

Categorização de tesouros e enriquecimento semântico: um estudo aplicado no Thesa

Kamila de Andrade Moura (UFSC)

Rita do Carmo Ferreira Laipelt (UFRGS)

Thiago Henrique Bragato Barros (UFRGS/UFSC)

Rene Faustino Gabriel Junior (UFRGS)



- **Tesauros:** instrumentos essenciais e consolidados na Organização do Conhecimento (OC).
- **Normas ISO, NISO e modelo SKOS** orientam seu desenvolvimento e adaptação às demandas contemporâneas.



- **Ampliação de seu papel:** de ferramenta de normalização terminológica para modelo conceitual.
- **Foco evolutivo:** enriquecimento semântico por meio das relações associativas.

- **Cenário do estudo:** Sistema Thesa – Tesouro semântico aplicado – Estudo de viabilidade para migração para versão 2.

Questão norteadora: *como categorizar os tesouros registrados no Thesa a fim de definir quais estão aptos a compor a base do Thesa 2.0?*

Objetivo: estabelecer critérios para a categorização dos tesouros armazenados na base do Thesa, visando a migração de dados para a nova versão do sistema.



- Linguagens terminológicas estruturadas, compostas por conceitos, termos e suas relações — **hierárquicas, associativas e de equivalência** — que vinculam os conceitos pelo significado. Essas relações são consideradas essenciais para sua existência e funcionamento (Gomes, 1990; Currás, 1995; Clarke, 2001; Laipelt, 2015; Hjørland, 2015; Moreira, 2019b).



- Instrumentos **interpretativos e mediadores** (Mazzochi, 2018), cuja construção enriquece a modelagem conceitual do domínio, contribuindo para a geração de conhecimento (Moura, 2024).

- Software para **elaboração de tesouros**, desenvolvido pelo ORCALAB e PPGCIN/UFRGS.
- Baseado em normas **ISO 25964** (1 e 2), **NISO Z39.50** e modelo **SKOS**.
- *Open source*, flexível e interoperável, com suporte a RDF e Linked Data.



- **Ferramentas de exportação:** glossários, índices, mapas conceituais, RDF, XML, JSON.
- Até **janeiro/2025**: 1.128 registros, 62.471 conceitos e 79.040 termos; utilizado por 41 instituições de ensino.

Thesa (continuação)



VIII ISKO BRASIL 2025



Tipos de relações de **equivalência** (10) e **associativas** (14) oferecidas pelo Thesa



Relações de equivalência	Relações associativas
Plural de	Ação ou objeto/contra agente
Extenso de	Ação/paciente ou objetivo
Abreviatura de	Ação/propriedade
Notação (código)	Ação/Produto
Sigla	Afinidade
variação de	Campo de estudo/objeto ou fenômeno
É traduzido por	Causa/Efeito
É gerúndio de	Características do produto
É feminino de	Conceito ou objeto/propriedade
É flexão verbal	Coordenação
	Matéria-prima/Produtos
	Objeto/Suas partes
	Oposição/Antonímia
	Processo/Parte de

Fonte: Thesa, 2025.



- Criação de **propriedades personalizadas** para descrição de domínios.
 - **Ligação entre tesauros** (internos e externos) via RDF.
 - **Importação** de dados de sistemas como TemaTres, SKOSMOS (XML/RDF).
 - Novos formatos de **exportação**: XML, RDF, TURTLE, CSV, JSON.
-
- 
- Implementação de **API REST** para **integração** com outros sistemas.
 - Inclusão de **múltiplos termos preferidos** (idiomas diversos).
 - Suporte a **elementos multimídia**: imagens e vídeos.
 - Potencial para **inovação e novas perspectivas** em pesquisa e ensino (Gabriel Júnior; Laipelt, 2023).

- Caracterização: pesquisa **aplicada**, de abordagem **quali-quantitativa**, **exploratória** e **descritiva**.
- **Análise documental** baseada em planilha extraída do banco de dados do Thesa – identificação de propriedades nos registros para categorização.
- Período de análise: **abril de 2024 a janeiro de 2025**.
- Total de registros analisados: **1.128**.



Propriedades de análise

- 1) **Tipo de registro:** Controle de autoridade de pessoas e instituições, Dicionário, Glossário, Ontologia, Tesouro ou Tesouro semântico;
- 2) **Situação:** Completo, Incompleto ou Vazio;
- 3) **Definições:** Sim/Não;
- 4) **Relações hierárquicas:** Sim/Não;
- 5) **Relações de equivalência:** Sim/Não;
- 6) **Relações associativas:** Sim/Não;
- 7) **Termos:** número de termos;
- 8) **Conceitos:** número de conceitos;
- 9) **ID Tesouro:** código de identificação junto ao Thesa;
- 10) **Nome Tesouro:** título do tesouro;
- 11) **Autoria:** nome do autor principal;
- 12) **Email:** contato do autor principal.



- **Acesso direto** aos 1.128 registros disponíveis no Thesa.
- Coleta dos dados via **navegação e verificação dos campos** (configurações e listagem alfabética).
- Exportação de **relatórios**: grafos, apresentações sistemática e alfabética, glossário e fichas terminológicas.
- Análise da **estrutura**: presença de definições terminológicas e uso das relações semânticas.



- **Preenchimento e categorização** dos registros em planilha Excel - Uso de **filtros do Excel** para análise quantitativa e categorização.
- Recorte focado em “**Tesouro**” e “**Tesouro semântico**”.
- **Tesauros completos**: termos + conceitos + definições + relações hierárquicas, associativas e de equivalência - **apenas um critério metodológico** — a utilidade do tesouro depende de seu contexto e finalidade.

- Tesouros classificados em 3 categorias:

Completo: definições + relações hierárquicas, associativas e de equivalência.

Incompleto: possui alguns dos atributos, mas não todos.

Vazio: sem atributos essenciais e/ou sem termos e conceitos; não apto à migração.



- Análise voltada para o uso das **relações associativas** sob a perspectiva da quantidade de uso e diversidade de tipos empregados na construção dos tesouros.
- Verificação de **temáticas semelhantes** nos tesouros para identificar a possibilidade de interligação entre os instrumentos.
- **Não foi avaliada**, neste estágio, a **correção das definições e das relações** — foco principal foi a identificação dos tesouros aptos à migração.

Situação X Ocorrência

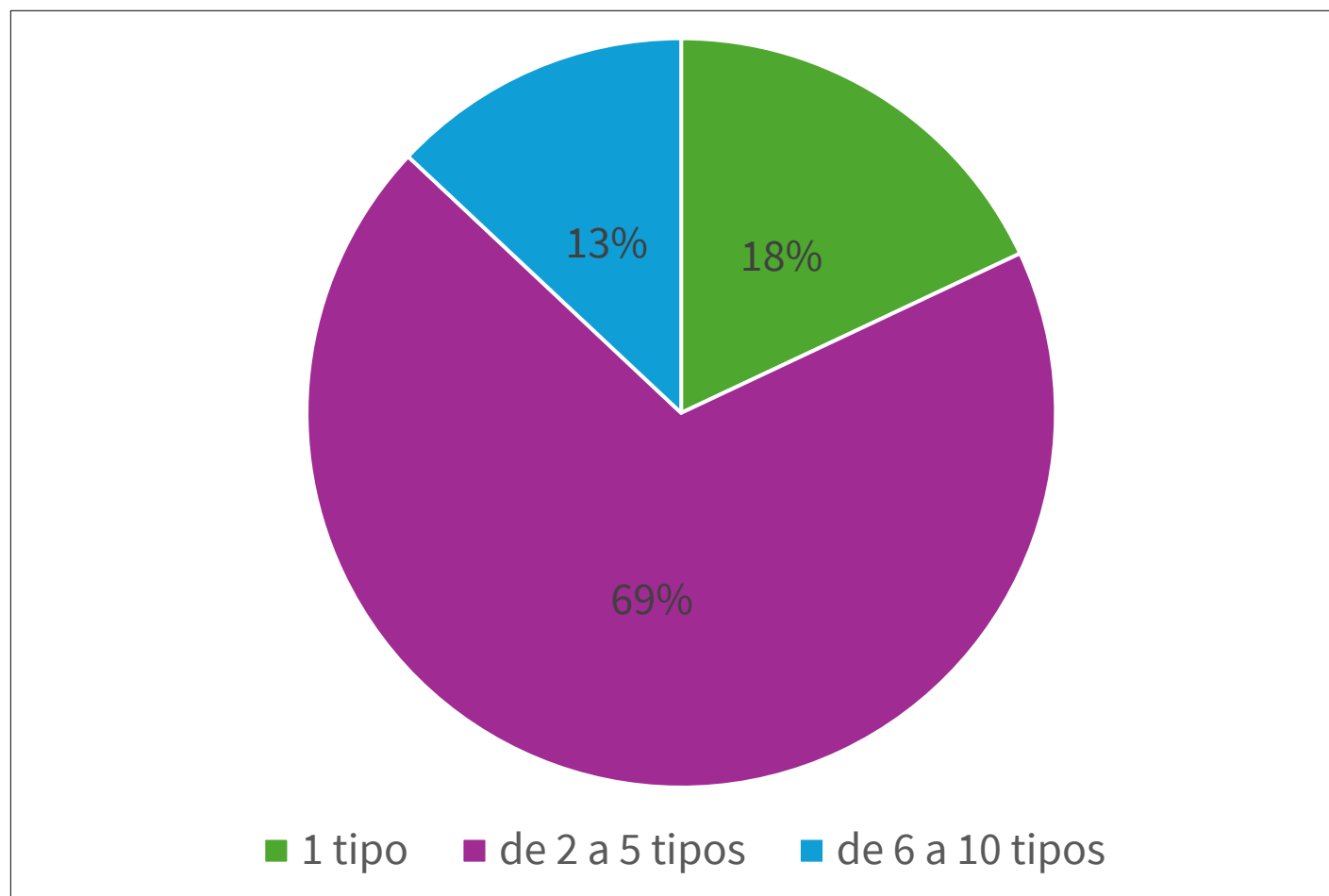
190 tesouros completos



Tipo de registro	Situação	Ocorrência
Tesouro	Completo	80 registros
	Incompleto	197 registros
	Vazio	224 registros
Tesouro semântico	Completo	110 registros
	Incompleto	243 registros
	Vazio	121 registros

Fonte: Elaborado pelos autores, 2025.

Uso de diversos tipos de relações associativas



Fonte: Elaborado pelos autores, 2025.

- 69% dos tesouros analisados utilizam de **2 a 5 tipos de relações associativas**; 13% usam de 6 a 10 tipos e 18% utiliza 1 tipo.
- Resultados **sugerem boa utilização do sistema**, com potencial consistência, **a ser confirmada em análises futuras**.



- Uso consistente das relações associativas **fortalece a semântica** dos tesouros.
- Entretanto, múltiplas relações não garantem, sozinhas, consistência conceitual — é necessária **avaliação da qualidade** das definições e das relações.

Relações associativas
mais utilizadas



Relações associativas	Frequência de uso
Afinidade	76%
Processo/Parte de	57%
Oposição/Antonímia	41%
Coordenação	31%
Ação/Produto	24%
Causa/Efeito	23%
Objeto/Suas partes	19%
Características do produto	19%
Ação/propriedade	18%
Conceito ou objeto/propriedade	12%
Ação/paciente ou objetivo	8%
Matéria-prima/Produtos	7%
Ação ou objeto/contra agente	5%
Campo de estudo/objeto ou fenômeno	5%

Fonte: Elaborado pelos autores, 2025.

- **Todos os tipos de relações associativas** do Thesa foram aplicados nos tesauros analisados.
- Predomínio das relações: *afinidade, processo/parte de, oposição/antonímia*.
 - Tesauros completos no Thesa **indicam um caminho evolutivo promissor**, mas confirmação exige análises qualitativas mais aprofundadas.
 - A evolução dos tesauros como SOC exige o enriquecimento e a explicitação das relações semânticas, **ancoradas na ontologia do domínio e nas diretrizes dos SOC**, para evitar conexãoismo superficial (Clarke, 2001; Hjørland, 2016; Mazzochi, 2018; Moreira, 2019a).



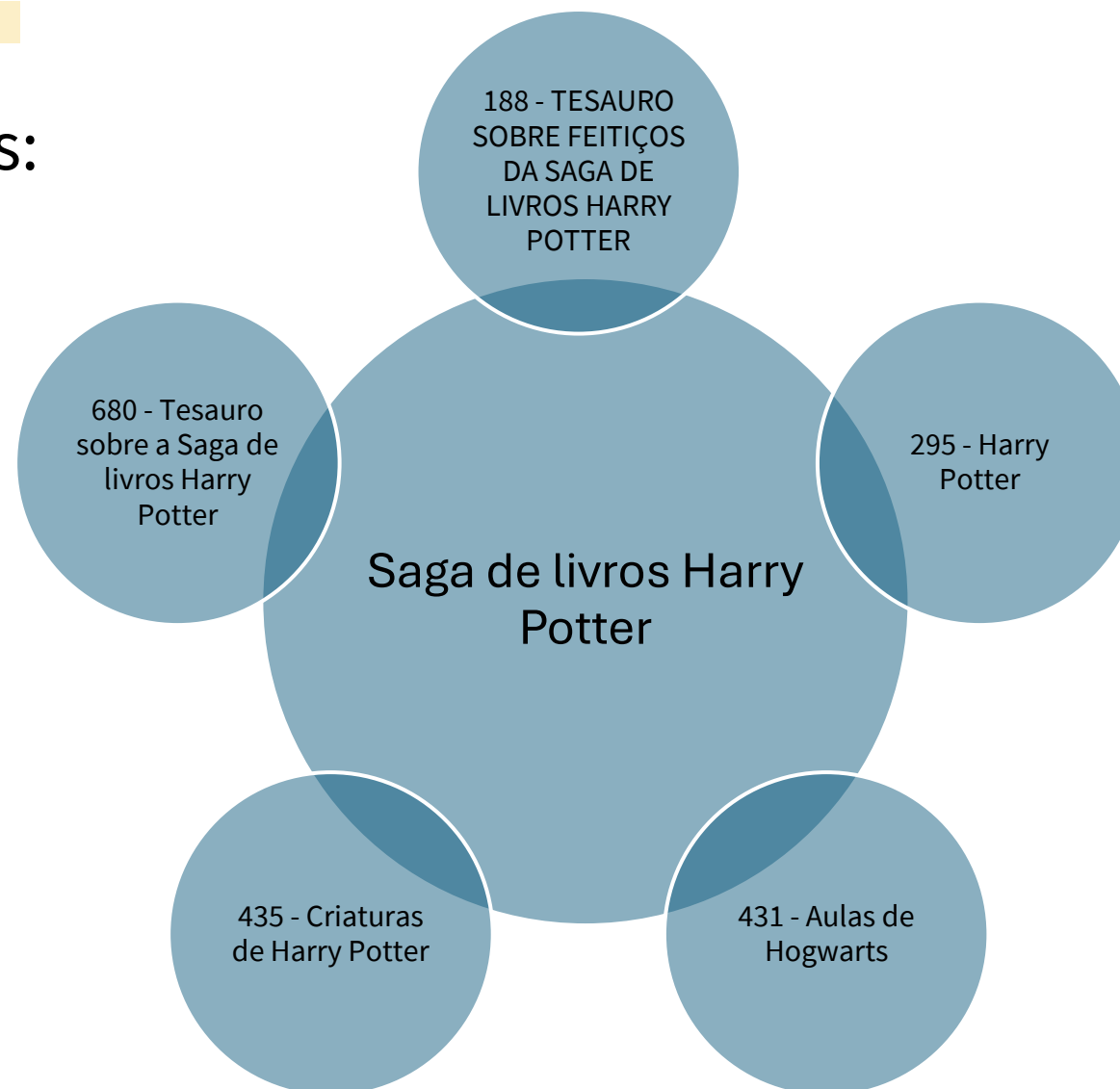
Resultados (continuação)



VIII ISKO BRASIL 2025



Tesouros com temáticas afins:
Saga Harry Potter



Fonte: Elaborado pelos autores, 2025.



Resultados (continuação)



VIII ISKO BRASIL 2025



Tesouros com temáticas afins: **Organização e Representação do Conhecimento**



Fonte: Elaborado pelos autores, 2025.



Outras temáticas em comum identificadas:

- **Mitologia;**
- **RPG;**
- **O Senhor dos Anéis;**
- **Chocolate;**
- **DC Comics;**
- **Dinossauros;**
- **Marvel/Vingadores.**



- O estudo revelou o **potencial de interligação entre tesouros**, criando **redes temáticas** que ampliam as possibilidades de associação e promovem novas perspectivas. A colaboração entre vocabulários gera um **ecossistema de significados** mais rico e fortalece a construção coletiva do conhecimento (Gabriel Júnior; Laipelt, 2023), impulsionando o desenvolvimento de objetos de fronteira (Campos, 2018).



O Thesa também se destaca como **ferramenta de ensino**, evidenciado pela presença de microtesouros produzidos em atividades acadêmicas.

Dos 190 tesouros completos, todos foram considerados aptos para migração ao Thesa 2.0, com possibilidades de aprofundamento futuro na análise das definições e das relações associativas.

- Analisados 1.128 registros; identificados **190 tesouros completos e aptos para migração**.
- Apenas uma **fração dos registros** analisados atendeu aos **critérios de completude**.
- **Desafios** identificados: **uso parcial** das funcionalidades e possível **baixa familiaridade dos usuários** com as diretrizes de construção.
- Necessidade de **estratégias de capacitação** (manuais, vídeos, FAQs).



- **Ampla uso das relações associativas** nos tesouros completos — destaque para: afinidade, processo/parte de, oposição/antonímia.
- Indícios de **amadurecimento** na concepção dos vocabulários; **necessidade de avaliações qualitativas futuras.**



- Identificação de temáticas afins — **potencial para criação de redes temáticas e interligação de tesouros.**
- Destaca a **importância dos tesouros como instrumentos dinâmicos de OC e o papel estratégico do Thesa** na promoção de práticas mais consistentes de construção terminológica e semântica.

Propostas de aprimoramento do **Thesa 2.0**:

- Melhoria da **interface** para facilitar uso de relações semânticas.
- **Diretrizes** mais detalhadas para construção de tesouros completos.
 - Estímulo ao **uso colaborativo** e construção de redes de tesouros interligados.
 - Aplicação de **inteligência artificial** para **avaliação** de qualidade, coerência e similaridade de **termos** e **relações semânticas** e o aprimoramento da **visualização gráfica** para a exploração de redes temáticas interligadas.



A continuidade dos esforços em pesquisa, inovação tecnológica e **capacitação** será essencial para consolidar o **Thesa 2.0** como uma plataforma cada vez mais **interoperável, adaptativa e alinhada** às exigências da OC e da web semântica.



CAMPOS, L. M. Classificação de objetos de fronteira na organização do conhecimento e o papel das ontologias. **Liinc Em Revista**, v. 14, n. 2, 2018. Disponível em: <https://doi.org/10.18617/liinc.v14i2.4314>. Acesso em: 05 nov. 2024.

CLARKE, S. G. Dextre. Thesaural relationships. In: BEAN, C. A.; GREEN, R. (ed.). **Relationships in the organization of knowledge**. Dordrecht: Kluwer, 2001. p. 37-52.

CURRÁS, E. **Tesauros**: linguagens terminológicas. Brasília: IBICT, 1995.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. F. Thesa 2.0: a interação dos tesauros em tempos de dados ligados. In: TOGNOLI, Natália Bolfarini; ALBUQUERQUE, A. C.; CERVANTES, B. M. N. (org.). **Organização e representação do conhecimento em diferentes contextos**: desafios e perspectivas na era da datificação. Londrina: ISKO-Brasil: PPGCI-UEL, 2023. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 15 set. 2023.

GOMES, H E. **Manual de elaboração de tesauros monolíngues**. Brasília: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

HJORLAND, B. Are relations in Thesauri “Context-Free, Definitional, and True in All Possible Worlds”? **Journal of the Association for Information Science and Technology**, v. 66, n. 7, p. 1367-1373, 2015. Disponível em: https://www.researchgate.net/publication/264805042_Are_relations_in_thesauri_contextfree_definitional_and_true_in_all_possible_worlds_Are_Relations_in_Thesauri_Context-Free_Definitional_and_True_in_All_Possible_Worlds. Acesso em: 20 ago. 2024.

HJORLAND, B. Knowledge organization. **Knowledge Organization**, n.43, v. 6, p. 475-483, 2016. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2016-6-475/knowledge-organization-ko-jahrgang-43-2016-heft-6?page=1>. Acesso em: 22 jan. 2025.

LAIPELT, R. C. F. **Metodologia para seleção de termos equivalentes e descritores de tesauros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. 213 f. Tese (Doutorado em Linguística Aplicada) -Programa de Pós-graduação em Linguística Aplicada, UNISINOS, São Leopoldo, 2015. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/4853>. Acesso em: 30 jan. 2022.

MAZZOCCHI, F. Knowledge Organization System (KOS): an introductory critical account. **Knowledge Organization**, 45, n. 1, 2018.

MOREIRA, W. Relações conceituais como elementos constitutivos essenciais dos sistemas de organização do conhecimento. **Informação & Informação**, v. 24, n.2, p. 1–30, 2019a. Disponível em: <https://doi.org/10.5433/1981-8920.2019v24n2p1>. Acesso em: 13 nov. 2023.

MOREIRA, W. Tesauros e ontologias como modelos de sistemas de organização do conhecimento. **Brazilian Journal of Information Science**, v. 13, n. 1, p. 15-20, 2019b.

MOURA, K. A. **Estudo do domínio documental jurídico para construção de tesouro de tipos documentais na ferramenta Thesa**. 2024. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, UFRGS, Porto Alegre, 2024. Disponível em: <https://lume.ufrgs.br/handle/10183/277306>. Acesso em 12 dez. 2024.

Agradecimentos



VIII ISKO BRASIL 2025



Os autores gostariam de agradecer ao CNPQ pelo financiamento do projeto 421849/2023-1 por meio do edital Universal/2023.

Obrigada a todos pela atenção!!!



kamoura15@gmail.com
ritacarmo@yahoo.com.br
bragato.barros@ufrgs.br
renefgj@gmail.com

