

DESCRIÇÃO DAS RELAÇÕES SEMÂNTICAS PARA APLICAÇÃO EM KOS: USO DO TESAURO SEMÂNTICO APLICADO (THESA)¹

Rene Faustino Gabriel Junior²

Universidade Federal do Rio Grande do Sul
renefgj@gmail.com

Rita do Carmo Ferreira Laipelt³

Universidade Federal do Rio Grande do Sul
ritacarmo@yahoo.com.br

Resumo

Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em KOS. Busca especificamente identificar tipologias de relações semânticas classificadas como relações associativas para aplicação no software Thesa - Tesouro 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 fazer isso é investir no aperfeiçoamento das relações semânticas dos Sistemas de Organização do Conhecimento (SOCs). A inserção das metacategorias semânticas identificadas possibilitam a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento como: Biologia vegetal, Eletromagnetismo, Comunicação Científica e Literatura. A consolidação dos resultados desta pesquisa possibilitará o desenvolvimento de tesouros semanticamente fortalecidos, numa transposição natural entre a teoria e a prática. 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 base em outros tesouros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

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

DESCRIPTION OF SEMANTIC RELATIONS FOR KOS APPLICATION: USE OF APPLIED SEMANTIC THESAURUS (THESA)

Abstract

This work aims to develop a model of semantic metacategories for application in SKOS. It specifically seeks to identify typologies of semantic relations classified as associative relations for application in the Thesa - Thesaurus Applied Semantic software. In the current context, the Information Retrieval Systems (SRI) search interface, with its different search possibilities, plays an intermediate role between the user and the documentary collection of an institution. Thus, it is necessary that these systems be designed in order to facilitate the retrieval of information by users. One way to do this is to invest in improving the semantic relationships of Knowledge Organization Systems (SOCs). The insertion of the semantic metacategories identified allowed the observation and description of their functioning in different domains of knowledge such as: Plant Biology, Electromagnetism, Scientific Communication and Literature. The consolidation of the results of this research will allow the development of semantically strengthened thesauri, in a natural transposition between theory and practice. From the future results of the use of Thesa, there is the possibility of establishing thesaurus between different domains and themes, as well as generating inferences based on other thesauri, facilitating the operationalization of their managers with the use of Artificial Intelligence.

Keywords: Knowledge Organization. Semantic relations. KOS. SKOS. Thesa.

¹ AGRADECIMENTOS - CNPq (439277/2018-3) UFRGS (IC).

²Prof. Dr. do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul

³Professora. Dra. do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul



1 INTRODUÇÃO

O *Knowledge Organization System* (KOS), ou Sistema de Organização do Conhecimento (SOC) em português, são mecanismos de organização de informação centrados principalmente em bibliotecas, museus e arquivos. O qual desempenha um papel fundamental em todas as pesquisas científicas voltadas à criação de conhecimento. A Ciência da Informação (CI) busca melhorar as metodologias e instrumentalizar do KOS de forma a melhorar o processo de recuperação da informação.

Os glossários, vocabulários controlados e tesouros são instrumentos amplamente utilizados pela CI para melhorar a precisão de revocação dos Sistemas de Recuperação de Informação (SRI). Entretanto com o avanço das tecnologias está se discutindo se as descrições das relações dos conceitos dos tesouros e KOS são suficientes para a interpretação tanto para seres humanos, como para interoperabilidade semântica dos computadores.

Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em KOS. 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 KOS será possível sistematizar Metacategorias de relações semânticas para aplicação no software Thesa - Tesouro Semântico Aplicado, com futuras aplicações nos SRIs.

No contexto atual, os (SRI) possibilitam definir diferentes estratégias de busca, e exerce 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 e, também, possibilite a descoberta por outras máquinas, como na mineração de texto (*text mining*), indexação automática (*automatic index*) ou aprendizado por máquina (*machine learning*). E uma das formas de enriquecer a descrição das relações conceituais é pelo fortalecimento das relações semânticas, tanto das descrições como das relações e a criação de microtesouros semânticos em vários domínios e subdomínios.

O uso de técnicas convencionais em SRI, tais como a indexação baseada em palavras-chave ou simplesmente palavras (descritores), não resolvem o problema de RI. Smeaton (1991) destaca que nos SRI podem ter termos com significados ou sentidos diferentes, quando aplicados em outros domínios; ou termos que podem alterar sua definição ou conceito com a alteração de sua ordem ou ligação, como no exemplo de análise estatística da informação e análise da informação estatística; ou ainda, termos que têm representações completamente

diferentes e podem ser utilizadas para expressar um mesmo conceito, como terremoto e abalo sísmico.

Para a recuperação da informação (RI), Hjørland (2007) destaca que as relações semânticas têm a função básica de contribuir para a otimização da precisão e da revocação, enquanto Green (2001) enfatiza que 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. Portanto, é necessário identificar e sistematizar os relacionamentos semânticos para que os mesmos possam ser utilizados de maneira inequívoca em KOSs.

Na literatura encontramos algumas prototipações para melhorias dos SRI para os usuários, como a proposta de Kuramoto (1999), onde identifica o problema de que as interfaces não são amigáveis e existe uma fraca precisão dos resultados de uma busca de informação, principalmente, dificuldades de interação entre o usuário e a interface do SRI. O autor propõe o uso de interfaces com facilidades gráficas, e a possibilidade de apresentação de termos gerais ao utilizado pelo usuário, de forma a orientar ao mesmo o domínio em que está pesquisando, bem como possibilitar melhor interação entre usuário e SRI. Sistemas como este, são possíveis com a integração dos instrumentos de KOS dentro dos SRIs.

A integração desses instrumentos, segundo Green (2001), com a explicitação dos relacionamentos entre os termos, não só traz a compreensão intuitiva para os usuários (seres humanos), mas também possibilita o compartilhamento com dispositivos computacionais. A autora, considera ainda que a explicitação desses relacionamentos “[...] talvez sejam a nossa melhor esperança para infundir qualidade superior em nossos sistemas de recuperação” (GREEN, 2001, p.14).

A indexação tradicional, estruturada na representação sintática do termo, não possibilita uma conexão da linguagem do indexador e do usuário, principalmente pelo usuário não ter acesso ao instrumento utilizado pelo indexador no momento da recuperação nos SRIs, o buscador utiliza apenas os termos exatos do usuário. Caso o usuário queira melhorar os resultados de sua busca, precisa conhecer os índices para ter uma recuperação eficiente e precisa, ou seja, esta recuperação de informação está no nível sintático (CESARINO, 1985; LEIVA, 1999).

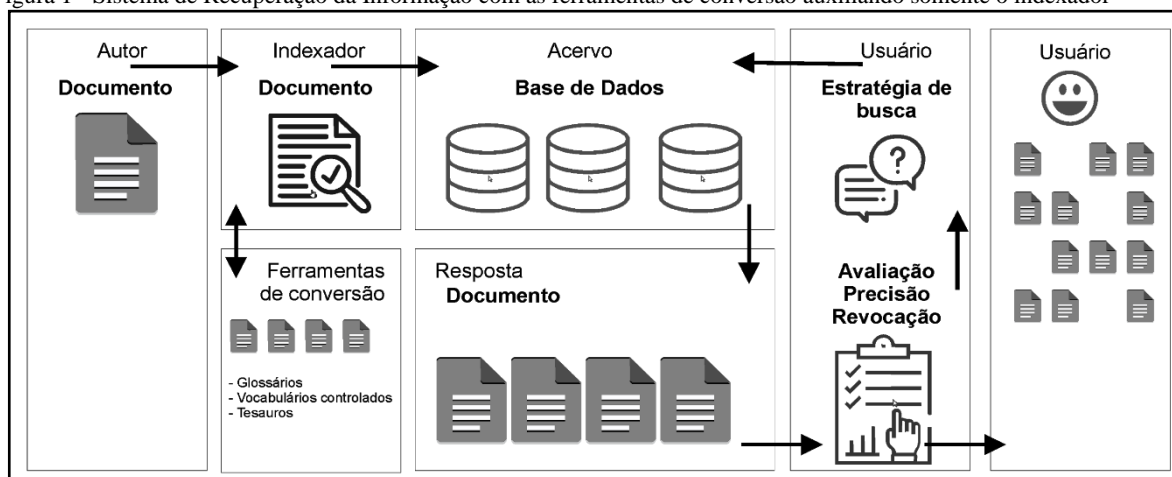
Para Orbst (2011) os sistemas de SRI precisam evoluir da interoperabilidade sintática e estrutural, e buscar a interoperabilidade semântica, com o enriquecimento das descrições das relações tanto para humanos como para computadores. Em relação aos resultados de uma busca, a concepção de um novo SRI deve procurar um novo enfoque de indexação de maneira a

diminuir os problemas relacionados com os procedimentos tradicionais de indexação automática (indexação baseada nas palavras), conforme discussão na seção precedente (KURAMORO, 1999).

Fujita (1989) descreve que o tratamento temático generalizado, com utilização de vocabulários padronizados, produz índices de assuntos gerais que não conseguem revelar o potencial de conhecimentos e ideais contidos em um acervo formado e desenvolvido em função de interesses educacionais e informacionais de uma comunidade de usuários que possui o direito à informação e deve acessá-la, sendo responsabilidade do indexador o desenvolvimento de uma metodologia de indexação capaz de propiciar amplo acesso à informação assegurando a uniformidade de tratamento dos documentos sem perder de vista a flexibilidade do vocabulário. Porém, a indexação é uma operação delicada que lida com ideias a serem transmitidas por termos que as representem, e depende de variáveis subjetivas submetidas à análise de um indexador humano que também possui idéias próprias. Cada termo ou palavra de um texto carrega um significado e uma função dentro de um contexto e uma mesma palavra ou termo empregada em diferentes contextos terá diferentes significados. Quando uma palavra ou termo, significativos, são extraídos isoladamente de um texto os mesmos estarão sendo dissociados de um contexto que lhe impõe um significado e uma função. Uma palavra ou termo usados ao acaso, sem obedecer a um rigor contextual, pode transformar a ideia original do autor do documento e produzir uma outra.

As linguagens documentárias são construídas para serem instrumentos que auxiliam os indexadores nos desafios de representar e disseminar informações contidas nos suportes de informação através de descritores. Este conjunto de descritores “[...] é utilizado para representar sem ambiguidades os assuntos temáticos contidos em documentos (indexação) e em consultas (pesquisas) para recuperar essa informação” (TESAURO ELETRÔNICO DO MINISTÉRIO DA SAÚDE, 201?). A análise conceitual do conteúdo de um documento exige que o indexador fixe uma política de indexação e critérios para a representação destes documentos, pois há diversas possibilidades de representação e interpretação do conteúdo.

Figura 1 - Sistema de Recuperação da Informação com as ferramentas de conversão auxiliando somente o indexador



Fonte: autores (2019).

A Figura 1 apresenta o modelo tradicional dos SRIs, onde as ferramentas de conversão auxiliam somente na seleção dos termos para o indexador, com base no documento analisado. Ou seja, o usuário para conhecer as linguagens utilizadas na indexação precisa ter acesso aos vocabulários, tesouros e glossários utilizados, sendo necessário um conhecimento destes instrumentos. Não possibilitando uma interação direta entre o usuário e o SRI em sua experiência de busca pela informação.

121

Buscando melhorar a interação, Kuramoto (1996) propôs o desenvolvimento de um SRI com o uso de sintagmas nominais e com interações entre o sistema e o usuário, apresentando além dos documentos recuperados, outros sintagmas nominais de nível imediatamente superior (TG), de forma a refinar ou expandir as possibilidades de recuperação do usuário, sem que o usuário conheça as ferramentas de conversão do sistema.

Na literatura encontramos algumas metodologias para ampliar o nível semântico da indexação, como o PRECIS ou PREserved Context Indexing System, criado por Derek Austin em 1968 para construir automaticamente os índices de assunto em cadeia da British National Bibliography (BNB), o qual preocupa-se essencialmente em preservar o contexto do documento através de uma metodologia própria de indexação.

A explicitação de relações semânticas, sobretudo as relações associativas, podem contribuir tanto para o aperfeiçoamento de KOS, no caso específico deste trabalho, do Thesa, como para a recuperação da informação e descoberta de novas fontes (nesse caso, quando tivermos um KOS 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 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 que um SOC abrange diferentes tipos de instrumentos de representação da informação tais como: tesouros, ontologias, taxonomias, redes semânticas entre outros.

Os SOCs produzem maior suporte semântico para os SRI. Considerando que as expressões utilizadas para representar documentos ou a pesquisa de usuários podem ter significados diferentes, sem atributos e links semânticos que contextualizam 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 (aproximação sintática) o que não é suficiente para a recuperação de resultados relevantes (MACULAN; LIMA; OLIVEIRA, 2017), principalmente quando o termo utilizado na busca for uma variação denominativa (sinônimos) ou conceitual (polissemia, um termo com dois significados).

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 tesouros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, folksonomia, e outros tipos de vocabulários controlados. Sua estrutura possibilita o compartilhamento e ligação dos SOCs por meio da Web (MILES; BECHHOFER, 2009).

O SKOS compartilha muito dos princípios de elaboração de vocabulários controlados, taxonomias, tesouros, esquemas de classificação, bem como possibilita capturar muito dessa semelhança e torná-la explícita, para que possa ser reutilizada por outras aplicações (MILES; BECHHOFER, 2009). O modelo é composto por termos em um vocabulário denominado *SKOS Core Vocabulary* (MILES; BRICKLEY, 2005), composto por um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF. Conforme Catarino (2014), o modelo SKOS baseia-se em classes e propriedades para representar o conjunto de dados e tem como elemento central o conceito.

Um tesouro, por exemplo, pode representar o conhecimento de um domínio por meio 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 conseqüentemente 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).

No entanto, Café e Brascher (2011, p. 26) destacam que a seleção dos relacionamentos não é aleatória, tendo sempre uma intencionalidade na 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, além das tradicionais descritas nos tesouros, como TG, TE e TR. Em evento da área, Hjørland (2007, p. 393) questiona se as relações que normalmente são utilizadas nos tesouros são suficientes para descrevê-las, e reflete se “... devemos explicar essa demanda por um conjunto muito mais rico de relações do que as normalmente usadas em tesouros?”. Ainda para o autor, uma função adicional que as relações semânticas em um SOC poderiam desempenhar seria “tornar 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, Quadro 1. As relações paradigmáticas são constituídas por sinonímia, antonímia (oposição conceitual), meronímia (parte de), hiperonímia e hiponímia (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 tesouros 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.

Diante do exposto, é possível afirmar que em tesouros e ontologias teremos relações paradigmáticas nas relações hierárquicas, associativas e de equivalência. Já entre as relações sintagmáticas teremos apenas relações associativas.

Quadro 1. Tipos de Relações Paradigmáticas e Sintagmáticas

Relações paradigmáticas	As relações sintagmáticas
<p>Equivalência:</p> <ul style="list-style-type: none"> ● sinonímia, ● variação denominativa (abreviatura, sigla, termo oculto, flexão verbal, etc) <p>Hierárquica:</p> <ul style="list-style-type: none"> ● hiperonímia (TG), ● hiponímia (TE) <p>Associativa:</p> <ul style="list-style-type: none"> ● meronímia (parte de), ● antonímia (oposição conceitual) ● coordenação (termos subordinados a um mesmo conceito - TG) 	<p>Associativa:</p> <ul style="list-style-type: none"> ● causa/efeito, ● afinidade - parentesco (*), ● características do produto(*), ● processo/agente, ● ação/produto da ação, ● ação/paciente ou objetivo, ● conceito ou coisa/propriedades, ● ação ou coisa/contra agente, ● coisa/suas partes (se não ocorre a relação hierárquica todo/parte), ● matéria prima/produto, ● ação/propriedade, ● campo de estudo/objetos ou fenômenos estudados

(*) não descrito na IFLA

Fonte: Autores (adaptado de IFLA, 2012)

3 THESA – TESAURO SEMÂNTICO APLICADO

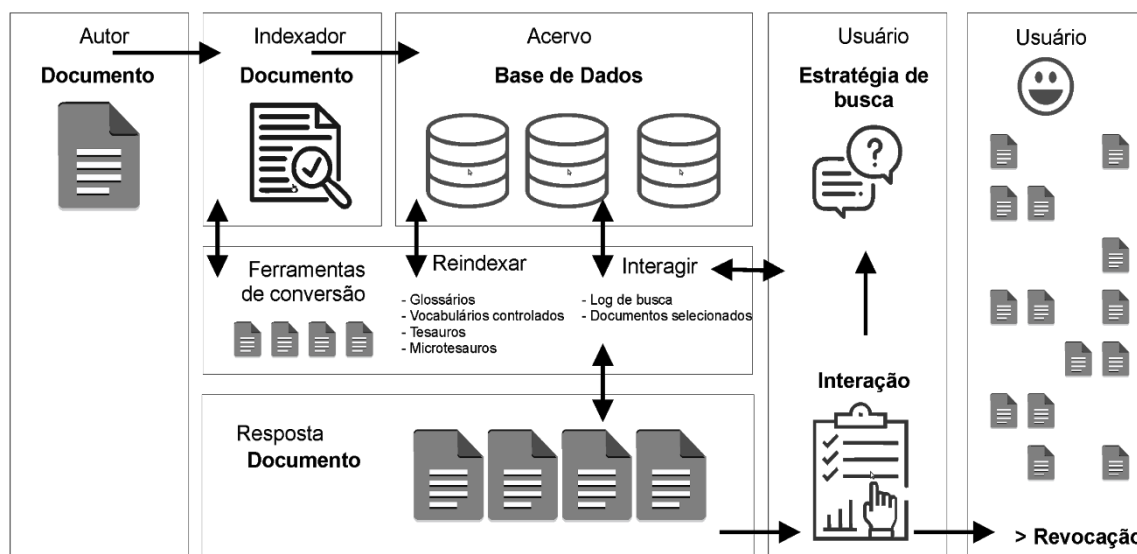
O Tesouro Semântico Aplicado (THESA) foi desenvolvido com o objetivo de disponibilizar um instrumento de elaboração de tesouros para os estudantes de graduação em Biblioteconomia da UFRGS utilizarem na disciplina de Sistema de Classificação III, tornando-se também um objeto de pesquisa.

O objetivo do Thesa é reduzir o trabalho operacional e possibilitar maior atenção ao trabalho de desenvolvimento cognitivo e conceitual referente a modelagem do domínio e na atribuição das relações paradigmáticas e sintagmáticas. Seu desenvolvimento baseou-se nas normas ISO 25.964 (1 e 2) e NISO Z39.50 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 tesouros da ISO e NISO vigentes foram identificados os elementos necessários para o desenvolvimento do protótipo, que atualmente está em sua versão 1.3. O software opera em ambiente Web e pode ser utilizado

gratuitamente no servidor da UFRGS⁴ (GABRIEL JUNIOR; LAIPELT, 2017). O Thesa é distribuído em licença Creative Commons (CCBY) no Github (RENEFGJr/Thesa), podendo ser distribuído e instanciado em servidores de outras instituições, o software é livre para utilização, sendo objeto para fins didáticos em disciplinas dos cursos de graduação, para pesquisa na pós-graduação ou para instrumentalização de bibliotecas e uso profissional.

Figura 1 - Integração da ferramenta (Thesa) no Sistemas de Recuperação de Informação (SRI).



Fonte: autores (2019).

Para tornar o Thesa uma ferramenta mais completa e integrada ao SRI estão sendo desenvolvidas algumas Interfaces de programação de aplicações (API) para identificação de candidatos a termos a partir da análise de textos completos de uma base de dados. Essas ferramentas de conversão devem atuar não somente como instrumentos do indexador, mas como instrumento que deva interagir com o usuário em seu processo de busca de informação, bem como ser um instrumento que analisa as interações e a linguagem do usuário, por meio de seus *logs* de busca e dos documentos que ele selecionou e visualizou, com bases nessas interações. A análise dessas interações, bem como a incorporação de novos vocabulários, microtesouros e de tesouros devem viabilizar um processo de reindexação dos documentos na própria base de dados.

O Thesa foi projetado não somente como ferramenta para descrever relações específicas, mas viabilizando o compartilhando de estrutura e dados em sistemas complexos, como o *Linked data*, mas também mantendo a compatibilidade com modelos simplificados como da ISO e NISO.

⁴ <http://www.ufrgs.br/tesouros>

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 KOS, 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 realizamos a implementação das mesmas no software Thesa.

O Thesa já é utilizado por 277 usuários e faz a curadoria de 194 tesouros. No entanto, muitos deles não estão visíveis no site do software, visto que a publicação ou não dos tesouros prontos é uma opção dos seus autores. A inserção das metacategorias semânticas identificadas até o momento possibilitou a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento. Para demonstração das relações semânticas inseridas no software utilizamos o microtesauros de Eletromagnetismo – Elaborado desenvolvido por Viviane Marques, estudante do curso de biblioteconomia.

O Thesa possibilita a elaboração de microtesauros, que são pequenos tesouros específicos de um domínio. A união de vários destes viabiliza a construção de uma tesouro em um domínio, com uma base relevante e especializada de conhecimento, proporcionando o relacionamento de múltiplas áreas do conhecimento.

De forma conceitual, o tesouro pode ser dividido em 2 tipos: em função da língua e em função do nível de especificidade. Quanto à língua podem ser: Monolíngues (de uma única língua); Bilingues (em duas línguas); ou Multilíngues (em duas ou mais línguas). Os tesouros bilingues e multilíngues, para além da sua função de tratamento e recuperação da informação, servem também como auxiliares nas tarefas de tradução (DODEBEI, 2002).

Quanto à especificidade, dividem-se em: Microtesauros, onde os descritores denotam conceitos com maior nível de especificidade, referindo-se a um domínio mais restrito. Macrotesauros, onde os termos representam conceitos mais ou menos amplos (ARAÚJO, 2012), sendo maior o número de descritores e pertencendo a um mesmo domínio. A junção de vários microtesauros, possibilita a criação ou edição separadamente, formando partes de um

tesauro com base em vários microtesauros, para posteriormente pode efetuar-se a fusão e verificar a integridade do tesauro final.

Um exemplo de microtesauro foi desenvolvido na subárea de Gestão Ambiental (MAIA; VASCONCELLOS SOBRINHO; CONDURÚ, 2016), ele foi estruturado com termos preferidos e ou não preferidos, organizados em ordem alfabética, sendo os termos selecionados apresentados em letras maiúsculas e em negrito. Na obra, a relação conceitual dos termos utilizada nesse microtesauro foi a de forma hierárquica e associativa, sendo a primeira representada pelos termos geral (TG) e específico (TE), enquanto a associativa apresenta o termo relacionado, representado pela sigla TR.

5 RESULTADOS

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.

Por utilizar os conceitos da Web Semântica, os conceitos criados no Thesa são persistentes, ou seja, é atribuído um indicador persistente Handle individual para cada conceito, e esses recebem suas URI. Por exemplo, o conceito “Ciência da Informação” no tesauro semântico “Brapci – FRBR” recebe a identificação de “thesa:c6290”. Este rótulo acompanha o conceito e suas definições de forma definitiva, não mais se alterando, sendo referenciado pelo identificador permanente <http://hdl.handle.net/10.500.11959/c/6290>. Os indicadores persistentes facilitam a localização das descrições do conceito, pois mesmo que o link do servidor seja modificado, existe a garantia do provedor de manter o Handle ativo.

O Thesa possibilita interoperabilidade, comunicação com outros programas, utilizando os *Application Programming Interface* (API) ou Interface de programação de aplicações, que por meio de um *Token* (dispositivo eletrônico gerador de senhas) possibilita acesso de outras aplicações sem a necessidade de intervenção humana.

Como descrito anteriormente, as relações simplistas do TG, TE e TR não são o suficiente para enriquecer as descrições das relações semânticas, desta forma foram incorporados no Thesa as Metacategorias, no Quadro 2 e 3 são apresentados as relações associativas e de equivalência incorporados no Thesa até o momento.

Quadro 2 - Apresentação das Metacategorias Semânticas Hierárquicas e de Equivalências nos tesouros tradicionais e no Thesa

<i>Tesouro</i>	<i>Tesouro Semântico</i>	<i>Exemplos de aplicações</i>
TG	TG (é tipo de ...)	“Biblioteca Universitária” é um tipo de “Biblioteca”
	TG (é parte de ...)	“Porto Alegre” é parte do “Rio Grande do Sul”*
USE	UP (abreviatura)	“Pag.” é abreviatura de “Página”
	UP (sigla)*	“AACR2” é a sigla de “Código de Catalogação Anglo Americano, segunda edição”
	UP (acrônimo)	“UFRGS” é a acrônimo de “Universidade Federal do Rio Grande do Sul”
	UP (é variação de)	“Catalogação” é variação de “Representação Descritiva” (Sinônimo)
	UP (é um gerúndio de)	“Catalogando” é um gerúndio de “Catalogar”
	UP (é plural de)*	“Livros” é plural de “Livro”
	UP (é singular de)*	“Livro” é singular de “Livros”
	UP (é masculino de)*	“Bibliotecário” é o masculino de “Bibliotecária”
	UP (é o feminino de)*	“Bibliotecária” é o feminino de “Bibliotecário”
	UP (é o extenso de)* (é o termo completo)	“Cinco” é o extenso de “5”
	UP (é a notação de)*	CDU02 é a notação de “Biblioteconomia”
	UP (é termo oculto)*	“Ranganatam” é o termo oculto de “Ranganathan”
	UP (é flexão verbal de)*	“Pesquisaram” é flexão verbal de “Pesquisar”

* Proposição dos autores

Fonte: Autores (2019).

A relação hierárquica do tipo todo-parte abrange um número limitado de situações em que uma parte de uma entidade única ou sistema pertence a um determinado conjunto. É uma relação que continua a ser aplicada a quatro classes principais de termos, são eles: a) sistemas e órgãos do corpo; b) localidades geográficas; c) disciplinas ou áreas do conhecimento e d) estruturas sociais hierarquizadas. A principal diferença entre a norma anterior e a norma vigente, sob esse aspecto, é que a anterior era mais flexível quanto à utilização da relação todo-parte em outras classes de termos, além das quatro principais. A norma vigente sugere o uso apenas das quatro classes principais e desencoraja sua utilização como relação hierárquica em outros casos possíveis do relacionamento todo-parte. De acordo com a ISO 25964-1,

A maioria dos outros casos do relacionamento todo-parte não é elegível para uma ligação hierárquica, porque a parte pode pertencer a mais de um todo. Por exemplo, não deve ser estabelecida uma relação TG / TE entre bicicletas e rodas porque uma roda poderia ser parte de um automóvel, um carrinho de mão ou de muitos outros objetos. (INTERNATIONAL..., p. 60, tradução nossa).

As relações de equivalência, foram propostas com base na norma ISO25964-1, e a partir da análise de logs do Portal LexMI realizada por Laipelt (2015) onde percebeu-se que, além das tipologias sugeridas pela norma (sinônimos, quase-sinônimos, abreviaturas, acrônimos e termos completos) os usuários também utilizam siglas, termos no singular, plural, masculino, feminino, flexão verbal, números e termos com erro de grafia (termo oculto).

O Quadro 3 apresenta as metacategorias selecionadas para descrição no Thesa selecionada:

Quadro 3 - Apresentação das Metacategorias Semânticas Associativas nos tesauros tradicionais e no Thesa

<i>Tesouro</i>	<i>Tesouro Semântico</i>	<i>Exemplos de aplicações</i>
TR	TR (causa/efeito)	Pouca Leitura (<i>causa</i>) Ignorância (<i>efeito</i>) Inconsistência (<i>causa</i>) Baixa precisão (<i>efeito</i>) Consistência da indexação (<i>causa</i>) Precisão na recuperação da informação (<i>efeito</i>)
	TR (afinidade)	“Arquivista” <i>tem afinidade com o</i> “Bibliotecário” (profissional da informação)
	TR (características do produto)	“Livro” <i>é de</i> “Papel”
	TR (agente/processo)	“Livro” (agente) <i>é</i> “Catalogado” (<i>processo</i>)
	TR (ação/produto da ação)	“Indexação”(<i>ação</i>) <i>produz</i> “Índices” (<i>produto da ação</i>)
	TR (ação/paciente ou objetivo)	“Restauração” (<i>ação</i>) <i>é realizada no</i> “Documento” (<i>paciente</i>)
	TR (ação/propriedade)	“Tombamento” (<i>ação</i>) <i>gera um</i> “Número de Patrimônio” (<i>propriedade</i>)
	TR (campo de estudo/objetos ou fenômenos estudados)	“Ciência da Informação” <i>estuda a</i> “Comunicação Científica”
	TR (ação ou coisa/contra agente)	“Deterioração por umidade” (<i>ação</i>) <i>é combatida com</i> “Ventilação” (<i>contra agente</i>)
	TR (conceito ou coisa/propriedades)	“Livro” (conceito) <i>tem no mínimo</i> “50 páginas” (<i>propriedade</i>)
	TR (coisa/suas partes)	“Estante” (<i>coisa</i>) <i>é composta de</i> “Prateleiras” (<i>suas partes</i>) (<i>quando não for relação hierárquica TG</i>)
	TR (matéria prima/produto)	“Aço” <i>é matéria prima das</i> “Estantes” (<i>produto</i>)
	TR (antonímia, oposição conceitual)	“Acervo Aberto” <i>é antonímia de</i> “Acervo Fechado”

	TR (coordenação)	“Biblioteca Universitária” (<i>conceito 1</i>) e “Biblioteca Especializada” (<i>conceito 2</i>) são “Tipos de Bibliotecas”
--	------------------	--

No caso das relações associativas, a norma ressalta que é possível subdividi-la em diferentes tipos. Desta forma, ao invés de simplesmente utilizar a convenção TR (termo relacionado) pode-se utilizar uma convenção que especifique a natureza dessa relação associativa. Por exemplo, no caso de uma relação de causa e efeito, a norma sugere que se utilize como rótulo CAUSA/EFEITO. De acordo com a norma vigente, especificar dessa maneira as relações entre os termos é opcional, tendo em vista o trabalho extra que isso representaria. Porém, se houver o desejo de disponibilizar o tesauro elaborado em rede, visando sua interoperabilidade, a realização dessa atividade extra se torna fundamental.

Figura 2 - Exemplo de representação do conceito “Campo Magnético” no Thesa

The image shows a screenshot of a Thesa entry for "Campo Magnético". At the top, it displays the title "Campo Magnético" and the URL "https://www.ufrgs.br/tesauros/index.php/thesauro/12334". Below the title, there is a list of related terms (TR) with their respective relationships (TG: Magnetismo). The list includes: "Corrente induzida", "Energia Magnética", "Força de Lorentz", "Força Magnética", "Lei de Ampère", "Lei de Biot e Savart", "Lei de Lenz", "Linha de Indução", "Carga Elétrica", and "Corrente Elétrica". To the right of the list, there are two definitions of the concept: "Todo campo magnético se estabelece por cargas elétricas em movimento." and "Cargas elétricas em movimento (corrente elétrica) criam, no espaço em torno delas, um campo magnético." Below the definitions, there is a diagram showing magnetic field lines (B) and a moving charge (q) with velocity (v). At the bottom of the screenshot, there is a semantic network diagram with "Campo Magnético" at the center, connected to various related terms like "Magnetismo", "Força Magnética", "Corrente Elétrica", "Lei de Lenz", "Força de Lorentz", "Corrente induzida", "Energia Magnética", "Linha de Indução", "Lei de Ampère", "Carga Elétrica", "Lei de Biot e Savart", and "Lei de Ampère".

Fonte: Marques (2018).

Na Figura 2 tem a representação da entrada do termo “Campo Magnético” retirado do Microtesauro de “Eletromagnetismo” elaborado no Thesa. Pode observar-se que a explicitação das relações semânticas enriquece a compreensão do significado do termo e suas relações com

outros termos do domínio. Como observado nas associações do termo de entrada com o tipo “Coordenação”, indicando “Corrente induzida”, “Lei de Ampère” e “Carga Elétrica” por estarem subordinados ao mesmo TG “magnetismo”. As relações do tipo “TR (Ação/Produto)” podem ser visualizados nas relações com os termos “Força de Lorentz”, “Força Magnética”, “Lei de Biot e Savart” e “Linha de Indução”. Descrevendo as relações de “TR (Causa/Efeito)” tem-se os termos “Energia Magnética”, “Lei de Lenz” e “Corrente Elétrica”.

O microtesauro é uma representação do domínio que se materializa a partir do ponto de vista de quem o constrói. A união de vários microtesauros (interoperabilidade) possibilita a construção de um tesauro mais complexo, em função do aumento do número de termos, remissivas e relações semânticas representadas em cada microtesauro. Corroborando Hjørland (2007) quando destaca que as explicações das relações semânticas tornam visíveis os diferentes interesses e paradigmas dentro do domínio, ampliando as opções para que o usuário possa fazer uma escolha melhor embasada.

6 CONSIDERAÇÕES FINAIS

O estudo propõe descrever as relações semânticas para aplicação em um sistema de organização do conhecimento. Para concretizar a proposta, foram inseridas no Thesa as metacategorias possibilitando a explicitação das relações entre os termos em microtesauros. Considera-se que o objetivo deste estudo foi atingido, com a apresentação das relações em uma realidade concreta.

Espera-se que essas atualizações e novas funcionalidades que estão sendo desenvolvidas para o Thesa estimulem os SRIs a incorporem sua utilização. O enriquecimento das relações semânticas entre os termos permite a contextualização dos mesmos, tanto por parte do usuário como do profissional da informação. A integração do Thesa irá gerar como resultado prático a recuperação de informação mais precisa, assim como a descoberta de informações a partir da recomendação de novos termos de busca, principalmente por proximidade do campo semântico. Outro resultado do uso do Thesa, está a possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas (microtesauros), bem como gerar inferências com base em outros tesauros, possibilitando estudos futuros com o uso de Inteligência Artificial.

Finalmente, a interligação do Thesa com bases de dados, como a Brapci por exemplo, possibilitará a troca de informações entre ambos viabilizando a construção de vocabulários especializados das áreas de conhecimento para as quais se tenham tesauros elaborados. Esses vocabulários por sua vez servirão de sustentação léxica para o reconhecimento de candidatos a

termos que podem ser extraídos automaticamente dos textos completos e apresentados aos bibliotecários e/ou gestores do Thesa para a tomada de decisão quanto a escolha dos termos a serem inseridos como descritores ou como remissivas em um tesouro. Esta funcionalidade irá tornar o desenvolvimento de tesouros mais rápido bem como linguística e semanticamente mais rica.

Para estudos futuros, propõem-se análise da integração de microtesouros da área de Ciência da Informação com os anais dos eventos da ISKO-Brasil, buscando analisar as funcionalidades técnicas e seu impacto para a representação e recuperação da informação no contexto dos usuários e dos indexadores.

REFERENCIAS

ARAÚJO, E. C. C. A. Criação de um microtesauros multilíngue na área da enfermagem. **Dissertação...** Universidade Lusófona de Humanidades e Tecnologias. Lisboa, 2012.

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. **Informação e Informação**, v. 16. n. 3. p. 25-51.

CANÇADO, M. **Manual de semântica: noções básicas e exercícios**. Belo Horizonte: UFMG, 2008.

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 Informação e Documentação. Brasília: Universidade de Brasília.

DODEBEI, V. **Tesouro: linguagem de representação da memória documentária**. Rio de Janeiro: Interciência: Niterói: Intertexto, 2002.

FUJITA, M. S. L. Avaliação da eficácia de recuperação do sistema de indexação precis. **Ciência da Informação**, v. 18, n. 2, 1989. DOI: 10.18225/ci.inf.v18i2.304 Acesso em: 09 set. 2019.

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

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

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

HODGE, G. **Systems of knowledge organization for digital libraries** : beyond traditional authority files. Washington, D.C.: The Digital Library Federation Council on Library Information Resources, 2000.

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

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

IFLA. **Requisitos Funcionales para Datos de Autoridad de Materia (FRSAD)**: un modelo conceptual. Netherlands: IFLA, 2012. 66 p. Disponível em: <<http://www.ifla.org/news/now-available-frsad-in-spanish>>. Acesso em: 07 Ago. 2012.

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. Annual General Meeting, 2015. Disponível em: <http://www.iskouk.org/content/great-debate#EventProgramme> Acesso em: 10 jun. 2017.

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

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.

KLESS, D.; SIMON MILTON, E. K.; JUTTA, L.. 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.

KURAMOTO, H. Proposta de um sistema de recuperação de informação assistido por computador - sriac. **Revista de Biblioteconomia de Brasília**, v. 21, n. 2, 1997. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/75896>>. Acesso em: 09 set. 2019.

KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual : os sintagmas nominais. **Ciência da Informação**, v. 25, n. 2, 1996. DOI: 10.18225/ci.inf..v25i2.655 Acesso em: 09 set. 2019.

LAIPELT, R. C. F. Metodologia para seleção de termos equivalentes e descritores de tesouros: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. **Tese (Doutorado)** - Programa de Pós-Graduação em Linguística Aplicada, Escola da Indústria Criativa : comunicação, design e linguagens, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2015.

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

MAIA, P. C. M; VASCONCELLOS SOBRINHO, M.; CONDURÚ, M. T. **Microtesauro de Gestão Ambiental**. Numa/UFGA, Belém, 2016.

MARQUES, V. (2018). Thesa: Eletromagnetismo. Porto Alegre : UFRGS. Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa/terms/75>. Acesso em: 10 jul. 2019.

MILES, A.; BECHHOFFER, S. (2009). SKOS Simple 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:

NISO National Information Standards Organization. **ANSI/NISO Z39.19-2005 (R2010)** - Guidelines for the construction, format, 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 multidisciplinary 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.

SMEATON Alan F. Prospects for intelligent, language-based information retrieval. **Online Review**, v. 15, n. 6, p. 373-382, 1991.

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