**Metacategorias semânticas para aplicação em SKOS**

Rita do Carmo Ferreira Laipelt[[1]](#footnote-1) , Luciana Monteiro Krebs[[2]](#footnote-2) , Renê Faustino Gabriel Júnior**[[3]](#footnote-3)**

**Resumo:** Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em SKOS. Busca especificamente, sistematizar as diferentes correntes teóricas sobre relações semânticas da área de Organização do conhecimento e seus respectivos métodos na Ciência da Informação. No contexto atual, a interface de pesquisa de Sistemas de Recuperação da Informação (SRI), com suas diferentes possibilidades de busca, exercem um papel intermediário entre o usuário e o acervo documental de uma instituição. Assim, é necessário que esses sistemas sejam pensados de forma a facilitar a recuperação da informação pelos usuários. Uma das formas de tornar os SRI mais acessíveis é investir no aperfeiçoamento das relações semânticas dos Sistemas de Organização do Conhecimento (SOCs). A partir da sistematização e descrição de tipologias de relações semânticas disponíveis na literatura da área de organização do conhecimento realizaremos a implementação das mesmas no software Thesa. A inserção das metacategorias semânticas identificadas possibilitará a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento, tendo em vista que o Thesa já é utilizado por 182 usuários e faz a curadoria de 142 tesauros. A consolidação dos resultados desta pesquisa possibilitará o desenvolvimento de tesauros semanticamente fortalecidos, numa transposição natural entre a teoria e a prática. Espera-se que os resultados gerados com a atualização do software façam com que SRI incorporem a utilização do mesmo, e assim seja possível futuramente a contextualização da informação, tanto por parte do usuário como do profissional da informação. Dos futuros resultados do uso do Thesa, está a possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas, bem como gerar inferências com bases em outros tesauros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

**Palavras-chave**: Organização do Conhecimento. Relações Semânticas. SKOS. Thesa.

**1. Introdução**

Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em SKOS. Busca especificamente, Sistematizar as diferentes correntes teóricas sobre relações semânticas da área de Organização do conhecimento e seus respectivos métodos na Ciência da Informação. Entendemos que a partir desse mapeamento das relações semânticas, especialmente aquelas classificadas como relações associativas em SKOS será possível sistematizar Metacategorias de relações semânticas para aplicação no software Thesa - Tesauro Semântico Aplicado.

No contexto atual, a interface de pesquisa de Sistemas de Recuperação da Informação (SRI), com suas diferentes possibilidades de busca, exercem um papel intermediário entre o usuário e o acervo documental de uma instituição. Assim, é necessário que esses sistemas sejam pensados de forma a facilitar a recuperação da informação pelos usuários. Uma das formas de tornar os SRI mais acessíveis aos usuários é investir no aperfeiçoamento das relações semânticas dos Sistemas de Organização do Conhecimento (SOCs). Tendo em vista que na recuperação da informação, de acordo com Hjørland (2007, p. 392) “[...] a função básica das relações semânticas é contribuir para a otimização da precisão e da revocação.” No entanto, de acordo com Green (2001), a magnitude e complexidade dos relacionamentos semânticos na área de organização do conhecimento dificultam o uso consistente dos mesmos tanto por profissionais da informação como por usuários finais. Logo, é necessário identificar e sistematizar esses relacionamentos semânticos para que os mesmos possam ser utilizados de maneira inequívoca em SOCs. Para Green ( 2001, p.14) “[...] embora a utilidade de relacionamentos explícitos seja mínima para usuários finais [...] a compreensão intuitiva que os humanos trazem para os relacionamentos não é compartilhada por dispositivos computacionais.” Por isso a explicitação desses relacionamentos “[...] talvez sejam a nossa melhor esperança para infundir qualidade superior em nossos sistemas de recuperação.” (GREEN, 2001, p.14). Visto que, segundo Green (2001), a qualidade superior dos sistemas de recuperação da informação também pode vir a partir da possibilidade de identificação de fontes relevantes, as quais, sem o uso de relacionamentos semânticos, não seriam recuperadas.

Acreditamos que a explicitação de relações semânticas, sobretudo as relações associativas, podem contribuir tanto para o aperfeiçoamento de SKOS, no caso específico deste trabalho, do Thesa, como para a recuperação da informação e descoberta de novas fontes (nesse caso, quando tivermos um SKOS como o Thesa interligado com bases de dados como a Brapci, por exemplo, isso será possível.).

**2 Relações Semânticas em Sistemas de Organização do Conhecimento**

Os diferentes tipos de instrumentos de representação do conhecimento existentes atualmente são designados pelo termo “Sistemas de Organização do Conhecimento” ( SOCs). Esta é uma nova denominação para o que conhecemos por linguagens documentárias, seu diferencial, no entanto está na incorporação de elementos de inovação tecnológica da era digital (CARLAN, 2010). Logo, podemos considerar como um SOC diferentes tipos de instrumentos de representação da informação tais como: tesauros, ontologias, taxonomias, redes semânticas entre outros.

Os SOC produzem maior suporte semântico para sistemas de recuperação de informações (SRI). Considerando que as expressões utilizadas para representar documentos ou a pesquisa de usuários podem ter significados diferentes, sem atributos e links semânticos que contextualizem o domínio, é impossível para a máquina interpretar o seu significado. De modo que na ausência de propriedades semânticas, um SRI poderá apenas medir a semelhança entre as palavras usadas em um documento com as da consulta do usuário o que não é suficiente para a recuperação de resultados relevantes (MACULAN; LIMA; OLIVEIRA, 2017).

O modelo Simple Knowledge Organization System (SKOS), de acordo com Isaac e Summers (2009) é utilizado para expressar vários tipos de esquemas conceituais tais como tesauros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, Folksonomia, e outros tipos de vocabulários controlados Sua estrutura possibilita o compartilhamento e ligação dos Sistemas de Organização do Conhecimento (SOC) por meio da Web (MILES; BECHHOFER, 2009). O SKOS compartilha muito dos princípios de elaboração de tesauros, taxonomias, esquemas de classificação e sistemas de cabeçalhos de assuntos, bem como possibilita capturar muito dessa semelhança e torna-la explicita, para que possa ser reutilizada por outras aplicações (MILES; BECHHOFER, 2009). O modelo é composto por termos num vocabulário denominado SKOS Core Vocabulary (MILES; BRICKLEY, 2005) que é um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF (RDFSchema). Conforme Catarino (2014), o modelo SKOS baseia-se em classes e propriedades e para representar o conjunto de dados tem como elemento central o conceito.

Um tesauro, por exemplo, pode representar o conhecimento de um domínio através de um conjunto de conceitos que apresentam relações semânticas hierárquicas, associativas e de equivalência. Essas três tipologias de relações podem ser consideradas limitadas e em alguns casos causar ambiguidade entre conceitos e consequentemente problemas de recuperação da informação. Por isso, é fundamental aprimorar as conexões entre conceitos para que, através de sua explicitação, tenhamos uma semântica mais forte conforme ocorre nas ontologias. Um SOC seria capaz de contemplar, por meio de uma espécie de hospitalidade infinita, múltiplas perspectivas (pontos de vista) em uma estrutura mais flexível (GNOLI, 2008).

Porém, Café e Brascher (2011, p. 26) destacam que

A arbitrariedade da seleção dos agrupamentos e relacionamentos não pode ser vista, no entanto, como algo totalmente aleatório. Ela se deve aos aspectos que desejamos destacar numa d eterminada representação do conhecimento.”

Deve, portanto, haver correspondência entre o sistema e a realidade que este representa, tendo em vista o contexto onde as expressões ocorrem para a inferência dos significados. (WEISS; BRASCHER, 2013).

Verifica-se na literatura da área de organização do conhecimento o incentivo ao uso de outros tipos de relações semânticas. Assim, Hjørland (2007, p. 393) questiona “Como devemos explicar essa demanda por um conjunto muito mais rico de relações do que as normalmente usadas em tesauros?”, e explica que uma função adicional que as relações semânticas em um SOC poderiam desempenhar seria “fazer os diferentes interesses e paradigmas visíveis para que o usuário possa fazer uma escolha informado” (HJØRLAND, 2007, p. 389). Para tanto, o autor destaca a análise da literatura da área a ser representada pelo SOC como essencial para a identificação de diferentes pontos de vista.

A partir da análise da literatura dos domínios é possível identificar relações paradigmáticas e sintagmáticas. As relações paradigmáticas são constituídas por sinonímia, antonímia, meronímia, hiperonímia e hiponímia. Nesse caso as relações conceituais são por exemplo: oposição conceitual, parte de, é um tipo de. Já as relações sintagmáticas ocorrem entre entidades de diferente natureza, seus itens co-ocorrem. Essas relações incluem objetos e agentes, eventos e processos (PREVOT, 2010).

As relações sintagmáticas podem ser tratadas nos tesauros a partir da explicitação das relações associativas. No entanto, de acordo com Maculan, Lima e Oliveira (2017) as relações associativas são consideradas as mais difíceis de definir e sobre as quais ainda não existe pesquisa suficiente para determinar suas bases teóricas. Portanto, de acordo com as autoras as relações

associativas devem ser estabelecidas principalmente a partir da análise da literatura, de modo que seja possível identificar os diferentes pontos de vista ali presentes.

**3 THESA – Tesauro Semântico Aplicado**

O Tesauro Semântico Aplicado (THESA) foi desenvolvido com o objetivo de disponibilizar um instrumento de elaboração de tesauros para os estudantes de graduação em Biblioteconomia da UFRGS utilizarem na disciplina de Linguagens Documentárias III. Deseja-se com isso reduzir o trabalho operacional dos alunos e dar maior atenção ao trabalho de desenvolvimento cognitivo e conceitual referente a modelagem do domínio. Seu desenvolvimento baseou-se nas normas ISO e NISO vigentes, de forma a compatibilizar suas diretrizes com os requisitos semânticos prementes nas novas demandas dos SOCs. Com base na literatura disponível, nas normas de construção de tesauros da ISO e NISO foram identificados os elementos necessários para o desenvolvimento do protótipo, principalmente no que tange ao levantamento das propriedades de ligação entre os conceitos. O software funciona em ambiente Web e pode ser utilizado gratuitamente, para fins didáticos em disciplinas dos cursos de graduação e pós-graduação ou para uso profissional. (GABRIEL JUNIOR; LAIPELT, 2017)

A versão beta (de teste) do Thesa pode ser acessada no endereço <http://www.ufrgs.br/tesauros.>

Destacamos que com o Thesa é possível descrever relações específicas, viabilizando desta forma o compartilhando de estrutura e dados em sistemas complexos, como o Linked data, e ainda manter a compatibilidade com modelos simplificados como da ISO e NISO. Sua estrutura é baseada na concepção das relações entre os conceitos, partindo do pressuposto que um conceito pode ser representado por um termo, uma imagem, um som, um link ou qualquer outra forma que possa ser explicitada. Desta maneira, o conceito é perene, enquanto a sua representação pode variar conforme o contexto histórico ou social, sendo definida uma forma preferencial, e inúmeras formas alternativas e ocultas.

**4 Metodologia**

Em consonância com nosso objetivo de desenvolver um modelo de metacategorias semânticas para aplicação em SKOS, este trabalho parte de pressupostos teóricos visando uma aplicação. Assim, a partir da sistematização e descrição de tipologias de relações semânticas

disponíveis na literatura da área de organização do conhecimento realizaremos a implementação das mesmas no software Thesa.

A inserção das metacategorias semânticas identificadas possibilitará a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento, tendo em vista que o Thesa já é utilizado por 182 usuários e faz a curadoria de 142 tesauros.

**5 Resultados Esperados**

O Thesa foi desenvolvido inicialmente para fins acadêmicos, principalmente para utilização dos estudantes do curso de Biblioteconomia da UFRGS. Porém, tendo em vista o potencial do sistema para a gestão de vocabulários percebeu-se a importância de aperfeiçoá-lo para utilização tanto com fins acadêmicos como da comunidade profissional de bibliotecários. A consolidação dos resultados desta pesquisa possibilitará o desenvolvimento de tesauros semanticamente fortalecidos, numa transposição natural entre a teoria e a prática.

Espera-se que os resultados gerados com a atualização do software façam com que SRI incorporem a utilização do mesmo, e assim seja possível futuramente a contextualização da informação, tanto por parte do usuário como do profissional da informação. Dos futuros resultados do uso do Thesa, está a possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas, bem como gerar inferências com bases em outros tesauros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

**Referencias**

BIZER, C.; HEATH, T.; BERNERS-LEE, T. (2009). Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, v.5, n.3, p.1-22.

CAFÉ, L.; BRASCHER, M. (2011). Organização do Conhecimento: Teorias Semânticas como base para estudo e representação de conceitos. Inf. Inf., Londrina, v. 16. n. 3. p. 25-51.

CARLAN, E. (2010). Sistemas de organização do conhecimento: uma reflexão no contexto da Ciência da Informação. 2010. 195 f. Dissertação. (Mestrado em Ciência da Informação) –Faculdade de Economia, Administração, Contabilidade e Ciência da Informamação e Documentação. Brasília: Universidade de Brasília.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. F. (2017) Thesa: ferramenta para construçÃo de tesauro semântico aplicado interoperável. *Revista P2P e INOVAÇÃO*, v. 3, n. 2, p.124-145.

GNOLI, C. (2008). Tem long-term research questions in knowledge organization. *Knowledge Organization.* v. 35, n. 2/3, p.137-149.

GREEN, R. Relationships in the organization of knowledge: an overview. In: BEAN, A; GREEN, R. (Orgs.). **Relationships in the Organization of Knowledge** . Boston/Dordrecht/London: Kluwer Academic Publishers, 2001. v. 2. p. 3-18.

HJORLAND, B. Semantic and Knowledge organization. **ARIST**, v.41, p. 367-405, 2007.

HJØRLAND, Birger. Theories are Knowledge Organizing Systems (KOS **). Knowledge Organization**, v. 42, n. 2, p. 113-128, 2015.

HODGE, Gail. (2000). Systems of knowledge organization for digital libraries: beyond traditional autority files. Washington, D.C.: The Digital Library Federation Council on Library Information Resources.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). Functional Requirements for Bibliographic Records(FRBR), final report. 2009. Disponível em: < [http://www.ifla.org/VII/s13/frbr/>.](http://www.ifla.org/VII/s13/frbr/) Acesso em: 24 maio de 2017.

ISKO UK AGM. 2015. The Great Debate: This House believes that the traditional thesaurus has no place in modern information retrieval. 19 February 2015 London.

ISKO, UKChapter.(2015) Annual General Meeting.Disponível em: [http://www.iskouk.](http://www.iskouk/) org/content/great-debate#EventProgramme Acesso em: 10 jun. 2017.

ISO. International Organization For Standardization. (2011) ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 1.

ISO. International Organization For Standardization. (2013) ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 2.

ISO. International Organization For Standardization. (1986). ISO 2788:1986. Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri.

KLESS, Daniel, SIMON MILTON, Edmund Kazmierczak, JUTTA, Lindenthal. 2015. Thesauri and Ontology Structure: Formal and Pragmatic Differences and Similarities. *Journal of the Association for Information Science and Technology* , v. 66, n. 7, p.1348-1366.

LIMA, G. A.; MACULAN, B. C. M. S. ( 2017). Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. *Ciência da Informação*, Brasilia, v. 46, n. 1, p. 60-72, 2017.

MILES, A.; BECHHOFER, S. (2009). SKOS Simples Knowledge Organization System Reference. *W3C Recommendation.* Disponível em: <http://www.w3.org/TR/skos-reference/.>Acesso em: 08 set. 2018.

MILES, A.; BRICKLEY, D. (2005). SKOS Core. **Guide W3C**: 2005. Disponível em:

<https://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/.>Acesso em: 05 ago. 2018

NISO National Information Standards Organization. ANSI/NISO Z39.19-2005 (R2010) - Guidelines for the construction, formar, and management of monolingual controlled vocabularies. 2010.

ORBST. (2011). The need for ontologies: Bridging the barriers of terminology and data structure. *Special Paper of the Geological Society of America*, n. 482, p.99-123.

PREVOT, L. et al. (2010). Ontology and the lexicon: a multisciplinary perspective. In: HUANG, C. et al. *Ontology and the lexicon: a natural language processing perspective.*New York: Cambridge University Press.

RAMALHO, R. R. A. S. (2015) Análise do modelo de dados skos: sistema de organização do conhecimento simples para a web. *Informação & Tecnologia*, v. 2, n. 1, p. 66-79.

WEISS, Leila Cristina; BRASCHER, Marisa. ( 2013). Princípios teóricos para o estabelecimento de relações semânticas em tesauros. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14. *Anais*... Florianópolis: UFSC.

1. https://orcid.org/0000-0002-7429-8490 Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS), Rio Grande do Sul. rita.laipelt@ufrgs.br | ritacarmo@yahoo.com.br [↑](#footnote-ref-1)
2. https://orcid.org/0000-0002-2882-6803 Doutoranda do Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul. Rio Grande do Sul e do Doctoral Programme Social Sciences at Katholieke Universiteit Leuven (KU Leuven) luciana.monteiro@ufrgs.br [↑](#footnote-ref-2)
3. https://orcid.org/0000-0003-1021-3360 Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS), Rio Grande do Sul.. rene.gabriel@ufrgs.br | renefgj@gmail.com [↑](#footnote-ref-3)