**Representação de um microtesauro em SKOS: um estudo metodológico**

Marilia Winkler de Morais1, Janailton Lopes Sousa2, Rogério Aparecido Sá Ramalho3

1 Graduanda do curso de Biblioteconomia e Ciência da Informação Universidade Federal de São Carlos – UFSCar, 0000-0002-5234-7239, winklermariliam@gmail.com

2 Mestrando do Programa de Pós Graduação em Ciência da Informação da Informação Universidade Federal de São Carlos – UFSCar, 0000-0002-2923-2441, janailtonlopes20@gmail.com

3 Professor Associado Universidade Federal de São Carlos – UFSCar, 0000-0002-8491-3514, ramalho@ufscar.br

**Resumo:** Com o aumento exponencial da quantidade de recursos disponibilizados em ambientes digitais, os instrumentos e métodos de representação e organização da informação são de expressiva importância para garantir a recuperação de recursos informacionais ultrapassando fronteiras culturais e geográficas de acesso à informação. Entre os novos instrumentos de representação destaca-se o *Simple Knowledge Organization System* – SKOS, um modelo de dados que tem como objetivo representar a estrutura básica e o conteúdo dos tradicionais sistemas de organização do conhecimentoem ambientes computacionais. O presente trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelo CNPq e pela FAPESP e possui como objetivo avaliar o uso do padrão SKOS para a representação de um microtesauro baseado na categoria ‘Informação e Conhecimento Estratégicos nas Organizações’ do Tesauro Brasileiro de Ciência da Informação. Trata-se de uma pesquisa de cunho teórico e metodológico que se propõe a abordar os temas de representação, organização e disponibilização de vocabulários em ambientes digitais. Caracteriza-se, também, como pesquisa de natureza aplicada e exploratória, uma vez que objetiva gerar conhecimentos para aplicação prática. A partir do estudo realizado foram identificadas as potencialidades e desafios da representação de vocabulários controlados em SKOS, destacando a necessidade de um maior aprofundamento em estudos que possibilitem mapear diretrizes para o estabelecimento dos diferentes tipos de relações previstas em SKOS.

**Palavras-Chave**: Representação da Informação; Tecnologias da Informação e Comunicação; Vocabulários Controlados; Sistemas de Organização do Conhecimento; Tecnologias Semânticas;

**Abstract:** With the exponential increase in the amount of resources available in digital environments, the instruments and methods of representation and organization of information are of significant importance to ensure the recovery of information resources across cultural and geographical borders of access to information. Among the new tools of representation, the Simple Knowledge Organization System (SKOS) stands out as a data model that aims to represent the basic structure and content of traditional knowledge organization systems in computational environments. The present work is linked to an ongoing research project, funded by CNPq and FAPESP and aims to evaluate the use of the SKOS for the representation of a microtesauro based on the category 'Information and Strategic Knowledge in Organizations' of the Brazilian Thesaurus of Information Science. It is a theoretical and methodological research that proposes to approach the themes of representation, organization and availability of vocabularies in digital environments. It is also characterized as a research of an applied and exploratory nature, since it aims to generate knowledge for practical application. From this study, the potentialities and challenges of the representation of SKOS controlled vocabularies were identified, highlighting the need for further study in order to map out guidelines for the establishment of the different types of predicted relations in SKOS.

**Keywords:** Information Representation; Information and Communication Technologies; Controlled Vocabularies; Knowledge Organization Systems; Semantic Technologies.

**1 Introdução**

A popularização dos meios de comunicação, através do avanço das tecnologias, vem desafiando a linearidade de acesso à informação principalmente em ambientes digitais (Souza & Alvarenga, 2004). Essa realidade expõe a necessidade dos profissionais da informação se atualizarem constantemente e explorarem ferramentas que contribuam para o crescimento ordenado da Web sob o viés dos usuários e de suas necessidades. Os métodos de organização da informação são de expressiva importância não apenas para fomentar o avanço dos processos de recuperação da mesma, mas principalmente para promover a transposição das fronteiras culturais e geográficas de acesso à informação e ao conhecimento.

Nessa perspectiva, ressaltam-se os Sistemas de Organização do Conhecimento (*Knowledge Organization Systems – KOS*) como ferramentas que englobam diversos tipos de esquemas que visam promover a organização do conhecimento, como esquemas de classificação, cabeçalhos de assunto, tesauros, ontologias, entre outros (Hodge, 2000). Ao longo dos últimos anos, por meio da incorporação das tecnologias computacionais nos processos de classificação tornou-se possível a compatibilização de vocabulários controlados com auxílio do computador, favorecendo a evolução das linguagens documentárias e a integração de tesauros em Sistemas de Recuperação de Informação (Alvite Díez, Pérez-Léon, Martínez-Gonzáles & Vicente-Blanco, 2010).

É nesse contexto de avanços tecnológicos e crescente preocupação em encontrar uma forma de representar os sistemas de organização do conhecimento como dados interoperáveis entre sistemas de computador que a Word Wilde Web Consortium apresentou em 2009 o SKOS – *Simple Knowledge Organization System* (Sistema de Organização do Conhecimento Simples) (Miles & Bechhofer, 2009). Embora o modelo esteja se destacando como importante meio de padronização dos recursos informacionais na Web, no Brasil ainda existem poucos estudos sobre esse padrão, o que reforça sua necessidade de compreensão para ampliar sua utilização na comunidade científica (Ramalho, 2015).

Este trabalho visa apresentar um microtesauro em SKOS elaborado a partir da categoria ‘Informação e Conhecimentos Estratégicos nas Organizações’ do Tesauro Brasileiro de Ciência da Informação, examinar os fundamentos que norteiam o padrão SKOS e identificar aproximações e divergências em relação aos instrumentos de representação tradicionalmente utilizados no campo da Ciência da Informação.

Trata-se de uma pesquisa de cunho teórico e metodológico que se propõe a abordar os temas de representação, organização e disponibilização de vocabulários em ambientes digitais. Possui abordagem qualitativa, pois busca apresentar características voltadas à descrição, compreensão e explicação de relações de um determinado fenômeno. Caracteriza-se, também, como pesquisa de natureza aplicada, uma vez que objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos (Silveira & Córdova, 2009).

**2 O Modelo de Dados SKOS e suas Propriedades**

O modelo SKOS alcançou o status de recomendação da W3C em 2009 e desde então tem sido utilizado como uma possibilidade de representação de vocabulários em ambientes digitais, pois permite que as representações sejam legíveis por máquinas e que haja o intercâmbio entre softwares. Destacamos a importância do SKOS, pois ele se apresenta como uma ferramenta de baixo custo que é capaz de compreender a semântica nativa de uma variedade de KOS através de um conjunto de propriedades e, assim, realizar a migração destes materiais para a Web Semântica (Baker et. al, 2013).

O elemento fundamental do SKOS é a definição de um conceito e sua identificação através de um URI, sendo que a representação lexical é expressa em RDF (Isaac & Summers, 2009). Assim, o SKOS inclui a classe *skos:Concept*. As principais características dos conceitos são as expressões utilizadas para se referir a eles na linguagem natural. No SKOS, tais expressões são chamadas de etiquetas: sendo elas *skos:prefLabel, skos:altLabel* e *skos:hiddenLabel*. A propriedade *skos:prefLabel* permite atribuir a uma etiqueta um recurso preferente. A propriedade skos:*altLabel* atribui a um conceito uma etiqueta léxica alternativa, sendo bastante útil para a representação de sinônimos, acrônimos e abreviaturas, enquanto que a propriedade skos:*hiddenLabel* é utilizada para incluir variantes de erros ortográficos.

É preciso considerar que para um sistema de organização do conhecimento, as relações semânticas vão além de etiquetas atribuídas a conceitos, mas abrangem também como os conceitos vinculam-se com outros conceitos. Tais relacionamentos determinam o tipo de ligação que os conceitos estabelecem entre si ou entre seus termos descritores, incorporando semântica à estrutura do esquema (Ramalho, 2017).

Para essas relações semânticas, o SKOS oferece três opções padrão: *skos:broader*, *skos:related* e *skos:narrower*. A propriedade *skos:broader* permite estabelecer que um conceito tem um significado mais amplo do que outro, sendo que a propriedade *skos:narrower* é utilizada com um sentido mutuamente inverso, ou seja, quando um conceito tem um significado mais concreto e específico do que outro. *Skos:related* permite a representação associativa. Relações do tipo associativa não indicam uma hierarquia, logo, nenhum conceito é mais geral ou mais específico que outro (Miles & Bechhofer, 2009).

Um conceito SKOS pode se conectar com outros conceitos mais amplos ao mesmo tempo, criando uma rede de conceitos através da classe *skos:ConceptScheme*. Uma vez criado um esquema de conceitos, é possível vincula-lo com outros conceitos que constituem o sistema através da propriedade *skos:inScheme*, o que permite também a possibilidade de afirmar que dois conceitos de diferentes esquemas possam ter um significado similar. Com a finalidade de proporcionar um acesso eficaz aos pontos de entrada das hierarquias de conceitos, SKOS define também a propriedade *skos:hasTopConcept* que permite vincular um esquema de conceitos a uma ou várias entradas de cabeçalho.

Dentre vários estudos que apontam para os benefícios da conversão de tesauros para ambientes digitais, a seguir iremos explorar alguns exemplos pontuais que demonstram a usabilidade de algumas propriedades do modelo SKOS que são fundamentais para esse processo.

**3 Representação de um Microtesauro em SKOS**

O tesauro é uma das ferramentas cujo uso é mais difundido devido à sua capacidade de ajustar-se às necessidades de gerenciamento e organização nos sistemas de informação (Fujita, Cruz & Patrício, 2017). Aliar esse material às tecnologias semânticas garante inteligência contribui para o avanço do controle de vocabulário em contexto digital, reduzindo as inconsistências na recuperação da informação.

Soergel et. al (2004) apresenta um exemplo da dificuldade que os usuários enfrentam em ambientes digitais quando supõe que um fazendeiro inicie sua pesquisa sobre arroz inserindo no campo de busca apenas o termo “arroz” (*rice* em inglês) e os resultados retornados apontam para uma variedade de assuntos que se relacionam com o próprio grão de arroz, com a *Rice University* em Houston ou com pessoas que possuam esse termo no nome, entre outros casos. Isto evidencia que os sistemas de recuperação se baseiam no texto inserido pelo usuário e não no significado do que foi escrito. É nítido como o desafio na recuperação de informação está na capacidade do sistema em identificar um conceito em um domínio específico de interesse para, asism, ser capaz de oferecer resultados que realmente sejam úteis para o usuário.

É preciso, então, que os tradicionais sistemas de organização do conhecimento se casem com estruturas conceituais que consigam garantir uma rede de relações mais vastas, mas ainda assim precisas. A figura a seguir ilustra uma comparação entre as relações existentes entre os termos leite, vaca e queijo cheddar extraídos do Tesauro AGROVOC no lado esquerdo da tabela, enquanto que no lado direito as mesmas relações são expressas por uma ontologia hipotética.

Figura 01: Relações do Tesauro AGROVOC em comparação com relações de uma ontologia hipotética



Fonte: Soergel at. al, 2004.

De acordo com a Figura 01, com base nos relacoinamentos da ontologia, um sistma pode inferir que o queijo cheddar *<contémSubstância>* gordura de leite. A partir desta informação, o sistema consegue identificar que em uma situação em que as vacas se alimentam com uma ração que esteja contaminada com mercúrio, por exemplo, o queijo cheddar possivelmente também conterá traços do metal. Assimilar tal informação não seria possível, entretanto, levando-se em consideração somente as relações do Tesauro AGROVOC que são restritas à *Broader Term* e *Narrorer Term*, correspondentes a Termo Geral e Termo Específico, respectivamente (Soergel et. al, 2004).

Esta comparação denuncia que as ontologias potencializam as representações de conceitos que não poderiam ser obtidas a partir de simples descrições textuais, garantindo novas formas de acesso e evidenciando a potencialidade da conversão de tesauros para ambientes digitais (Ramalho, 2016).

Outros estudos abordam a conversão de tesauros para ambientes digitais através do modelo SKOS, como o Tesauro alemão para Ciências Socias, TheSoz. Tal procedimento demandou atenção especial não apenas aos termos do tesauro e as relações existentes entre eles, “(...) mas também à estrutura geral e às questões de design do tesauro, por exemplo a existência de um sistema de classificação adicional ou até que ponto o tesauro está em conformidade com as normas ISO estabelecidas” (Zapilko et. al, 2012, p.2). Logo, um mapeamento inicial pode ser de grande complexidade.

Ramalho (2017) apresentou a conversão da categoria “Tecnologias da Informação e Comunicação – TICs” do Tesauro Brasileiro de Ciência da Informação para o modelo SKOS. Neste estudo, evidenciou-se que a propriedade *skos:Collection* apresenta limitações que podem comprometer sua utilização para a representação de tesauros. Por definição, *skos:Collection* é a propriedade que representa uma coleção, sendo que uma coleção pode ser entendida como um conjunto de conceitos (*skos:Concept)* que tenham algo em comum. Como o modelo SKOS estabelece que as coleções sejam disjuntas dos conceitos, é impossível usar os relacionamentos semânticos do SKOS (*skos:broader, skos:narrower e skos:related*) para localizar a coleção em si.

Ou seja, a agrupação de conceitos em coleções acaba não substituindo as instruções sobre a localização de conceitos em um esquema conceitual e impede que recursos declarados como skos:Collection possam fazer parte de uma hierarquia conceitual, uma vez que representam diferentes tipos de relacionamento semânticos (Ramalho, 2017). No entanto, Isaac e Summers (2010) consideram como uma opção viável para resolver este problema manter a representação dos conceitos com a propriedade *skos:Concept* e usar relações semânticas normais para vinculá-los a outros conceitos quando da representação de sistemas destinados principalmente a estruturas hierárquicas de navegação.

 Desta forma, ao realizarmos a conversão da categoria “Informação e Conhecimentos Estratégicos nas Organizações”, conforme figura abaixo.

Figura 02: Informação e Conhecimentos Estratégicos nas Organizações em SKOS



Fonte: elaborado pelos autores.

A figura 02 ilustra uma das representações possíveis em SKOS do TBCI, em que este é considerado um esquema de conceitos que se vincula a outros conceitos, como a categoria “Informação e Conhecimentos Estratégicos nas Organizações”. Dessa foram, um novo esquema conceitual pode reutilizar conceitos já existentes através da propriedade *skos:inScheme*. Ou seja, um esquema referente a outro domínio pode reutilizar livremente os conceitos existentes nesta categoria.

Uma rede de esquemas de conceitos pode servir de base para novas aplicações que permitam uma navegação eficaz entre diferentes sistemas de organização. Além disso, o verdadeiro potencial da Web Semântica só acontece quando os dados se inter-relacionam uns com os outros.

Além das propriedades do modelo em questão, é possível notar a presença da propriedade *dct:title* do Dublin Core, um modelo de metadados para descrição de objetos digitais. Tal fator evidencia a flexibilidade do SKOS em permitir que outros vocabulários integrem o modelo e potencializem suas funcionalidades.

**4 Considerações Finais**

A partir da análise realizada foi constatado que SKOS favorece maior interoperabilidade entre vocabulários a partir da integração de diferentes conjuntos de dados, fato que corrobora sua importância, uma vez que os sistemas de organização do conhecimento tradicionais apresentam aplicações limitadas e pouca interoperabilidade, conforme destacam Pastor-Sánchez, Martínez-Méndez & Rodrígues-Muñoz (2012).

A análise do modelo SKOS evidencia a importância de compreender as potencialidades oferecidas pelas tecnologias semânticas na conversão de significativos materiais para o ambiente da Web Semântica. Tal proposição vai ao encontro com a necessidade crescente de sistematização de fundamentos teóricos que possibilitem elaborar vocabulários controlados mais adequados aos ambientes digitais contemporâneos, e que possuam maior interoperabilidade semântica, favorecendo que os avanços tecnológicos possam ser devidamente sedimentados no campo teórico.

A transposição do Tesauro Brasileiro de Ciência da Informação é de grande valor para o desenvolvimento da área de Ciência da Informação e representa ainda mais um avanço para a área não apenas no Brasil, mas também para todos os países lusófonos.

**REFERÊNCIAS**

Alvite Díez, L., Pérez-Léon, B., Martínez-Gonzáles, M. M. & Vicente Blanco, D. J. (2010). Propuesta de representación del tesauro EuroVoc en SKOS para su integración en sistemas de información jurídica. Scire, 2(16), 47-51.

Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G. & Summers, E. (2013). Key choices in the design of Simple Knowledge Organization System (SKOS). Web Semantics: Science, Services and Agents on the World Wide Web, (20), 35-49.

Boccato, V. R. C., Ramalho, R. A. S. & Fujita, M. S. L. (2008). A contribuição dos tesauros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. IBERSID, pp. 199-209.

Fujita, M. S. L.; Cuz, M. C. A e. & Patrício, B. O. M. (2017). A construção de tesauros na perspectiva dos manuais de indexação. In XVIII Encontro Nacional de Pesquisa em Ciência da Informação, 2017. São Paulo.

Garcia Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de documentación* (Vol. 7, pp. 79-95). Facultad de Comunicación y Documentación y Servicio de Publicaciones de la Universidad de Murcia. Recuperado de http://www.um.es/fccd/anales/ad07/ad0706.pdf

García-Marco, F. J. (2007). Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. *El profesional de la información*, 16(6), 541-550.

Hjorland, B. (2003). Fundamentals of knowledge organization. *Knowledge organization*, 30(2), 87-111.

Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries*: Beyond Traditional Authority Files. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources. Retrieved from http://old.diglib.org/pubs/dlf090/dlf090.pdf

International Organization for Standardization. (2011). *ISO 25964-1:2011*: Information and documentation – thesauri and interoperability with other vocabularies – part 1: Thesauri for information retrieval. Genebra.

Isaac, A., & Summers, E. (2009). *SKOS Simple Knowledge Organization System Primer*. World Wide Web Consortium (W3C) Working Group Note 18 August 2009. Retrieved from http://www.w3.org/TR/skos-primer/

Isaac, A. & Summers, E. (2010). Manual de SKOS. Anales de Documentación, nº 13, 285-320.

Lara, M. L. G. (2015). Propostas de tipologias de KOS: uma análise das referências de formas dominantes de organização do conhecimento. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 20 (n. esp.), 89-107. Recuperado de http://dx.doi.org/10.5007/1518-2924.2015v20nesp1p89

Miles, A., & Bechhofer, S. (2009). *SKOS simple knowledge organization system reference*. World Wide Web Consortium (W3C) recommendation, 18. Retrieved from http://www.w3.org/TR/skos-reference/

Miles, A., & Brickley, D. (2005). *SKOS Core Guide*. World Wide Web Consortium (W3C), November. Retrieved from http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/

Mustafa El Hadi, W. M. (2015). Cultural Interoperability and Knowledge Organization Systems. In Guimarães, J. A. C., & Dodebei, V. (Org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil; FUNDEPE, 575-606.

Pastor-Sanchez, J. A., Martínez-Mendez, F. J., & Rodríguez-Muñoz, J. V. R. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14(4), 10.

Pastor-Sánchez, J. A., Martínez-Méndez, F. J., & Rodríguez-Muñoz, J. V. (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. *El profesional de la información*, 21(3), 245-253.

Pinheiro, L. V. R. & Ferrez, H. D. (2014). Tesauro Brasileiro de Ciência da Informação. (p. 384). Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

Ramalho, R. A. S. (2015). Análise do modelo de dados SKOS: Sistema de Organização do Conhecimento Simples para web. Informação & Tecnologia (Itec), 2(1), 66-79.

Ramalho, A. S. R. (2016). O modelo de dados SKOS: novas perspectivas no âmbito da representação do conhecimento. ENANCIB: Encontro Nacional de Pesquisa em Ciência da Informação, Bahia, Brasil, 17.

Ramalho, R. A. S. (2017). Representação SKOS da categoria Tecnologia da Informação e Comunicação do Tesauro Brasileiro de Ciência da Informação: um estudo preliminar. ENANCIB: Encontro Nacional de Pesquisa em Ciência da Informação, São Paulo, Brasil, 18.

Silveira, D. T. & Córdova, F. P. (2009). A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. Métodos de pesquisa. Porto Alegre: Editora da UFRGS.

Soergel, D. et al. (2004). Reengineering thesauri for new applications: the AGROVOC example. In Digital Inf.,4, p1-23. Retrieved from https://journals.tdl.org/jodi/index.php/jodi/article/view/112/111.

Souza, R. R. & Alvarenga, L. (2004). A web semântica e suas contribuições para a Ciência da Informação. Ciência da Informação, 33(1), 132-141.

Zapilko, B. et al. (2012). TheSoz: A SKOS Representation of the Thesaurus for the Social Sciences. Retrieved from <http://www.semantic-web-journal.net/content/thesoz-skos-representation-thesaurus-social-sciences>.