhos que compartilhem o mesmo conjunto de autores. Além disso, o grafo permite a navegação intuitiva entre as diferentes entidades, facilitando a descoberta de novas relações e *insights*.

2.8 INTEROPERABILIDADE E EXPANSIBILIDADE

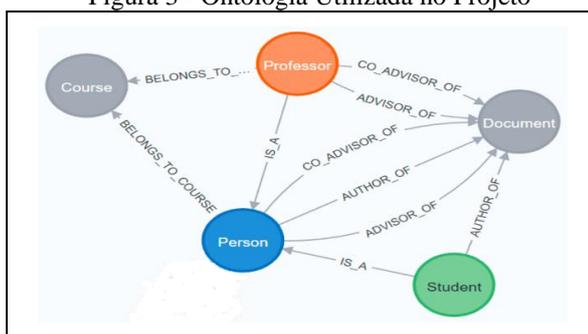
A estrutura do grafo foi projetada para ser interoperável com outros sistemas e bases de dados, utilizando padrões abertos como RDF (*Resource Description Framework*) e SPARQL (*SPARQL Protocol and RDF Query Language*). Isso garante que o grafo possa ser integrado a outras plataformas e utilizado em conjunto com outros sistemas de gestão de informação. Além disso, a estrutura é expansível, permitindo a adição de novos nós e relações à medida que a base de dados cresce ou que novas necessidades surgem.

3 APLICAÇÃO BASE DE DADOS DE DISSERTAÇÕES E TESES DO PPGE GC

A implementação prática da ontologia e do grafo de conhecimento proposto para a base de dados de dissertações e teses do PPGE GC envolve várias etapas críticas, cada uma delas projetada para garantir que o sistema atenda às necessidades específicas da instituição enquanto maximiza a eficiência e a usabilidade.

- **Data de Publicação:** Este campo indica a data em que o trabalho foi oficialmente publicado, o que é crucial para a organização cronológica da base de dados e para a análise temporal das produções acadêmicas.
- **URL:** Este campo fornece um link direto para o documento completo, permitindo acesso imediato ao trabalho original. A inclusão de URLs facilita a disseminação e o acesso aos trabalhos acadêmicos, promovendo uma maior visibilidade e impacto da pesquisa.
- **Resumo:** Este campo contém um breve resumo do conteúdo da dissertação ou tese, replicando o texto conforme encontrado no documento original. O resumo é um dos elementos mais importantes na descrição do conteúdo do trabalho, oferecendo uma visão geral que pode ser usada para avaliar a relevância do documento para diferentes consultas.

Figura 3 - Ontologia Utilizada no Projeto



Fonte: elaborada pelos autores.

A ontologia proposta foi projetada com flexibilidade para permitir futuras expansões e adaptações. À medida que novas necessidades surgirem ou que a base de dados evolua, novos elementos podem ser adicionados à ontologia, garantindo que ela continue atendendo às demandas do PPGEGC.

3.2 DESENVOLVIMENTO DO BACKEND

O primeiro passo na aplicação prática é o desenvolvimento do *backend*, que será responsável pela gestão do grafo de conhecimento. Esta camada do sistema foi construída utilizando tecnologias robustas e escaláveis, através do Neo4j (*Neo4j Graph Database*, 2024) bancos de dados orientados a grafos, que oferece suporte nativo para o armazenamento e consulta de grafos. O *backend* é responsável por gerenciar as operações ETL (*Extract, Transform e Load*) sobre o grafo de conhecimento, além de garantir a integridade e a consistência dos dados armazenados, oferecendo suporte as consultas através de *Cypher*.

3.3 INTEGRAÇÃO COM FONTES DE DADOS

Uma etapa crucial no desenvolvimento do sistema é a integração com as fontes de dados existentes. Analisando as fontes de dados disponíveis para obtenção dos registros referentes as informações sobre as teses e dissertações do PPGEGC disponíveis na UFSC, foi selecionada a que suporta a aplicação disponível em <https://repositorio.ufsc.br> e também a que armazena as informações referentes aos estudantes e professores da pós-graduação.

Dados de alunos e professores: informações obtidas através do Sistema de Controle Acadêmico da Pós-Graduação (CAPG). Os *scripts* para extração das informações estão disponíveis nas figuras 4 e 5. Os dados de professores foram extraídos e tratados manualmente, em função do pequeno número de registros. As figuras 6 e 7 apresentam alguns registros brutos e após o tratamento. A figura 8 apresenta o *script* de importação do arquivo diretamente para o Neo4j.

Figura 4 - *Script* de extração de professores

```

---
--- Professores
select distinct prs.nm_profes_prs as Nome, prs.nm_instituicao_prs as Instituicao,
prs.cd_situac_prs as CodSituacao, sit.cd_situac_pro as Situacao,
prs.cd_serpro_prs as Identificador
from capg..disciplinaPeriodo per per
join capg..vi_professor_prs_web2 prs on per.cd_serpro_prs = prs.cd_serpro_prs
where per.cd_curso_cur = 42001010 /* EGC */
and per.cd_sitdis_odi <> 3 /* cancelada */
order by prs.nm_profes_prs

```

Fonte: elaborada pelos autores. Comando SQL³

Figura 5 - *Script* de extração de alunos

```

---
--- Alunos
select distinct stu.nm_aluno as Nome, stu.nm_instituicao as Instituicao,
stu.cd_situac as CodSituacao, sit.cd_situac_pro as Situacao,
stu.cd_serpro as Identificador
from capg..disciplinaPeriodo per per
join capg..vi_aluno_prs_web2 stu on per.cd_serpro = stu.cd_serpro
where per.cd_curso_cur = 42001010 /* EGC */
and per.cd_sitdis_odi <> 3 /* cancelada */
order by stu.nm_aluno

```

Fonte: elaborada pelos autores. Comando SQL³

Figura 6 - Dados de professores sem tratamento

Nome	Instituicao	CodSituacao	Situacao
1 AIDES JOSÉ ROVER	UFSC	10	Estatut. RJUJ
2 ALDO VON WANGENHEIM	UFSC	10	Estatut. RJUJ
3 ALEXANDRE MARTINS RODRIGUEZ	UFSC	20	Agente Adm. Nomeado
4 Alexandra Augusta Pereira Klein	UFSC	6	Pós-Doutorando
5 Alexandre Augusto Biz	UFSC	10	Estatut. RJUJ
6 ALEXANDRE DE AVILA LEITE	UFSC	20	Agente Adm. Nomeado
7 ALEXANDRE LEOPOLDO GONCALVES	UFSC	20	Agente Adm. Nomeado
8 ALEXANDRE LEOPOLDO GONCALVES	UFSC	20	Agente Adm. Nomeado
9 ALICE THERESINHA CYRUS PEREIRA	UFSC	80	Inativo
10 ALINE FRANCA DE ABREU	UFSC	80	Inativo
11 ANA MARIA BENCIOVENI FRANZONI	UFSC	10	Estatut. RJUJ
12 Andrea Valéria Steil	UFSC	2	Participante
13 Andrea Valéria Steil	UFSC	10	Estatut. RJUJ
14 ANDREA VALERIA STEIL	UFSC	20	Agente Adm. Nomeado
15 ANDREA LUCAS SOARES	Universidade do Porto	5	Comediado
16 ARACI HACK CATAPIAN	UFSC	80	Inativo
17 ARACI HACK CATAPIAN	UFSC	60	Substituto Lei 8745/93
18 ARIAN REY T. MORALES	UFSC	2	Participante
19 ARIAN REY T. MORALES	UFSC	80	Inativo
20 BRUNO HARTMUT KOPITZKE	UFSC	80	Inativo
21 BRUNO HARTMUT KOPITZKE	UFSC	80	Inativo

Fonte: elaborada pelos autores.

Figura 7 - Dados de professores com tratamento

Nome	Instituicao	CodSituacao	Situacao
1 Aires Jose Rover	UFSC	10	Estatut. RJUJ
2 Aldo Von Wangenheim	UFSC	10	Estatut. RJUJ
3 Alejandro Martins Rodriguez	UFSC	20	Agente Adm. Nomeado
4 Alexandra Augusta Pereira Klein	UFSC	6	Pós-Doutorando
5 Alexandre Augusto Biz	UFSC	10	Estatut. RJUJ
6 Alexandre De Avila Leite	UFSC	20	Agente Adm. Nomeado
7 Alexandre Leopoldo Gonçalves	UFSC	80	Inativo
8 Alexandre Leopoldo Gonçalves	UFSC	10	Estatut. RJUJ
9 Alice Theresinha Cyrus Pereira	UFSC	80	Inativo
10 Aline Franca De Abreu	UFSC	80	Inativo
11 Ana Maria Bencioveni Franzoni	UFSC	10	Estatut. RJUJ
12 Andrea Valéria Steil	UFSC	10	Estatut. RJUJ
13 Antônio Lucas Soares	UFSC	5	Comediado
14 Araci Hack Catapan	Universidade do Porto	20	Agente Adm. Nomeado
15 Arac Hack Catapan	UFSC	80	Inativo
16 Arac Hack Catapan	UFSC	60	Substituto Lei 8745/93
17 Bruno Hartmut Kopitzke	UFSC	80	Inativo
18 Bruno Requião Da Cunha	UFSC	5	Comediado
19 Carlos Augusto Monguilhott Remor	UFSC	2	Participante
20 Carlos Henrique Berg	UFSC	80	Inativo
21 Christianne Coelho De Souza Reimisch	Universidade Federal de Santa Catarina	5	Comediado
22 Cláudia Stefani Teixeira	UFSC	2	Participante
23 Cláudia Stefani Teixeira	UFSC	10	Estatut. RJUJ

Fonte: elaborada pelos autores.

Dados de dissertações e teses: os dados foram importados diretamente do database do Dspace que suporta a aplicação disponível em <https://repositorio.ufsc.br> para o Neo4j. O *script* de extração e carga está disponível na figura 9, sendo esta a etapa com menores dificuldades a serem superadas. Os metadados foram analisados, e aqueles com maior relevância para o projeto selecionados.

³ Cedido por Euclides Pinheiro de Melo Analista de Tecnologia da Informação da Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação da UFSC, desenvolvedor do CAPG



especialmente úteis para gestores e pesquisadores que precisam de insights detalhados sobre a produção científica do programa.

4 CONCLUSÕES E TRABALHOS FUTUROS

A implementação de grafos de conhecimento para a restauração e aprimoramento da base de dados de dissertações e teses do PPGE GC representa um avanço significativo na forma como os dados acadêmicos são organizados, acessados e utilizados. Ao adotar ontologias padrão como FOAF e *Dublin Core*, juntamente com uma ontologia customizada, o projeto não apenas garante a interoperabilidade dos dados, mas também promove uma estrutura flexível que pode ser adaptada conforme novas necessidades surgirem.

A aplicação prática deste projeto, conforme discutido nas seções anteriores, demonstra como a combinação de tecnologias emergentes, como bancos de dados orientados a grafos, com práticas de engenharia do conhecimento, pode resultar em sistemas mais robustos e eficientes para a gestão de informações acadêmicas. A integração dessas tecnologias com sistemas existentes, como o Pergamum, e a adoção de padrões abertos, como RDF e SPARQL, asseguram que o sistema será tanto escalável quanto interoperável, capaz de crescer junto com as necessidades do PPGE GC.

No entanto, a implementação de um sistema como este não está isenta de desafios. Um dos principais desafios é a curadoria inicial dos dados, que requer um esforço significativo para garantir que todos os metadados sejam capturados de maneira completa e precisa. Além disso, a adoção de novas tecnologias pode exigir treinamento e capacitação para os usuários finais, garantindo que eles possam utilizar o sistema de forma eficaz.

Outro ponto a ser considerado é a sustentabilidade do projeto a longo prazo. Isso inclui a manutenção contínua do grafo de conhecimento, atualização da ontologia conforme novas categorias ou relações sejam identificadas, e a integração de novas ferramentas ou funcionalidades à medida que a tecnologia evolui. A colaboração entre a equipe técnica e os pesquisadores do PPGE GC será essencial para garantir que o sistema continue a atender às necessidades acadêmicas e administrativas de maneira eficaz.

A implementação deste projeto no PPGE GC servirá como um modelo para outras instituições de ensino superior que enfrentam desafios semelhantes na gestão de suas bases de dados acadêmicas. Ao compartilhar as lições aprendidas e as melhores práticas desenvolvidas ao longo deste projeto, esperamos contribuir para o avanço do conhecimento na área de gestão de dados acadêmicos e incentivar a adoção de tecnologias inovadoras em outros contextos.



Como sugestão de trabalhos futuros, alguns pontos foram identificados durante o desenvolvimento do trabalho, destacando-se os seguintes:

- Os dados migrados apresentaram inconsistência, destacando-se que uma obra pertence a outro programa de Pós-Graduação e outra não está catalogada como tese ou dissertação, além de inconsistência nos nomes dos autores, o que requer uma análise da origem dos dados que alimentaram o repositório (<https://repositorio.ufsc.br>) e o consequente processo de carga das informações;
- O desenvolvimento de uma camada de aplicação, que permita explorar os recursos disponibilizados pelas consultas SPARQL e Cypher;
- Incluir novos tipos de documentos no modelo, principalmente os artigos dos autores e professores;
- Expandir as definições das relações entre os documentos para identificar aqueles com temas, assuntos e outras características em comum, facilitando o processo de revisão bibliográfica e outros *insights*.
- Ampliar o uso do sistema para todos os programas de Pós-Graduação da UFSC.

AGRADECIMENTOS

Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Capes) – Código financeiro 001.

REFERÊNCIAS

- Albreiki, B., Habuza, T., Palakkal, N., & Zaki, N. (2024). Clustering-based knowledge graphs and entity-relation representation improves the detection of at risk students. *Education and Information Technologies*, 29(6), 6791–6820. <https://doi.org/10.1007/s10639-023-11938-8>
- Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Bittencourt, O. T., Carvalho, R. L. de, Barbosa, G. V., & Santos, G. F. do. (2024). SEGURANÇA CIBERNÉTICA NAS UNIVERSIDADES: UMA REVISÃO SISTEMÁTICA DA LITERATURA SOBRE A GESTÃO DE SEGURANÇA DA



- INFORMAÇÃO NO ENSINO SUPERIOR. *InterSciencePlace*, 19.
<https://www.interscienceplace.org/index.php/isp/article/view/723>
- Brickley, D., & Miller, L. (2010). *FOAF Vocabulary Specification 0.98*.
<http://xmlns.com/foaf/spec/>
- Duda, R. F., Barcelos, B. O., & Gauthier, F. A. O. (2022). ONTOLOGIAS NO DOMÍNIO DE LEARNING ANALYTICS: CLASSIFICAÇÃO E ÁREAS DE APLICAÇÃO. *Anais do Congresso Internacional de Conhecimento e Inovação – ciki*, 1(1), Artigo 1.
<https://doi.org/10.48090/ciki.v1i1.1325>
- Expanding APIs with Knowledge Graphs – Assault By Logic*. (2021, abril 26).
<http://johnsteinmetz.net/expanding-apis-with-knowledge-graphs/>
- Filho, J. F. P. (2023). Information Management Applied to Police Sciences: A perspective of organizational culture and innovation for public safety. *Advanced Notes in Information Science*, 4. <https://doi.org/10.47909/anis>
- Goulart, E. S. de Á., Monteiro, R. L. S., Carneiro, T. K. G., & Santos, B. B. L. dos. (2023). Modelagem de dados abertos governamentais para identificação de redes de escolas similares. *Anais do Workshop de Informação, Dados e Tecnologia - WIDaT*, 6.
<https://doi.org/10.22477/vi.widat.55>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge Graphs. *ACM Comput. Surv.*, 54(4), 71:1-71:37.
<https://doi.org/10.1145/3447772>
- Miller, E. J. (2001). An Introduction to the Resource Description Framework. *Journal of Library Administration*, 34(3–4), 245–255. https://doi.org/10.1300/J111v34n03_04
- Neo4j Graph Database*. (2024, novembro 6). Graph Database & Analytics.
<https://neo4j.com/product/neo4j-graph-database/>



- Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1), 11–33. Proceedings of the IEEE. <https://doi.org/10.1109/JPROC.2015.2483592>
- Silva, A. M. de F., Viegas, C. M. de A. R., Campos, D., Silva, L. M. S., Aragão, L. T. A., & Neiva, R. (2024). As inteligências artificiais aplicadas à educação híbrida. *Cuadernos de Educación y Desarrollo*, 16(5), e4133–e4133. <https://doi.org/10.55905/cuadv16n5-016>
- Silva, E. O., & Bax, M. P. (2023). Produção de Grafos de Conhecimento e Estratificação de Riscos em Saúde Mental a partir de Dicionários de Dados Semânticos. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 16. <https://revistas.ancib.org/index.php/tpbci/article/view/618>
- Sun, K., Liu, Y., Guo, Z., & Wang, C. (2016). *Visualization for Knowledge Graph Based on Education Data*. 10(3).
- Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). *Dublin core metadata for resource discovery*. <https://www.rfc-editor.org/rfc/rfc2413>
- Wolf, M., Kunze, J. A., Lagoze, C., & Weibel, S. (1998). *Dublin Core Metadata for Resource Discovery* (Request for Comments RFC 2413). Internet Engineering Task Force. <https://doi.org/10.17487/RFC2413>