

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-2 – Organização e Representação do Conhecimento

REQUISITOS SEMÂNTICOS PARA A MODELAGEM E APERFEIÇOAMENTO DE TESAUROS

Rene Faustino Gabriel Junior - Universidade Federal do Rio Grande do Sul (UFRGS)

Rita do Carmo Laipelt - Universidade Federal do Rio Grande do Sul (UFRGS)

SEMANTIC REQUIREMENTS FOR MODELING AND IMPROVING OF THE THESAURUS

Modalidade da Apresentação: Pôster

Resumo: A Ciência da Informação está sendo demandada para repensar alguns de seus instrumentos de organização da informação, entre eles estão os tesauros. O estudo tem como objetivo buscar uma reflexão sobre a aplicabilidade de requisitos semânticos de contextualização de domínio na modelagem de sistemas de organização do conhecimento, principalmente nos tesauros. Buscou-se na literatura subsídios para compor a reflexão acerca das relações semânticas entre os conceitos, e a utilização do Thesa como instrumento de aplicação concreta das relações semânticas identificadas. Mesmo o estudo estando em fase inicial, os resultados já apontam que o estabelecimento dos tipos de relações existentes entre os conceitos demanda muita reflexão, principalmente pela complexidade em modelar e explicitar algumas tipologias de relações.

Palavras-Chave: Organização do Conhecimento. Tesauros. Relações Semânticas. Thesa.

Abstract: Information Science is being demanded to rethink some of its instruments of information organization, among them are the thesauri. The aim of this study is to investigate the applicability of semantic requirements of domain contextualization in the modeling of knowledge organization systems, especially in thesauri. In the literature, we sought the support of composing the reflection about the semantic relations between the concepts, and the use of Thesa as an instrument of concrete application of the semantic relations identified. Even though the study is in its initial phase, the results already indicate that the establishment of the types of relations existing between the concepts demands a lot of reflection, mainly due to the complexity in modeling and explaining some types of relations.

Keywords: Organization of Knowledge. Thesauri. Semantic Relationships. Thesa.

1 INTRODUÇÃO

Na área da Organização da Informação está emergindo discussões e questionamentos sobre o quanto é apropriado o uso dos tesauros como instrumento para a organização e

recuperação de informações, sobretudo os tesouros tradicionais com base nas normas ISO e NISO. Estes questionamentos são oriundos principalmente dos modelos propostos pela Ontologia formal, que baseiam seus questionamentos na proposição universal de conhecimento e interoperabilidade.

Em 2015, mais especificamente na Inglaterra, o Capítulo Inglês da ISKO promoveu um evento incitando a comunidade da Ciência da Informação para um “Grande Debate”, cujo tema era “Esta casa acredita que os tesouros tradicionais não têm nenhum lugar na recuperação moderna da informação” (ISKO UK AGM, 2015, tradução nossa), influenciados principalmente pelas novas ferramentas da internet como a Web Semântica (BERNERS-LEE, HENDLER, LASSILA, 2001) e os dados Ligados (BIZER; HEATH; BERNERS-LEE).

Nessa discussão, Hjørland (2016b) defende a prevalência dos tesouros em detrimento das ontologias nos Sistemas de Recuperação de Informação (SOC), citando Kless e outros autores (2014) que apontaram doze diferenças fundamentais entre os tesouros e as ontologias. Ponderando que as ontologias formais não são adequadas para os SOCs, e que os tesouros tradicionais precisam evoluir principalmente nas descrições de suas relações.

Dessa forma, Hjørland (2016b) ressalta que a "relação associativa" nos tesouros, como feita hoje, é particularmente infrutífera. Para o autor devem existir agrupamentos de relações bem definidas, como antônimos, causa-efeito, relações sequencias, entre outros, e esses agrupamentos se feitos de forma explícita ficariam melhor do que a forma utilizada nos tesouros tradicionais.

Esse trabalho tem como objetivo buscar uma reflexão sobre a aplicabilidade de requisitos semânticos de contextualização de domínio na modelagem de sistemas de organização do conhecimento, principalmente nos tesouros.

2 RELAÇÕES SEMANTICAS

Com a ampliação da capacidade de armazenamento e processamentos dos computadores a forma como organizamos e recuperamos a informação está se alterando. Antes, os catálogos eram limitados a alguns pontos de acesso reduzindo as opções do usuário. Na recuperação de Informação Moderna já não existem mais essas limitações, sendo possível realizar buscas mais elaboradas além dos campos tradicionais com autor, título e assunto; ampliando a busca para os conteúdos dos resumos, textos completos e citações, ampliando as possibilidades de recuperação (HJORLAND, 2016b). Desta forma os usuários estão diante

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

de uma nova forma de aquisição de informação, que de acordo com Hjørland (2016a), possibilita aos mesmos apreenderem sobre os gêneros, as terminologias e as relações de citações entre os trabalhos, e a distinguirem entre as diferentes "vozes" que fazem parte de um domínio, adequando suas estratégias de busca durante sua interação com o SRI.

Em um contexto histórico, lembramos que a Ciência da Informação, a partir de 1940, passou a utilizar os tesouros, em especial no processo de recuperação da informação, como um instrumento capaz de transportar conceitos e suas relações mútuas, como expressas nos documentos em uma linguagem regular, com controle de sinônimo e estruturas sintáticas simplificadas (DODEBEI, 2002).

Dessa forma, um tesouro pode representar o conhecimento de um domínio por meio de um conjunto de 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. Nos tesouros, “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 determinada representação do conhecimento” (CAFÉ; BRASCHER, 2011, p. 26). 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, 2016).

Desta forma, 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 tesouros a partir da explicitação das relações associativas. No entanto, de acordo com Maculan, Lima e Oliveira (2016) as relações associativas são consideradas as mais difíceis de definir e sobre as quais ainda não existe pesquisa suficiente para determinar suas bases teóricas.

Soergel (1974) destaca que as relações associativas devem explicitar de que forma um conceito A possibilite lembrar do conceito B, porém não existindo equivalência entre eles. Entretanto Shirky e outros argumentam que não há hierarquia do conhecimento, apenas links,

mas se você deseja gerenciar terminologia e se você quiser usá-lo para encontrar o material, é necessário impor alguma estrutura ordenada.

De acordo com Lancaster (1986) não é possível estabelecer regras precisas o bastante para este relacionamento do mesmo modo que se pode fazer para as relações hierárquicas. Hutchins (1975) tem um ponto de vista similar, e destaca que a dificuldade em definir esse tipo de relacionamento leva a resultados arbitrários e não sistemáticos. A norma NISO Z39.19 (2010) também destaca a importância da explicitação da relação associativa entre os descritores vinculados, de forma a evitar julgamentos subjetivos e ou inconsistentes. Tudhope, Alani e Jones (2005) complementam a visão dos autores anteriormente mencionados, alertando que muitas vezes um TR é utilizado para representar nada mais do que uma conexão vaga, do tipo “ver também”, que pode gerar uma expansão incontrolada de relações pouco significativas, podendo gerar perda de precisão.

Marroni (2006) e Olson (2007) observam que a norma de elaboração de tesouros, ANSI / NISO Z39.19, prevê um conjunto limitado de relações de termos relacionados (TR) permitidos: processo / agente, ação / propriedade, ação / produto, ação / alvo, causa / efeito, conceito de objeto / propriedade, conceito ou objeto / origem, conceito ou objeto / medidas, matéria-prima / produto e disciplina ou campo / objeto ou praticante, bem como de oposição (antônimos).

Os diferentes tesouros tendem a desenvolver de modo singular sua terminologia e relações preferidas, assim, produzem outra "voz" que o usuário deve considerar para formular sua estratégia de busca. Os tesouros devem fornecer sistemas mais descritivos e menos prescritivos, justificando cuidadosamente as escolhas prescritivas. Nesse sentido, a área de organização do conhecimento questiona se os tesouros tradicionais têm condições de atender às demandas geradas em função dos recursos disponíveis atualmente para a recuperação de informação.

Verifica-se na literatura da área de organização do conhecimento o incentivo ao uso de outros tipos de relações semânticas. Em vista disto, Hjørland (2007, p. 393) questiona “Como devemos explicar essa demanda por um conjunto muito mais rico de relações do que as normalmente usadas em tesouros?”. 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 sua escolha informado” (HJORLAND, 2007, p. 389, tradução nossa).

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

Nessa perspectiva de ampliar a descrição das relações semânticas nos tesouros, foi proposto o desenvolvimento de um instrumento que possibilite a explicitação das relações entre os conceitos de modo que seja possível para o usuários e computadores, distinguirem as diferentes tipologias de relações existentes. Nos tesouros tradicionais é possível estabelecer relações entre os conceitos, entretanto as especificidades dessas relações não são explicitadas suficientemente para compor uma semântica forte, e viabilizar a contextualização dos termos para que os usuários possam identificar o tipo específico de relação estabelecida.

Embora a utilização das relações hierárquica dentro dos tesouros possibilite a identificação do Termo Geral (TG) e Termo Específico (TE) de uma relação, isso não é suficiente para estabelecer uma semântica forte visto que o uso padronizado de TG e TE não contextualiza nem explicita a relação que de fato ocorre entre os termos. É possível verificar um exemplo disto, a partir da análise do termo “biblioteca” extraído do Tesouro em Ciência da Informação de Pinheiro e Ferrez (2014), de acordo com Gabriel Junior e Laipelt (2017) o termo “biblioteca” é apresentado como Termo Geral para “bibliotecas centrais”, “bibliotecas departamentais” e “bibliotecas públicas”, que são termos semanticamente vinculadas devido a estrutura da instituição. Entretanto, o termo “biblioteca” também é superordenado para “bibliotecas híbridas” e “bibliotecas digitais”, que são termos semanticamente vinculados devido ao tipo de acervo disponibilizado. Logo, temos duas tipologias de relações diferentes tratadas de modo generalizado como se fossem iguais.

De forma experimental está sendo desenvolvido pelos pesquisadores o Thesa, uma ferramenta flexível, que possibilita a explicitação e consequente contextualização das relações entre os conceitos (GABRIEL JUNIOR; LAIPELT, 2017), para isso utiliza-se a modelagem RDF. Segundo Pastor-Sanchez, Martínez-Mendez e Rodríguez-Muñoz (2012) uma das vantagens da aplicação RDF, está na possibilidade de expandir os tipos de relacionamento com OWL ou simplesmente com RDF, permitindo a criação de novas relações, tais como transitividade, simetria, relação inversa ou reflexividade, declarando rigorosamente as restrições semânticas necessárias.

Para reforçar a semântica forte as notas gerais, definidas pelas normas ISO (2011) e NISO (2010), foram desmembrados em notas mais específicas como a “garantia literária”, “exemplo de aplicação”, “cronologia de atualização do termo” e a inserção das “referências”, de modo que o usuário possa identificar as características específicas do termo no domínio. Outros tipos de notas podem ser criados no Thesa com a incorporação de novas propriedades

RDF. O detalhamento das notas é importante, pois definem seu contexto histórico, aplicações e formas de uso, além de auxiliar os SRI baseados na Web.

Nos tesouros tradicionais, como mostrado no Quadro 1, a representação da relação pelo padrão TG, TE e TR não qualificam o tipo de relação, entretanto, no Thesa é possível qualificar as relações por meios da explicitação das propriedades que caracterizam as relações estabelecidas, sem deixar de contemplar o padrão tradicional TG, TE e TR.

Quadro 1: Exemplos de relações semânticas dos tesouros e do Thesa.

Tesauro			Thesa		
Termo	Relação	Termo	Conceito	Propriedade	Conceito
Biblioteca Escolar	TG	Biblioteca	Biblioteca Escolar	é uma tipologia de	Biblioteca
Biblioteca	TE	Biblioteca Escolar	Biblioteca	tem como tipo	Biblioteca Escolar
Biblioteca digital	TG	Biblioteca	Biblioteca digital	tipo de acervo da	Biblioteca
Biblioteca	TE	Biblioteca digital	Biblioteca	tem acervo do tipo	Biblioteca Digital
Biblioteca Escolar	TR	Centros de Multimeios	Biblioteca Escolar	tem como parte (todo / parte)	Centros de Multimeios
Centros de Multimeios	TR	Biblioteca Escolar	Centros de Multimeios	faz parte da (todo / parte)	Biblioteca Escolar

Fonte: Autores (2017).

3 CONSIDERAÇÕES FINAIS

Atualmente está sendo questionado a relevância dos tesouros tradicionais na recuperação de informação moderna (HJORLAND, 2016a). Neste contexto, este estudo buscou refletir sobre a qualificação das relações semânticas, e nos levou a pensar em alternativas para o aperfeiçoamento dos tesouros tradicionais. Nesse sentido está sendo desenvolvido o Thesa como instrumento concreto para viabilizar a construção de tesouros semânticos aplicados em diversos domínios. Ainda em caráter experimental já foi possível utilizar o Thesa para trabalhar com o desmembramento de notas para contextualização do domínio dos termos. As próximas etapas da pesquisa partem da identificação das relações prototípicas em diversos domínios, a partir da construção de vários tesouros, de forma a identificar e descrever propriedades das relações sintagmáticas.

Embora o estudo ainda esteja em execução, já é possível perceber que o estabelecimento dos tipos de relações existentes entre os conceitos demanda muita reflexão, pois nem sempre é simples modelar e explicitar de maneira clara alguns tipos de relações, porém depois de estabelecida a tipologia, a aplicação pelo usuário se torna mais fácil.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

REFERÊNCIAS

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. **Scientific American**, May, 2001. Disponível em: <<http://www.scian.com/2001/0501issue/0501bernerslee.html>>. Acesso em: 24 maio 2017.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked Data - The Story So Far. **International Journal on Semantic Web and Information Systems (IJSWIS)**, v. 5, n. 3, p. 1-22, 2009. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswislinked-data.pdf>>. Acesso em 26 maio 2017.

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

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

HJORLAND, B. Does the Traditional Thesaurus Have a Place in Modern Information Retrieval? **Knowledge Organization**, v. 43, n. 3, 2016a.

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

HJORLAND, Birger. Theories are Knowledge Organizing Systems (KOS). **Knowledge Organization**, v. 42, n. 2, p. 113-128, 2016b.

MARRONI, G. N. B. **Identificação e delimitação de relações associativas em tesouros**: um estudo de caso na área do Direito do Trabalho. 2006. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2006.

HUTCHINS, W.J. **Languages of indexing and classification**: a linguistics study of structures and functions. Herts (England): Peter Peregrinus, 1975.

IFLA Study Group on the Functional Requirements for Bibliographic Records. **Functional Requirements for Bibliographic Records (FRBR)**, final report. 2009.

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**. Disponível em: <<http://www.iskouk.org/content/great-debate#EventProgramme>>. Acesso em 15 maio 2017.

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

LANCASTER, F. W. **Vocabulary control for information retrieval**. 2.ed. Arlington, Virginia: 1986. 270p.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. F. Thesa: ferramenta para construção de tesouro semântico aplicado interoperável. **Revista P2P e INOVAÇÃO**, v. 3, n. 2, 2017.10.21721/p2p.2017v3n2.p124-145. DOI:10.21721/p2p.2017v3n2.p124-145.

KLESS, D.; SIMON MILTON, E. K.; LINDENTHAL, J. Thesauri and Ontology Structure: Formal and Pragmatic Differences and Similarities. **Journal of the Association for Information Science and Technology**, v. 66, n. 7, 2014. DOI: 10.1002/asi.23268.

MACULAN, B. C. M. S.; LIMA, G. B. O.; OLIVEIRA, E. D. Conversion Methods from Thesaurus to Ontologies: A Review. **Advances in Knowledge Organization**, v.15, p. 300-307, 2016.

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

PASTOR SANCHEZ, J.A.; MARTINEZ MENDEZ, F.J.; RODRIGUEZ MUÑOZ, J.V. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. **El Profesional de la Información**, v.21, n.3, p.245-253, 2012.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), 2014.

OLSON, H. A. How we construct subjects: A feminist analysis. **Library Trends**, v. 56, n. 2, 2007, 509–54.1

PREVOT, L. et al. 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, 2010.

SOERGEL, D. **Indexing languages and thesauri**: construction and maintenance. New York: Wiley, 1974.

TUDHOPE, Douglas; ALANI, Harith; JONES, Christopher. Augmenting thesaurus relationships: possibilities for retrieval. **Journal of Digital Information**, v. 1, n. 8, 2001.

WEISS, L. C.; BRÄSCHER, M. Relações semânticas em tesouros: contribuições da abordagem pragmática. InCID: **Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 7, n. 2, p. 136-155, oct. 2016.