

Análise dos tipos de relações do modelo SKOS: perspectivas de representação de recursos audiovisuais

Rogério Aparecido Sá Ramalho¹, Brígida Maria Nogueira Cervantes²

¹Professor Associado Universidade Federal de São Carlos – UFSCar, 0000-0002-8491-3514,
ramalho@ufscar.br

²Professora Associada Universidade Estadual de Londrina – UEL, 0000-0001-7356-1798,
brigidacervantes@gmail.com

Resumo: Com o aumento exponencial da quantidade de bases de dados para o armazenamento de recursos audiovisuais, um novo desafio tem sido imposto para a área de Organização do Conhecimento, o de representar recursos audiovisuais. 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 conteúdo de *Knowledge Organization System* – KOS em ambientes computacionais, favorecendo a incorporação de recursos audiovisuais nos processos de representação. Este trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelo CNPq, e objetiva analisar os tipos de relações previstas no modelo de dados SKOS e as perspectivas de utilização deste modelo para a representação de recursos audiovisuais. Apresenta caráter teórico-metodológico e abordagem qualitativa, de natureza aplicada. Reverbera-se que no modelo SKOS as relações constituem a ‘espinha dorsal’ para a modelagem de KOS, podendo ser classificadas em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização. Assim, o modelo SKOS apresenta novas perspectivas para a representação de recursos audiovisuais, a partir do fornecimento de uma estrutura simplificada, porém flexível, que contribui para uma maior interoperabilidade e alcance das relações estabelecidas, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

Palavras chave: *Knowledge Organization Systems*; Vocabulários Controlados; Modelagem de Dados; Tecnologias Semânticas; Ontologias.

Abstract: With the exponential increase in the number of databases for the storage of audiovisual resources, a new challenge has been imposed for the area of Organization of Knowledge, that is, to represent audiovisual resources. Among the new representational instruments, we highlight the Simple Knowledge Organization System (SKOS), a data model that aims to represent the basic structure and content of Knowledge Organization System (KOS) in computational environments, favoring the incorporation of audiovisual resources in representation processes. This work is linked to an ongoing research project, funded by CNPq, and aims to analyze the types of relationships planned in the SKOS data model and the perspectives of using this model for the representation of audiovisual resources. It is an applied-natured work with a theoretical-methodological character and a qualitative approach. It is pointed out that in the SKOS model the relationships constitute the 'backbone'

for the KOS modeling, being classified in three macro categories: Structural, Semantic and Compatibility Relationships. Thus, the SKOS model presents new perspectives for the representation of audiovisual resources, from the provision of a simplified but flexible structure that contributes to greater interoperability and reach of established relationships, providing a semantic dimension practically unexplored previously.

Keywords: *Knowledge Organization Systems; Controlled Vocabularies; Data Modeling; Semantic Technologies; Ontology.*

1 Introdução

A partir da incorporação de tecnologias computacionais nos processos de representação e organização tornou-se possível a compatibilização de vocabulários controlados de maneira mais eficiente, favorecendo a integração de tesouros em Sistemas de Recuperação de Informação e conseqüente novas possibilidades de representação (Alvite Díez, Pérez León, Martínez González & Vicente Blanco, 2010).

Ao longo das últimas décadas, com o aumento exponencial da quantidade de bases de dados para o armazenamento de recursos audiovisuais, um novo desafio tem sido imposto para a Organização do Conhecimento o de representar recursos audiovisuais. 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 conteúdo de *Knowledge Organization System* – KOS em ambientes computacionais, apresentando novas possibilidades aos processos de representação.

O presente trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelo CNPq, e possui como objetivo analisar os tipos de relações previstas no modelo de dados SKOS, de modo a identificar novas perspectivas de utilização deste modelo para a representação de recursos audiovisuais.

Fundamentando-se em uma pesquisa de caráter teórico-metodológico, pretende oferecer respostas aos problemas de pesquisa apresentados. Possui abordagem descritiva, pois busca apresentar características voltadas à descrição, compreensão e explicação de relações de um determinado fenômeno de modo a considerar o caráter interativo entre os objetivos definidos e as orientações teóricas da pesquisa. 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, p. 35).

2 Evolução dos Sistemas de Representação

Uma discussão recorrente alusiva aos modelos e sistemas de representação é o consenso sobre o que representar e como fazê-lo. Ao longo dos anos, diferentes áreas (Ciência da Informação, Ciência da Computação, Linguística, Ciência Cognitivas, dentre outras) têm apresentado diversos modelos de representação, considerando propósitos e características distintas.

No âmbito da área de Ciência da Informação para compreender a real importância da Representação torna-se necessário considerar seu próprio significado e abrangência, a partir dos estudos realizados que resultaram em grande número de padrões e diretrizes, bem como em desenvolvimentos teóricos como os princípios de Cutter, análise facetada e avanços da tecnologia da informação (Hjørland, 2003).

As linguagens documentárias podem ser consideradas como um dos instrumentos de representação mais difundidos no domínio da Ciência da Informação, caracterizando-se como sistemas de signos que visam à uniformização da linguagem utilizada em um Sistema de Informação, proporcionando uma representação padronizada do conteúdo informacional (Bocato, 2008).

Segundo Hodge (2000) o termo *Knowledge Organization Systems (KOS)* foi proposto em 1998 pelo *Networked Knowledge Organization Systems Working Group* para englobar sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, redes semânticas e ontologias. Na taxonomia de tipos de KOS proposta por Hodge (2000) também são incluídos, entre outros, dicionários, glossários, taxonomias e tesouros.

Lara (2015, p. 92) argumenta que “no Brasil, não há consenso sobre a utilização de um termo que abranja o conjunto de instrumentos de organização da informação e do conhecimento”. Apesar das variações de denominação identificadas na literatura nacional para designar tais instrumentos, observa-se um crescente aumento no número de pesquisas relacionadas a esta temática.

A partir da popularização do ambiente Web, novos instrumentos de representação organização e representação têm sido propostos, provocando convergências entre aportes teóricos e metodológicos de diversos campos científicos, no intuito de atender as demandas informacionais de uma sociedade cada vez mais baseada em interações virtuais.

Na última década foi constatado um crescente interesse de pesquisadores no estudo e desenvolvimento de ontologias, envolvendo pesquisas empenhadas, principalmente, na consolidação das definições de tal instrumento de representação e análise de possibilidades efetivas de uso (Santos, Correa & Silveira, 2013).

García-Marco (2007) ressalta que o crescente aumento do número de pesquisas relacionadas a temática de ontologias na área de Ciência da Informação constitui a “ponta do iceberg” de um processo de reconfiguração disciplinar, como resultado da integração de diversas Ciências.

Para García Jiménez (2004), apesar das similaridades existentes entre tesouros e ontologias, ambos os instrumentos possuem vinculações operativas e teóricas distintas, sendo desenvolvidos a partir de diferentes trajetórias históricas. Além disso, é importante ressaltar que tradicionalmente os modelos clássicos de representação eram desenvolvidos para serem utilizados diretamente por usuários humanos, enquanto os novos instrumentos de representação desenvolvidos para serem utilizados em ambientes digitais, como as ontologias, tem como foco possibilitar a realização de inferências automáticas a partir de sistemas computacionais (Ramalho, 2015a).

Apesar das potencialidades oferecidas pelas ontologias, sua utilização ainda tem sido limitada, devido principalmente a complexidade e custos inerentes ao processo de modelagem. Por outro lado, Pastor-Sanchez, Martinez-Rodriguez e Mendez-Muñoz (2012) destacam que a exploração dos tesouros na Web é muito limitada, pois as diferentes formas de recuperação na Internet exigem vocabulários interoperáveis e

multilíngues, abandonando o paradigma lexical em favor de um paradigma conceitual, de modo que o próprio conceito de tesouro precisa ser adaptado às novas necessidades informacionais emergentes. Segundo Sanchez-Jimenez e Gil-Urdiciain (2007) esta adaptação permite superar as deficiências das relações fortemente ancoradas ao nível léxico que caracterizam os tesouros tradicionais.

Clarke e Zeng (2012) lembram que os princípios para a construção de tesouros mudaram radicalmente da norma ISO 2788 para a ISO 25964, publicada em 2011, tais mudanças podem ser justificadas devido ao avanço das tecnologias digitais que demandam uma necessidade de interoperabilidade semântica, em que a distinção entre termo e conceito é fundamental.

Nesta perspectiva, o *Simple Knowledge Organization System* – SKOS tem ganhado destaque ao longo dos últimos anos como uma alternativa intermediária, que permite a transição de KOS tradicionais para o ambiente Web, fornecendo um elo entre o formalismo lógico rigoroso das linguagens de ontologias e o mundo caótico, informal e pouco estruturado de ferramentas de colaboração baseadas na Web.

3 O Modelo de Dados SKOS

No ano de 2009 o Word Wide Web Consortium (W3C) formalizou como Recomendação o padrão *Simple Knowledge Organization System* – SKOS (Sistema de Organização do Conhecimento Simples) como um modelo de dados para a representação de Sistemas de Organização do Conhecimento no ambiente Web, buscando atender às novas demandas informacionais relacionadas ao compartilhamento e interoperabilidade de vocabulários no ambiente Web (Miles & Bechhofer, 2009; Ramalho, 2015b).

O modelo SKOS proporciona um cânone para a representação formal da estrutura básica e conteúdo de Sistemas de Organização: como tesouros, esquemas de classificação, listas de cabeçalho de assunto, taxonomias, entre outros tipos similares de vocabulários controlados (Ramalho, Vidotti & Fujita, 2007). Sua codificação é baseada no padrão RDF (*Resource Description Framework*), favorecendo a representação de informações de maneira estruturada, assim como a integração de diferentes esquemas conceituais.

Conforme os relatos de Pastor-Sanchez, Martínez-Mendez e Rodriguez-Muñoz (2009) o desenvolvimento do modelo SKOS começou em 2002, com as atividades do grupo de trabalho *Semantic Web Advanced Development for Europe* (SWAD-E) e foi difundido publicamente em novembro de 2005, com a publicação do documento *SKOS Core Guide* (Miles & Brickley, 2005). Contudo, apenas em agosto de 2009 com a publicação do *SKOS Simple Knowledge Organization System Reference* o SKOS passou a ser considerado como uma recomendação oficial do W3C, um modelo de dados para a disponibilização e compartilhamento de sistemas de organização do conhecimento na Web.

Tendo como principal objetivo favorecer a representação de vocabulários no ambiente Web, como base em um modelo simplificado, o SKOS não tem a função de substituir vocabulários controlados em seu contexto original de uso, e sim favorecer uma maior reutilização e interoperabilidade entre os vocabulários existentes (Ramalho, 2015a). Segundo Isaac e Summer (2009) o padrão SKOS possibilita que conceitos possam ser identificados utilizando *Uniform Resource Identifiers* (URI),

rotulados com seqüências textuais em uma ou mais línguas naturais, documentados por meio de diferentes tipos de notas, relacionados semanticamente entre si, a partir de hierarquias informais e redes associativas; e agregados em diferentes esquemas conceituais. Mustafa El Hadi (2015) argumenta que desde a padronização do modelo SKOS muitos vocabulários passaram a ser protegidos por direitos autorais para serem publicados como dados ligados na Web, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

Dextre Clarke e Zeng (2012, p. 23) destacam a importância do padrão SKOS

As Recomendações SKOS do W3C são projetadas para suportar publicações de vocabulários, como tesouros na Web. E na sua essência é um modelo de dados que distingue expressamente entre os conceitos e etiquetas para representar conceitos. [...] Além disso, que é veiculado pelo SKOS Core para as relações entre conceitos, a extensão fornece suporte adicional para identificar, descrever e linkar entidades lexicais.

De acordo com a norma ISO 25964-1 (International Organization for Standardization, 2011) tradicionalmente os KOS estão fundamentados em três tipos principais de relações: relações hierárquicas, as relações associativas e relações de equivalência. Contudo, apesar do crescente número de pesquisas com foco no desenvolvimento de vocabulários SKOS, verifica-se que os novos tipos de relações previstas por este modelo têm sido pouco explorados na literatura.

Como o modelo SKOS tem como objetivo representar formalmente a estrutura e conteúdo de KOS, sua implementação não se limita a representação de Conceitos, incorporando também a definição de Coleções Conceituais e Esquemas de Conceitos.

As Coleções Conceituais (*skos:Collection*) são utilizadas para agrupar conceitos que compartilham determinadas características ou que necessitam de ordenação específica (*skos:OrderedCollection*). Em relação aos Esquemas de Conceitos (*skos:ConceptScheme*), estes são utilizadas para favorecer o estabelecimento de relações entre diferentes KOS em um mesmo vocabulário SKOS, possibilitando interoperabilidade semântica.

As relações existentes entre Conceitos, Coleções e Esquemas Conceituais podem ser definidas como ‘Relações Estruturais’ e possibilitando a representação adequada dos diferentes elementos que compõem o modelo SKOS. A figura 01 apresenta graficamente as Relações Estruturais previstas no modelo SKOS.

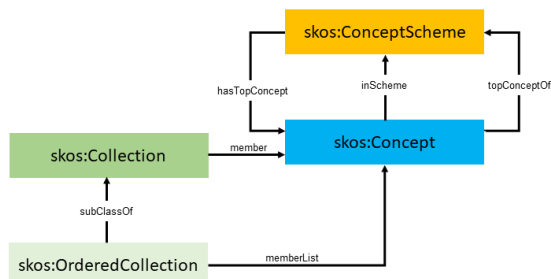


Figura 01. Relações Estruturais em SKOS.

Conforme pode ser observado na Figura 01 as Relações Estruturais previstas em SKOS são representadas a partir de diferentes propriedades previstas no modelo, à medida que os elementos Conceitos, Coleções e Esquemas Conceituais representam conjuntos disjuntos e não podem fazer parte de uma mesma hierarquia conceitual.

As Relações Semânticas são inegavelmente os tipos de relações mais explorados para a elaboração de KOS, do mesmo modo, em SKOS podem ser categorizadas como Relações Hierárquicas (*skos:broader* e *skos:narrower*), Relações Associativas (*skos:related*) e Relações de Mapeamento (*skos:mappingRelation*).

Relações de Mapeamento são tradicionalmente denominadas na literatura como Relações de Equivalência. Assim, traçando um paralelo das possibilidades de mapeamento entre conceitos oferecidas em SKOS com a literatura no campo de Organização do Conhecimento é possível estabelecer aproximações com a Matriz de Compatibilização Conceitual proposta por Dahlberg (1981, 1983).

Nesta perspectiva, considera-se adequado denominar como Relações de Compatibilização as relações previstas no modelo SKOS (*skos:exactMatch*, *closeMatch*, *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch*) que possuem como objetivo favorecer a compatibilização entre os conceitos existentes em diferentes esquemas conceituais. A figura 02 apresenta graficamente uma síntese da classificação proposta para os diferentes tipos de Relações previstas em SKOS.

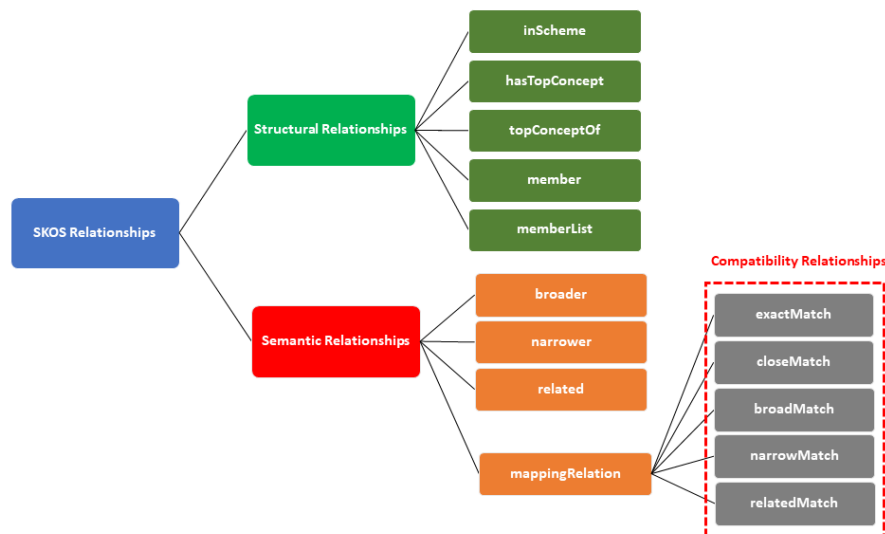


Figura 02. Categorização dos Tipos de Relações previstas em SKOS.

A figura 02 apresenta uma categorização das Relações previstas em SKOS em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização. Além das Relações Semânticas tradicionalmente exploradas em KOS convencionais, o modelo SKOS apresenta um novo subconjunto de Relações Semânticas que pode ser denominada como Relações de Compatibilização, pois favorece a compatibilização formal e o alinhamento de conceitos existentes em diferentes esquemas conceituais.

4 Considerações Finais

Este artigo teve como objetivo analisar os tipos de relações previstas no modelo de dados SKOS, a partir da análise realizada verificou-se que as relações constituem a ‘espinha dorsal’ para a modelagem de KOS, podendo ser classificadas em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização.

A partir das Relações de Compatibilização é possível estabelecer aproximações com estudos relacionados à Matriz de Compatibilização Conceitual proposta por Dahlberg (1981, 1983). Destaca-se que tais relações favorecem ir além de representações puramente conceituais, apresentando novas perspectivas para a representação a partir da compatibilização de recursos audiovisuais, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

As Relações Estruturais possibilitam o estabelecimento de uma estrutura simplificada, porém flexível, que contribui para uma maior interoperabilidade e alcance das relações estabelecidas. Assim o presente estudo corrobora a crescente necessidade de sistematização de fundamentos teóricos que possibilitem validar o desenvolvimento de novos instrumentos e métodos, à medida que avanços tecnológicos possam ser devidamente sedimentados no campo teórico, contribuindo para o fortalecimento de um arcabouço teórico que favoreça um melhor entendimento e utilização das tecnologias digitais contemporâneas nos processos de representação de recursos audiovisuais.

Referências

- Alvite Díez, M. L., Pérez León, B., Martínez González, M. M., & Vicente Blanco, D. J. (2010). Propuesta de representación del tesouro Eurovoc en SKOS para su integración en sistemas de información jurídica. *Scire: Representación y Organización del Conocimiento*, 16(2), 47-51.
- Bocato, V. (2008). A linguagem documentária como instrumento de organização e recuperação da informação. In Hoffmann, W. A. M., & Furnival, A. C. (Orgs.) *Olhar: ciência, tecnologia e sociedade* (p. 269-278). São Paulo: CECH-UFSCar.
- Clarke, S. G., & Zeng, M. L. (2012). From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. *Information Standards Quarterly (ISQ)*, 24(1).
- Dahlberg, I. (1981). Conceptual definitions for interconcept. *International Classification*, 8(1).
- Dahlberg, I. (1983). Terminological definitions: characteristics and demands. In: *Problèmes de la définition et de la synonymie en terminologie*. Québec, GIRSTERM, 13-51.
- Dextre Clarke, S. G., & Zeng, M. L. (2012). From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling. *Information Standards Quarterly (ISQ)*, 24(1).
- García 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.

- Hjørland, 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*. Geneva.
- 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/>
- 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-Sánchez, 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.
- Ramalho, R. A. S., Vidotti, S. A. B. G., & Fujita, M. S. L. (2007). Web semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*. 8(6).
- Ramalho, R. A. S. (2015a). Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a web. *Informação & Tecnologia*, 2(1), 66-79.
- Ramalho, R. A. S. (2015b). Ontologias e Simple Knowledge Organization System (SKOS): aproximações e diferenças. Guimarães, J. A. C., & Dodebei, V. (Org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil; FUNDEPE, 100-107.
- Sánchez-Jiménez, R., & Gil-Urdiciain, B. (2007). Lenguajes documentales y ontologías. *El profesional de la información*, 16(6), 551-560.
- Santos, M. T., Correa, R. F., & Silveira, M. A. A. (2013). Estudos brasileiros sobre Ontologia na Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*, 14(1). Recuperado de http://www.dgz.org.br/fev13/Art_05.htm
- Silveira, D. T., & Córdova, F. P. (2009). A pesquisa científica. In: Gerhardt, T. E., & Silveira, D. T. (Orgs.). *Métodos de pesquisa*. Porto Alegre: Editora da UFRGS.