

# RELAÇÕES SEMÂNTICAS E INTEROPERABILIDADE EM TESAUROS REPRESENTADOS EM SKOS

## SEMANTIC RELATIONS AND INTEROPERABILITY IN THESAURUS REPRESENTED IN SKOS

Everton Rodrigues Barbosa<sup>a</sup>  
Angel Freddy Godoy Viera<sup>b</sup>

### RESUMO

**Objetivo:** Este estudo tem como objetivo compreender como o modelo de dados *Simple Knowledge Organization System* e seus modelos de extensão tem sido utilizado para promover a interoperabilidade com outros vocabulários e refinar as relações semânticas em tesouros na web. **Metodologia:** Utiliza a pesquisa documental nos guias de referência dos modelos de dados utilizados para representar os tesouros na web. **Resultados:** os modelos de dados têm sido utilizados para representar os termos e suas variações linguísticas, os relacionamentos entre grupos e subgrupos de conceitos, numa perspectiva intra-vocabulários, e os relacionamentos entre conceitos de vocabulários distintos, numa perspectiva inter-vocabulários. **Conclusões:** O uso do *Simple Knowledge Organization System*, e dos seus modelos de extensão contribuem para uma melhor estruturação dos conceitos em tesouros. Os modelos de extensão são apropriados para a representação dos relacionamentos de equivalência compostos, ou para a estruturação de grupos e subgrupos de conceitos em tesouros.

**Descritores:** Tesouros. *Linked Data*. Interoperabilidade. Relacionamentos semânticos. *Simple Knowledge Organization System*.

### 1 INTRODUÇÃO

No cenário das tecnologias digitais, o tesouro tem sido um instrumento utilizado para representar e organizar os recursos informacionais na web. Estes oferecem suporte ao processamento inteligente de informação, possibilitando a interoperabilidade com outros vocabulários controlados, inclusive, no âmbito dos

---

<sup>a</sup> Doutor em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina. E-mail: evertonpos@gmail.com

<sup>b</sup> Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina. Professor da Universidade Federal de Santa Catarina. E-mail: a.godoy@ufsc.br

dados conectados.

A permeabilidade dos vocabulários na *web* contribui para a evolução de uma *web* de dados. Uma vez modelados e interligados, os vocabulários ampliam não apenas a capacidade de troca semântica entre sistemas de computador, como também, permitem que os usuários ampliem suas possibilidades de busca e de recuperação da informação.

O desenvolvimento do modelo *Simple Knowledge Organization System* (SKOS) em 2009 foi um marco importante para a publicação de vocabulários controlados como *Linked Data*, pois permitiu que, os denominados Sistemas de Organização do Conhecimento (SOC) pudessem ser representados em um esquema de dados legível por computador (ZOGHLAMI; KERHERVÉ; GERBÉ, 2011). Desde então, inúmeras organizações passaram a concentrar esforços no sentido de disponibilizar seus vocabulários, favorecendo a reutilização destes por outras aplicações na *web*.

Nesta oportunidade, esta pesquisa concentra suas discussões em torno dos vocabulários controlados do tipo tesouro. Esta escolha fundamenta-se nos estudos de Mochón-Bezares, Méndez-Rodríguez e La Fuente (2020), no qual, foi constatado que os tesauros são os vocabulários controlados com maior adesão ao modelo de dados SKOS. No entanto, apesar dos ganhos com a adesão do modelo SKOS, algumas limitações persistem, principalmente no que diz respeito à necessidade de refinar os relacionamentos semânticos em tesauros, estruturar hierarquicamente os conceitos em vocabulários multidisciplinares e promover a interoperabilidade com outros conjuntos de dados.

Partindo do pressuposto de que essas limitações precisam ser melhor compreendidas, e que os modelos de dados são recursos imprescindíveis para sanar tais dificuldades, este estudo tem como objetivo investigar como o SKOS e seus modelos de extensão tem colaborado com aprimoramento das relações semânticas e com a interoperabilidade em tesauros na *web*. Apresenta como metodologia a pesquisa documental, na qual buscou analisar a documentação ou os guias de referências do SKOS, dos modelos de extensão ISO-THES e UNESKOS. Foram analisadas também amostras de representações em

tesauros publicados na *web* para ilustrar suas possibilidades de aplicação.

## 2 TESAUROS REPRESENTADOS EM SKOS

Os tesauros começaram a ser utilizados no campo da Ciência da Informação na década de 1940, “[...] sendo um instrumento capaz de transportar conceitos e suas relações mútuas, tal como expressos nas linguagens dos documentos, em uma língua regular, com controle de sinônimos e estruturas sintáticas simplificadas” (DODEBEI, 2002, p. 66). Segundo Zeng (2008) o tesouro é um tipo de SOC onde os conceitos que cobrem um domínio são representados por termos organizados de forma sistemática em três níveis de relacionamentos: hierárquico, associativo e de equivalência.

Dessa forma, o tesouro é capaz de conduzir o processo de indexação, e, do mesmo modo “[...] funciona como ferramenta para a recuperação da informação, auxiliando no processo de busca e expansão de consultas.” (HANANDEH, 2013, p. 676). Além de contribuir historicamente com a organização de grandes coleções de documentos, especialmente em arquivos, museus e bibliotecas, Carlan (2010, p. 32) afirma que os tesauros.

[...] têm sido utilizados também para organizar recursos digitais na Internet. Com a *web* semântica, as ferramentas para desenvolvimento de SOC estão se popularizando, devido à necessidade de compartilhamento com uso de padrões orientados ontologicamente.

A *web* tem impactado o uso dos tesauros, tradicionalmente eficazes em ambientes controlados, com estruturas claramente definidas e uteis para acessar informações. Porém, com a heterogeneidade de formatos e estruturas da *web*, os tesauros precisam incorporar modelos de dados flexíveis e que possam ser estendidos para melhor representar seu conteúdo.

Com efeito, para expandir a capacidade dos relacionamentos em tesauros, tornando-os mais adaptáveis e reutilizáveis em diferentes domínios, seria necessário formalizar a estrutura do tesouro em um conjunto de axiomas, numa solução mais próxima das ontologias, pois as ontologias tem maior capacidade expressiva de modelagem de sistemas de informação (GARCÍA-TORRES; PAREJA-LORA; PRADANA-LÓPEZ, 2008). Não obstante, Matthews,

Miles e Wilson (2009) concordam que projetar o tesauro na Web Semântica envolve necessariamente a permeabilidade com outras propostas (ontologias e mapas de conhecimento) ou seu uso combinado, no estabelecimento de uma mudança de paradigma na criação de tesauros, enfocando os tesauros conceituais.

O grupo de trabalho do W3C, denominado *SWAD-Europe Thesaurus Activity*, ao longo dos anos 1990 e 2000, registra o desenvolvimento e aplicação de modelos de dados voltados para a publicação de tesauros na *web*. Entre as iniciativas para representar os tesauros neste ambiente, destacam-se alguns modelos de dados e linguagens, como a linguagem *Extensible Markup Language* (XML) o modelo de dados *Resource Description Framework* (RDF), e a linguagem para descrever ontologias *Ontology Web Language* (OWL). Um dos projetos de maior destaque é o AGROVOC, desenvolvido e mantido pela Organização das Nações Unidas para Agricultura e Alimentação (FAO). O AGROVOC serviu como um estudo de caso para explorar a reengenharia de um tesauro tradicional em uma ontologia OWL (SOERGEL *et al.*, 2004), escolhida devido à possibilidade de especificar um domínio a partir de objetos e classes de objetos.

Mais adiante, ganha destaque o modelo *Simple Knowledge Organization System* (SKOS), apresentado pelo W3C como solução para o problema da baixa integração dos SOC na *web*. O SKOS foi desenvolvido com o objetivo de oferecer suporte efetivo às ações de organização da informação na *web*, permitindo que os tradicionais SOC pudessem ser expressos e compartilhados como dados legíveis por computador, tornando-os interoperáveis com outras sistemas (W3C, 2009a).

O modelo começou a ser desenvolvido em 2002, pelo grupo de trabalho *Semantic Web Advanced Development for Europe* (SWAD-E). Em 2005 foi difundido publicamente por meio da publicação do SKOS *core Guide*, porém, apenas em 2009 veio a se tornar uma recomendação oficial do W3C (PASTOR-SANCHEZ; MARTÍNEZ-MENDEZ; RODRIGUEZ-MUÑOZ, 2009).

A codificação do SKOS é baseada no modelo *RDF*, este último, segundo Ferreira e Santos (2013, p. 13), se utilizada da linguagem XML “como sintaxe

para o intercâmbio e o processamento de metadados”, colaborando para a interoperabilidade entre os sistemas de informação e de descrição existentes, pois é capaz de expressar o relacionamento de um objeto com suas propriedades e valores.

O SKOS é uma recomendação oficial do W3C e a ontologia mais usada para representar sistemas interoperáveis de organização do conhecimento na Web Semântica (MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020). Porém, sua versão inicial foi elaborada considerando os padrões ISO-5964:1985 e ISO-2788:1986, normas que foram atualizadas, principalmente, em função da necessidade de estabelecer diretrizes quanto a interoperabilidade dos tesauros com outros vocabulários.

## 2 EXTENSÕES DO MODELO SKOS

Na tentativa de oferecer soluções às limitações do SKOS Core original, algumas extensões do modelo foram desenvolvidas ao longo do tempo, como o SKOS *eXtension for Labels* (SKOS-XL). O SKOS-XL foi lançado em 2009 com o intuito de fornecer suporte adicional para identificar, descrever e vincular entidades que apresentam certa complexidade lexical (W3C, 2009b). O SKOS-XL dispõem de uma classe especial de entidades lexicais, chamada *skosxl:Label*, e cada instância desta classe possui uma única forma literal simples RDF. Ademais, três propriedades de rotulagem são usadas para rotular conceitos com instâncias de *skosxl:Label*, denominadas *skosxl:prefLabel*, *skosxl:altLabel* e *skosxl:hiddenLabel*. Estas possuem funções semelhantes às propriedades *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel* definidas no modelo SKOS convencional (W3C, 2009a). O Quadro 1 apresenta as principais propriedades do SKOS-XL.

**Quadro 1 – Propriedades do vocabulário SKOS-XL**

Propriedades	Descrição
<i>skosxl:Label</i>	Classe especial de entidades lexicais
<i>skosxl:literalForm</i>	Forma literal de uma entidade lexical
<i>skosxl:prefLabel</i>	Etiquetas preferidas
<i>skosxl:altLabel</i>	Etiquetas alternativas

<i>skosxl:hiddenLabel</i>	Etiquetas ocultas
<i>skosxl:labelRelation</i>	Serve como um ponto de extensão, que pode ser refinado para cenários de rotulagem mais específicos.

**Fonte:** W3C (2009a).

Apesar dos benefícios do SKOS-XL para representar entidades lexicais complexas, a extensão não cobre alguns relacionamentos de equivalência compostos. Uma equivalência simples se estabelece entre dois conceitos, porém, quando um conceito em um tesouro é representado pela combinação de vários conceitos, essa relação é considerada como um tipo de equivalência composta. Esse cenário ocorre devido à complexidade de representar os relacionamentos de equivalência entre um termo não preferido em uma determinada língua com pelo menos dois termos preferenciais (PASTOR-SÁNCHEZ, 2015a).

Nesse caso, é necessário adaptar o SKOS de acordo com as normas ISO-25964 (2011) e a ISO-25964 (2013), pois essas discorrem sobre a necessidade de representar os conceitos considerando outros tipos de relacionamentos, os quais, podem exigir níveis de correspondências lexical ou semânticas mais difíceis de realizar (MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019). O principal desafio se concentra na representação de diferentes níveis de relacionamento e na interoperabilidade com outros SOC (PASTOR-SÁNCHEZ, 2015a).

Desse modo, a extensão UNESKOS e a ontologia ISO-THES, desenvolvidas pelo grupo de trabalho ligado ao UNESCO Thesaurus, foi proposta para mitigar a limitação de manutenção dos diferentes tipos de equivalências e agrupamento de diferentes domínios em um único esquema de conceitos. O Quadro 2 mostra alguns exemplos de classes do vocabulário ISO-THES.

**Quadro 2 – Exemplos de classes do vocabulário ISO-THES**

<b>Principais classes</b>	<b>Descrição</b>
iso-thes:CompoundEquivalence	Para representar a relação de equivalência composta entre um termo preferido e um termo não preferido.
Iso-thes:ConceptGroup	Para representar um grupo de conceitos que constituem um subconjunto do tesouro.

iso-thes:microThesaurusof	Para vincular um termo a um grupo de conceitos do tesauro, reconhecida como um microtesauro.
Iso-thes:supergroup	Para vincular um termo a um grupo superior de conceitos do microtesauro.
iso-thes:subgroup	Para vincular um termo a um grupo subordinado de conceitos do microtesauro.
iso-thes:ThesaurusArray	Para vincular grupos de conceitos irmãos.

**Fonte:** Isaac e Tudhope (2015).

Um dos gatilhos para o desenvolvimento do UNESKOS foi a dificuldade em representar áreas de conhecimento que precisavam ser estruturadas em grupos e subgrupos de conceitos, reconhecidos como microtesauros. O desenvolvimento do UNESKOS se tornou necessário uma vez que o SKOS não fornece nenhum elemento nativo para uma representação direta dessas estruturas. Por esse motivo, o UNESKOS busca complementar o SKOS cumprindo as especificações da ISO-25964, na medida em que facilita a representação de qualquer tipo de tesauro, estruturando-os segundo grupos, subgrupos e supergrupos de classes de um domínio específico (ISAAC; TUDHOPE, 2015; PASTOR-SÁNCHEZ, 2015b). O Quadro 3 ilustra algumas propriedades do UNESKOS.

**Quadro 3 - UNESKOS Vocabulary Specification**

Propriedades	Descrição
uneskos:contains	propriedade inversa do <i>skos:inScheme</i> que permite definir os elementos que estão contidos em um esquema de conceito.
uneskos:hasMainConcept	propriedade que permite definir certos conceitos como pontos de acesso às coleções de conceito.
uneskos:hasMicroThesaurus	propriedade proposta como um mecanismo para descobrir o(s) grupo(s) de conceito(s) que conformam um esquema de conceito. Define o caminho para o grupo de conceitos, ao qual, o conceito pertence.
uneskos:mainConceptOf	Conecta um conceito a uma coleção.
uneskos:memberOf	propriedade inversa do <i>skos:member</i> para identificar a(s) Coleção (ões) às quais um conceito pertence. Esta propriedade permite definir links dos conceitos para suas coleções.

**Fonte:** Pastor-Sánchez (2015b).

A adesão do SKOS e dos modelos de extensão dependem de como os termos serão estruturados no tesouro. No entanto, entende-se que o principal motivo dos projetos de publicação dos tesouros na web a adotarem os modelos de extensão é a necessidade de especificar os tipos de relacionamentos ou de melhor estruturar classes de conceitos, principalmente, quando se trata de vocabulários que representam mais de uma área do conhecimento. Assim, acredita-se que os vocabulários buscam mitigar os problemas elucidados a partir de duas perspectivas: intra-vocabulários e inter-vocabulários.

O uso dos conceitos intra-vocabulários e inter-vocabulários estão ancorados, respectivamente, na noção de relações semânticas entre os termos de um mesmo vocabulário e na abordagem de mapeamento entre vocabulários distintos. Na literatura é possível identificar o uso de termos como: “*intra-thesaurus*” ou “*inter-thesaurus*” (BINDING; TUDHOPE, 2016, p. 8); “*intra-KOS*” ou “*inter-KOS*” (Baker *et al.*, 2013, p. 41); “*intra-language*” e “*inter-language*” em tesouros multilíngues (IFLA, 2009; INTERNATIONAL STANDARD ORGANIZATION, 2011); ou até mesmo “*Intra vocabulary*” e “*Inter vocabulary*” (INTERNATIONAL STANDARD ORGANIZATION, 2013; CLARKE, 2019; COX *et al.*, 2021). Optou-se, portanto, utilizar estes últimos para se referir a tesouros multilíngues ou monolíngues, identificados inclusive nas normas ISO.

A perspectiva inter-vocabulários é considerada uma abordagem de interoperabilidade realizado por meio do “mapeamento entre diferentes esquemas de vocabulários” (INTERNATIONAL STANDARD ORGANIZATION, 2013, p. vi). Para Clarke (2019), as pesquisas em bancos de dados seriam relativamente simples se todos usassem uma mesma linguagem natural, os mesmos protocolos de máquinas e a mesma linguagem de indexação e que para superar as disparidades encontradas, duas abordagens de interoperabilidade são especialmente relevantes para os SOC: mapeamentos inter-vocabulários e dados vinculados.

A perspectiva intra-vocabulários se fundamenta nas relações semânticas entre termos que estão no mesmo esquema de vocabulário. Cox *et al.* (2021) considera nestes casos as relações hierárquicas ou de equivalência. Por exemplo, quando dois ou mais termos que representam um mesmo conceito



estão no mesmo tesouro, um deles é designado como termo preferido e o(s) outro(s) termos não-preferenciais(s), esse tipo de relação é conhecida como de equivalência intra-vocabulário.

### **3 RELAÇÕES SEMÂNTICAS E INTEROPERABILIDADE: UMA PERSPECTIVA INTRA E INTER-VOCABULÁRIOS**

Os relacionamentos semânticos são relações baseadas no significado dos conceitos ou termos do vocabulário. No que se refere os relacionamentos entre conceitos, os vocabulários são estruturados numa perspectiva hierárquica e associativa. Já nas relações entre termos, os relacionamentos são estabelecidos entre termos preferenciais e não preferenciais, numa perspectiva de equivalência.

Aprimorar os relacionamentos semânticos em tesauros perpassa pelos relacionamentos entre grupos e subgrupos de conceitos, numa perspectiva intra-vocabulários, bem como questões relacionadas a interoperabilidade, numa perspectiva inter-vocabulários. Este último considerado um requisito para promoção de vínculos entre vocabulários controlados e outras fontes de dados na *web*.

No que diz respeito a perspectiva intra-vocabulários, alguns tesauros podem cobrir domínios complexos, nos quais, possuem uma grande quantidade de conhecimento compilado. Estes podem compreender subdomínios específicos, fazendo surgir a necessidade de estruturá-los em grupos de conceitos. Um exemplo clássico de um tesouro estruturado a partir de microtesauros é o Tesouro multidisciplinar da UNESCO. Nele, os conceitos estão agrupados em sete grandes áreas temáticas que se dividem em microtesauros.

No esforço para difundir os tesauros no ambiente dos dados conectados, além da preocupação com a estrutura interna dos conceitos, os tesauros devem concentrar esforços no sentido de promover a interoperabilidade com outros recursos na *web*. A interoperabilidade está associada a capacidade de mapear conceitos correspondentes em diferentes tesauros e demais vocabulários em dois níveis: lexical e semânticos.

A interoperabilidade se refere a capacidade de um sistema trocar ou

compartilhar informação, considera essencialmente suas características sintáticas ou estruturais. Quando se trata de interoperabilidade semântica, a atenção se volta não apenas para o intercâmbio, mas, principalmente, para o significado desses dados em diferentes sistemas de informação. No âmbito dos tesauros, a interoperabilidade é viabilizada quando os vínculos são estabelecidos entre conceitos de diferentes vocabulários. No contexto dos dados conectados, estabelecer esses relacionamentos com recursos externos, além de facilitar a identificação de correspondentes semânticos em conjuntos de dados heterogêneos, tem o potencial de estabelecer uma base lógica e conceitual consistente que ajuda os usuários a explorar e recuperar dados contextuais diretamente relacionados aos domínios representados no tesauro.

As correspondências entre termos podem acontecer no nível dos rótulos e / ou no nível dos conceitos. No entanto, para que a interoperabilidade seja assegurada, é necessário realizar comparativos entre vocabulários controlados no nível conceitual. Esses mapeamentos podem ser unidirecionais, ou seja, de um tesauro para o outro, ou recíprocos, quando são verificadas as correspondências mútuas entre os dois vocabulários.

A principal dificuldade em realizar os mapeamentos inter-vocabulários está na maneira de fazê-lo. Os mapeamentos podem ser realizados de maneira manual por especialistas, de maneira automática ou semiautomática. Apesar de a marcação manual oferecer uma precisão muito alta devido a participação de especialistas do domínio, ela não pode ser aplicada a grandes volumes de dados.

O mapeamento pode ocorrer entre mesmos tipos de vocabulários ou para diferentes tipos de bases de conhecimento, como por exemplo, o *Wikidata*<sup>1</sup>. Neubert (2017, p. 12) ressalta a importância de utilizar o *Wikidata* como um vocabulário de “*hub*”, ou seja, um conjunto de dados central, conectado a diversos outros conjuntos de dados na web. Os referidos autores afirmam que “devido à natureza do *Wikidata*, a rede de conceitos de muitos

---

<sup>1</sup> Lançado pela fundação Wikimedia em 2012, a fim de criar uma base de conhecimento compartilhada e fornecer dados estruturados para as Wikipédias em diferentes idiomas, Wikimedia Commons e outros projetos da fundação (WIKIMEDIA PROJECT, 2019).

domínios é acessível e mantida como uma única base de conhecimento aberta e colaborativa”. Assim, mapear um tesouro para vocabulários de “hub” é desejável por dois motivos: em primeiro lugar, o tesouro estará interligado a vários outros recursos em diferentes idiomas, o que pode ajudar a estender as explicações e contextos de conceitos; em segundo lugar, é possível explorar os *links* para outros SOC que já foram mapeados para os vocabulários de “hub”.

A tarefa de encontrar correspondentes semânticos entre o tesouro e os recursos de conjuntos de dados externos pode ser realizada por meio de serviços de pesquisa que consulta a nuvem LOD para então derivar um rótulo semântico correspondente no tesouro local. Segundo Wetz *et al.* (2012) essa abordagem pode ser chamada de interligação ou alinhamento de entidades que trazem um benefício para o tesouro, pois “a estratégia de vincular entidades LOD a um tesouro local usa os conceitos de dados vinculados para expandir e enriquecer as informações armazenadas no tesouro, levando a uma base de conhecimento mais valiosa.” (WETZ *et al.*, 2012, p. 7, tradução nossa).

Binding e Tudhope (2016) considera que, embora haja um vínculo conceitual bastante rico intra tesouros, pode não haver um *link* semântico correspondente de um conceito para diferentes tesouros, ou seja, um conceito em um tesouro local talvez não possua um rótulo correspondente em vocabulários externos, o que tornaria a participação de especialistas humanos ainda mais necessária. Nesse contexto, ressalta-se como o SKOS e seus modelos de extensão são aplicados no sentido de sanar tais dificuldades, algumas delas, elucidadas com a análise e discussão dos exemplos a seguir.

#### **4 ANÁLISE E DISCUSÃO**

De modo geral, as relações semânticas podem ser categorizadas em três tipos: equivalência, hierárquica e associativa. Estes podem incluir os relacionamentos entre grupos e subgrupos de conceitos, relacionamentos entre termos e suas variações linguísticas e os relacionamentos entre vocabulários

distintos (INTERNATIONAL STANDARD ORGANIZATION, 2011; 2013). Na sequência, serão apresentados os tipos de relacionamentos mais frequentes encontrados em tesouros publicados na *web*, e exemplos de como o SKOS e os modelos de extensão podem ser utilizados.

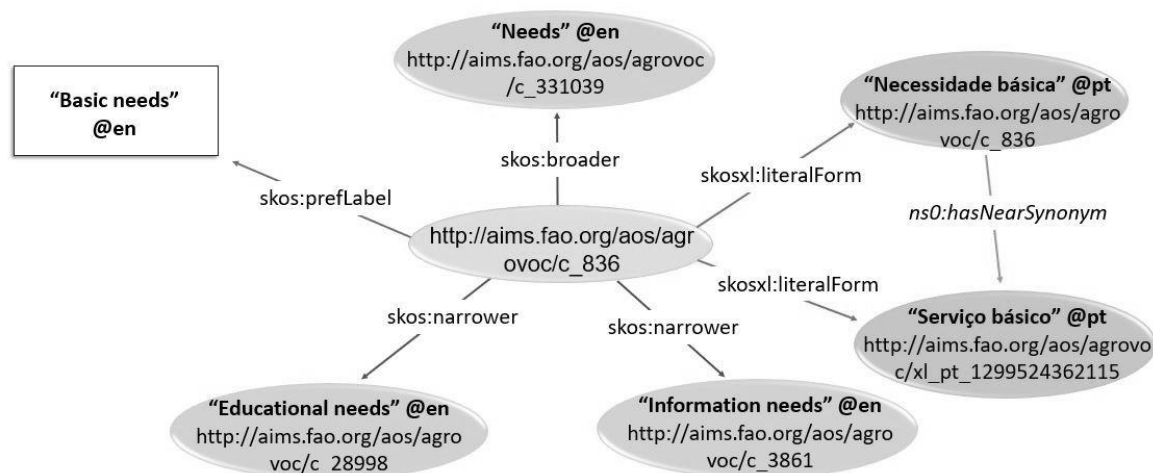
### **a) Termos e suas variações linguísticas**

As variações linguísticas em tesouros demandam um esforço de estruturação da linguagem conforme o domínio de aplicação, idiomas distintos, e as referências regionais ou usos locais. São representadas pelas relações de equivalência entre termos que possuem variações lexicais, porém com o mesmo significado. Não obstante, as variações semânticas também devem ser levadas em consideração, pois pode haver palavras semelhantes com significados diferentes. Nesses casos, os tesouros na *web* precisam utilizar vocabulários com propriedades capazes de sistematizar os relacionamentos entre as variações linguísticas. Assim, Caracciolo *et al.* (2013), Stellato *et al.* (2015) e Bandholtz *et al.* (2010) destacam o uso do SKOS-XL e ontologias próprias.

Enquanto os rótulos SKOS representam propriedades simples, os rótulos SKOS-XL são classes que podem ser vinculadas a outros conceitos por propriedades de objetos (BANDHOLTZ *et al.*, 2010), ou seja, no SKOS *Core* apenas os conceitos podem ser usados como recursos, as entidades alternativas se apresentam apenas vinculadas aos conceitos e não podem ser descritas com mais detalhes. De outra forma, o SKOS-XL permite que cada entidade seja representada como um recurso, que pode incluir atributos específicos, como por exemplo a representação de um termo em outro idioma. Ademais, os vocabulários podem enriquecer os relacionamentos entre variações linguísticas no mesmo idioma aplicando ontologias auxiliares, como é o caso do AGROVOC.

A Figura 1 ilustra o conceito “*Basic needs*” representado no AGROVOC no idioma inglês, a propriedade *skosxl:literalForm* é utilizada para expressar duas variantes linguísticas no idioma português, “Necessidade básica” e “Serviço básico”.

Figura 1 - Utilização das propriedades do SKOS-XL no AGROVOC



Fonte: Elaborado pelo autor baseado no AGROVOC.

Neste exemplo, é possível perceber que a relação de aproximação, ou de quase sinonímia, entre os dois termos em português é especificada por meio da propriedade *ns0:hasNearSynonym* da *Agrontology*<sup>2</sup>. Este último consiste em um vocabulário baseado na linguagem Web Ontology Language (OWL) que fornece um conjunto de propriedades de domínio para o AGRVOC Thesaurus.

Portanto, além da necessidade de representar a relação entre um conceito de um idioma com suas formas literais em outros idiomas, os tesouros na *web*, especialmente os multilíngues, podem utilizar alternativas como o desenvolvimento de ontologias para refinar os relacionamentos entre conceitos do vocabulário que não podem ser especificados utilizados as propriedades convencionais do SKOS ou SKOS-XL.

### b) Relacionamento entre grupos e subgrupos de conceitos (intra-vocabulários)

A habitual estruturação temática dos tesouros leva em consideração as principais disciplinas do domínio, no entanto, quanto se pretende representar

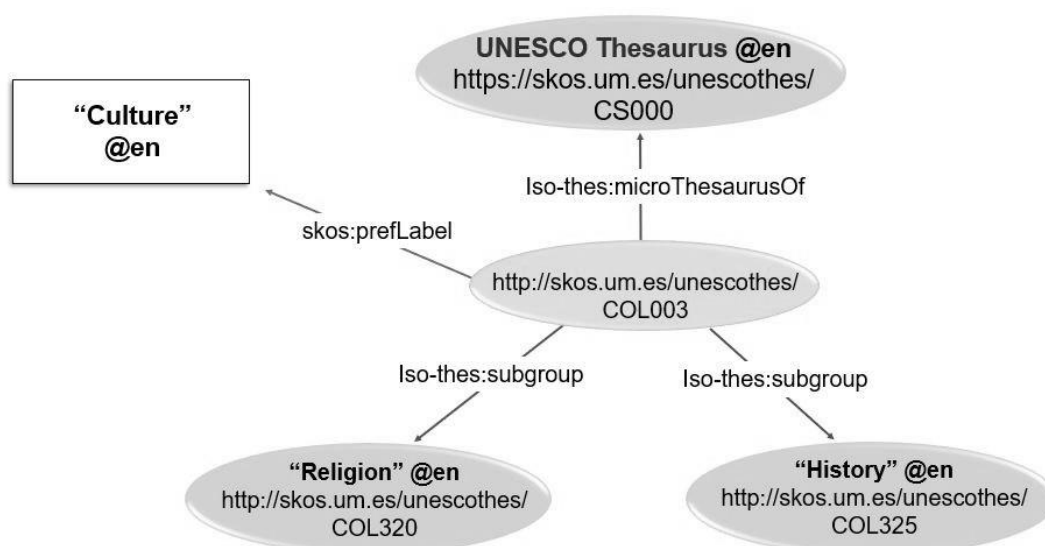
<sup>2</sup> <http://aims.fao.org/agrovoc/agrontology>

domínios complexos com elevado número de termos é preciso considerar os grupos e subgrupos de conceitos, bem como suas interrelações. Os subgrupos de conceitos são capazes de especificar a relação entre os termos das diversas subdisciplinas com as respectivas disciplinas de um domínio, fazendo com que todos os termos estejam vinculados a uma classe superior de conceitos. Nessa oportunidade, Isaac e Tudhope (2015) e Pastor-Sánchez (2015) recomendam o uso do ISO-THES.

O ISO-THES não só permite uma melhor visualização dos conceitos na medida em que os agrupa por áreas de conhecimento, como também, promove a organização e vinculação dos triplos RDF de maneira mais adequada, facilitando assim, a manutenção do conjunto de dados do tesouro.

Tomemos como exemplo a estruturação dos grupos e subgrupos de conceitos no Tesouro da Unesco da versão espanhola disponibilizada em SKOS (Figura 2). O termo UNESCO Thesaurus é tratado como um rótulo principal do esquema de conceitos, subordinado a ele existem diversos grupos de conceitos tratados como micro tesouros. Um desses grupos é representado pelo conceito “Culture”, que por sua vez, está relacionado a outros conceitos tratados como subgrupos de conceitos, como “Religion” e “History”.

**Figura 2 - Utilização das propriedades do ISO-THES no UNESCO Thesaurus**



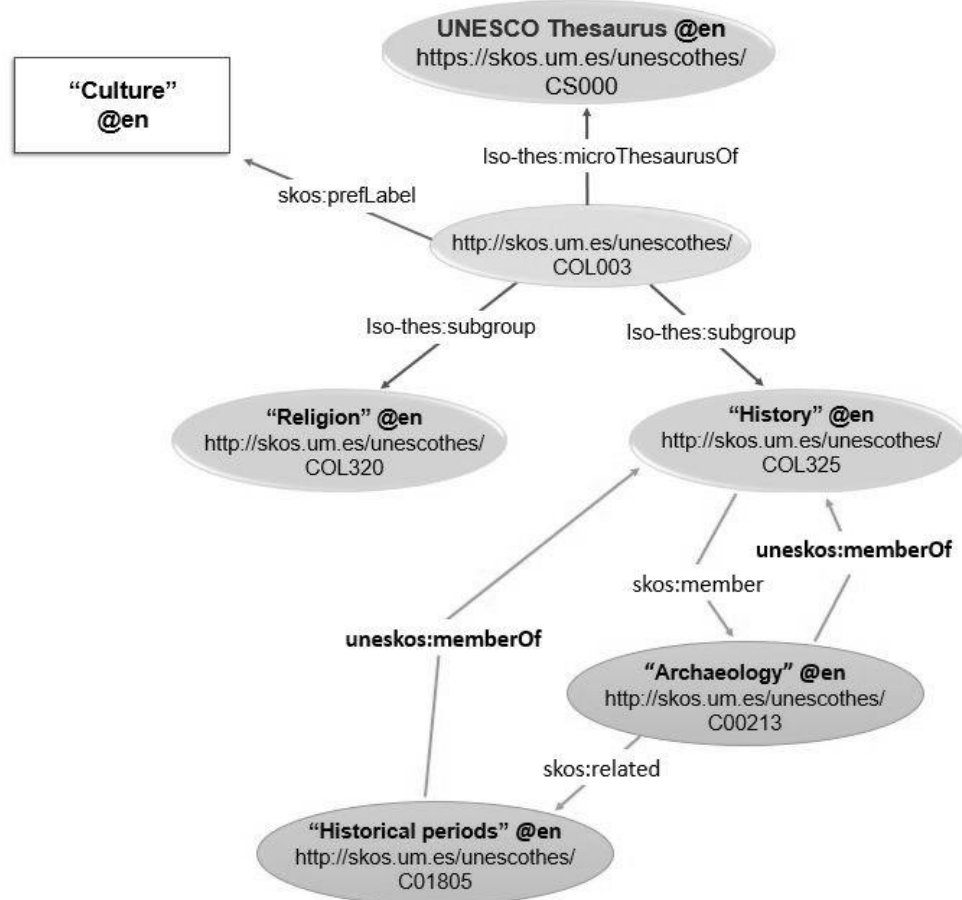
**Fonte:** Elaborado pelo autor baseado no UNESCO Thesaurus.

Os microtesouros, supergrupos e subgrupos de conceitos não são

estruturados como termos equivalentes, abrangentes ou específicos, as propriedades são utilizadas apenas para organizar as áreas do tesouro em termos macroestruturais. Embora os conceitos “*Religion*” e “*History*” constituam um subgrupo do conceito “*Culture*”, estes são representados como domínios a parte, para os quais, podem existir outros conceitos vinculados que assumem os tipos de relacionamentos convencionais. Foi então que surgiu a necessidade de desenvolver o modelo de extensão UNESKOS.

Para melhor ilustrar a aplicação da extensão UNESKOS, a Figura 3 apresenta o exemplo de uso da propriedade uneskos:memberOf no UNESCO Thesaurus. O conceito “*History*” é um grupo de conceitos, o qual “*Archaeology*” se apresenta como um conceito membro. Este último, por sua vez, está vinculado a outros conceitos mais abrangentes, relacionados ou mais específicos. Desse modo, facilita a manutenção da estrutura lógico-semântica do tesouro.

Figura 3 - Exemplo de aplicação do UNESKOS no UNESCO Thesaurus



Fonte: Elaborado pelo autor baseado no UNESCO Thesaurus.

A aplicação do UNESKOS amplia as possibilidades de estruturação dos conceitos na macroestrutura do tesouro, na perspectiva de que um termo pode estar relacionado a diversas classes de conceitos. Assim, os tesauros passam de estruturas convencionais planas a uma rede semântica de conceitos mais complexas, na qual, podem existir outros níveis de relacionamentos além dos tipos tradicionalmente aplicados.

### **c) Relacionamentos entre conceitos de vocabulários distintos (Inter-vocabulários)**

Para promover a interoperabilidade entre tesauros é preciso realizar o mapeamento entre diferentes esquemas de vocabulários. Os mapeamentos buscam identificar um conceito em um tesouro e vinculá-lo a outro conceito correspondente em um ou diversos vocabulários distintos na *web*. Os mapeamentos podem ocorrer entre conceitos no nível de equivalência exata ou aproximada.

Os conceitos em tesauros são representados pelos termos preferidos, no entanto, as representações podem variar conforme o tipo de vocabulário que se pretende mapear. Por exemplo um termo equivalente ou aproximado pode estar representado como uma notação em um sistema de classificação, como uma categoria em taxonomias, como termo tópico em esquemas de cabeçalho de assuntos, ou como rótulos em ontologias. Assim, os modelos de dados usados para representar os tesauros na *web* precisam utilizar propriedades que efetuem os mapeamentos entre vocabulários com características diferentes (NEUBERT, 2017; WETZ *et al.* 2012).

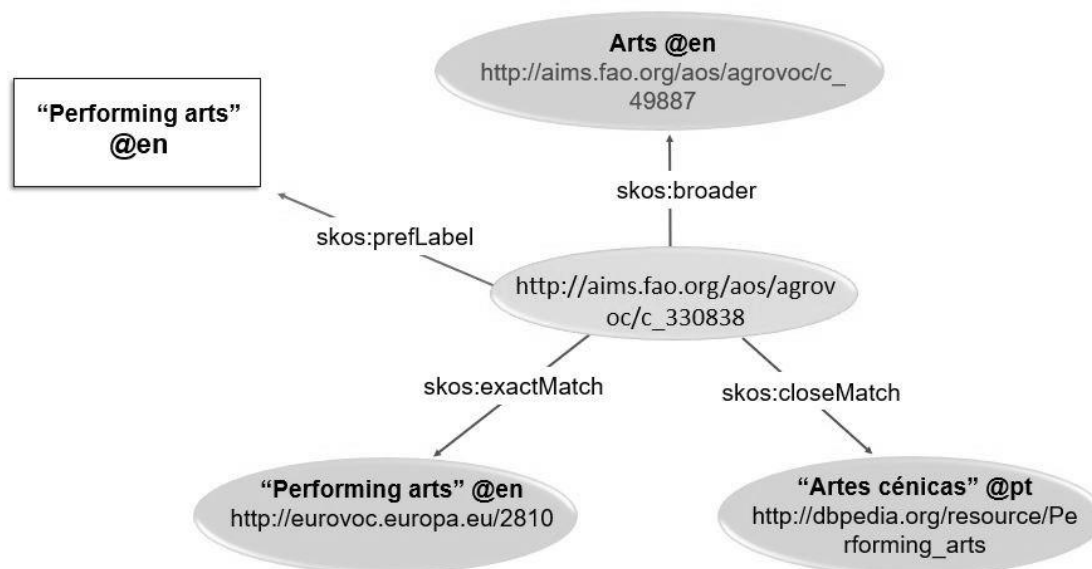
Neste aspecto, os relacionamentos entre os termos são especificados pelas propriedades do SKOS core, como o *skos:closeMatch*, *skos:exactMatch*, *skos:relatedMatch*. O SKOS e o SKOS-XL têm se mostrado suficientes para estabelecer vínculos com vocabulários externos. O exemplo ilustrado na Figura 4, mostra o uso das propriedades *skos:exactMatch* e *skos:closeMatch*. A primeira utilizada para estabelecer um vínculo entre conceitos exatos representados no vocabulário AGROVOC e no Eurovoc, e a segunda propriedade expressa a relação aproximada entre o conceito no AGROVOC e o



conceito “Artes cénicas”, no idioma português, representado como um rótulo do DBpédia<sup>3</sup>.

Tais propriedades podem ser utilizadas em casos de equivalência exatas, quando o conceito de um vocabulário de origem possui o mesmo significado e escopo do conceito no vocabulário de destino. De outra forma, se os conceitos possuírem algumas diferenças de conotação ou contexto, ou se possuírem significados aproximados, embora que um seja mais abrangente que o outro, as propriedades podem ser utilizadas para representar uma equivalência inexata ou parcial.

**Figura 4 - Exemplo de aplicação de propriedades do SKOS para estabelecer vínculos inter-vocabulários**



**Fonte:** Elaborado pelo autor baseado nos vocabulários AGROVOC e DBpédia.

Desse modo, é possível executar mapeamentos recíprocos entre tesauros e vocabulários distintos, principalmente àqueles que estão diretamente relacionadas ao domínio. A interligação de dados, como é chamada a atividade no âmbito dos projetos de LOD, está relacionada à possibilidade de o tesauro local encontrar entidades exatas ou aproximadas em conjuntos de dados mantidos por outras instituições na *web*, permitindo desempenhar o maior

---

<sup>3</sup> Projeto que extrai conteúdo estruturado da plataforma Wikipédia para criar uma rede de dados concetados, permitindo que os usuarios possam realizar consultas sobre esse conteúdo, semelhante as consultas realizadas em bancos de dados. (DBPEDIA ASSOCIATION, 2020).

potencial de reconciliação com a nuvem de dados.

## 5 CONSIDERAÇÕES FINAIS

Os subsídios teóricos apresentados nesta pesquisa permitiram revelar a permeabilidade dos SOC, especialmente os tesauros, no contexto da Web Semântica. Apesar de tradicionalmente serem construídos sob uma estrutura léxica rígida, o desenvolvimento de padrões, tecnologias e recomendações do W3C permitiram que os tesauros ampliassem sua capacidade de auxiliar na busca e recuperação da informação em ambientes digitais.

Percebe-se que alguns projetos estão reduzidos à interligação de conceitos em tesauros, ao passo que outros oferecem a possibilidade de explorar informações contextuais e fornecer acesso a outros conjuntos de dados, como o *Wikidata* e o *DBpédia*. Estes projetos, usam as propriedades do tipo *skos:Match* para ligar conceitos a recursos externos na *web*, ampliar a reconciliação com a nuvem de dados e expandir as possibilidades de busca e recuperação da informação. Desse modo, os usuários podem encontrar não apenas a informação que necessitam, mas também recursos complementares diretamente relacionados ao domínio do conhecimento, podendo enriquecer significativamente suas pesquisas.

Nessa perspectiva, por meio da análise documental e das amostras em tesauros publicados na *web*, foi possível identificar que os modelos de dados e ontologias tem sido utilizado para representar os termos e suas variações linguísticas, os relacionamentos entre grupos e subgrupos de conceitos, numa perspectiva intra-vocabulários, e os relacionamentos entre conceitos de vocabulários distintos, numa perspectiva inter-vocabulários.

Tanto na perspectiva intra-vocabulários como na perspectiva inter-vocabulários, Caracciolo *et al.* (2013), Stellato *et al.* (2015) e Bandholtz *et al.* (2010) consideram a abordagem multicultural e multilíngue da representação do conhecimento um ponto a ser dirimido. Uma vez que a linguagem pode apresentar variações lexicais ou semânticas que mudam conforme o público, o idioma, a região ou o local em que ela será aplicada. Com isso, o tesauro deve considerar o princípio da garantia cultural, pois indivíduos de diferentes culturas

podem interpretar o conhecimento e ter necessidades informacionais distintas.

Sendo a *web* uma rede global de informações, e uma vez que os tesouros estejam interligados com outros conjuntos de dados, inevitavelmente, o tesouro precisa estar alinhado semanticamente a conceitos em outros idiomas. As equivalências entre os conceitos precisam conferir o significado, o escopo, contexto de uso de um termo nos idiomas em que cobrem o domínio do tesouro. Assim, os mapeamentos entre termos de vocabulários precisam ser realizados considerando não apenas o nível lexical, mas também, o nível conceitual.

Havendo a possibilidade de promover a integração dos dados por meio de conceitos, e sendo o tesouro é um instrumento importante na recuperação da informação em plataformas digitais, recomenda-se em futuros estudos os desdobramentos da pesquisa e desenvolver melhor a aplicação demonstrada, de maneira que explore a importância de desenvolver ontologias ou modelos de extensão para refinar os relacionamentos semânticos em tesouros, e assim, melhorar recuperação de da informação na *web*.

## REFERÊNCIAS

BANDHOLTZ, T.; SCHULTE-COERNE, T.; GLASER, R.; FOCK, J.; KELLER, T. iQvoc – Open Source SKOS (XL) Maintenance and Publishing Tool. *In: EXTENDED SEMANTIC WEB CONFERENCE*, 7., 2010. **Proceedings** [...]. Heraklion, Crete, Greece, 2010. p. 1-12. Disponível em: <http://ceur-ws.org/Vol-699/Paper2.pdf>. Acesso em: 11 dez. 2020.

BAKER, T.; BECHHOFFER, S.; ISAAC, A.; MILES, A.; SCHREIBER, G.; SUMMERS, E. Key choices in the design of Simple Knowledge Organization System (SKOS). **Journal of Web Semantics**, [S.l.], v. 20, p. 35-49, 2013. Disponível em: <http://dx.doi.org/10.1016/j.websem.2013.05.001>. Acesso em: 11 dez. 2020.

BINDING, C.; TUDHOPE, D. Improving interoperability using vocabulary linked data. **International Journal on Digital Libraries**, [S.l.], v. 17, n. 1, p. 5-21, 2016. Disponível em: <https://doi.org/10.1007/s00799-015-0166-y>. Acesso em: 07 abr. 2020.

CARLAN, E. **Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação**. 2010. 195 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, 2010. Disponível em:

<http://eprints.rclis.org/14519/1/Carlan-Eliana-Dissertacao.pdf>. Acesso em: 10 jan. 2020.

CARACCILO, C.; STELLATO, A.; MORSHED, A.; JOHANNSEN, G.; RAJBHANDARI, S.; JAQUES, Y.; KEIZER, J. The AGROVOC Linked Dataset. **Semantic Web**, v. 4, n. 3, p. 341-348, 2013. Disponível em: <https://dl.acm.org/doi/10.5555/2786071.2786087>. Acesso em: 07 abr. 2020.

CLARKE, S. G. D. Thesaurus (for information retrieval). In: **Encyclopedia of Knowledge Organization**. ISKO, 2019. Disponível em: <https://www.isko.org/cyclo/thesaurus>. Acesso em: 08 fev. 2022.

COX, S. J. D.; GONZALEZ-BELTRAN, A. N.; MAGAGNA, B.; MARINESCU, M.-C. Ten simple rules for making a vocabulary FAIR. **Plos Computational Biology**, [S.l.], v. 17, n. 6, jun. 2021. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009041>. Acesso em: 08 fev. 2022.

DBPEDIA ASSOCIATION. **DBpedia**. [2020]. Disponível em: <https://wiki.dbpedia.org/>. Acesso em: 09 ago. 2020.

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

FERREIRA, J. A.; SANTOS, P. L. V. A. da C. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade: Estudos**, João Pessoa, v. 23, n. 2, p. 13-23, 2013. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/15436>. Acesso em: 10 jan. 2019.

GARCÍA-TORRES, A.; PAREJA-LORA, A.; PRADANA-LÓPEZ, D. Reutilización de tesouros: el documentalista frente al reto de la web semántica. **El profesional de la información**, v. 17, n. 1, p. 8-21, 2008. Disponível em: <https://doi.org/10.3145/epi.2008.ene.02>. Acesso em: 10 jan. 2019.

HANANDEH, E. Building an Automatic Thesaurus to Enhance Information Retrieval Information Retrieval. **International journal of computer science**, [S.l.], v. 10, n. 3, p. 22-29, 2013. Disponível em: <https://ijcsi.org/papers/IJCSI-10-1-3-676-686.pdf>. Acesso em: 10 jan. 2019.

IFLA. **Guidelines for multilingual thesauri**. The Hague, IFLA Headquarters, 2009. Disponível em: <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>. Acesso em: 08 fev. 2022.

ISAAC, A.; TUDHOPE, D. S. **[ISO-THES]**. 2015. Disponível em: <http://pub.tenforce.com/schemas/iso25964/skos-thes>. Acesso em: 18 out. 2019.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964-1**: Information and Documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. Geneva, Switzerland: ISO, 2011.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964-2**: Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies. Geneva, Switzerland: ISO, 2013.

MARTÍNEZ-GONZÁLEZ, M. M.; ALVITE-DÍEZ, M. The support of constructs in thesaurus tools from a Semantic Web perspective: framework to assess standard conformance. **Computer Standards & Interfaces**, [S.l.], v. 65, p. 79-91, jul. 2019a. Disponível em: <https://doi.org/10.1016/j.csi.2019.02.003>. Acesso em: 04 nov. 2019.

MATTHEWS, B.; MILES, A.; WILSON, M. **Modelling thesauri for the Semantic Web**: Semantic Web Advanced Development for Europe (SWAD-Europe), 2009. Disponível em: <http://www.webcitation.org/5m2lmCyQY>. Acesso em: 1 maio 2018.

MOCHÓN-BEZARES, G.; MÉNDEZ-RODRÍGUEZ, E. M.; LA FUENTE, G. B. de. Conversión normalizada (SKOS) de sistemas de organización del conocimiento interoperables en la web. **Investigación Bibliotecológica: archivonomía, bibliotecología e información**, Ciudad de México, v. 34, n. 82, p. 53-86, 2020. Disponível em: <http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58071>. Acesso em: 07 abr. 2020.

NEUBERT, J. Wikidata as a linking hub for knowledge organization systems? Integrating an authority mapping into Wikidata and learning lessons for KOS mappings. *In*: EUROPEAN NETWORKED KNOWLEDGE ORGANIZATION SYSTEMS (NKOS) WORKSHOP, 17., 2017. **Proceedings** [...] Thessaloniki, Greece: CEUR, 2017. p. 1-12. Disponível em: <http://ceur-ws.org/Vol-1937/paper2.pdf>. Acesso em: 07 abr. 2020.

PASTOR-SÁNCHEZ, J.-A. Proposal to represent the UNESCO Thesaurus for the semantic web applying ISO-25964. **Brazilian Journal of Information Science**, v. 9, n. 2, p. 1-8, 2015a. Disponível em: <https://doi.org/10.36311/1981-1640.2015.v9n2.01.p1>. Acesso em: 12 jan. 2019.

PASTOR-SÁNCHEZ, J.-A. **UNESKOS Vocabulary**. 2015b. Disponível em: <http://skos.um.es/TR/uneskos/#UNESKOS>. Acesso em: 10 jan. 2020.

PASTOR-SANCHEZ, J.-A.; MARTÍNEZ-MENDEZ, F. J.; RODRIGUEZ-MUÑOZ, J. V. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. **Information Research**, [S.l.], v. 14, n. 4, p. 1-15, 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em: 12 jan. 2019.

SOERGEL, D.; LAUSER, B.; LIANG, A.; FISSEHA, F.; KEIZER, J.; KATZ, S. Reengineering thesauri for new applications: The AGROVOC example. **Journal of Digital Information**, [S.l.], v. 4, n. 4, p. 1-23, 2004. Disponível em: <http://www.fao.org/3/af234e/af234e00.htm>. Acesso em: 12 jan. 2019.

STELLATO, A.; RAJBHANDARI, S.; TURBATI, A.; FIORELLI, M.; CARACCILO, C.; LORENZETTI, T.; KEIZER, J.; PAZIENZA, M. T. VocBench: A Web Application for Collaborative Development of Multilingual Thesauri. *In*: GANDON, F.; SABOU, M.; SACK, H.; D'AMATO, C.; CUDRÉ-MAUROUX, P.; ZIMMERMANN, A. (ed.). **The Semantic Web, Latest Advances and New Domains, ESWC 2015**: Lecture Notes in Computer Science, v. 9088. Cham: Springer, 2015a. p. 38-53. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-18818-8\\_3](https://link.springer.com/chapter/10.1007/978-3-319-18818-8_3). Acesso em: 04 abr. 2020.

W3C. **SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant**. 2009a. Disponível em: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/skos-xl.html>. Acesso em: 13 maio 2018.

W3C. **SKOS Simple Knowledge Organization System Reference**. 2009b. Disponível em: <https://www.w3.org/TR/skos-reference/>. Acesso em: 13 maio 2018.

WETZ, P.; STERN, H.; JAKOBITSCH, J.; PAMMER, V. Matching linked open data entities to local thesaurus concepts. *In*: SEMANTICS, I., Posters & Demonstrations Track. Graz, Austria, p. 5-7, 2012. **Proceedings** [...] Europe: CEUR Workshop Proceedings, 2012. p. 6-11. Disponível em: <http://ceur-ws.org/Vol-932/paper2.pdf>. Acesso em: 04 abr. 2020.

WIKIMEDIA PROJECT. **Wikidata**. 2019. Disponível em: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page). Acesso em: 04 abr. 2020.

ZENG, M. L. Knowledge organization systems (KOS). **KO Knowledge Organization**, [S.l.], v. 35, n. 2-3, p. 160-182, 2008. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-160.pdf>. Acesso em: 12 jan. 2019.

ZOGLHAMI, K.; KERHERVÉ, B.; GERBÉ, O. Using a SKOS engine to create, share and transfer terminology data sets. *In*: INTERNATIONAL CONFERENCE ON SIGNAL IMAGE TECHNOLOGY AND INTERNET-BASED SYSTEMS, 7., Dijon, France, 2011. **Proceedings** [...] New York: IEEE, 2011. Disponível em: <https://doi.org/10.1109/SITIS.2011.59>. Acesso em: 12 jan. 2019.

## SEMANTIC RELATIONS AND INTEROPERABILITY IN THESAURUS REPRESENTED IN SKOS

### ABSTRACT

**Objective:** This study aims to understand how SKOS and its extension models have been used to promote interoperability with other vocabularies and to refine semantic relationships in thesaurus on the web. **Methodology:** Uses documentary research in the reference guides of the data models used to represent thesaurus on the web. **Results:** data models have been used to represent terms and their linguistic variations, the relationships between groups and subgroups of concepts, in an intra-vocabulary perspective, and the relationships between different vocabulary concepts, in an inter-vocabulary perspective. **Conclusions:** The use of SKOS, and its extension models, contribute to a better structuring of concepts in thesaurus. Extension models are appropriate for representing compound equivalence relationships, or for structuring groups and subgroups of concepts in thesaurus.

**Descriptors:** Thesaurus. Linked Data. Interoperability. Semantic relationships. SKOS.

## RELACIONES SEMÁNTICAS E INTEROPERABILIDAD EN TESAuros REPRESENTADOS EN SKOS

### RESUMEN

**Objetivo:** Este estudio tiene como objetivo comprender cómo se han utilizado SKOS y sus modelos de extensión para promover la interoperabilidad con otros vocabularios y para refinar las relaciones semánticas en los tesauros en la web. **Metodología:** Utiliza la investigación documental en las guías de referencia de los modelos de datos utilizados para representar el tesoro en la web. **Resultados:** se han utilizado modelos de datos para representar términos y sus variaciones lingüísticas, las relaciones entre grupos y subgrupos de conceptos, en una perspectiva intra-vocabulario, y las relaciones entre diferentes conceptos de vocabulario, en una perspectiva inter-vocabulario. **Conclusiones:** El uso de SKOS y sus modelos de extensión contribuyen a una mejor estructuración de conceptos en tesoro. Los modelos de extensión son apropiados para representar relaciones de equivalencia compuestas o para estructurar grupos y subgrupos de conceptos en un tesoro.

**Descriptores:** Tesoro. Datos vinculados. Interoperabilidad. Relaciones semánticas. SKOS.

**Recebido em:** 29.09.2021

**Aceito em:** 14.02.2022