

THESA: FERRAMENTA PARA CONSTRUÇÃO DE TESAURO SEMÂNTICO APLICADO INTEROPERÁVEL

Rene Faustino Gabriel Junior

Faculdade de Biblioteconomia e Comunicação
Universidade Federal do Rio Grande do Sul (UFRGS)

Rita do Carmo Ferreira Laipelt

Faculdade de Biblioteconomia e Comunicação
Universidade Federal do Rio Grande do Sul (UFRGS)

Resumo

Apresenta para a comunidade acadêmica e de profissionais da informação o software para elaboração de tesouros THESA (Tesouro Semântico Aplicado). Descreve os conceitos e a estrutura do aplicativo para o desenvolvimento de múltiplos tesouros e multiusuários bem como as facilidades para o estabelecimento de relações semânticas propiciadas pelo sistema. Sua metodologia é baseada nas normas ISO e NIZO vigentes compatibilizadas com os requisitos semânticos prementes nas novas demandas dos sistemas de organização do conhecimento. O software funciona em ambiente web, pode ser baixado gratuitamente, e por tratar-se de um *open source* (código aberto), pode ser modificado ou aperfeiçoado desde que sejam mantidos os créditos. Entre suas vantagens está a expansão dos tipos de relacionamentos com OWL ou simplesmente com RDF. Foi concebido para ser uma ferramenta flexível e passível de incorporações futuras de novas propriedades de relacionamento entre termos, conceitos e esquemas. Com interoperabilidade entre tesouros de outros autores ou de outros domínios, possibilitará a construção de dicionários de sinônimos e inferências de tesouros já concebidos.

Palavras-chave: Thesa. Tesouro. Vocabulário controlado. Semântica.

THESA: SEMANTIC THESAURUS CONSTRUCTION TOOL APPLIED INTEROPERABLE

Abstract

It presents to the academic community and professionals of the information the THESA thesaurus software (Thesaurus Semantic Applied). Describes the concepts and structure of the application for the development of multiple thesauri and multiusers as well as the facilities for the establishment of semantic relations provided by the system. Its methodology is based on the ISO and NIZO standards in force, compatible with the semantic requirements of the new demands of knowledge organization systems. The software works in a web environment, can be downloaded for free, and because it is an open source, it can be modified or improved as long as the credits are kept. Among its advantages is the expansion of the types of relationships with OWL or simply with RDF. It is designed to be a flexible and future-capable tool for new properties of relationships between terms, concepts, and schemas. With interoperability between thesauri of other authors or other domains, it will allow the construction of thesaurus dictionaries and inferences already conceived.

Keywords: THESA. Thesaurus. Controlled vocabulary. Semantic.



INTRODUÇÃO

Os tesouros são construídos principalmente para atuarem como instrumentos de auxílio aos bibliotecários na realização da indexação. Porém, o formato tradicional desses instrumentos (impresso ou eletrônico) geralmente é desvinculado dos Sistemas de Recuperação da Informação (SRI), sobretudo no que tange ao ato de recuperação da informação em si, fator esse que os torna limitados. Aliado a esta dificuldade, destaca-se que a construção de tesouros ainda segue um padrão clássico de elaboração, optando sempre por um modelo simplificado das relações entre os termos, ou seja, baseando-se em uma relação minimalista, e por consequência proporcional de baixa relação semântica entre os conceitos. Exemplo disto são as relações sintagmáticas, representadas pelos termos relacionados, que historicamente não qualificam as relações que de fato ocorrem entre os termos, apenas indicam que existe uma relação entre eles sem qualificar o tipo.

Tradicionalmente, para a construção de tesouros, obedece-se a duas regulamentações: as Normas ANSI/NISO Z39.19 e ISO25964. A norma ISO25964, conhecida como ISO 2788:1986 (ISO, 1986) até 2011, foi elaborada pela *International Organization for Standardization* (ISO), com sede em Genebra, e é composta de duas partes, a primeira publicada em 2011 e descreve os padrões para construção de tesouros e a interoperabilidade com outros vocabulários, fornecendo recomendações para o desenvolvimento e manutenção de tesouros destinados a aplicações em Recuperação de Informações (RI). A segunda parte, publicada em 2013, apresenta as recomendações para o estabelecimento e manutenção de mapeamentos entre vários tesouros, ou entre tesouros e outros tipos de vocabulários (ISO, 2011; ISO, 2013).

O padrão ANSI/NISO Z39.19, publicado pelo *National Information Standards Organization* (NISO) em Baltimore nos Estados Unidos da América, apresenta as diretrizes e convenções para o conteúdo, exibição, construção, testes, manutenção e gerenciamento de vocabulários controlados monolíngues. Ela se concentra em vocabulários controlados que são usados para a representação de objetos de conteúdo em sistemas de organização do conhecimento, incluindo listas, anéis de sinônimo, taxonomias e tesouros (NISO, 2010).

Entretanto, para otimizar a representação de dicionários, esquemas de classificação, taxonomia, vocabulários controlados e tesouros, focado principalmente nos padrões da Web Semântica constituídos pelos padrões RDF e RDFS, surge em meados da década de 2010 o padrão *Simple Knowledge Organization System* (SKOS), como modelo para representação

tanto para seres humanos como para computadores. Modelo este recomendado pela *World Wide Web Consortium* (W3C) para compartilhamento de dados vinculados (*Linked data*). Com o SKOS (MILES; BECHHOFFER, 2009) é possível estabelecer relações semânticas fortes entre os conceitos a partir da ampliação da descrição de propriedades específicas de suas relações.

Diferente das normas ISO e NISO, que não especificam de forma detalhada o estabelecimento das relações e possibilitam apenas o uso de relações simplificadas, ou seja, relações semânticas fracas, o SKOS possibilita o detalhamento dessas relações, e ainda a incorporação de novas relações por meio de novas classes de propriedades, com o uso do RDF.

Na tentativa de localizar aplicativos para o desenvolvimento de tesouros e vocabulários controlados, foram estudados os softwares disponíveis na Internet que atendessem os critérios de serem multilíngues, multitema, colaborativos, atendessem as normas ISO e NISO, de acordo com as diretrizes do SKOS (RAMALHO, 2015), acessíveis pela web e ainda que disponibilizassem uma interface em português de com licença livre e semântica forte. Como resultado dessa busca exploratória, foram localizados diversos softwares para o desenvolvimento de vocabulários controlados. Entretanto, em uma análise dos mesmos, observou-se de modo geral que os softwares que poderiam atender os critérios eram pagos, com versões demos limitadas, ou softwares que foram desenvolvidos utilizando como modelo as normas ISO ou NISO, não atendendo aos padrões do modelo SKOS.

Buscando uma solução que atendesse a todos os critérios estabelecidos, alguns membros da linha Estudos em Organização do Conhecimento do Grupo de Pesquisa Comunicação Científica da UFRGS se propuseram a desenvolver um aplicativo para esse fim, que pudesse ser compartilhado com bibliotecas e instituições de ensino de Biblioteconomia, de forma gratuita e que possibilitasse o detalhamento das associações entre os termos e conceitos e dos relacionamentos entre os conceitos, estabelecendo semântica forte.

O objetivo deste artigo é apresentar para a comunidade acadêmica e de profissionais da informação o software Thesa que está sendo desenvolvido a partir das diretrizes do SKOS e compatibilidade com as normas ISO e NISO. Como desmembramento do objetivo principal, descreve-se: apresentar os conceitos e a estrutura do THESA para desenvolvimento de múltiplos tesouros e multiusuários; descrever o THESA e as facilidades para o estabelecimento das relações semânticas que o sistema propicia.

REFERENCIAL TEÓRICO

Na computação, Orbst (2011) escalona as relações semânticas em fraca (*weak*) até forte (*strong*). Esse escalonamento, segundo o autor, é caracterizado em três níveis: a primeira considerada com sintática, representada pelos glossários e os vocabulários controlados; o segundo estrutural, representadas pelas taxonomias e pelos tesouros; e pôr fim a semântica, com as teorias lógicas, mapas conceituais e a definição das ontologias. Quanto mais forte for a relação semântica, mais fácil será para gerar novos conhecimentos, seja pelos humanos ou pela “inteligência” das máquinas.

A CI já incorpora em suas teorias muitas das relações da semântica, o tesouro e todo instrumental para seu desenvolvimento são bons exemplos. Para Khoo e Na (2006) os estudos sobre Sistemas de Organização do Conhecimento (SOC) centram-se nos conceitos e termos, porém existe a necessidade de melhorar a eficácia e refinamento da metodologia de elaboração dos tesouros, com o aprofundamento dos estudos para a identificação, processamento e gestão das relações entre conceitos.

Na interpretação de Pastor-Sanchez, Martínez-Mendez e Rodríguez-Muñoz (2012) existem, atualmente, duas abordagens para a construção de tesouros no qual seus elementos são estruturados: uma considera um tesouro como um conjunto de **termos** interligados estabelecendo a relação semântica a partir deles, essa representação é a clássica adotada pelos padrões ISO e ANSI/NISO. A segunda alternativa, apontada pelos autores, destaca que os tesouros são relacionamentos semânticos de **conceitos**, sendo esses ligados aos termos sob a forma de etiquetas lexicais, que podem ou não ter relações lexicais estabelecidas entre eles, esse modelo é encontrado no SKOS.

Para fins desse estudo, e por haver diferenciações entre termos e conceitos, adota-se a definição de **conceito** definida por Dahlberg (1978) como uma unidade de conhecimento, e o **termo** como um elemento de representação do conceito. (CATARINO, 2014). Ainda para Dahlberg (1979, p. 361):

[. . .] os conceitos são sínteses rotuladas de enunciados verdadeiros sobre objetos do pensamento: esses enunciados – asserções – levam ao reconhecimento ou à separação das características dos conceitos, que também podem ser consideradas como os elementos dos conceitos.

Para o estabelecimento semântico é necessário definir esquemas de conceitos para um domínio, que, para Miles e Brickley (2005), são um conjunto de conceitos, que,

opcionalmente podem incluir declarações sobre os relacionamentos semânticos entre eles. Essas relações na CI são descritas nas relações associativas.

Relações associativas

Para Catarino (2014), a partir da explicitação das relações associativas, é possível melhorar a contextualização entre os conceitos. Dessa forma, é possível estabelecer uma semântica forte que possibilite seu compartilhamento tanto entre computadores como entre os profissionais da informação e seus usuários de modo que o tesauro deixe de ser um instrumento desconectado da realidade do SRI. Essa nova concepção amplia o uso e reuso dos tesouros no desenvolvimento dos novos sistemas de informação que estão sendo desenvolvidos. Visto que, o uso de relações simplificadas não possibilita sua compreensão pelos sistemas automatizados computacionais.

Nos modelos de tesouros semânticos, destaca-se que os significados, tanto das relações hierárquicas como das relações de equivalência, variam de acordo com o contexto. Já nos modelos tradicionais, as hierarquias são estabelecidas apenas a partir de premissas universalistas o que pode torná-las demasiadamente rígidas.

Uma grande dificuldade é a generalização feita em muitos tesouros quando se usa o termo mais amplo (TG) e o termo mais específico (TE), não só para relações lógicas que envolvem superordenação e subordinação (gênero / espécie), mas também para relações envolvendo membros Todo / parte). Quando podemos falar de subordinação per se, a hierarquia - em seu sentido lógico - é gênero / espécie (herança de propriedade). Embora as relações de gênero / espécie e de todo / parte sejam diferentes na maioria dos tesouros, os mesmos símbolos são usados para indicar dois tipos de hierarquia. O mesmo ocorre com as relações associativas que, embora possam ser distinguidas pelo tipo de vínculo que é estabelecido entre um conceito e outro, acabam sendo representadas por um único símbolo (TR). Essas relações devem indicar o tipo de associação em jogo (processo e consequência, atividade e agente, ação e paciente etc.). (LARA, 2013).

Olson (2007) observa que o padrão de construção do dicionário de sinônimos, ANSI / NISO Z39.19, prevê um conjunto limitado de relações de termos relacionados (TR) permitidos: processo / agente, processo / agente, ação / propriedade, ação / produto, ação / alvo, Causa / efeito, conceito de objeto / propriedade, conceito ou objeto / origem, conceito ou

objeto / medidas, matéria-prima / produto e disciplina ou campo / objeto ou praticante, bem como de oposição (antônimos).

Embora, a utilização das relações hierárquica dentro dos tesouros possibilite a identificação do Termo Geral (TG) e Termo Específico (TE) de uma relação, essa ainda é caracterizada como uma semântica fraca, por não possibilitar a contextualização das relações. Para demonstrar, utilizamos como exemplo, o termo “biblioteca” extraído do Tesouro em Ciência da Informação de Pinheiro e Ferrez (2014), onde “biblioteca” é apresentada como Termo Geral para “bibliotecas centrais”, “bibliotecas departamentais” e “bibliotecas públicas”, semanticamente vinculadas a estrutura da instituição; e a “bibliotecas híbridas” e “bibliotecas digitais” que se diferenciam pelo tipo de acervo disponibilizado. Pode-se observar que no tesouro, as relações são tratadas como iguais.

Fenômeno similar ocorre nas relações de equivalência (USE e UP), onde são apresentadas relações entre sinônimos como “resenhas de livros” e “recensões”; e entre termo completo e acrônimo, como “Requisitos Funcionais para Registros Bibliográficos” e “FRBR”, ambos exemplos extraídos do Tesouro de Pinheiro e Ferrez (2014).

Nas relações associativas é onde encontramos a maior diversidade de relações entre os conceitos, ou seja, para se estabelecer uma semântica forte nos tesouros é necessário descrever as relações associativas entre os conceitos. Szostak, Gnoli e López-Huertas (2016) reafirmam o que é proposto pela ISO, quanto a necessidade de maior explicitação a partir da especificação das tipologias de Termo Relacionado (TR). O desenvolvedor do tesouro provavelmente já tem uma ideia do tipo de relação associativa em questão, quando a mesma é identificada. No contexto da era digital, é necessário um olhar mais acurado para atribuir os TR, tendo em vista os benefícios que o encorajamento desta prática trará para o desenvolvimento de tesouros com semântica forte. Destaca-se que também podem existir relações associativas entre as propriedades, porém esse não é o foco deste artigo.

Os Tesouros tendem a se concentrar em substantivos, ignorar alguns verbos e traduzir outros verbos para substantivo (exemplo: aprendendo ao invés de aprendizagem), e tratar advérbios e adjetivos na melhor das hipóteses como constituintes em frases nominais em vez de separadamente. A análise acima implica que verbos, adjetivos e advérbios merecem mais atenção se quisermos buscar uma abordagem sintética à classificação na qual esses elementos distintos devem ser livremente combinados. Uma classificação de relações seria uma solução particularmente útil na construção de tesouros (ZENG, 2008; GREEN; BEAN; MYAENG, 2002). Em suma, um tesouro compreensivo que distingue diferentes tipos de relações associativas, de equivalência e hierárquica que englobe substantivos, verbos, advérbios e

adjetivos proporcionaria um melhor ponto de entrada para uma classificação abrangente baseada em fenômenos conforme a sugerida acima.

Neste sentido, pesquisas e práticas que envolvam as questões da organização da informação na Web são essenciais para a área. Pressupõe-se, portanto, que investir na formação dos profissionais da área de Ciência da Informação, inserindo nos currículos disciplinas e conteúdo que abordem as tecnologias e recomendações do W3C para a Web Semântica e Dados Lincados, é imprescindível. Contemplem-se as questões dos modelos de descrição, de construção de esquemas de conceito e os metadados e linguagens envolvidas (CATARINO, 2014).

Poucos estudos contextualizado essas relações são realizadas no Brasil, como destaque observa-se a pesquisa de Weiss e Bräscher (2016), que analisou a abordagem semântica pragmática, os estudos indicaram três tipos de relações semânticas dos tesouros: hierárquica, equivalência e associativa. Outro estudo teórico de Lara (2015) destaca que a CI exigiria novas lentes, bem como novas ferramentas para os mapeamentos semânticos, e apesar de ainda serem construídos tendo como referência princípios clássicos de organização, permitem trazer à tona uma série de SOC cujo conhecimento construído pode ser consultado, escolhido, combinado e re combinado sem se prender aos vértices que estão em sua origem.

130

Relações Semânticas dos Tesouros

As Normas da ISO e NIZO para tesouros e vocabulários controlados ainda não contemplam formas específicas para se estabelecer relações associativas fortes, e que na literatura ainda não existe pesquisa suficiente para determinar suas bases teóricas, conforme apontam Lancaster (1986), Austin (1993) e Marroni (2006).

Nessa abordagem, as relações associativas devem ser estabelecidas principalmente a partir da análise da literatura do tema que se propõe em verificar o tesouro, de modo a levar em conta a existência de diferentes pontos de vista ali presentes, e servir como um guia para o usuário localizar informações possivelmente inesperadas e adicionais, mas ainda assim úteis para que alcance seu objetivo.

Nos SRI tradicionais, a busca é realizada pela palavra-chave, resultando em diversos registros onde existam a incidência do termo conforme a estratégia de busca estabelecida. Na utilização de um sistema respaldado por um tesouro semântico, o SRI considera as propriedades estabelecidas entre os conceitos, de forma a oferecer buscas mais precisas,

baseadas em significados (ZAPICO; VIVAS, 2014), havendo um diálogo entre o usuário e o SRI.

Muitas tecnologias, linguagens e metadados têm sido desenvolvidos e recomendados para dar suporte às ações de organização da informação (CATARINO, 2014). Nesta perspectiva de fortalecer as relações semântica, dentre os modelos de desenvolvimento de vocabulários controlados e tesouros, destaca-se o *Simple Knowledge Organization System* (SKOS).

Simple Knowledge Organization System (SKOS)

O Simple Knowledge Organization System (SKOS) é um modelo para expressar a estrutura básica e conteúdo de sistemas de organização do conhecimento. SKOS é, como o nome representa, um sistema simples para a organização do conhecimento. Segundo Isaac e Summers (2009), o modelo serve para expressar vários tipos de esquemas de conceitos tais como tesouros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, Folksonomia, e outros tipos de vocabulários controlados.

131

Sua estrutura possibilita o compartilhamento e ligação dos Sistemas de Organização do Conhecimento (SOC) por meio da Web (MILES; BECHHOFER, 2009). O SKOS compartilha muito dos princípios de elaboração de tesouros, taxonomias, esquemas de classificação e sistemas de cabeçalhos de assuntos, bem como possibilita capturar muito dessa semelhança e torná-la explícita, para que possa ser reutilizada por outras aplicações (MILES; BECHHOFER, 2009). O modelo é composto por termos no vocabulário denominado SKOS *Core Vocabulary* (MILES; BRICKLEY, 2005) que é um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF (RDFSchema).

O SKOS foi adotado, em 2009, pelo W3C como um padrão para dar suporte a representação, uso e interoperabilidade de diferentes tipos de vocabulários controlados (ZOGHLAMI; KERHERVÉ; GERBÉ, 2011). Importante ressaltar que o SKOS também pode ser usado na conversão de tesouros já existentes, facilitando a integração, sendo que o modelo apresenta uma estrutura muito similar aos tradicionais utilizados nos SOC. O modelo SKOS baseia-se em classes e propriedades e para representar o conjunto de dados tem como elemento central o conceito. (CATARINO, 2014).

O SKOS pode ser entendido como um modelo para expressar esquemas de conceitos na organização da informação. Contudo, pouco se tem publicado no Brasil a respeito deste modelo de construção de esquemas de conceito; pressupõe-se que isto ocorra por existirem poucas pesquisas sobre esta temática (CATARINO, 2014).

O SKOS pode ser utilizado dentro dos Sistemas de Organização do Conhecimento (SOC), no entanto, a modelagem destes, seguindo o padrão da Web Semântica com base no RDF, aparentemente não tem sido difundido o suficiente na formação dos profissionais que atuam em unidades de informação, nas quais os SKOS são ferramentas imprescindíveis para as práticas da organização da informação. Portanto, entender o SKOS no âmbito da Ciência da Informação é necessário para disseminar este modelo entre os profissionais e pesquisadores da CI (CATARINO, 2014).

Entretanto, a conversão de tesouros e outros vocabulários em formatos compatíveis com a Web Semântica como o SKOS requer cautela, dado que a operação não é apenas instrumental e é neste momento que se tornam evidentes problemas de construção originalmente relacionados com a explicitação de conceitos e suas relações. A origem desses problemas é a falha em observar os princípios da organização lógico-semântica dos vocabulários (LARA, 2013).

132

Ferramentas para construção de tesouros

Os tesouros são sistemas valiosos de Organização do Conhecimento (OC) que suportam a recuperação avançada de informações. A Web Semântica trouxe um interesse renovado nos tesouros como suporte para buscas semânticas e outros serviços de valor agregado. Ferramentas que gerenciam tesouros permitem que eles sejam criados, editados e consultados. As restrições de integridade sobre os tesouros também devem ser controladas por essas ferramentas. No entanto, existe também a possibilidade de troca de tesouro, que se torna relevante no contexto da Web Semântica, onde o intercâmbio de informações é uma facilidade crucial. De fato, a interoperabilidade no nível de informação recebeu um impulso importante com a estabilização do padrão SKOS como uma Recomendação do World Wide Web Consortium (W3C). Além disso, a viabilidade da integração de software é um recurso valioso para desenvolvedores de software (MOYA-MARTÍNEZ; GIL LEIVA, 2001).

Buscando na literatura estudos que tivessem identificadas ferramentas de construção de tesouros na Web, foi localizado o estudo de Martínez-González e Alvite-Díez (2014) que

avaliaram diversas funcionalidades das ferramentas para construção de tesouros disponíveis, que incluiu as questões de suporte de construção, integridade, interoperabilidade de informação e viabilidade de integração de software. Com base nesse estudo, foram reanalisados os softwares identificados pelos autores, porém analisando o contexto de seu uso para ensino e aprendizagem nas universidades, fácil e gratuito, e concluiu-se que o MultiTes (MULTITES, 2017), desenvolvido desde 1983 em Java, tem uma versão em inglês gratuita com muitas limitações, e para acesso é necessário ter uma versão “PRO” paga.

O One-2-One ou 121 (TERMTREE, 2017) possibilita a construção de múltiplos tesouros e estruturas, porém a edição do tesouro é realizada em um programa fora da web, tem sua interface em inglês e a versão completa é paga, disponibiliza somente uma versão gratuita para testes por tempo limitado. O PoolParty (POOLPARTY, 2017; KOLLER, 2009) não foi possível testar, o software além de ser pago e em Inglês agregou outros serviços desde a avaliação de Martínez-González e Alvite-Díez (2014), transformando-se em uma companhia de serviços Web Semântica.

O SKOS Editor ou SKOSed (PRÓTÉGÉ, 2017) é um *plugin* do Protégé, que tem como principal função o desenvolvimento de Ontologias (OWL), pouco prático para construção de tesouros. ThManager (LACASTA; et al., 2007) é uma ferramenta de código aberto para a criação e visualização de vocabulários SKOS RDF uma iniciativa do W3C, a ferramenta foi implementada em Java, não utiliza a web como editor e está disponível somente em inglês. O TemaTres (GONZALES-AGUILAR; RAMÍREZ-POSADA; FERREYRA, 2012, GONZALES-AGUILAR, 2013) é o software que mais se enquadra no perfil desejado, funciona na web, tem interface em português, atende as normas ISO e NISO, porém dentro de suas limitações é a impossibilidade da construção de múltiplos tesouros em uma única instalação, e o estabelecimento específico das relações entre termos e conceito e entre conceitos, tratando tudo como termo.

Foram realizadas outras buscas na tentativa de identificar um software para construção de tesouros passível de ser utilizado em sala de aula, para aprendizagem, na qual pudessem ser desenvolvidos múltiplos tesouros, de forma colaborativa e ainda disponível em português. O único aplicativo localizado foi o VisualThesaurus (THINKMAP, 2017), que utiliza muitos conceitos do SKOS de forma visual, porém além da ferramenta ser paga, está disponível somente em Inglês.

Tesouro Semântico Aplicado (THESA)

O Thesa foi desenvolvido objetivando disponibilizar um instrumento para os estudantes de graduação de biblioteconomia na disciplina de Linguagens Documentárias para a elaboração de tesouros, de modo que possibilite reduzir o trabalho operacional e dar maior atenção ao trabalho de desenvolvimento cognitivo e conceitual referente a modelagem do domínio.

Como norteador do aplicativo, baseou-se nas normas ISO e NISO vigentes, de forma a compatibilizar suas diretrizes com os requisitos semânticos prementes nas novas demandas dos SOCs. Com base na literatura disponível, nas normas de construção de tesouros da ISO e NISO foram identificados os elementos necessários para o desenvolvimento do protótipo, principalmente no que tange ao levantamento das propriedades de ligação entre os conceitos. O protótipo foi desenvolvido utilizando a linguagem *php* e banco de dados *MySQL*, de forma a possibilitar o compartilhamento e desenvolvimento colaborativo da ferramenta.

O software funciona em ambiente Web e pode ser baixado gratuitamente, podendo ser utilizado para fins didáticos em disciplinas dos cursos de graduação e pós-graduação ou para uso profissional. O Thesa foi desenvolvido com o princípio de multi-idioma, podendo ser traduzido para qualquer idioma, entretanto sua versão de teste está somente em português, as traduções vão depender de se estabelecerem convênios com instituições nativas de outros idiomas, que demonstrem interesse pelo uso do software.

O Thesa utiliza uma concepção de múltiplos tesouros, ou múltiplos esquemas, ou seja, o usuário pode criar um número ilimitado de tesouros em diferentes áreas do conhecimento, os usuários/elaboradores desses tesouros, podem deixá-los para uso público ou privado, possibilitando o acesso de outros usuários. No Thesa partiu-se da concepção de URI, empregada pelo SKOS e sistemas baseados na Web Semântica, ou seja, cada conceito é associado a um endereço permanente na Internet e a um identificador único do conceito, e esse representado por termos por meio de propriedades.

A versão beta (de teste) do Thesa está pode ser acessada no endereço <http://www.ufrgs.br/tesouros>, o software também está pode ser baixado no GitHub em <https://github.com/ReneFGJr/Thesa>. Ressalta-se que o Thesa é um *open source* (código aberto), podendo ser modificado ou aperfeiçoado, desde que mantendo os créditos, e ainda aceita contribuições de melhoramentos pela comunidade.

É redundante descrever o Thesa como um “tesauro semântico aplicado”, pois todos os tesouros apresentam as relações entre os termos (semântica) sobre algum tema (aplicado). Porém, o que se pretende destacar no Thesa é a possibilidade de descrever as relações específicas, viabilizando desta forma o compartilhando de estrutura e dados em sistemas complexos, como o *Linked data*, e ainda manter a compatibilidade com modelos simplificados como da ISO e NISO.

A estrutura do Thesa é baseada na concepção das relações entre os conceitos, partindo do pressuposto que um conceito pode ser representado por um termo, uma imagem, um som, um link ou qualquer outra forma que possa ser explicitada. Nessa abordagem, o conceito é perene, enquanto a sua representação pode variar conforme o contexto histórico ou social, sendo definida uma forma preferencial, e inúmeras formas alternativas e ocultas.

A Figura 1 apresenta o modelo da estrutura do Thesa. Nela é possível observar que o seu núcleo principal é totalmente compatível com o SKOS, com a incorporação de algumas propriedades nas descrições das relações entre o conceito e o termo, como singular, gênero, abreviação, notação e variações léxicas.

A norma ISO 25.964:2011 (ISO, 2011) trata as relações semânticas de formas bem simples, utilizando somente TG, TE e TR. Opção equivalente é encontrada no SKOS, com as propriedades *narrower*, *broader*, e *related*. Porém como preconizado na literatura, existem outras explicitações que precisam ser feitas para relacionar os conceitos, desta forma, no Thesa foi necessário incorporar especificidades como “parte de”, “causa efeito”, “ação propriedade” entre outras, sendo o TG, TE, TR, *narrower*, *broader* e *related* somente suas simplificações para fins de compatibilidade.

A mesma lógica é utilizada nos descritores para representar os conceitos, a relação da “variação léxica” do Thesa, atribui-se a propriedade “altLabel” do SKOS e o “USE” na norma ISO. Desta forma, pode-se então partir de um grau de detalhamento (semântica forte) para representações simplificadas conforme o nível de interoperabilidade que se precise do SOC ou do SRI.

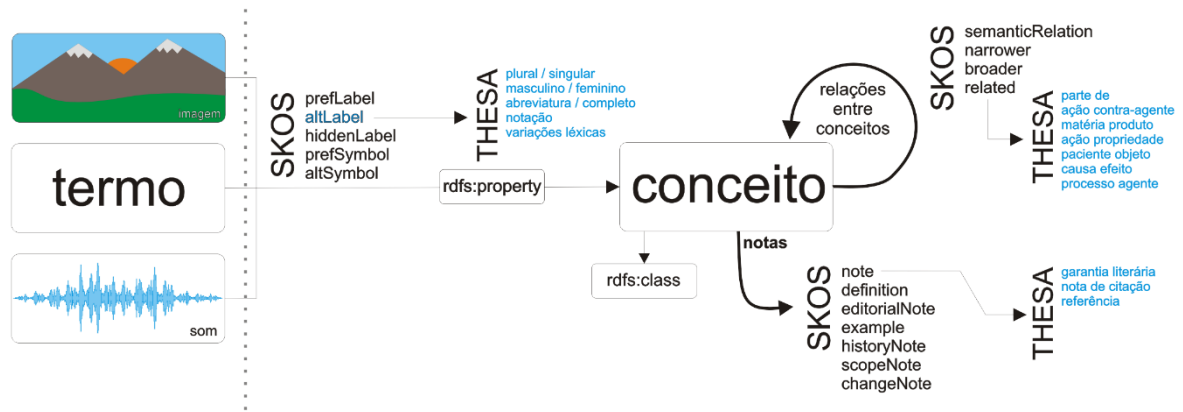


Figura 1 - Modelo da estrutura do Thesa/ Fonte: autores (2017).

Na concepção e construção de tesouros é fundamental a elaboração de Grafos que auxiliem a visualização das relações existentes, nesse sentido o Thesa gera automaticamente o mapa conceitual das relações, conforme estabelecidas, a Figura 2 apresenta um exemplo das relações em dois níveis do conceito “FRBR” e seus rótulos (termos preferidos) gerados pelo aplicativo.

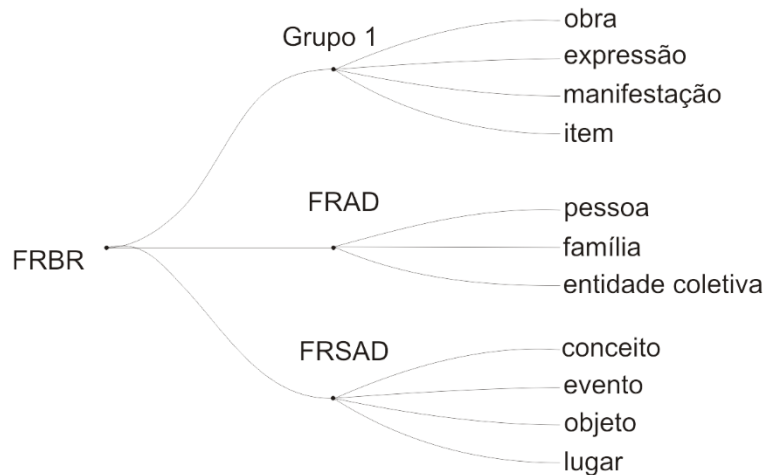


Figura 2 - Modelo de Mapa conceitual gerado pelo THESA/ Fonte: Autores (2017).

O Thesa foi desenhado para ser uma ferramenta flexível e passível atualizações, com a incorporações de novas propriedades de relacionamento entre os termos, conceitos e esquemas (domínios), bem como a interoperabilidade entre tesouros de outros autores ou de outros domínios, possibilitando futuramente a construção de dicionários de sinônimos e inferências de tesouros já concebidos.

Para a descrição dos recursos, o Thesa utiliza o formato RDF, composto de triplas para representar um recurso, propriedade e recurso. Segundo, Pastor-Sanchez, Martínez-Mendez e Rodríguez-Muñoz (2012) uma das vantagens da aplicação RDF, está na possibilidade de expandir os tipos de relacionamento com OWL ou simplesmente com RDF / RDFS, permitindo a criação de novas relações, tais como transitividade, simetria, relação inversa ou reflexividade, declarando rigorosamente as restrições semânticas necessárias.

A modelagem do Thesa, seguiu as diretrizes do SKOS, o que possibilitou descrição de um conceito com mais de um termo preferencial (perfTerm), desde que esteja em idiomas diferentes. Desta forma todo termo inserido no sistema deve estar relacionado a um idioma, para isso é incorporando o sufixo “@” e a notação do idioma em cada termo, um exemplo é o próprio termo “tesauro” que em português é representado por “tesauro@pt”, enquanto em inglês é utiliza-se como “Thesauri@en”.

Para se estabelecer uma semântica forte foi necessário desmembrar algumas notas gerais definidas pelas normas ISO e NISO, esse desmembramento possibilitou a incorporação de notas mais específicas como a “garantia literária” e a inserção das “referências”. Outros tipos de notas podem ser criados no Thesa com a incorporação de novas propriedades RDF. O detalhamento das notas é importante pois definem seu contexto histórico, aplicações e formas de uso, além de auxiliar os SRI baseados na Web.

Um dos requisitos do aplicativo veio da necessidade de se desenvolver vários vocabulários temáticos em uma única plataforma, possível de compartilhamento e reuso, ou seja, em uma única instalação ter vários tesouros com autores diferentes. No Thesa é necessário definir que é o “autor principal” do tema, esse perfil permite gerenciar colaboradores para desenvolvimento do tesouro e disponibilizar para acesso público. Cabe ressaltar que toda ação dentro do Thesa é registrada no sistema de rastreamento (log), possibilitando ao gestor analisar movimentações de seu tema.

Com a criação de diversos temas será a possibilidade de identificação de co-ocorrência de termos em diferentes domínios, gerando como resultado anéis de sinônimos. Para Olson (2002), os sistemas sobrepostos permitem a identificação de "objetos de fronteira", que podem ser mais apropriados para atender às necessidades de comunidades de prática específicas. Vários tipos de vocabulários podem ser ligados para mostrar significantes diferentes que têm significados comuns, sem dar prioridade a um vocabulário principal ou um vocabulário em detrimento de outro, estabelecendo relações quase-hierárquicas. (LARA, 2013).

Em um mundo conectado, não se pode mais conceber o desenvolvimento de SRI sem pensar em Dados Lincados (*Linked Data*), dessa forma o aplicativo deve ter funcionalidades que permitam a importação, exportação e interoperabilidade de seus conteúdos. O Thesa foi desenvolvido com a visão nessa nova realidade da informação, com a disponibilização de vários formatos de apresentação de dados, tanto para impressão como para compartilhamento, que pode ser na versão completa ou simplificada, exemplos dessas saídas são o Glossário, com a relação dos termos e remissivas; Índice; Índice Hierárquico (apresentação sistemática), mapa conceitual, arquivos Excel, XML, RDF e páginas Web para navegar *online* no conteúdo.

A Figura 3 apresenta a saída para impressão de um conceito, destaca-se a notação da URI, representada por “c155”, essa notação tem como termo preferencial o “FRBR”. É possível também visualizar seus termos específicos, as remissivas (usado por) e as notas de definição e garantia de literária.

The screenshot shows a web interface for the concept 'FRBR' (URI 'c155'). It features a search bar at the top with 'c155' and 'FRBR' entered. Below the search bar, there are three main sections: 'Termo(s) específico(s)' with a bulleted list of terms like 'TE Frad', 'TE Frsad', and 'TE Grupo 1'; 'Termo Equivalente (usado por)' with a bulleted list including 'UP Functional requirements for bibliographic records'; and 'Notas' which contains a detailed definition of FRBR and a 'skos:garantiaLiteraria' note about its efficiency in information recovery.

Figura 3 - Exemplo de impressão do conceito FRBR no Thesa./ **Fonte:** Autores (2017).

Quando se busca o reuso da informação, as ferramentas de **interoperabilidade** do sistema são de suma importância, seja para a troca de informação com outros SRI, ou seja, para uso do vocabulário pelo usuário em outros sistemas. Nesse contexto o Thesa foi desenhado para ser uma ferramenta para possibilitar esse compartilhamento. A Figura 4 apresenta um exemplo de exportação em formato XML, esse arquivo é compreensível somente por computadores.

```
1. <rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core">
2. <skos:Concept rdf:about="http://www.ufrgs.br/tesauros/c/v/155#frbr">
```

```

3. <skos:inScheme rdf:resource="http://www.ufrgs.br/tesauros/schema"/>
4. <skos:prefLabel lang="en">FRBR</skos:prefLabel>
5. <skos:altLabel lang="en">Functional requirements ....</skos:altLabel>
6. <skos:narrower>thesa:170</skos:narrower>
6. <skos:narrower>thesa:177</skos:narrower>
6. <skos:narrower>thesa:190</skos:narrower>
7. <skosxl:garantiaLiteraria>"... eficiência na recuperação da informação ....
    </skos:narrowed>
10. </skos:Concept>
11. </rdf:RDF>

```

Figura 4 - Arquivo de exportação XML do conceito FRBR.

Fonte: Autores (2017).

Das ferramentas de interoperabilidade do Thesa, destaca-se a possibilidade de compartilhamento de conceitos entre temas diferentes, ou ainda entre instalações diferentes. O Thesa utiliza acrônimos para cada instalação, o que possibilita a criação de uma única ID para cada sistema associada à sua URL (endereço na internet). Dessa forma é possível o compartilhamento e reuso de um conceito descrito em outro tesauro, importando todas suas definições, relacionamentos e notações, podendo ainda ser notificado quando o conceito original sofrer alguma atualização. Na Figura 5 é apresentado o conceito “FRBR” e ao lado direito as ferramentas de interoperabilidade, observa-se que a URI é composta de seu acrônimo “thesa” seguida da ID do conceito “c155”.

139

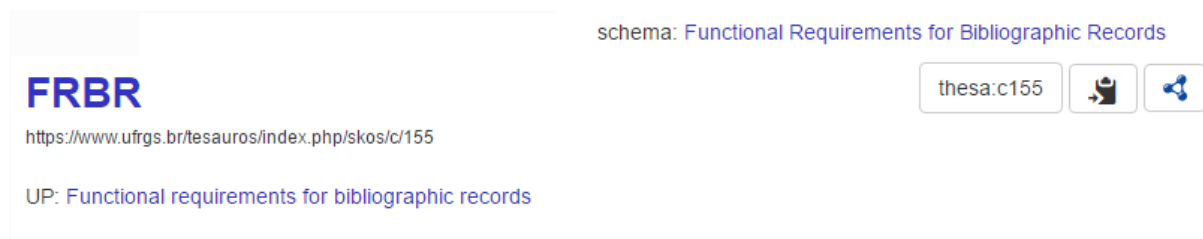


Figura 5 - Ferramentas de interoperabilidade do Thesa para o conceito FRBR.

Fonte: Autores (2017).

Nas Figuras 6 e 7 são apresentadas a interface Web do Thesa, com as telas de entrada e da seleção do tesauro. O sistema foi desenvolvido utilizando o framework *Codeigniter*, compatível com *Module, View and Controller* (MVC) e os estilos do *Bootstrap*,

compatibilizando o sistema com aplicativos de celulares. As cores do sistema podem ser personalizadas.

A interface do Thesa

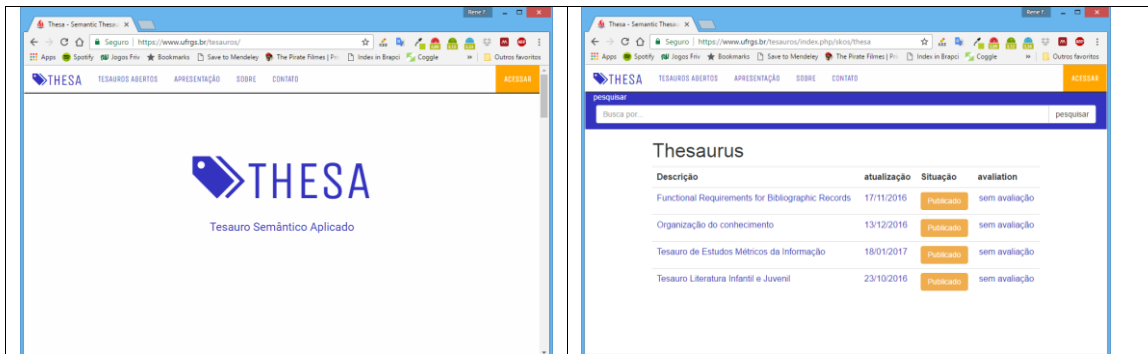


Figura 6 - Tela de entrada

Figura 7 - Tela de seleção do tesauro aberto

Considerações finais

O uso de vocabulários controlados e tesauros nos SRI sempre foi necessário para garantir a precisão. Porém, sempre foram limitados dentro dos sistemas de gestão de acervos de bibliotecas, deixando esses instrumentos alienados dos mecanismos de busca para o usuário, ou seja, instrumentos coadjuvantes desenvolvidos e utilizados pelos bibliotecários de forma manual sem o vínculo com os SRI.

O Thesa foi desenvolvido inicialmente para fins acadêmicos, principalmente no aprendizado dos estudantes de biblioteconomia da UFRGS, porém, vendo o potencial do sistema de garantir um gerenciamento mais eficiente, com a minimização de seus gestores das ações operacionais e de regras de validação, focando no estabelecimento das relações lógicas entre os conceitos, viu-se na ferramenta um grande potencial para a comunidade profissional de informação.

O aplicativo seguiu a diretrizes do SKOS e das normas ISO e NISO de tesauros e vocabulários controlados, buscando uma forma de garantir a compatibilidade com os diversos sistemas existentes e ainda os que poderão ser desenvolvidos. Partindo de uma revisão da literatura da área, pode-se observar que o muitos dos conceitos estabelecidos pelo SKOS já

estão incorporados pela comunidade acadêmica, porém poucos estudos utilizaram concretamente o modelo em uma realidade concreta. Desta forma, o Thesa vem a ser uma opção o desenvolvimento de vocabulários e tesouros, sendo um facilitador para o reuso de informação, alinhado a proposta de Dados Lincados do W3C e a Web Semântica.

A proposta desse estudo foi apresentar para a comunidade acadêmica e de profissionais da informação o software Thesa, aplicativo este que está sendo desenvolvido a partir das diretrizes do SKOS e compatibilidade com as normas ISO e NISO. Espera-se que os resultados gerados a partir do software sejam incorporados aos SRI e que a comunidade participe com críticas e sugestões de melhoramento, bem como desenvolvedores de SRI compatibilizem seus sistemas com os conceitos preconizados pelo SKOS, de forma a facilitar futuramente a contextualização da informação, tanto por parte do usuário como do profissional da informação.

Dos futuros resultados do uso do Thesa, está na possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas, bem como gerar inferências com bases em outros tesouros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

141

Agradecimentos

À Universidade Federal do Rio Grande do Sul, mais especificamente à Central de Processamento de Dados (CPD/UFRGS), pela disponibilização de um ambiente para execução dos protótipo, testes e disponibilização para a comunidade.

Referências

AUSTIN, Derek; DALE, Peter. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües**. Trad. Bianca Amaro de Melo. Brasília: IBICT; SENAI, 1993.

CATARINO, M. E. Simple knowledge organization system: construindo sistemas de organização do conhecimento no contexto da web semântica. **Informação & Tecnologia**, v. 1, n. 1, p. 17-28, 2014. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/v/a/15110>>. Acesso em: 04 jan. 2017.

DAHLBERG, I. Teoria da classificação, ontem e hoje. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1972, Rio de Janeiro. Anais... Rio de Janeiro: IBICT/ABDF, 1979. v.1, p.352-370.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, v. 7, n. 2, p. 101-107, 1978.

GONZALES-AGUILAR A.; RAMÍREZ-POSADA M.; D. FERREYRA. Tematres: software para gestionar tesauros. *El Profesional de la Información*, v. 21, n. 3, 2012. DOI: 10.3145/epi.2012.may.14

GONZALES-AGUILAR, A. Creación y visualización de tesauros a partir de un vocabulario de técnicas de la información y de la comunicación. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 18, n. 37, 2013. DOI:10.5007/1518-2924.2013v18n37p303.

GREEN, R.; BEAN, C. A.; MYAENG, S.H. **The semantic of relationships: an interdisciplinary approach**. Dordrecht: Kluwer Academic, 2002.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 1, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 2, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 2788:1986. Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri, 1986.

ISAAC, A.; SUMMERS, E. SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. Disponível em: <<http://www.w3.org/TR/skos-primer/>> Acesso em: 2 fev. 2017.

KHOO, C.S.G.; NA, J. Semantic relations in Information Science. *Annual Review of Information Science and Technology*, v.40, p.157-228, 2006.

KOLLER A. SKOS Thesaurus management based on linked data with PoolParty. In: Third annual European semantic technology conference, Vienna, 2-3 December 2009. Disponível em: <<https://pdfs.semanticscholar.org/69a0/78a5b1c7a4091cfbff30039bf177b429d14c.pdf>>. Acesso em: 21 fev. 2017.

LACASTA, J. et al. ThManager: an open source tool for creating and visualizing SKOS. **Information Technology and Library**, v. 26, n. 3, 2007. DOI: 10.6017/ital.v26i3.3274

LANCASTER, F.W. Vocabulary control for information retrieval. Arlington: Information Resources, 1986.

LARA, M. L. G. Documentary languages and knowledge organization systems in the context of the semantic web. **Transinformação**, v. 25, n. 2, 2013.

LARA, M. L. G. 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**, v. 20, n. esp.1, 2015. DOI:10.5007/1518-2924.2015v20nesp1p89.

MARRONI, G. N. B. Identificação e delimitação de relações associativas em tesauros: um estudo de caso na área de direito do trabalho. 2006. 142 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2006.

MARTÍNEZ-GONZÁLEZ, M.M.; ALVITE-DÍEZ, M.-L. On the evaluation of Thesaurus tools compatible with the Semantic Web. **Journal of Information Science**, v. 40, n. 6, 12, dez. 2014, p. 711-722.

MILES A.; BECHHOFFER S. SKOS Simple Knowledge Organization System reference. W3C Recommendation 18 August 2009. Disponível em: < <https://www.w3.org/2004/02/skos/specs> >. Acesso em: 19 fev. 2017.

MILES, A.; BRICKLEY, D. SKOS Core Guide. W3C: 2005a. Disponível em:<<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>>. Acesso em: 2 fev. 2017.

MOYA-MARTÍNEZ, G; GIL LEIVA, I. Evaluación de software de gestión de tesauros. **Ciencias de la Información**, v.3, n. 32, 2001.

MULTITES. Thesaurus Authoring and Publishing Solutions . Disponível em: <<http://www.multites.com/productsWCS.htm>>. Acesso em 19 fev. 2017.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). ANSI/NISO Z39.19-2005 (R2010) - Guidelines for the construction, format, and management of monolingual controlled vocabularies. 2010.

OLSON, H. A. Classification and universality: Application and construction. **Semiótica**. n. 139, p. 377–391, 2002. DOI:10.1515/semi.2002.031

OLSON, H. A. How we construct subjects: A feminist analysis. **Library Trends**, v. 56, n. 2, 2007, 509–541.

ORBST. The need for ontologies: Bridging the barriers of terminology and data structure. Special Paper of the Geological Society of America, n. 482, p.99-123, September 2011. DOI: 10.1130/2011.2482(10)

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

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

POOLPARTY. Semantic Web Company. Disponível em: <<https://www.poolparty.biz/>>. Acesso em: 19 fev. 2017.

PRÓTÉGÉ. Skos Editor. Disponível em: <http://protegewiki.stanford.edu/wiki/SKOS_Editor>. Acesso em: 19 de fev. 2017.

144

RAMALHO, R. R. A. S. Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a web. **Informação & Tecnologia**, v. 2, n. 1, p. 66-79, 2015.

SZOSTAK, R.; GNOLI, C.; LÓPEZ-HUERTAS, M. **Interdisciplinary Knowledge Organization**. Base: Springer, 2016.

TERMTREE. Business Classification Scheme and Thesaurus Management Software. Disponível em: <<http://www.termtree.com.au/>>. Acesso em 19 fev. 2017.

THINKMAP. Disponível em: < <http://www.thinkmap.com/>>. Acesso em 3 fev. 2017.

WEISS, L. C.; BRÄSCHER, M. Relações semânticas em tesouros: contribuições da abordagem pragmática. InCID: Revista de Ciência da Informação e Documentação, v. 7, n. 2, 2016.10.11606/issn.2178-2075.v7i2p136-155. DOI:10.11606/issn.2178-2075.v7i2p136-155.

ZAPICO, M.; VIVAS, J. La sinonimia como caso particular de distância semântica. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 19, n. 40, 2014, p. 253-266. DOI:10.5007/1518-2924.2014v19n40p253.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization:** international journal devoted to concept theory, classification, indexing, and knowledge representation, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

ZOGLAMI, K.; KERHERVÉ, B.; GERBÉ, O. Using a SKOS engine to create, share and transfer terminology data sets. In: INTERNATIONAL CONFERENCE ON SIGNAL IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS, 7th, 2011, Dijon, France. Proceedings... Los Alamitos, California: IEEE Computing Society, 2011. p.46-53.