

Parâmetros para avaliação de vocabulários SKOS

Parameters for evaluating SKOS vocabularies

Janailton Lopes Sousa (1), Rogério Aparecido Sá Ramalho (2), Brígida Maria Nogueira Cervantes (3)

(1) (2) Universidade Federal de São Carlos, Rod. Washington Luiz, Km 235, São Carlos - SP, janailtonlopes20@gmail.com, ramalho@ufscar.br. (3) Universidade Estadual de Londrina, Rodovia Celso Garcia Cid - Pr 445 Km 380, Campus Universitário, Londrina - PR, brigidacervantes@gmail.com

Resumo

O avanço das tecnologias digitais de informação e comunicação permitiu o surgimento de novos padrões que incorporam os fundamentos dos Knowledge Organization Systems (KOS). Esse alicerce teórico influenciou o surgimento do Simple Knowledge Organization System (SKOS), que admite a representação de vocabulários controlados enquanto tipo de KOS. Diante do aumento da publicação de vocabulários SKOS, a avaliação surge como atividade inerente à manutenção da qualidade desses vocabulários. O objetivo deste estudo é apresentar parâmetros para avaliação de vocabulários codificados em SKOS. Os procedimentos metodológicos fundamentam-se na pesquisa exploratória, revisão bibliográfica e análise teórica descritiva com uma abordagem estritamente qualitativa. Como resultado é apresentado um referencial de métodos que podem ser utilizados na análise de diferentes tipos de vocabulários SKOS, levando em consideração as características fundamentais inerentes aos vocabulários controlados. Com isso, é possível identificar uma série de erros recorrentes na construção desse tipo de vocabulário. Constatou-se que a avaliação é uma etapa necessária para os diferentes tipos de KOS e, por isso no âmbito da Ciência da Informação existem meios distintos de avaliação, que incluem métodos com abordagens objetivas, subjetivas, automática ou manual, podendo ou não enfatizar sua usabilidade, estrutura e relevância. A qualidade de um vocabulário controlado exige a análise de requisitos específicos dos recursos oferecidos, todavia se percebe certa dificuldade de avaliação de forma holística, de modo que, nesse sentido, os parâmetros apresentados contribuem para o avanço dessa discussão na Ciência da Informação.

Palavras-chave: Organização do Conhecimento; Vocabulários controlados; SKOS.

Abstract

The advancement of digital information and communication technologies has enabled the emergence of new standards that incorporate the foundations of Knowledge Organization Systems (KOS). This theoretical foundation influenced the emergence of the Simple Knowledge Organization System (SKOS), which allows the representation of controlled vocabulary as a type of KOS. In view of the increase in the publication of SKOS vocabularies, evaluation comes as an inherent activity in maintaining the quality of these vocabularies. The aim of this study is to present parameters for the evaluation of coded vocabularies in SKOS. The methodological procedures are based on exploratory research, bibliographic review and descriptive theoretical analysis with a strictly qualitative approach. As a result, it is presented a reference of methods that can be used in the analysis of different types of SKOS vocabularies, taking into account the fundamental characteristics inherent to the controlled vocabularies, with which it is possible to identify a series of recurrent errors in the construction of this type of vocabulary. It was verified that the evaluation is a necessary stage for the different types of KOS, so in the scope of the Information Science there are distinct means of evaluation, which include methods with objective, subjective approaches, automatic or manual, and may or may not emphasize its usability, structure and relevance. The quality of a controlled vocabulary requires the analysis of specific requirements of the resources offered, but it is possible to perceive a certain difficulty of evaluation in a holistic way, in this sense, the presented parameters contribute to the advancement of this discussion in Information Science.

Keywords: Organization of Knowledge; Vocabularies controlled; SKOS.

1 Introdução

O avanço das tecnologias digitais de informação e comunicação proporcionou o surgimento de padrões que incorporaram os fundamentos de *Knowledge Organization Systems* (KOS). Os vocabulários vinculados são os principais tipos de KOS que aderem a essa tendência, por meio de tecnologias como o *Resource Description Framework* (RDF), que se baseia na premissa, sujeito, predicado e objeto para declarar seus relacionamentos. Isso alicerçou, sobretudo, a origem do *Simple Knowledge Organization System* (SKOS).

SKOS é um padrão que surge da necessidade de codificar um conjunto de relacionamentos de vocabulários usando um esquema RDF (BAKER et al, 2013). Os vocabulários SKOS possibilitam que recursos informacionais sejam contextualizados e recebam identificadores únicos, que dinamizam os relacionamento entre termos e lhes atribuem características que equivalem a um conceito.

Este trabalho é um recorte de pesquisa em andamento financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e pelo Conselho Nacional de Desenvolvimento Científico e

Tecnológico (CNPq). O objetivo deste estudo é apresentar parâmetros para avaliação de vocabulários, com ênfase para aqueles que estão codificados em SKOS.

Os procedimentos metodológicos assumem características de pesquisa exploratória, pois, segundo Köche (2002, p. 126), estudos exploratórios desencadeiam “[...] um processo de investigação que identificam a natureza do fenômeno e aponta as características essenciais das variáveis que se quer estudar”. Além de revisão bibliográfica e análise teórica descritiva, porque o caráter descritivo permite descrever as características do objeto de estudo investigado (COSTA, COSTA, 2001), possui também uma abordagem estritamente qualitativa.

No escopo deste trabalho há uma breve introdução, seguida da apresentação dos fundamentos básicos SKOS para referenciar os procedimentos que são identificados como parâmetros de avaliação de vocabulários SKOS e, por fim, alguns apontamentos decorrentes dos resultados deste estudo.

2 Fundamentos do SKOS

SKOS favorece uma aproximação entre instrumentos e métodos de construção de vocabulários controlados utilizados tradicionalmente no campo da Ciência da Informação. SKOS agrega o formalismo lógico das linguagens de marcação, como a *Extensible Markup Language* (XML) e a contextualização de dados em RDF como meio de representação formal para o ambiente Web.

As Recomendações SKOS do *World Wide Web Consortium* (W3C) são projetadas para suportar publicações de vocabulários controlados na web, em sua essência seria um tipo de modelo de dados que distingue conceitos e etiquetas.

Broughton (2010) ressalta a proximidade entre SKOS e as classificações facetadas, ressaltando que algumas aplicações tecnológicas apresentam semelhanças com os tradicionais sistemas de organização facetados, de modo que a teoria da classificação facetada pode fornecer um modelo para os elementos utilizados nos sistemas de representação.

De acordo com Isaac e Summers (2009), SKOS possibilita que conceitos possam ser identificados utilizando *Uniform Resource Identifier* (URIs), etiquetados com sequências textuais em uma ou mais línguas naturais, documentados por meio de diferentes tipos de notas, relacionados semanticamente entre si a partir de hierarquias e redes associativas, agregando diferentes esquemas conceituais.

SKOS foi projetado como uma alternativa de baixo custo para permitir a transição de KOS tradicionais para o ambiente Web, favorecendo a representação de aspectos semânticos dos recursos informacionais

(RAMALHO, 2015). Além disso, possibilita a inserção de outros esquemas de metadados, como Dublin Core (DC), ou linguagens formais de representação do conhecimento, como a *Ontology Web Language* (OWL).

Segundo Maculan (2015), SKOS possui uma extensão XL (*eXtension for Labels*), que possibilita a formalização de declarações em RDF, representando relações que podem ser interpretadas (legíveis) por máquinas, auxiliando na interoperabilidade entre diferentes vocabulários.

Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012) complementam também que SKOS possibilita estabelecer vínculos entre conceitos de diferentes esquemas, favorecendo uma aproximação com as novas tendências de pesquisa computacionais.

Para entender como um vocabulário SKOS se estrutura é necessário identificar as etiquetas que prefixam as categorias de uso de cada conceito. Essas categorias de uso e seus prefixos estão definidos nas recomendações *SKOS Primer* (ISAAC E SUMMERS, 2009); *SKOS Reference* (MILES, BECHHOFFER, 2009a) e *SKOS-XL Namespace Document - HTML Variant* (MILES, BECHHOFFER, 2009b).

SKOS beneficia-se de uma codificação RDF, que permite que conceitos compostos publicados na web sejam vinculados e formem esquemas conceituais. Cada conceito é apontado por uma URI e etiquetado por uma *tag* ou *string* em linguagem natural e, além disso, fornece uma linguagem de modelagem conceitual leve e intuitiva para desenvolver e compartilhar novos KOS.

Por incorporar um esquema RDF, seu uso pode ser facilmente estendido para atender a requisitos específicos. Segundo Isaac e Summers (2009), SKOS introduz a classe `<skos:Concept>`, que permite aos implementadores afirmar que um determinado recurso é um conceito. Isso pode ser feito em duas etapas: criando (ou reutilizando) URIs para identificar exclusivamente o conceito; e afirmando no RDF, por meio da propriedade `<rdf:type>`, que o recurso identificado por esse URI é do tipo `<skos:Concept>`.

SKOS fornece três propriedades para anexar etiquetas de recursos conceituais: `<skos:prefLabel>`, `<skos:altLabel>` e `<skos:hiddenLabel>`. O `<skos:prefLabel>` possibilita atribuir um termo preferido a um recurso, enquanto o `<skos:altLabel>` permite atribuir um termo alternativo.

O `<skos:hiddenLabel>` é um recurso acessível para aplicativos de indexação e pesquisa baseadas em texto, utilizado para incluir variantes com erros ortográficos ou variações de termos, ocultos para o usuário comum. (ISAAC; SUMMERS, 2009).

As relações semânticas são umas das principais

características do SKOS. Nesse sentido, evoca as categorias fundamentais de relações que são usadas em vocabulários de acordo com a ISO 25964 por meio de três propriedades padrão de relacionamentos entre termos.

Para relações hierárquicas, amplas e estreitas, utiliza-se, respectivamente, `<skos:broader>` e `<skos:narrower>`, que permitem a representação de links hierárquicos, como a relação entre um gênero e suas espécies mais específicas, ou, dependendo de interpretações, a relação entre um todo e suas partes (ISAAC; SUMMERS 2009).

O `<skos:related>` permite a representação de links associativos (não hierárquicos), que também podem ser usados para representar links de partes inteiras que não são entendidos como relacionamentos hierárquicos. Nenhuma etiqueta é mais geral ou mais específica quando se utiliza essa propriedade (ISAAC; SUMMERS, 2009).

SKOS é uma aplicação de vocabulários controlados em ambientes digitais, fato evidenciado pela própria norma ISO 25964, conforme aponta Clarke e Zeng (2012), em que o SKOS é reconhecido como padrão interoperável. Apesar desse *background* normativo, as propriedades SKOS não equivalem a uma conversão literal, mas aproximada dos modelos de representação de vocabulários.

De acordo com Schilling (2012), SKOS objetiva fornecer uma ponte entre diferentes comunidades da Ciências da Informação e de práticas dentro da biblioteca envolvidas no projeto e aplicação dos sistemas de organização do conhecimento. Um dos elos possíveis é a publicação e vocabulários vinculados.

De acordo com Gracy, Zeng e Skirvin (2013), no universo *Linked Data* (LD) os padrões de dados de biblioteca são frequentemente ignorados em favor de estruturas de dados abertas baseadas nos padrões da Web Semântica, como o RDF, SKOS e OWL. Nesse sentido é importante buscar aproximações conceituais entre esses padrões com diferentes tipos de KOS.

Segundo Mader, Haslhofer e Isaac (2012), SKOS possui uma abordagem de compromisso mínimo, o que implica dizer que esse padrão se abstém de definir restrições que imponham um domínio ou requisitos de outro padrão. Essa liberalidade com a qualidade dos dados não se preocupa com estruturas complexas e alto grau de detalhamento de suas propriedades.

Oposto a OWL, entendida como um tipo de representação complexa, SKOS simplifica a concepção de um vocabulário, por isso em alguns momentos é entendido como ontologia leve. Todavia, essa característica não impede a adoção de outros esquemas de metadados compatíveis com SKOS.

O enriquecimento semântico pode ser uma alternativa

para essa liberalidade do SKOS, pois na criação de vocabulários vinculados existe a possibilidade de inserir propriedades do RDF, OWL, DC, assim como outros padrões. Esse enriquecimento descritivo auxilia na representação de recursos informacionais, além proporcionar a estruturação de metadados bibliográficos com propriedades do *Linked Data*.

Diante destas propriedades do SKOS e da relevância normativa da ISO 25964, é necessário considerar os aspectos utilizados na avaliação de vocabulários tradicionais (vocabulários que não estão em formatos digitais ou diferentes de SKOS) para realçar os distintos tipos de avaliação existentes até os parâmetros aplicados aos vocabulários publicados em SKOS.

3 Parâmetros de avaliação de vocabulários

A avaliação é uma etapa necessária aos diferentes tipos de KOS, por isso no âmbito da Ciência da Informação existem distintos tipos de avaliação, que são propostas para atender essa demanda. Incluem-se métodos com abordagens objetivas, subjetivas, automática ou manual, podendo enfatizar a usabilidade, estrutura, relevância e a qualidade geral de um vocabulário.

Quando se aplica ao ambiente digital, Pastor-Sanchez, Martinez-Rodriguez e Mendez-Muñoz (2009) afirmam que a exploração dos vocabulários na Web é muito limitada. Tais autores insistem que o próprio conceito de tesouro tem evoluído para se adaptar aos novos modelos de representação da Web, abandonando o paradigma lexical em favor de um paradigma conceitual. Segundo Sanchez-Jimenez e Gil-Urdiciain (2007) essa adaptação permite superar as deficiências das relações fortemente ancoradas ao nível léxico que caracterizam os tesouros tradicionais.

Apesar de existirem diversos manuais para elaboração de vocabulários controlados, os métodos de avaliação dos mesmo não tem sido evidenciados. Tal afirmação é reforçada quando se percebe os escassos tipos de avaliação para vocabulários controlados.

A ANSI/NISO Z39.19 (2010) destaca que há pouca pesquisa ou literatura lidando especificamente com teste e avaliação de vocabulários controlados, o que pode ser devido em parte à dificuldade de isolar a qualidade de um vocabulário a partir do desempenho de outros elementos de um sistema de informação, como tecnologia, mecanismo de pesquisa e interface.

Para embasar essa perspectiva são apresentados os principais autores e entidade que abordaram sobre avaliação de vocabulários tradicionais no Ciência da Informação, entre eles a ANSI/NISO (2010), Soergel (2002), Lancaster (2002) com indicações mais gerais e Gil-Leiva (2008), Soler-Monreal (2009), Soler-Monreal e Gil-Leiva (2011), propondo métodos quantitativos e qualitativos.

Quando aplicados aos vocabulários SKOS, as avaliações recorrem às verificações estruturais mínimas especificadas nas recomendações da W3C para SKOS e propriedades comuns aos vocabulários vinculados. Nesse contexto se destacam Manaf, Bechhofer e Stevens (2012), Mastora, Peponakis e Kapidakis (2017), Mader, Haslhofer e Isaac (2012), Suominen e Mader, (2014) e Suominen e Hyvönen (2012) com abordagens de análises automáticas.

Os parâmetros apresentados estão divididos em dois tipos, o primeiro para vocabulários tradicionais e o segundo para vocabulários SKOS, (Tabela 1) com os respectivos autores supracitados. Desse modo explora os principais procedimentos adotados para os distintos tipos de vocabulário.

A partir da análise das orientações para avaliação de vocabulários identificados na literatura foi possível simplificar as propostas destes parâmetros. Quer sejam em infraestruturas tradicionais (impresso ou publicados em padrões incompatíveis com formatos interoperáveis) ou digitalmente interoperável como SKOS.

Tabela 1. Orientações para avaliação de vocabulários

Tipo de Vocabulário	Referência
Tradicional	ANSI/NISO (2010)
	Soergel (2002)
	Lancaster (2002)
	Gil-Leiva (2008), Soler-Monreal (2009), Soler-Monreal; Gil-Leiva (2011)
	Manaf; Bechhofer; Stevens (2012)
SKOS	Mastora; Peponakis; Kapidakis (2017)
	Mader; Haslhofer; Isaac (2012)
	Suominen; Mader, (2014)
	Suominen; Hyvönen (2012)

A ANSI/NISO Z39.19 (2010) alerta que os seguintes métodos devem ser considerados quando confrontados com o desafio de medir a qualidade ou eficácia de um vocabulário controlado. Refere-se, portanto, a três tipos de avaliação:

A avaliação heurística, realizada por um perito ou um quórum de especialistas. Tal avaliação pode ser informal e qualitativa ou formal e quantitativa. Esse tipo de avaliação se difundiu no campo de experiência do usuário (*User eXperience*) devido aos estudos de

Nilsen e Molich (1990), que consideram a avaliação heurística uma inspeção sistemática realizada por especialistas com base em um conjunto de critérios que definem a Interação Humano-Computador (IHC).

A modelagem de afinidade, que reúne uma amostra representativa de usuários para avaliar um vocabulário, solicitando uma classificação desses termos. Os resultados são analisados contra a hierarquia de termos existente e os usuários também podem ser convidados a pontuar termos equivalentes e relacionados de acordo com o nível de similaridade.

O teste de usabilidade, que é uma espécie de avaliação holística do sistema de informação, podendo fornecer informações sobre a eficácia do vocabulário controlado. Esse tipo de avaliação possui um certo vínculo com a avaliação heurística, porque também propõe o estabelecimento de métricas de uso (NIELSEN, 1993).

Essas modalidades de avaliação transmitem aspectos gerais sobre cada tipo de avaliação para vocabulários controlados, por isso a própria ANSI cita o trabalho de Soergel (2002), que apresenta uma série de critérios que podem ser adotados neste processo, com perguntas que podem ser feitas para englobar aspectos sintáticos, semânticos e conceituais dos termos, assim como os seus relacionamentos, podendo conter critérios adicionais aplicáveis aos vocabulários acessíveis eletronicamente, se são hiperlinks ativos de referências cruzadas, se existe um índice de termos online ou pesquisa por palavra-chave, etc. (SOERGEL, 2002).

Para Lancaster (2002) algumas medidas podem ser utilizadas no processo de avaliação, como razão de equivalência, razão de reciprocidade, definição, flexibilidade, nível de pré-coordenação e tamanho do grupo de termos.

Gil-Leiva (2008) e Soler-Monreal (2009), estudando métodos de avaliação de vocabulários controlados do tipo tesouros, apresentam uma divisão em três grupos principais, que denominam de avaliação qualitativa intrínseca e avaliação intrínseca quantitativa ou estatística extrínseca. Para Soler-Monreal e Gil-Leiva (2011) a avaliação intrínseca pode ser realizada com o objetivo de analisar os próprios vocabulários controlados de modo a estudar sua estrutura, os campos temáticos ou facetas, notas de escopo, relações semânticas, grau de especificidade etc.

Por outro lado, avaliação extrínseca estuda o impacto nos sistemas de informação que os utilizam tanto na indexação quanto na recuperação (SOLER-MONREAL; GIL-LEIVA, 2011). Apesar de Soler-Monreal (2009) apresentar esses critérios de avaliação e apresentar o SKOS como linguagem ontológica, não há um direcionamento exato para avaliar esse tipo de vocabulário.

Pesquisas como de Manaf, Bechhofer e Stevens (2012)

são bem mais específicas, pois avaliam diretamente vocabulários que foram publicados em SKOS para compreender os estilos de modelagem usados nesses vocabulários publicados na web, verificando números de conceitos, relacionamentos e suas URIs.

O método de avaliação de Manaf, Bechhofer e Stevens (2012) demonstra que apesar de existir um grande número de vocabulários publicados, muitos não conseguem se adequar ao mínimo requerido em sua estruturação, expressos nos documentos de referência SKOS, chamando a atenção de vocabulários publicados na web que podem não seguir essas recomendações no seu processo de elaboração.

No estudo realizado por Mastora, Peponakis e Kapidakis (2017) que tinha por objetivo identificar até que ponto cada KOS é apropriado para migração para o SKOS, foram selecionados três tipos de KOS, nomeadamente o *thesaurus multilingue Eurovoc*, o *Library of Congress Subject Headings* (LCSH) e o *Dewey Decimal Classification* (DDC) sistema, justificados pelo fato de tesouros, cabeçalhos de assunto e esquemas de classificação são os três tipos mais frequentes de KOS usados no contexto das bibliotecas.

A análise desses vocabulários foi realizada por um *parser* desenvolvido pelo *Greek Institute for Language* para lidar somente com textos grego, portanto, reforça a escolha de um tesouro multilíngue. No caso do *Eurovoc*, os outros vocabulários possuíam traduções na língua grega, considerada pelos autores como a língua natural com menor índice de dupla interpretação.

Mastora, Peponakis e Kapidakis (2017) concluem ao final da análise que a conversão de dados tradicionais para RDF não é uma condição suficiente para adquirir todo o potencial da Web Semântica, ainda mais em casos de KOS tradicionais, em que há certa falta de semântica bem definida e, portanto, torna-se necessário realizar algum tipo de reengenharia na construção ou conversão de vocabulários em SKOS.

Dentre os métodos de avaliação apresentado, os trabalhos apresentados por Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac (2012) e Suominen e Mader (2014) são os que mais se aproximam de uma análise precisa que consegue aliar fundamentos normativos e métricas da infraestrutura conceitual do SKOS, logo é o principal parâmetro destacado.

Por meio desses respectivos trabalhos é possível apresentar uma compilação de parâmetros aplicáveis aos vocabulários SKOS. Pois conseguem englobar grande parte dos requisitos essenciais dispostos nas recomendações W3C e da ISO 25964.

Como avaliar um vocabulário não é um procedimento orientado por uma ação comum, Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac (2012) e Suominen e Mader, (2014) identificam erros recorrentes em

vocabulários SKOS que podem ser analisados por meio de verificação automática.

A incidência de tais erros podem comprometer a qualidade de um vocabulário, portanto indicam o que não deve conter em um vocabulário. Tais parâmetros endossam o processo de avaliação e indicam os passos para adoção de diretrizes que garantam a qualidade de um vocabulário SKOS.

Suominen e Hyvönen (2012) identificaram cerca de 14 erros comuns em vocabulários SKOS, categorizados como: URIs válidas, *tags* de idioma ausentes, etiquetas ausentes, conceitos soltos, classes owl disjuntas, uso consistente de etiquetas, uso consistente de propriedades de mapeamento, uso consistente das relações semânticas, valores ambíguos de etiquetas preferidas, sobreposição disjuntiva, propriedades de etiqueta, relações semânticas disjuntivas, hierarquia mais amplas em ciclos e espaço em branco extra (SUOMINEN; HYVÖNEN, 2012).

Mader; Haslhofer e Isaac (2012) classificam os problemas recorrentes em vocabulário SKOS em três categorias mais amplas, especificando os problemas contidos em cada uma delas. A primeira categoria é classificada como etiquetagem e documentação, que inclui: *tags* de idiomas omitidos ou inválidos, cobertura incompleta de idiomas, conceitos não documentados e conflitos de etiquetas.

A segunda categoria refere-se a questões estruturais, que incluem: conceitos órfãos, componentes fracamente conectados, relações hierárquicas cíclicas, relações associativas sem valor, conceitos únicos transitivamente relacionados, conceitos superiores omitidos e conceito superior com conceitos mais amplos (MADER; HASLHOFER e ISAAC, 2012).

A terceira aplica-se aos problemas específicos de dados vinculados, que incluem: links ausentes, links externos ausentes, links quebrados e recursos SKOS indefinidos. Essa categoria se refere notadamente a problemas de recursos informacionais que podem estar desvinculados dos seus endereços de localização (MADER; HASLHOFER e ISAAC, 2012).

Suominen e Mader (2014) mantêm as 3 categorias principais de problemas, mas expandem sua cobertura, resultando na seguinte classificação:

Etiquetagem e problemas de documentação, abrangendo *tags* de idiomas omitidos ou inválidos, cobertura incompleta de idiomas, conceitos não documentados, etiquetas sobrepostas, etiquetas ausentes, rótulos preferidos inconsistentes, violação de rótulos não contíguos e espaços extras vazios em etiquetas (SUOMINEN; MADER, 2014).

Questões estruturais, incluindo conceitos órfãos, clusters de conceitos desconectados, relações hierárquicas cíclicas, relações associativas sem valor,

conceitos únicos transitivamente relacionados, conceitos superiores omitidos, conceitos principais não marcados, conceitos principais como conceitos mais amplos, conceitos relacionados unidirecionalmente, confrontos relacionais, mapeamento de confrontos e violação de classes disjuntas (SUOMINEN; MADER, 2014).

Entre os problemas específicos de dados vinculados estão links ausentes, links externos ausentes, links quebrados, recursos SKOS indefinidos, violação do esquema HTTP/URI e URIs inválidas (SUOMINEN; MADER, 2014).

Outros problemas foram identificados e incluídos no rol de categorias de análise dos problemas de etiquetagem e problemas de documentação. Nessa instância foram adicionados: caracteres não imprimíveis em etiquetas, etiquetas vazias e referências de notação ambígua. Na instância de questões estruturais, foram incluídos: redundância hierárquica e relações reflexivas.

As questões estruturais envolvem diretamente problemas de ordem lógica, que comprometem o entendimento contextual dos conceitos em um vocabulário. Esse é um dos principais problemas, porque tal falha implica diretamente em erro sintático, semântico e conceitual.

Os problemas específicos de dados vinculados retomam a discussão da manutenção semântica entre os recursos representados pela a violação ou quebra de relacionamentos que prejudicam a contextualização do objeto representado. Para Sousa, Martins e Ramalho (2018), os modelos de dados com capacidade semântica constituem-se como ricas fontes de informação, além de proporcionar a interoperabilidade na troca dessas informações, de maneira que sua manutenção é de suma importância.

Desse modo, evidenciam a necessidade de avaliação de vocabulários controlados, que assumem características de ordem semântica, que, sobretudo, precisam aderir às boas práticas de publicação de vocabulários interligados, assim como os fundamentos que subsidiam essa tecnologia, conforme destaca Ramalho, Martins e Sousa (2017). A partir dos fundamentos que estruturam os vocabulários controlados, quer sejam normas, recomendações, padrões ou modelos de representação, é possível identificar pontos de referência que sustentarão análises futuras.

4 Considerações finais

O resultado desta pesquisa se trata de uma fonte referencial de métodos que pode ser utilizada para analisar diferentes tipos de vocabulários SKOS. Levando em consideração características fundamentais inerentes aos vocabulários controlados.

Diante dos problemas recorrentes na estruturação de

um vocabulário que adota SKOS como linguagem codificada de representação, Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac (2012), Suominen e Mader (2014) fornecem subsídios para a elaboração de uma tabela (Tabela 2, em apêndice), que expõe os problemas mais comuns na avaliação de um vocabulário SKOS, contendo as categorias principais dos erros encontrados em vocabulário SKOS, com 24 problemas consistentes.

Os principais problemas de etiquetagem para vocabulários em SKOS se referem às inconsistências que podem decorrer do processo de inserção ou conversão dos termos, além de possíveis erros de tradução e informações incompletas inseridas como notas de documentação.

Uma avaliação que consiga detectar esses tipos de erros evidencia sua importância, principalmente quando aplicadas a vocabulários muito extensos e a validação de conversões para o padrão SKOS. A utilização de métricas automáticas auxilia na objetivação desse processo, pois simplificam os critérios de avaliação orientados pela própria recomendação que especifica o uso do SKOS.

Os problemas estruturais podem comprometer a semântica de um vocabulário, pois prejudicará a ordem conceitual de um vocabulário com redundâncias lógicas sem valor. Portanto, a identificação desse tipo de erro reduz a probabilidade de inconsistências de organização conceitual.

Os problemas de dados vinculados aplicados aos vocabulários SKOS estão diretamente relacionados com a conectividade entre os recursos informacionais representados e o endereço que indica a sua localização. Logo, trata-se de problemas de links quebrados, ausência de direcionamento e URIs inválidas. A identificação desse tipo de erro, aumenta as chances de recuperação de um recurso informacional e da qualidade dos dados vinculados, que são publicados na forma de vocabulários.

Os vocabulários controlados têm demonstrado grande importância enquanto tipo de KOS, pois incorporam fundamentos que facilmente se adaptam em novos suportes tecnológicos. O aumento na publicação de vocabulários vinculados no ambiente web revela um grande salto na propagação da interoperabilidade semântica entre KOS. Entretanto, quando não se observa as regras de construção, manutenção e avaliação, um vocabulário pode se tornar problemático do ponto de vista técnico e conceitual.

A qualidade de um vocabulário controlado exige a análise de requisitos específicos dos recursos oferecidos, como infraestrutura, disponibilidade, usabilidade, entre outros fatores, todavia se percebe uma dificuldade de avaliação de vocabulários de forma holística com métodos distintos. Portanto, a

apresentação dos parâmetros de avaliação para vocabulários SKOS destacados neste trabalho oferecem um passo que aponta para o avanço dessa discussão.

SKOS é recurso que potencializa a publicação de vocabulários, amparado por normas internacionais que validam a adoção de modelos simples de organização do conhecimento e, nesse sentido, encontra-se no escopo da Ciência da Informação e da Organização do Conhecimento. Logo, este trabalho apresenta de forma resumida esse padrão, com seus principais parâmetros de avaliação.

Referências

- BAKER, T. et al. Key Choices in the Design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics*, v.20, 2013. Disponível em: <http://www.cs.vu.nl/~guus/papers/Baker13a.pdf>. Acesso em 05 fev 2019.
- BROUGHTON, V. Concepts and Terms in the Faceted Classification: the Case of UDC. *Knowledge Organization*, v. 37, Issue 4, 2010. p.270-279.
- CLARKE, S. G. D.; ZENG, M.L. From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling, in *Information Standards Quality, Special Edition: Year in Review and State of the Art of the Standards*, p.20-26, 2012.
- COSTA, M. A. F. ; COSTA, M. de F. B. Metodologia da pesquisa: conceitos e técnicas. Rio de Janeiro: Interciência, 2001.
- GIL-LEIVA, I. Manual de indización: teoría y práctica. [Manual of indexing: theory and practice.] Gijón, Spain: Trea, 2008.
- GRACY, K. F.; ZENG, M. L.; SKIRVIN, L. Exploring methods to improve access to Music resources by aligning library Data with Linked Data: A report of methodologies and preliminary findings. *Journal of the American Society for Information Science and Technology* banner, v. 64, Issue 10. 07 junho 2013. Disponível em: <https://doi.org/10.1002/asi.22914>. Acesso em: 19 jan. 2019.
- ISAAC, A.; SUMMERS, E. (Ed.). SKOS Simple Knowledge Organization System Primer: W3C Working Group Note, 18 August 2009. Disponível em: <https://www.w3.org/TR/2009/>. Acesso em: 19 jan. 2019.
- KÖCHE, J. C. Fundamentos de metodologia científica: teoria da ciência e iniciação à pesquisa. Petrópolis, RJ: Vozes, 2002.
- LANCASTER, F.W. El control del vocabulario en la recuperación de la información. 2ª ed. Valencia: Universidad de Valencia. 2002.
- MACULAN, B. C. M. S. Estudo e aplicação de metodologia para reengenharia de tesauro: remodelagem do THESAGRO. 2015. 339f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte, 2015.
- MADER, C.; HASLHOFER, B.; ISAAC, A. "Finding Quality Issues in SKOS Vocabularies". TPD, 2012.
- MANAF, N.A.A.; BECHHOFFER, S.; STEVENS, R. The Current State of SKOS Vocabularies on the Web. In: E. Simperl et al. (Eds.): ESWC 2012, LNCS 7295, Springer-Verlag Berlin Heidelberg 2012.pp. 270–284.
- MASTORA, A.; PEONAKIS, M.; KAPIDAKIS, S. SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs. *Journal of Information Science*, v. 43, n. 4, 2017. p.492–508.
- MILES, A.; BECHHOFFER, S. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 2009a. Disponível em: <http://www.w3.org/TR/skos-reference/>. Acesso em: 20 jan 2019.
- MILES, A.; BECHHOFFER, S. SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant. 18th August 2009b. Disponível em: <http://www.w3.org/TR/skos-reference/skos-xl.html> Acesso em: 20 jan 2019.
- NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual thesauri. 2010.
- NIELSEN, J. Usability Engineering. Boston: Academic Press, Cambridge, MA, 1993.
- NIELSEN, J.; MOLICH, R. Heuristic evaluation of user interfaces - Disponível em <http://www.acm.org/dl>. Seattle, Washington, United States, 1990. Acesso em dez. 2002.
- PASTOR-SANCHEZ, J. A.; MARTINEZ-MENDEZ, F.J.; RODRIGUEZ-MUNOZ, 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, 2012.
- PASTOR-SANCHEZ, J.-A.; MENDEZ, F. J. M.; RODRIGUEZ-MUNOZ, J. V. Advantages of thesauri representation with the Simple Knowledge Organization System (SKOS) compared with other proposed alternatives. *Information Research*, vol. 14, dez. 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em 04 jun. 2018
- RAMALHO, R. A. S.; MARTINS, P. G. M.; SOUSA, J. L. Evolução das linguagens de marcação: um breve histórico à luz da área de Ciência da Informação. *Inf. Prof., Londrina*, v. 6, n. 2, jul./dez. 2017. p. 20 – 34. Disponível em: <http://www.uel.br/revistas/infoprof/>. Acesso em 10 jan 2019.
- RAMALHO, R.A.S. Análise do Modelo de Dados SKOS: Sistema de Organização do Conhecimento Simples para a Web. *Informação & Tecnologia (Itec)*, v. 2, 2015. p. 66-79.
- SANCHEZ-JIMÉNEZ, R; GIL-URDICIÁIN, B. Lenguajes documentales y ontologías. In: *El profesional de la información*. v. 16, n. 6, 2007. p. 551-560.
- SCHILLING, V. "Transforming Library Metadata into Linked Library Data", American Library Association, September 25, 2012. Disponível em:

<http://www.ala.org/alcts/resources/org/cat/research/linked-data>. Acesso em 10 jan 2019.

SOERGEL, D. Thesauri and Ontologies in Digital Libraries: Tutorial. In: Evaluation of thesauri. Joint Conference on Digital Libraries, Portland, Oregon, July 14, 2002. Disponível em: <<http://www.dsoergel.com/cv/B63.pdf>> Acesso em 08 set 2018.

SOLER-MONREAL, C; GIL-LEIVA, I. Evaluation of controlled vocabularies by inter-indexer consistency. Information Research: An International Electronic Journal, v. 16, n. 4, 2011. Disponível em: <http://www.informationr.net/ir/16-4/paper502.html>. Acesso em 10 jan 2019.

SOLER-MONREAL, M. C. Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores. Dpto. De Comunicación Audiovisual, Documentación e Historia del Arte (TESIS DOCTORAL). Universitat Politècnica de València. Valencia, 2009.

SOUSA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Modelos de representação semântica na era do Big Data. Brazilian Journal of Information Studies: Research Trends. v. 12, n. 3, 2018. p. 34-40.

SUOMINEN, O., MADER, C. Assessing and improving the quality of SKOS vocabularies. Journal on Data Semantics, v. 3, n. 1, 2014. p. 47–73.

SUOMINEN, O.; HYVÖNEN, E. Improving the quality of skos vocabularies with skosify. In Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012), Springer-Verlag, Gal-way, Ireland, October, 2012.

Apêndice – Categorias e erros

Tabela 2. Problemas comuns de avaliação

<i>Categoria</i>	<i>Problema</i>
Etiquetagem e problemas de documentação	Tag de idioma omitido ou inválido
	Cobertura incompleta de idiomas
	Ausência de idioma comum
	Conceitos não documentados
	Etiquetas sobrepostas
	Etiquetas ausentes
	Caracteres não imprimíveis em etiquetas
	Etiquetas vazias
Questões estruturais	Referências de notação ambígua
	Conceitos órfãos
	Clusters de conceito desconectados
	Relações hierárquicas cíclicas
	Relações associativas sem valor
	Conceitos únicos transitivamente Relacionados
	Conceitos relacionados unidirecionalmente
	Conceitos superiores omitidos,
	Conceitos principais como conceitos mais amplos
	Redundância hierárquica
Problemas específicos de dados vinculados	Relações reflexivas
	Links ausentes
	Links externos ausentes
	Links quebrados
	Recursos SKOS indefinidos
	Violação do esquema de URI/HTTP

Fonte: elaborado pelos autores