

## **Parajás: modelagem ontológica para base de dados em direito**

Eliandro Jeovane Natal Da Silva, Universidade Federal do Paraná (UFPR), Brasil, [eliandrosilva22@gmail.com](mailto:eliandrosilva22@gmail.com), <https://orcid.org/0009-0004-4794-4209>

Paula Carina De Araújo, Universidade Federal do Paraná (UFPR), Brasil, [paula.carina.a@gmail.com](mailto:paula.carina.a@gmail.com), <https://orcid.org/0000-0003-4608-752X>

Patrick Fernandes Rezende Ribeiro, Universidade Federal do Paraná (UFPR), Brasil, [patrick.ribeiro@ufpr.br](mailto:patrick.ribeiro@ufpr.br), <https://orcid.org/0000-0002-5973-1110>

Luciano Heitor Gallegos Marin, Universidade Federal do Paraná (UFPR), Brasil, [luciano.gallegos@ufpr.br](mailto:luciano.gallegos@ufpr.br), <https://orcid.org/0000-0002-4331-6588>

### **Abstract**

This article presents an initial ontological modeling developed for the Parajás - Law Database project. Its objective is to semantically structure the knowledge of Brazilian legal scientific production to enhance the organization, indexing, and retrieval of information. For the development of this ontology, the OntoForInfoScience methodology was adopted, formalizing the conceptual model in OWL language with implementation in the Protégé software. The main result is a semi-formal domain ontology, with classes, properties, and axioms that represent the entities and relationships of the field. This ontology will support the development of the Parajás project's database and will provide a basis to improve the management of scientific information in Law, promoting its greater interoperability and visibility.

### **Keywords**

Ontology; Knowledge Organization; Law Data Base; Open Science.

### **Resumo**

O presente artigo apresenta uma modelagem ontológica inicial desenvolvida para o projeto Parajás - Base de Dados em Direito tendo, por objetivo, estruturar semanticamente o conhecimento da produção científica jurídica brasileira para qualificar a organização, a indexação e a recuperação da informação. Adotou-se, para o desenvolvimento desta ontologia a metodologia OntoForInfoScience, formalizando o modelo conceitual em linguagem OWL com implementação no *software* Protégé. O principal resultado é uma ontologia de domínio, de natureza semiformal, com classes, propriedades e axiomas que representam as entidades e relações da área. Esta ontologia subsidiará o desenvolvimento da base de dados do projeto Parajás e oferecerá subsídios para aprimorar a gestão da informação científica em Direito, promovendo sua maior interoperabilidade e visibilidade.

### **Palavras-chave**

Ontologia; Organização do Conhecimento; Base de Dados em Direito; Ciência Aberta.

## **1. Introdução**

A publicidade da informação científica em acesso aberto vai além da disponibilização dos textos completos, somando-se a isso a padronização e a interoperabilidade de metadados que garantem a recuperação da informação para o reuso do conhecimento produzido. Este conjunto de princípios tem orientado políticas e iniciativas internacionais, como as diretrizes da UNESCO (2022, p. 24), que ressaltam a função organizacional das infraestruturas de informação abertas e comunitárias na promoção de uma ciência eficiente e inclusiva.

Nesse cenário, a organização do conhecimento estrutura conceitos e representa relações lógicas (regras, hierarquias, dependências) e semânticas (significados, contextos, sentidos) entre eles, aprovisionando

as bases teóricas e metodológicas para a criação de sistemas de organização do conhecimento, como taxonomias, tesouros e ontologias (Medeiros; Café, 2008). Entre essas ferramentas, as ontologias destacam-se por sua capacidade de representar domínios do conhecimento por meio de estruturas legíveis por máquinas – pois usam linguagens formais, como OWL, RDF que permitem representar conhecimento de forma estruturada e computacionalmente processável. Com isso, as ontologias permitem inferência lógica – os sistemas podem deduzir novas informações a partir das relações definidas, interoperabilidade – diferentes sistemas conseguem comunicar-se e entender os mesmos conceitos, e reutilização do conhecimento em sistemas computacionais, isto é, o conhecimento modelado pode ser reutilizado em outros contextos ou aplicações (Guarino; Oberle; Staab, 2009; Gruber, 1995).

O presente artigo, portanto, insere-se nesse campo de estudos da organização do conhecimento e tem como objetivo apresentar a modelagem ontológica desenvolvida para a *Parajás – Base de Dados em Direito*, iniciativa vinculada ao projeto de pesquisa Gestão da Informação Científica no Contexto da Ciência Aberta<sup>12</sup>. A proposta da Parajás busca oferecer uma infraestrutura informacional especializada para indexar periódicos científicos do domínio do direito no Brasil, segundo os princípios da ciência aberta, contribuindo para a padronização semântica, interoperabilidade e visibilidade do conhecimento jurídico nacional.

### 1.1 Organização do Conhecimento

A organização do conhecimento é um campo de estudo que estrutura conceitos para criar representações da realidade (Medeiros; Café, 2008). Ela inclui a representação do conhecimento, que utiliza modelos conceituais para organizar ideias a partir de suas relações lógicas e de significado, gerando os Sistemas de Organização do Conhecimento (SOC).

Os SOC funcionam como metalinguagens em sistemas de recuperação de informação, facilitando o acesso a objetos informacionais e suas descrições (Hjørland, 2007; 2016). Segundo Café e Bräscher (2011, p. 25), a organização do conhecimento baseia-se, principalmente, em análises semânticas.

Nesse processo, as propriedades dos conceitos definem suas relações semânticas, agrupando-os de modos distintos em cada área do saber. O resultado é a representação do conhecimento, que reflete uma visão consensual da realidade após análise de um domínio específico (Medeiros; Café, 2008, p. 6).

A representação do conhecimento (processo) e os SOC (instrumentos) organizam conceitos em diferentes áreas do conhecimento. Os SOC mais usados na Ciência da Informação são:

- Taxonomia: é um vocabulário controlado organizado hierarquicamente (ANSI/NISO, 2005, p. 9). Elimina ambiguidades, controla sinônimos e estabelece relações lógico-semânticas para melhorar armazenamento e recuperação da informação. Classifica elementos em subgrupos mutuamente exclusivos, seguindo critérios predeterminados (Miranda, 2014, p. 219).
- Tesouro: vocabulário controlado com termos (descritores) semanticamente relacionados, usado para organizar e recuperar informações (Sales; Café, 2009). É uma linguagem pós-coordenada e especializada, onde termos simples ou compostos se interligam sintática e semanticamente.
- Ontologia: resultado do estudo de categorias de entidades em um domínio, visando organizar e tornar elementos compreensíveis (Sowa, 2010). Diferencia-se de taxonomias e tesouros por usar relações semânticas formais rigorosas, representando conhecimento em formatos legíveis por máquinas. Requer linguagem formal com significados logicamente explícitos, expressos por estruturas matemáticas (Feitosa, 2006, p. 73)

---

<sup>1</sup> Link de informações do projeto: <http://dgp.cnpq.br/dgp/espelhogrupo/8079711387911310>

<sup>2</sup> Projeto financiado pelo Conselho de Desenvolvimento Científico e Tecnológico (CNPq).

## 1.2 Ontologia

Ontologias são modelos que representam formalmente um domínio específico, definindo seus conceitos, relações e regras. Esses modelos usam estruturas abstratas para descrever o significado de um conjunto de dados, facilitando o processamento por computadores (Guarino; Oberle; Staab, 2009; Gruber, 1995). Por serem estruturadas de forma lógica, permitem o compartilhamento e a reutilização do conhecimento. Além disso, ajudam a transformar conhecimento implícito em explícito, criando representações que podem ser interpretadas por máquinas (Urbanskyy, 2018).

O principal objetivo de uma ontologia é organizar o conhecimento por meio de variáveis e suas relações, estabelecendo um entendimento comum sobre um tema (Kalet *et al.*, 2017). Isso melhora a comunicação e traz benefícios como interoperabilidade e reutilização de dados. Uma ontologia armazena informações em uma base de conhecimento, que pode ser usada por diferentes aplicações. Ela também permite criar hierarquias, estabelecer relações entre conceitos e fazer inferências lógicas, sendo útil em diversas áreas (Kalet *et al.*, 2017).

A construção de ontologias exige métodos diferentes dos usados no desenvolvimento de sistemas tradicionais. Embora existam metodologias específicas, nenhuma delas é tão consolidada quanto as da engenharia de *software*, e ainda não há um padrão definido (Urbanskyy, 2018).

Atualmente, a linguagem mais usada para desenvolver ontologias é a Web Ontology Language (OWL), recomendada pelo World Wide Web Consortium (W3C) para a *Web Semântica* (Rasmussen *et al.*, 2021). O W3C busca melhorar a organização e o processamento de dados na internet, pois muitas informações são armazenadas sem padronização, dificultando sua interpretação por sistemas automatizados.

## 2. Metodologia

Visando a construção da ontologia “Parajás”, o presente trabalho adotou a metodologia OntoForInfoScience. Ela é proposta como uma ferramenta para auxiliar profissionais na organização do conhecimento, a superar desafios comuns, incluindo a complexidade terminológica e as barreiras relacionadas a aspectos lógicos e filosóficos envolvidos na criação de ontologias (Mendonça, 2015).

A metodologia OntoForInfoScience é estruturada em uma pré-etape, seguida por 8 etapas (Mendonça, 2015), as quais são:

1. Pré-etape: verifica se é necessário criar uma ontologia para resolver o problema. Se não for preciso, recomenda-se usar ferramentas alternativas, como tesouros ou vocabulários controlados. Se confirmada a necessidade da ontologia, avança-se para as próximas etapas.
2. Especificação: o desenvolvimento da ontologia tem início com a sua especificação, feita por meio de um *template* que estabelece seu domínio, objetivo e escopo. Nesta etapa, são determinados os usuários finais, os cenários de uso e o grau de formalidade. A ontologia é então categorizada conforme sua natureza, sendo definidos seus limites de atuação. Essas definições iniciais servem como base para todas as etapas subsequentes do projeto.
3. Aquisição e extração de conhecimento: esta etapa compreende a seleção de fontes de informação relevantes e a definição de métodos para coletar dados do domínio em questão. São usadas diferentes estratégias para isso, desde a leitura e análise manual de documentos até o uso de ferramentas semiautomáticas para identificar e extrair termos e conceitos. O resultado deste processo é a criação de glossários que organizam os conceitos, ações e relações identificadas.

4. Conceitualização: nesta etapa, o conhecimento é estruturado para a criação da ontologia. A partir dos glossários já definidos, identificam-se e analisam-se os conceitos que se tornarão as classes. Este processo gera um modelo gráfico para visualizar as conexões entre os conceitos. Dessa maneira, o conhecimento é organizado antes de ser formalizado.
5. Fundamentação ontológica: nesta etapa, o desenvolvedor identifica e seleciona uma ontologia de referência, que deve estar alinhada com os objetivos e a fundamentação teórica do projeto. Uma vez escolhida, essa ontologia de base é importada para um software de edição, onde servirá como alicerce para a construção da nova ontologia.
6. Formalização: a formalização transforma conceitos já definidos em um modelo estruturado. Nela, o desenvolvedor organiza as classes, define os atributos (com seus nomes, sinônimos e descrições) e estabelece definições formais. Essas definições, baseadas nas descrições textuais iniciais, especificam as propriedades das classes por meio de tipos de dados, restrições à repetição e operadores lógicos.
7. Avaliação: o objetivo desta etapa é analisar se a ontologia foi construída de forma correta e se representa fielmente a realidade. Para isso, são realizados dois processos: a validação, que verifica se os conceitos correspondem com o contexto real, e a verificação, que assegura a integridade e o correto funcionamento da ontologia.
8. Documentação: descreve todo o processo de construção ao longo das etapas, gerando um manual técnico.
9. Disponibilização: a ontologia finalizada é disponibilizada aos usuários, tanto em sua forma gráfica quanto formal.

A ontologia foi implementada no Protégé<sup>3</sup>, um *software* de código aberto projetado para criação e edição de ontologias. Essa ferramenta oferece uma interface gráfica intuitiva e arquitetura modular, que permite a expansão de funcionalidades através de plugins.

### 3. Resultados

Na pré-etapa do trabalho, realizou-se uma análise preliminar voltada à identificação da necessidade de desenvolver uma ontologia para o projeto Parajás. Essa análise teve como propósito verificar em que medida uma ontologia poderia contribuir para solucionar o problema de organização e recuperação da informação identificado na base de dados. Constatou-se que estruturas tradicionais de representação do conhecimento, como tesouros ou vocabulários controlados, seriam insuficientes para atender às demandas do projeto, uma vez que não permitiam representar adequadamente as relações conceituais existentes no domínio do direito, tampouco possibilitavam a inferência automática de novos conhecimentos.

Na etapa seguinte, de Especificação, foram definidos os objetivos e requisitos da ontologia por meio do preenchimento do *template* proposto pela metodologia. Esse processo resultou na criação do primeiro artefato do projeto: a especificação da ontologia (Quadro 1), que orientou as etapas seguintes.

Com base nessa especificação, procedeu-se à coleta de informações sobre o domínio do direito, utilizando-se como fontes primárias os periódicos científicos jurídicos nacionais e, como fontes secundárias, a literatura de domínio em organização do conhecimento e ontologias. A partir dessa análise, foram elaborados três glossários: de conceitos, de verbos e de relações<sup>4</sup>. Esses glossários serviram de base para a estruturação conceitual da ontologia.

---

<sup>3</sup> <https://protege.stanford.edu/>

<sup>4</sup> Os glossários de conceitos, verbos e relações estão disponíveis para acesso em: <https://doi.org/10.5281/zenodo.17488199>

Na etapa de Conceitualização, com base nos glossários, iniciou-se a modelagem conceitual do conhecimento, em que os conceitos levantados foram organizados em uma Tabela de Conceitos e Propriedades<sup>5</sup>.

No que se refere à fundamentação ontológica, optou-se por desenvolver uma estrutura original, sem reutilizar ontologias existentes no domínio, valendo-se exclusivamente dos conhecimentos consolidados nas etapas anteriores.

Atualmente, a ontologia encontra-se na fase de Formalização. O modelo conceitual<sup>6</sup> foi transformado em uma estrutura formal com 31 classes, 12 propriedades, 6 relações e 36 axiomas, utilizando o software Protégé na versão 5.6.5.

As etapas de avaliação e disponibilização estão previstas no cronograma de execução. Após a conclusão da formalização, a ontologia passará por testes antes de ser disponibilizada aos usuários. A verificação de consistência lógica será realizada diretamente no Protégé, utilizando seu mecanismo de inferência (*reasoner*), que analisará as regras e relações para identificar possíveis contradições ou inconsistências. Paralelamente, a documentação é realizada durante todo o processo de construção da ontologia.

A especificação desta ontologia descreve seu objetivo principal, que é organizar e representar o conhecimento no domínio do Direito, com foco na produção científica brasileira. Ela define quais usuários poderão se beneficiar e como poderão utilizá-la. Também apresenta sua classificação como uma ontologia de domínio com alto grau de formalidade, projetada para integrar-se a sistemas de informação.

---

<sup>5</sup> A Tabela de Conceitos e Propriedades está disponível para acesso em: <https://doi.org/10.5281/zenodo.17488199>

<sup>6</sup> O modelo conceitual está disponível para acesso em: <https://doi.org/10.5281/zenodo.17488199>

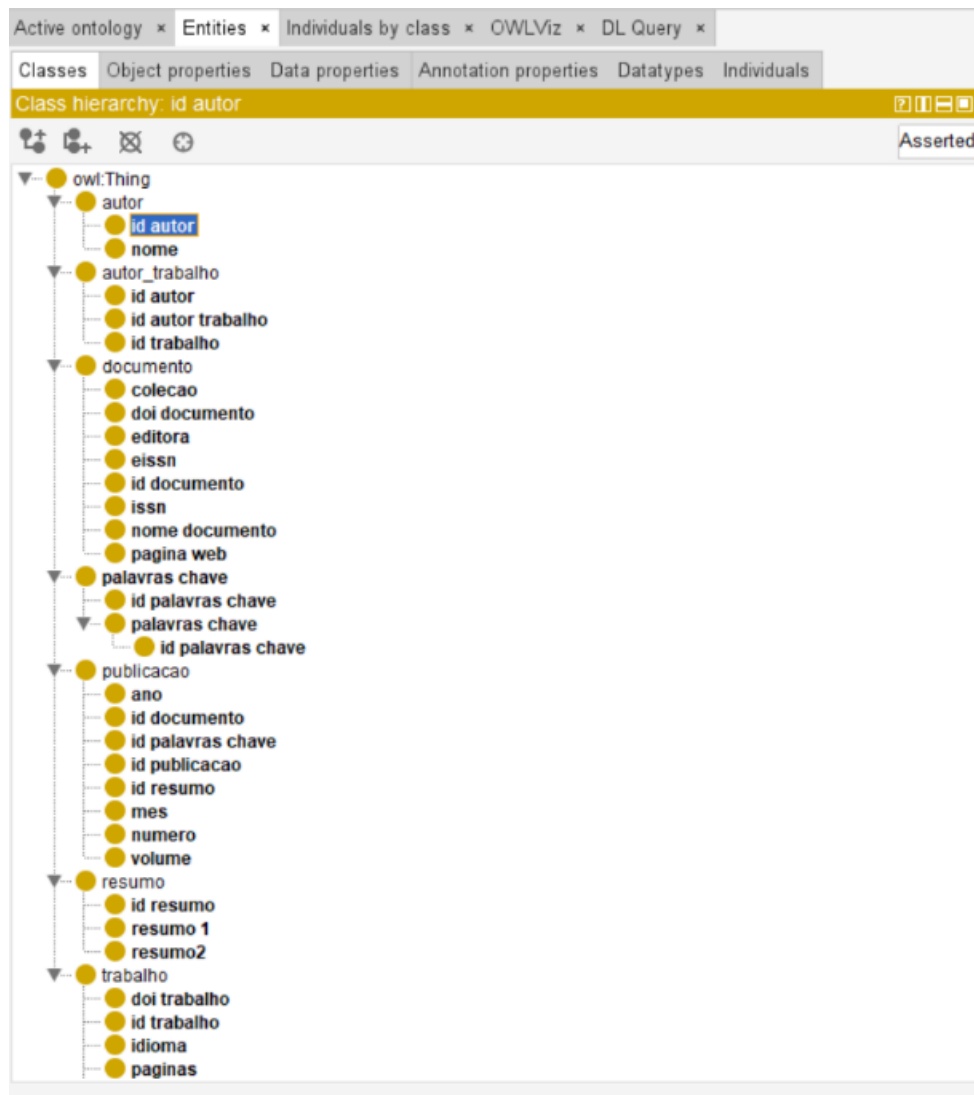
<b>Domínio</b>
- Direito (com foco na produção científica jurídica brasileira)
<b>Propósito Geral</b>
- Estruturar semanticamente os conceitos, entidades e relações envolvidas na organização, indexação, recuperação e análise da produção científica no domínio do Direito, de modo a subsidiar o desenvolvimento da Base de Dados em Direito (Parajás).
<b>Classes de usuários</b>
- Pesquisadores da área do Direito; - Estudantes; - Profissionais do campo jurídico; - Gestores de Políticas Públicas;
<b>Uso pretendido</b>
- Permitir a recuperação avançada de informações por meio de metadados semânticos
<b>Tipo da ontologia</b>
- Quanto à estrutura: Ontologia de domínio - Quanto ao grau de formalidade: Pesada
<b>Grau de formalidade</b>
- Semiformal
<b>Delimitação do escopo de cobertura</b>
- Ponto de Partida: Periódicos científicos vinculados a Programas de Pós-Graduação em Direito (PPGDs) do Brasil - Questões de competência: QC1. Como representar semanticamente um periódico científico da área do Direito e suas características?

#### **Quadro 1**

*Template* de Especificação

*Fonte:* Os autores

A Figura 1 mostra as classes da ontologia “Parajás”. No topo da hierarquia está a classe genérica “owl:Thing”, que é a superclasse de todas as outras classes. Abaixo dela, aparecem várias subclasses que representam conceitos relacionados à autoria e à publicação de trabalhos.



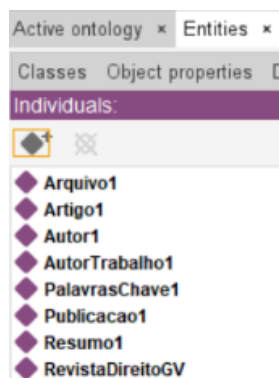
**Figura 1**

Hierarquia de Classes da Ontologia (CC-BY-4.0)

*Nota.* Captura de tela do software Protégé mostrando a hierarquia de classes da ontologia Parajás. Inclui as classes principais “autor”, “auto trabalho”, “documento”, “palavras\_chave”, “publicação”, “resumo” e “trabalho”, cada uma com suas propriedades e identificadores correspondentes. A estrutura ilustra a organização conceitual da ontologia.

*Fonte:* Os autores

A Figura 2 exibe as instâncias da ontologia, listando elementos que representam entidades específicas do domínio modelado. Cada item (como Arquivo1, Artigo1, Autor1 e PalavrasChave1) corresponde a uma instância de uma classe definida na estrutura ontológica. Essas instâncias permitem a representação de dados reais dentro da ontologia.



**Figura 2**

Visualização das Instâncias da Ontologia no Protégé (CC-BY-4.0)

*Nota.* Captura de tela do software Protégé mostrando a aba *Individuals*, na qual são exibidas as instâncias criadas a partir das classes da ontologia Parajás. Entre as instâncias apresentadas estão “Arquivo1”, “Artigo1”, “Autor1”, “AutorTrabalho1”, “PalavrasChave1”, “Publicacao1”, “Resumo1” e “RevistaDireitoGV”. A imagem ilustra a correspondência entre as classes e suas instâncias, evidenciando a aplicação prática do modelo ontológico.

*Fonte:* Os autores

#### 4. Conclusão

A construção da ontologia da *Parajás – Base de Dados em Direito* representa um esforço para fortalecer a infraestrutura de informação científica no domínio jurídico brasileiro, alinhando-se aos princípios da ciência aberta e da valorização da produção regional. Ao estruturar semanticamente os conceitos, entidades e relações envolvidos na produção científica em Direito, a ontologia oferece um modelo compartilhável que contribui para a qualificação da descrição, recuperação e análise de conteúdos informacionais do domínio do Direito.

A aplicação da metodologia OntoForInfoScience viabilizou a condução sistemática das etapas de construção ontológica, desde a especificação dos objetivos até a formalização computacional da estrutura conceitual, com uso da linguagem OWL e implementação no *software* Protégé. A especificação técnica da ontologia posiciona-a como uma ontologia de domínio, com grau de formalidade semiformal, voltada à organização da produção científica dos Programas de Pós-Graduação em Direito (PPGDs) no Brasil. Seu público-alvo inclui pesquisadores, estudantes, profissionais jurídicos e gestores de políticas públicas, que poderão utilizá-la para recuperação avançada de informações por meio de metadados semânticos, promovendo maior precisão e contextualização nas buscas.

O próximo passo será o de integrar a ontologia com o Handle System<sup>7</sup> - que é um sistema de informação projetado para identificadores persistentes de objetos digitais conhecidos como *handles*. Esta integração garantirá interoperabilidade e rastreabilidade dos conteúdos científicos, pois permitirá criar uma infraestrutura de recuperação de informações semântica e persistente. Após essa etapa, a ontologia estará pronta para a disponibilização.

Como contribuição prática, a ontologia proposta não só subsidia o desenvolvimento técnico da base de dados Parajás, mas também oferece subsídios para o aprimoramento da gestão da informação científica em Direito, considerando as exigências crescentes de interoperabilidade entre plataformas, como as que compõem os sistemas de avaliação da pós-graduação brasileira. Por fim, ao articular fundamentos da organização do conhecimento com tecnologias da *web* semântica, este trabalho reforça a relevância das ontologias como instrumentos estratégicos para garantir o acesso, a visibilidade e o impacto da produção científica nacional.

<sup>7</sup> <https://www.handle.net/>



## Declaração de uso de IA

Declaro que não foi utilizado recurso de inteligência artificial em nenhuma etapa da elaboração deste artigo.

## Agradecimentos

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)  
Coordenação de Formação de Pessoal de Nível Superior (CAPES)  
Universidade Federal do Paraná (UFPR)

## Referências

- Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (pp. 1–17). Springer.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5–6), 907–928.
- Urbanskyy, S. M. C. (2018). *Uma ontologia para representação de conhecimento sobre boas práticas nas infecções relacionadas à assistência à saúde* (Dissertação de mestrado, Universidade Federal de Ciências da Saúde de Porto Alegre).
- Kalet, A. M., Philips, M. H., Jones, T. L., & Wazer, D. E. (2017). Developing Bayesian networks from a dependency-layered ontology: A proof-of-concept in radiation oncology. *Medical Physics*, 44(8), 4350–4359.
- Rasmussen, M. H., Pauwels, P., Lefrançois, M., Schneider, G. F., Hvidtfeldt, R., & Karlshøj, J. (2021). BOT: The building topology ontology of the W3C linked building data group. *Semantic Web, Preprint*, 1–19.
- Mendonça, F. M. (2015). *Ontoforinfoscience: Metodologia para construção de ontologias pelos cientistas da informação – Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA)*. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUBDA35H3K>. Acessado em 13 de junho de 2025.
- Medeiros, M. B. B., & Café, L. M. A. (2008). Organização da informação ou organização do conhecimento? In *9º Encontro Nacional de Pesquisa em Ciência da Informação*, São Paulo. *Anais* (p. 6). São Paulo: USP.
- Café, L., & Bräscher, M. (2011). Organização do conhecimento: Teorias semânticas como base para estudo e representação de conceitos. *Informação & Informação*, 16(2), 25–51. <https://doi.org/10.5433/1981-8920.2011v16n2p25>
- Hjørland, B. (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41(1), 367–405. Disponível em: [https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/Semantics and KO K.pdf](https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/Semantics%20and%20KO%20K.pdf). Acessado em 4 de março de 2021.
- ANSI/NISO. (2005). *ANSI/NISO Z39.19-2005: Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO. Disponível em: [http://www.niso.org/kst/reports/standards/kfile\\_download?id%3Astring%3Aiso-8859-1=Z39-19-2005.pdf](http://www.niso.org/kst/reports/standards/kfile_download?id%3Astring%3Aiso-8859-1=Z39-19-2005.pdf). Acessado em 22 de agosto de 2020.
- Miranda, F. B. (2014). Da classificação de obras lexicográficas e seus problemas: Proposta de uma taxonomia. *Alfa: Revista de Linguística*, 58, 215–231. Disponível em: <https://www.scielo.br/j/alfa/a/sy3vj8trKCXXyGtgV8Nwc5R/?lang=pt>. Acessado em 15 de dezembro de 2021.
- Sales, R. de, & Café, L. (2009). Diferenças entre tesouros e ontologias. *Perspectivas em Ciência da Informação*, 14(1), 99–116. Disponível em:

<https://www.scielo.br/j/pci/a/qr7hthZ93SnkHfRybCBLBxd/abstract/?lang=pt>. Acessado em 11 de dezembro de 2021.

Sowa, J. F. (2010). *Ontology. KR Ontology*. Recuperado de: <https://www.jfsowa.com/ontology/>

UNESCO. (2022). *Recomendação da UNESCO sobre ciência aberta*. Paris: UNESCO. <https://doi.org/10.54677/XFFX3334>. Acesso em: 5 de fevereiro de 2023.

Feitosa, A. (2006). *Organização da informação na web: das tags à web semântica*. Brasília: Thesaurus.

### **Licença**

Este resumen ampliado se distribuye bajo licencia Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA 4.0).