

A REPRESENTAÇÃO TEMÁTICA NO CONTEXTO DA WEB SEMÂNTICA¹

Maria Elisabete Catarino*
Brígida Maria Nogueira Cervantes**
Ilza Almeida de Andrade***

relato de pesquisa

RESUMO

O artigo aborda, por meio de pesquisa bibliográfica e documental, como as práticas da representação temática estão relacionadas com as atividades do *Word Wide Web Consortium* (W3C) para a Web Semântica. Apresenta inicialmente os conceitos e práticas da representação temática no âmbito da Ciência da Informação e em seguida discorre sobre os conceitos da Web Semântica, com destaque para um de seus componentes, os vocabulários e, por fim, verifica como a representação temática tem sido utilizada nas aplicações da Web Semântica. O levantamento foi feito nos idiomas português e inglês, em periódicos e livros das áreas de Ciência da Informação, *Internet* e *Web*, bem como em documentos disponíveis no site do W3C. Com a pesquisa constatou-se que as práticas da Organização da Informação, especificamente as da Representação Temática são compatíveis e aplicáveis no âmbito das recomendações do W3C, como se pode verificar nos projetos estudados que estão em desenvolvimento, por exemplo, da CDD e da CDU.

Palavras-chave: Web Semântica. Representação Temática. Ontologias. OWL. SKOS.

* Doutora em Tecnologias e Sistemas de Informação pela Universidade do Minho, Portugal. Professora do Departamento de Ciência da Informação da Universidade Estadual de Londrina, Brasil. Diretora do Sistema de Bibliotecas da Universidade Estadual de Londrina, Brasil.
E-mail: beteca@uel.br.

** Doutora em Ciência da Informação pela Universidade Estadual Paulista Julio de Mesquita Filho, Brasil. Docente permanente no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, Brasil.
E-mail: brigidacervantes@gmail.com.

*** Mestre em Gestão da Informação pela Universidade Estadual de Londrina, Brasil. Bibliotecária na Biblioteca Central da Universidade Estadual de Londrina, Brasil.
E-mail: ilza@uel.br.

I INTRODUÇÃO

Este trabalho apresenta o resultado de uma pesquisa cujo objetivo foi averiguar como as práticas da Ciência da Informação (no âmbito da Organização da Informação, mais especificamente a Representação Temática) estão relacionadas com as atividades do W3C para a Web Semântica. Este objetivo foi dividido em três objetivos específicos: a) discorrer sobre os conceitos da Web Semântica para um de seus componentes, os vocabulários; b) abordar os conceitos e práticas da Representação Temática no âmbito da Ciência da Informação; e c) verificar como a Representação Temática tem sido utilizada nas aplicações de Web Semântica.

Optou-se pela realização de uma pesquisa descritiva/exploratória por meio de pesquisa bibliográfica (em fontes nacionais e internacionais: livros e periódicos, das áreas de ciência da informação e web) e documental (recomendações e normas do consórcio Web).

Pretende-se com este estudo publicar as relações percebidas entre as práticas da Ciência da Informação para a Representação Temática de recursos com as propostas de normas e recomendações da W3C, visando aproximar os profissionais da Ciência da Informação ao contexto da Web Semântica, iniciando uma discussão que nos leve às melhores práticas e uma formação voltada para a organização de recursos na Web.

Nas seções a seguir serão abordados os conceitos da Representação Temática e da Web Semântica, na sequência estes conceitos no âmbito da proposta do W3C e a apresentação de exemplos do que se tem feito destacando-se a relação entre a prática da Ciência da Informação.

¹ Artigo elaborado a partir da pesquisa intitulada "Contribuições da Ciência da Informação para a Web Semântica", coordenada por M.E.CATARINO.

2 A REPRESENTAÇÃO TEMÁTICA NO ÂMBITO DA CIÊNCIA DA INFORMAÇÃO

A representação do conteúdo de um documento é denominada representação temática e abarca os processos de classificação e indexação. No que se refere ao conteúdo a representação temática, auxilia significativamente na recuperação da informação a partir da determinação e disponibilização de pontos de acesso, apoiadas pelos recursos tecnológicos. Em razão disso, atentar às características do ambiente e de seus usuários são fatores determinantes no processo de organização da informação. Nesse contexto, faz-se necessário refletir a respeito de modelos metodológicos de organização da informação e do conhecimento que atendam as necessidades informacionais de usuários no âmbito de um sistema de informação especializado. Fujita (2003, p. 176), esclarece que “[...] é preciso conhecer os objetivos do sistema de informação e a demanda da comunidade usuária para a seleção dos conceitos”.

A linguagem traduz e transfere em enunciados sequenciais o que se manifesta como simultaneidade superposta no cérebro e no real. Assim, utiliza-se a língua e outros sistemas de significação socialmente construídos para elaborar os significados, as representações que dão sentido à nossa vida. Salienta-se, com isso, que é na linguagem que se constroem as culturas humanas, ou seja, que se elaboram os discursos e as narrativas que direcionam nossas ações.

Nessa direção, a Documentação tem dedicado especial atenção à Representação Temática, seja no que se refere às chamadas linguagens hierárquicas (ou notacionais) de indexação; seja quanto às linguagens alfabéticas, conhecidas como listas de cabeçalhos de assunto e tesouros, e aos estudos sobre a interface Terminologia e Documentação (GUIMARÃES, 1994).

Com isso, permite-se que o sentido do texto original, elaborado em linguagem natural e que se aplica a contextos e situações específicas, seja representado por meio de uma Linguagem Documentária. A linguagem documentária é considerada como uma linguagem intermediária, por ser um recurso fundamental e muito utilizado para a representação temática da informação. Sendo assim, a linguagem documentária vem a

ser um instrumento que se destina a estabelecer a ligação entre a linguagem construída para o tratamento dos conteúdos informacionais e a linguagem utilizada pelo usuário para a recuperação dos conteúdos no âmbito de um sistema de recuperação da informação (LARA, 2004).

A linguagem documentária é composta de vocabulário e sintaxe.

O vocabulário se refere à relação dos descritores usados para a identificação do conteúdo de um documento e a sintaxe se refere às regras utilizadas para a combinação dos descritores usados para a identificação do conteúdo de um documento. (FUJITA, 2004, p. 26).

No âmbito das linguagens documentárias, o vocabulário deve ser entendido como o conjunto de ocorrências provenientes de várias áreas, da terminologia (linguagens de especialidades) e das palavras utilizadas pelos usuários (linguagem natural). Pelo fato de serem linguagens construídas, o vocabulário utilizado para a sua concepção é originário de discursos diferentes, representado por palavras preferenciais, numa relação combinatória de significados e significantes (CINTRA *et al.*, 2002). Os termos preferidos são estabelecidos pela relação entre termos dispostos no vocabulário de especialidade a que pertencem os conteúdos a serem representados e entre os termos utilizados pelos usuários.

Lopes (2002, p. 47) define linguagem documentária e/ou vocabulário controlado como “[...] um conjunto de termos organizados de forma hierarquizada e/ou alfabética, com o objetivo de possibilitar a recuperação de informações temáticas reduzindo substancialmente a diversidade de terminologia”. O vocabulário controlado, conforme a norma ANSI/NISO Z39.19 (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p. 5) é “Uma lista de termos organizados explicitamente. Todos os termos devem ter uma definição clara e inequívoca e devem ser controlados por uma Autoridade Registrada”. Um vocabulário controlado pode ser tão simples como uma pequena lista de termos ou tão complexo como um tesouro contendo termos com uma estrutura hierárquica complexa e diferentes tipos de relações entre os termos.

Os vocabulários controlados visam à padronização da indexação da informação, para que isso ocorra com qualidade, o vocabulário deve ser muito bem estruturado com a área relacionada. Para Guinchat e Menou (1994, p. 145) “quanto mais organizado for o vocabulário, maior é a possibilidade de controlá-lo, isto é, de evitar os duplos empregos, [...] ou ainda o uso errôneo de descritores”. Cunha e Cavalcanti (2008, p. 378) definem vocabulário controlado como “[...] conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como na recuperação. A finalidade principal deste controle é fazer coincidir a linguagem do pesquisador com a do indexador.” Nessa perspectiva, é uma importante ferramenta para auxiliar o bibliotecário na indexação, sendo a indexação a “[...] representação do conteúdo temático de um documento por meio dos elementos de uma linguagem documentária ou de termos extraídos do próprio documento” (CUNHA; CAVALCANTI, 2008, p. 193).

Em concordância com autores supracitados, vocabulário controlado tem o intuito de colaborar com os profissionais da informação para que, em conjunto com o vocabulário livre e/ou linguagem natural, forme instrumentos como os tesouros, cabeçalhos de assunto, e que em combinações certas ocorram às recuperações na busca pela informação.

Dessa forma, os vocabulários controlados se apresentam nas diversas áreas do conhecimento e tornam possíveis as normalizações temáticas, contribuindo de forma efetiva para o armazenamento e a recuperação das informações. Independente da linguagem utilizada faz-se necessário estabelecer uma relação entre o vocabulário natural do usuário e o vocabulário controlado, utilizado pelo sistema de recuperação da informação. Portanto um vocabulário controlado pode ser utilizado tanto no processo de representação temática como para auxiliar na busca e recuperação da informação.

Nessa perspectiva, a utilização dos vocabulários controlados como instrumento de controle de vocabulário contribui de forma efetiva para que isso aconteça, pois, o “[...] vocabulário controlado não é congelado no tempo, com as alterações de linguagem, novos conceitos surgem, alguns termos saem de moda ou tornam-se obsoletos”. (AMERICAN NATIONAL STANDARDS INSTITUTE;

NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p. 5). Assim, existe uma necessidade contínua de manutenção do vocabulário controlado, uma vez que tenha sido construído. Os vocabulários controlados auxiliam o indexador na organização da informação.

Sob a luz da Norma ANSI/NISO Z39.19:

Vocabulário Controlado é usado para melhorar a eficácia dos sistemas de armazenamento e recuperação, sistemas de navegação Web, e outros ambientes que buscam tanto identificar e localizar o conteúdo desejado através de algum tipo de descrição usando a linguagem. O propósito principal de controle de vocabulário é alcançar consistência na descrição dos objetos de conteúdo para facilitar a recuperação. (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p. 1, tradução nossa)

Os termos representam os conceitos e constituem os vocabulários controlados utilizados em indexação. Para a Norma NBR 12676, o conceito é “qualquer unidade de pensamento. O conceito pode ter o seu conteúdo semântico reexpresso pela combinação de outros conceitos, que podem variar de uma língua ou de uma cultura para outra”. (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 1).

No contexto da Web, são vistos esquemas como: tesouros, taxonomias, mapas conceituais, redes semânticas, *folksonomias* e ontologias “[...] São sistemas de organização e representação do conhecimento KOS (*knowledge Organization Systems*) - e servem para fazer a indexação de recursos na Web com diferentes vocabulários (alguns deles controlados, outros não)” (MOREIRO GONZÁLEZ, 2011, p. 16-17). Esses sistemas permitem que haja a socialização da informação, em ambientes digitais, por meio da representação que realiza.

O tesouro constitui-se no exemplo mais característico de vocabulário controlado, limita-se, em princípio, ao vocabulário de especialidade. Oportuno salientar que considera-se “especialidade” toda ação humana que exige conhecimento específico. Enfatiza-se que o tesouro, ainda que se aproprie das terminologias e dos sistemas conceituais por ele mapeados, se propõe como modo de organização da informação, articulando temas, subtemas e termos orientados

para dar conta de conjuntos documentários e para transferir informação para determinados públicos segundo objetivos específicos (LARA, 2001).

As taxonomias são criadas e utilizadas, muitas vezes, em aplicações de indexação para navegação na Web. Devido a sua (normalmente simples) estrutura hierárquica, são eficazes na condução dos usuários até os termos mais específicos em um determinado domínio (CERVANTES, 2009). É uma ferramenta que tem forte aproximação com o conceito da web semântica, possibilita padronizar os conteúdos semânticos em *sites* e a navegação por meio de ordenação hierárquica. Pode ser entendida ainda como instrumento para gestão e organização do conhecimento em ambientes organizacionais.

Campos e Gomes (2008) apresentam algumas características das taxonomias sendo essas: conter uma lista estruturada de conceitos/termos de um domínio; incluir termos sem definição, somente com relações hierárquicas; possibilitar a organização e recuperação de informação por meio de navegação; permitir agregação de dados, além de evidenciar um modelo conceitual do domínio; ser um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em um sistema de recuperação de informação; ser um novo mecanismo de consulta em portais institucionais.

Nessa direção, de acordo com Moreira (2012) os mapas conceituais são esquemas de conceitos relacionados e representados por meio de diagramas estruturados hierarquicamente. As redes semânticas, segundo Moreira González (2011, p.107) “[...] São grafos orientados que proporcionam uma representação declarativa de objetos, propriedades e relações [...]”. Já a ontologia refere-se a combinação dos termos e suas relações. Possui uma linguagem previamente estabelecida por especialistas, cuja função, é de atuar com regras que dirigem as relações entre os termos, consentindo que a busca por informação seja realizada por meio de consultas de conceitos (ALMEIDA; BAX, 2003).

3 O CONTEXTO DA WEB SEMÂNTICA

A Web Semântica é um projeto da *World Wide Web Consortium* (W3C) cujo objetivo é tornar as informações legíveis por máquina, por meio do desenvolvimento de tecnologias, linguagens, padrões e recomendações.

O conceito passou a ser utilizado a partir de 2001, quando Berners-Lee, Hendler e Lassila abordaram o tema num artigo publicado na revista *Scientific American*. No entanto, posteriormente, Berners-Lee considerou que o título mais adequado para o projeto seria *Data Web* ou *Web of Data* (Web de Dados) (KING, 2007; W3C, 2010).

Web Semântica não é uma “nova” Web no sentido de ser uma outra rede separada. Trata-se de uma nova proposta para a Web, na qual a informação possui um significado bem definido, permitindo melhor interação entre computadores e pessoas (BERNERS-LEE, T.; HENDLER, J.; LASSILA, 2001). Para estes autores os computadores precisam ter acesso a coleções de informações estruturadas (dados e metadados) e regras de inferência que auxiliem no processo de dedução automática para a aplicação de raciocínio automatizado.

Segundo o W3C Brasil (2012), “a missão do W3C é levar a World Wide Web ao seu potencial máximo, desenvolvendo protocolos e diretrizes que garantam seu crescimento a longo prazo da web”. Os princípios que orientam os trabalhos da W3C são: a) Web para todos: neste princípio está contido o valor social do projeto da Web Semântica, pois pretende-se que todos tenham acesso ao conteúdo da Web, independente de tipo de *hardware*, *software*, infra-estrutura de rede, idioma, cultura, localização geográfica ou capacidade física e mental; b) Web em todas as coisas: este princípio significa tornar o conteúdo da Web disponível em qualquer tipo de dispositivo, tais como telefones celulares, *smartphones*, *tablets*, sistemas interativos de TV, eletrodomésticos, etc. (W3C BRASIL, 2012).

A visão do W3C é do compartilhamento do conhecimento em escala global (W3C BRASIL, 2012). No começo a Web era uma ferramenta de leitura, depois com a sua evolução (denominada por alguns de Web 2.0), tornou-se uma rede participativa, onde muitos usuários passaram de meros leitores a autores e editores e, principalmente, atores numa rede social global.

O projeto da Web Semântica concebe a ideia de uma Web de dados linkados (*linked data*), nesta concepção existem alguns conceitos básicos (dados linkados, busca, aplicações verticais, inferência e vocabulários) que serão apresentados a seguir.

Para tornar a Web Semântica uma realidade, os dados contidos na rede devem estar

em um formato padrão, acessível e gerenciável por tecnologias apropriadas. Além do acesso aos dados da Web, é necessário, ainda, criar relações entre os dados, ou seja, Dados Lincados, que é o nome que se dá às coleções de dados relacionados na Web (W3C, 2010). Para tanto é preciso adotar o *Resource Description Framework* (RDF) recomendado pelo W3C, que é um modelo padrão para intercâmbio de dados na Web (RDF WORKING GROUP, 2004).

O conceito busca (*Query*) refere-se, de forma genérica, às tecnologias e aos protocolos desenvolvidos para a recuperação da informação (W3C, 2010). Na Web Semântica, cujo modelo de dados é o RDF, existe uma linguagem própria de busca que é a SPARQL, diferente das existentes para as bases de dados tradicionais, como por exemplo a SQL para as bases de dados relacionais, e o XQuery para as bases de dados estruturadas em XML.

Aplicações verticais (*vertical applications*) é uma expressão usada para se referir às aplicações genéricas desenvolvidas por comunidades que exploram as tecnologias do W3C. Estas comunidades colaboram com os demais membros do consórcio cooperando para o aperfeiçoamento das recomendações, tecnologias e padrões desenvolvidos pelo W3C (2010).

Inferência, no contexto da Web Semântica, refere-se à descoberta de novas relações entre os recursos Web com base nos dados e em algumas informações adicionais sob a forma de um vocabulário ou conjunto de regras, sendo que ambas as formas recorrem a técnicas de representação do conhecimento (W3C, 2010). Os vocabulários definem as classes e subclasses de conceitos, e, ainda, como cada recurso pode ser associado a essas classes, caracterizando as relações entre as classes e suas instâncias. Já as regras objetivam a geração de novas relações baseadas nas já existentes.

Vocabulários são constituídos por termos que representam conceitos, possíveis relacionamentos entre eles e as limitações no seu uso. Os vocabulários podem também ser denominados "ontologias". Segundo o W3C (2010), na prática, não existe uma recomendação para o uso de um ou outro conceito, mas, comumente a tendência é usar o conceito "vocabulário" para coleções de termos mais simples e "ontologia" para as coleções mais complexas. O conceito de complexidade usado

aqui é no sentido de grau de formalidade, como, por exemplo, no âmbito das ontologias que, segundo a tipologia de Uschold e Gruninger, vão desde as Altamente Informais (as mais simples), até as Rigorosamente Formais (as mais complexas) (CATARINO, 2009).

Este estudo pretende destacar as relações prováveis entre a Web Semântica e a Representação Temática, neste sentido, a seguir serão abordados alguns temas já citados atrás: o conceito de Ontologia, linguagem OWL e o sistema de organização do conhecimento SKOS.

4 REPRESENTAÇÃO TEMÁTICA E A WEB SEMÂNTICA

Para abordar a Representação Temática no âmbito da Web Semântica, faremos uma apresentação sucinta sobre Ontologias, SKOS e algumas ações que estão sendo realizadas nesta direção.

A origem da palavra Ontologia remonta ao grego *Onto* (ser) + *Logos* (razão/palavra) e atualmente o termo tem sido usado em dois sentidos, um no campo da Filosofia e outro no contexto da computação.

Uma Ontologia para as Ciências da Computação e Informação é a especificação de uma conceitualização que é um conjunto de ideias, conceitos, relações ou outras abstrações que compõem o domínio de um modelo ou discurso. Uma Ontologia define um vocabulário representacional para a conceitualização, e especifica restrições no uso deste vocabulário de forma que os fatos sobre um determinado domínio podem ser compartilhados, comunicados e debatidos (GRUBER *et al.*, 2007). Não se pode deixar de referir aqui uma das definições mais citadas que é a de Gruber (1993): uma ontologia é uma especificação explícita de uma conceitualização.

conforme se observa na descrição do princípio fundamental da WS, um dos componentes fundamentais são os vocabulários controlados, e dentre as diversas camadas da Arquitetura da Web Semântica, é a camada de Ontologias a que suporta a evolução de vocabulários que podem definir relações entre diferentes conceitos. (CATARINO; BAPTISTA, 2008, p.35)

Diversas linguagens para a construção e compartilhamento de ontologias foram

desenvolvidas. Mais recentemente as linguagens passaram a ser concebidas para explorar as características da Web. Tais linguagens são denominadas *Web-Based Ontology Languages*.

Retomando o conceito de vocabulários, de acordo com a W3C Brasil (2012), estes são importantes ferramentas e valiosos instrumentos para organizar os dados de um domínio. Usando OWL (para construir vocabulários, ou 'ontologias') e SKOS (para projetar sistemas de gestão do conhecimento), é possível enriquecer os dados adicionando significados, permitindo que mais pessoas (e mais máquinas) possam reutilizá-los e fazerem mais com os dados.

W3C Brasil destaca, portanto, no conceito de vocabulários para Web Semântica, a linguagem OWL.

OWL é uma linguagem para a construção de Ontologia que foi criada pelo grupo de trabalho *Web-Ontology* da W3C que teve o objetivo de construir uma nova linguagem de marcação de ontologias para a Web Semântica, ela é baseada nas linguagens OIL e DALM-OIL e é uma recomendação da W3C. Três "sublinguagens" são derivadas: *OWL Lite*, *OWL DL* e *OWL Full* (MCGUINNESS; HARMELEN, 2004).

O *Simple Knowledge Organisation 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. Trata-se de um modelo proposto para expressar esquemas de conceitos. 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, terminologias, glossários. Miles e Brickley (2005) definem esquemas de conceitos como sendo um conjunto de conceitos, que, opcionalmente podem incluir declarações sobre os relacionamentos semânticos entre eles.

Trata-se de uma aplicação em RDF resultante de uma pesquisa do *Semantic Web Advanced Development for Europe Project* que está associado à *Semantic Web Activity* do W3C (W3C SWA).

O modelo é composto por termos num vocabulário denominado *SKOS Core Vocabulary* que é um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF.

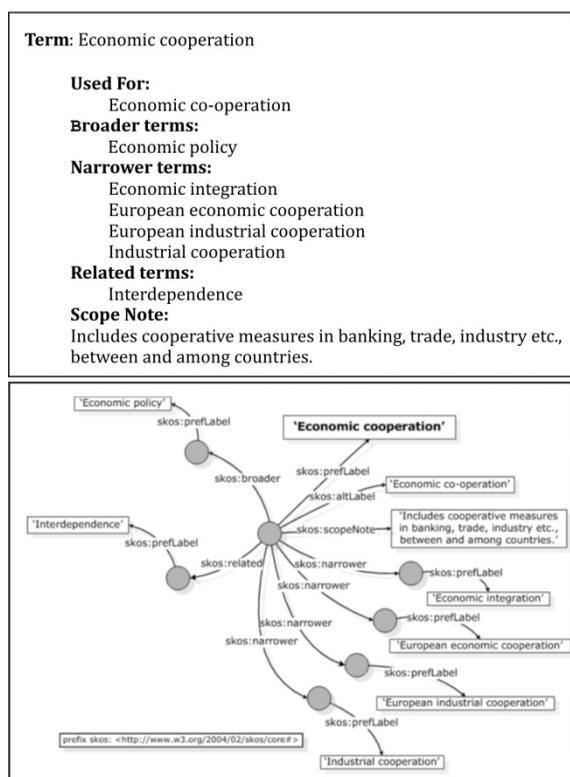
O RDF é o modelo padrão para intercâmbio de dados na Web recomendado pelo

W3C e, portanto, sendo o SKOS escrito em RDF, torna os esquemas de conceito expressos neste modelo, passíveis de serem lidos por agentes inteligentes conforme as recomendações da W3C.

Na figura 1, Miles e Brickley (2005a) exemplificam com um extrato do *UK Archival Thesaurus* (UKAT), a representação textual do termo e seus relacionamentos em gráfico RDF (RDF graph) usando o *SKOS Core Vocabulary*.

No *RDF graph* são representados os triplos RDF (*Subject/Predicate/Object*), ou seja, Recurso/Propriedade/Valor (CATARINO; SOUZA, 2012). Neste exemplo o SUBJECT/RECURSO é o termo adotado: *EconomicCooperation*. Suas propriedades (PREDICATE), representadas por termos do *SKOS Core Vocabulary: Term* (*skos:prefLabel*), *Used For* (*skos:altLabel*), *BroaderTerms* (*skos:broader*), *NarrowerTerms* (*skos:narrower*), *RelatedTerms* (*skos:related*) e *Scope Note* (*skos:scopeNote*). Os Valores (OBJECT) são os termos relacionados, por exemplo o termo *EconomicCo-operation*; este termo é o valor corresponde à propriedade *skos:altLabel* (*Used For*).

Figura 1: Extrato UKAT: Termo *EconomicCooperation* e o correspondente RDF graph



Fonte: Miles e Brickley (2005a)

Todavia, não é objetivo deste estudo descrever detalhadamente o SKOS e sim, apresentar o modelo, citando alguns exemplos de aplicação. Para compreensão do modelo como um todo é preciso entender os conceitos de propriedade, classes, relacionamentos entre os termos, e todas as possibilidades de uso dos termos do *SKOS Core Vocabulary* (MILES; BRICKLEY, 2005b).

O exemplo apresentado atrás representa uma aplicação do SKOS para um tesouro, no entanto, cabe ressaltar que SKOS pode ser usado para outros tantos esquemas de conceitos.

Citaremos aqui outros dois projetos SKOS: a CDD e a CDU nos moldes de dados lincados (*linked data*) da Web Semântica.

O primeiro projeto é uma ação do OCLC (2009). A equipe Dewey da OCLC trabalha no sentido de aplicar os princípios de Dados

Lincados na Classificação Decimal de Dewey, criando um “Serviço de Terminologia” legível por humanos e/ou por máquinas. Este serviço disponibiliza um URI para cada conceito do Dewey, conforme exemplos a seguir.

Na figura 2 pode-se ver o URI para a classe 640 da CDD (ver <http://dewey.info/class/64/2009-08/about.pt>). Esta página segue os princípios dos Dados Lincados na medida em que utiliza o modelo SKOS, que por sua vez é em RDF, este último, a base para a Web Semântica. Em uma parte do código fonte da página representada na figura 3 pode-se ver os termos do *SKOS Core Vocabulary* para a descrição dos conceitos. Neste código pode-se ver a representação da classe 640 (Gerenciamento da Casa e Família), relacionando-a com a classe mais geral, 600 (Tecnologia).

Figura 2: Classe 640 Dewey no *Dewey.info* e seu código fonte (URI classe 640 Dewey)

The image shows a screenshot of the Dewey Decimal Classification website. At the top, it says "Dewey Decimal Classification". Below that, it shows a hierarchy: 600 Tecnologia, and 640 Gerenciamento da casa e família. Under 640, there is a list of sub-classes: 640 Gerenciamento da casa e família, 641 Alimentos e bebidas, 642 Refeições e serviços de mesa, 643 Casa e equipamentos da casa, 644 Serviços da casa, 645 Mobiliários da casa, 646 Costura, vestuário e vida pessoal, 647 Gerenciamento de casas públicas, 648 Cuidados da casa, and 649 Criação de filhos e cuidados domiciliares de pessoas.

```

<div about="http://dewey.info/class/64/2009/08/about.pt" typeof="skos:Concept">
<span class="edition">http://dewey.info/scheme/e22/</span>
<div rel="skos:broader">
<span about="http://dewey.info/class/6/2009/08/about.pt" typeof="skos:Concept">
<a resource="http://dewey.info/class/6/2009/08/about.pt" property="skos:notation"
datatype="ddc:Notation" href="http://dewey.info/class/6/about">600</a>
<a resource="http://dewey.info/class/6/2009/08/about.pt" property="skos:prefLabel" xml:lang="pt"
href="http://dewey.info/class/6/2009-08/about.pt">Tecnologia</a>
</span>
</div>
<span class="notation" property="skos:notation" datatype="ddc:Notation">640</span>
<a class="at" resource="http://dewey.info/class/64/2009/08/about.pt" property="skos:prefLabel"
xml:lang="pt" href="http://dewey.info/class/64/2009-08/about.pt">Gerenciamento da casa e família</a>
</span>

```

Fonte: (OCLC, 2009)

O outro projeto que exemplifica o uso de SKOS para um sistema de classificação decimal é o projeto da *UDC Consortium*, denominado

UDC SummaryLinked Data. *UDC Summary* contém 2.400 classes principais (*top level*) da Classificação Decimal Universal (CDU).

Figura 3: *UDC SummaryLinked Data*: classe 02 Biblioteconomia.

UDC Summary Linked Data	
URI	http://udcdata.info/014810
Notation	02
Caption	Biblioteconomia
See also	002 Documentação. Livros. Escritos. Autoria 01 Bibliografia e bibliografias. Catálogos
Broader class	0 CIÊNCIA E CONHECIMENTO. ORGANIZAÇÃO. INFORMÁTICA. INFORMAÇÃO. DOCUMENTAÇÃO. BIBLIOTECONOMIA. INSTITUIÇÕES. PUBLICAÇÕES
Narrower classes	021 Função, valor, utilidade, criação, desenvolvimento de bibliotecas 022 Localização da biblioteca, edifícios. Equipamento 023 Administração da biblioteca. Pessoal 024 Relações da biblioteca com o público. Regulamentos sobre o uso da biblioteca 025 Departamentos administrativos das bibliotecas 026 Bibliotecas especializadas 027 Bibliotecas gerais

Fonte: *UDC Consortium*, (2011)

No site da *UDC Linked Data* é possível visualizar as Tabelas Auxiliares e a Principal. Cada uma das classes é representada num URI específico, conforme exemplo na figura a seguir, da classe 02 biblioteconomia, <http://udcdata.info/014810>.

No mesmo site ainda é disponibilizado um arquivo para *download* contendo a *UDC Linked Data* em SKOS. A seguir, uma parte deste arquivo contendo apenas a notação 02 Biblioteconomia.

Figura 4: *UDC Linked Data*, SKOS, 02 biblioteconomia.

```
<skos:Concept rdf:about="http://udcdata.info/014810">
<skos:inScheme rdf:resource="http://udcdata.info/ude-schema"/>
<skos:broader rdf:resource="http://udcdata.info/013358"/>
<skos:notation rdf:datatype="http://udcdata.info/UDCnotation">02</skos:notation>
.....
<skos:prefLabel xml:lang="pt">Biblioteconomia</skos:prefLabel>
.....
<skos:related rdf:resource="http://udcdata.info/000002"/>
<skos:related rdf:resource="http://udcdata.info/000001"/>
</skos:Concept>
```

Fonte: *UDC Consortium*, 2011

5 CONSIDERAÇÕES FINAIS

A representação da informação é subdividida em duas partes: a representação descritiva e a temática. A representação descritiva está focada na descrição bibliográfica, que permite a individualização de um documento a partir de suas características específicas. Nesta temática foi publicado um artigo que discorreu a respeito da Representação Descritiva no contexto da Web Semântica (CATARINO; SOUZA, 2012).

Já a Representação Temática, foco deste estudo, trata da representação dos assuntos, visando facilitar a recuperação de materiais relevantes, aproximando documentos com temas semelhantes. (MAIMONE; SILVEIRA; TÁLAMO, 2011, p.28).

Portanto, na área de Ciência da Informação, tem-se o conceito de Representação Temática para se referir à representação dos conceitos de recursos de informação, sendo os vocabulários, produtos resultantes do processo da Representação Temática. Já no contexto da Web Semântica, utiliza-se o conceito de vocabulários de uma forma mais geral: (vocabulários) são utilizados para definir termos (e seus relacionamentos) para descrever e representar uma área do conhecimento, ou para serem adotados numa aplicação específica (W3C, 2012).

Neste sentido, vimos, pela pesquisa realizada, que a proposta do W3C para a Web Semântica, tem como um dos pilares os vocabulários, que podem ter diferentes níveis de complexidade. Pressupõem-se, portanto, que a adoção das práticas da Organização da Informação, mais diretamente as da Representação Temática, oriundas da Ciência da Informação, são totalmente compatíveis e aplicáveis no âmbito das recomendações do consórcio WWW.

Destacamos a linguagem OWL como recomendação para o desenvolvimento de ontologias e o SKOS como uma proposta de modelo para a criação de sistemas de organização do conhecimento. Ambos, linguagem e modelo, formam uma base importante para a aplicação da proposta de Dados Lincados no âmbito da Representação Temática, pois possibilitam que os profissionais da informação, ao aplicarem suas técnicas de representação da informação o façam de forma a tornar os dados legíveis por humanos e por máquinas.

Como forma de ilustrar a pesquisa realizada, foram apresentados dois projetos que utilizam as propostas da Web Semântica para dois conhecidíssimos sistemas de organização do conhecimento: CDD e CDU. Nestes exemplos apresentados, pode-se visualizar o uso do SKOS, que segue o modelo RDF, a base da Web Semântica e dos Dados Lincados.

Outros exemplos tão importantes quanto os citados atrás podem ainda ser explorados. Um caso é o Tesaurus da UNESCO, disponível na Internet, atualmente em quatro idiomas (inglês, russo, espanhol e francês), cuja versão atual é resultado dos trabalhos do Grupo de Investigação de Tecnologias da Informação da Universidade de Murcia e da Cátedra em Gestão da Informação nas organizações da UNESCO. Este tesaurus também desenvolvido em SKOS.

Portanto, finalizando as considerações, há que se destacar que este tema merece muitos esforços de pesquisa pelos pesquisadores da Ciência da Informação, no sentido de trazer para a formação dos profissionais arquivistas, bibliotecários, museólogos e gestores da informação, os conhecimentos necessários para atuar neste campo.

Este estudo teve o intuito de demonstrar como as práticas da Representação Temática, dentro da linha de Organização de Informação na Ciência da Informação, estão relacionadas com as propostas do W3C para a Web Semântica. Averiguou-se que já existem ações neste sentido e que, portanto, existe explicitamente este relacionamento.

Espera-se que este artigo, em conjunto com o citado anteriormente, que abordou a relação da Descritiva com a Web Semântica (CATARINO; SOUZA, 2012) possa ser uma base para outros trabalhos que construam um caminho de atuação dos profissionais da Ciência da Informação no contexto da Organização da Informação na Web.

Artigo recebido em 24/06/2015 e aceito para publicação em 28/09/2015

THEMATIC REPRESENTATION IN THE SEMANTIC WEB CONTEXT

ABSTRACT *The paper discusses, via bibliographic and documental research, how practices of the thematic representation are related to the activities of the Word Wide Web Consortium (W3C) for Semantic Web. First, presents the concepts and practices of thematic representation within the Information Science. Then discusses the concepts of the Semantic Web, highlighting one of its components, the vocabularies. Finally, checks how the thematic representation has been used in Semantic Web applications. The research was developed in Portuguese and English languages, in papers and books in the information science as well as in documents available in the site of the W3C. With the research it was found that the practices of the Organization of Information, specifically the Thematic Representation are compatible with and applied within the framework of the recommendations of the W3C, as can be verified in the studied projects that are in development, for example, the CDD and CDU.*

Keywords: *Semantic Web. Thematic Representation. Ontology. OWL. SKOS.*

REFERÊNCIAS

ALMEIDA, M. B; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019.pdf>>. Acesso em: 18 jul. 2012.

AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma: NISO Press, 2005. 184p. Disponível em: <<http://www.niso.org/standards/resources/Z39-19-2005.pdf>>. Acesso em: 15 fev. 2011.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: Métodos para análise de documentos - Determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, New York, 17 May 2001.

CAMPOS, M. L. A.; GOMES, H.E. Taxonomia e classificação: o princípio de categorização. **DataGramZero: Revista de Ciência da Informação**, v. 9, n. 4, p. ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_01.htm>. Acesso em: 5 jan. 2012.

CATARINO, M. E.; BAPTISTA, A. A. Web Semântica e a qualidade no intercâmbio da informação. In: TOMAÉL, M.I. (Org.). **Fontes de Informação na Internet**. Londrina: EDUEL, 2008. p.31-52.

CATARINO, M. E. **Integração das folksonomias aos metadados**: identificação de novos elementos de metadados como contributo para a descrição de recursos em repositórios. 2009. Tese (Doutorado em Tecnologias e Sistemas de Informação) - Departamento de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, Guimarães, Portugal, 2009.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, v.24, n. 2, p.77-90, abr./ago. 2012.

CERVANTES, B. M. N. **A construção de tesouros e a integração de procedimentos terminográficos**. 2009. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

CINTRA, A. M. M. *et al.* **Para entender as linguagens documentárias**. 2.ed. São Paulo: Polis, 2002. (Coleção Palavra-Chave, 4).

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Briquet Lemos Livros, 2008.

FUJITA, M. S. L. **A leitura documentária do indexador**: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional. 2003. Tese (Livre Docência em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

FUJITA, M. S. L. A Representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Revista Brasileira de Educação Especial**, Marília, v. 10, n. 1, p. 257-272, set./dez. 2004. Disponível em: <http://www.abpee.net/homepageabpee04_06/artigos_em_pdf/revista10numero3pdf/1fujita.pdf>. Acesso em: 2 jun. 2012.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. 2, p. 199-220, 1993.

GRUBER, T. R. *et al.* Ontology Summit 2007 - Ontology framework draft statement: understanding the distinctions. **Applied Ontology**, v. 3, n. 3, p. 191-200, Aug. 2008.

GUIMARÃES, J. A. C. Análise documentária em jurisprudência: uma metodologia de indexação de acórdãos para instrumentos terminológicos. In: SIMPOSIO IBEROAMERICANO DE

- TERMINOLOGÍA - RITerm., 4., 1994, Buenos Aires. **Terminología y desarrollo**: actas. Buenos Aires, 1994. Disponível em: <http://www.ufrgs.br/riterm/esp/txt_simposios_anteriores_1994.html>. Acesso em: 20 ago. 2011.
- GUINCHAT, C.; MENOU, M. As linguagens documentais. In: _____. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. corr. aum. por Marie France Blanquet. Tradução de de Míriam Vieira da Cunha. Brasília: IBICT, 1994. p. 131-165.
- ISAAC, A.; SUMMERS, Ed. (Ed.). **SKOS Simple Knowledge Organization System Primer**: W3C Working Group Note, 18 August 2009. Disponível em: <<http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>>. Acesso em: 19 nov. 2012.
- KING, R. Q&A with Tim Berners-Lee. **BloombergBusinessweek: Technology**, 9 April 2007. Disponível em: <<http://www.businessweek.com/stories/2007-04-09/q-and-a-with-tim-berners-leebusinessweek-business-news-stock-market-and-financial-advice>>. Acesso em: 19 nov. 2012.
- LARA, M. L. G. O unicórnio (o rinoceronte, o ornitorrinco...), a análise documentária e a linguagem. **DataGramZero: Revista de Ciência da Informação**, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <http://www.datagramazero.org.br/dez01/F_I_art.htm>. Acesso em: 15 mar. 2012.
- LARA, M. L. G. de. Linguagem documentária e Terminologia. **Transinformação**, Campinas, v. 16, n. 3, p. 231-240, set./dez. 2004. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/>>. Acesso em: 20 ago. 2011.
- LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/172/151>>. Acesso em: 10 out. 2011.
- MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. F. G. M. Reflexões acerca das relações entre a Representação Temática e Descritiva. **Informação & Sociedade**: estudos, João Pessoa, v.21, n.1, p. 27-35, jan./abr. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/7367/5596>>. Acesso em: 15 nov. 2012.
- MCGUINNESS, D. L.; HAMELEN, F. (Ed.). **OWL Web Ontology Language Overview**: W3C Recommendation, 10 February 2004. Disponível em: <<http://www.w3.org/TR/owl-features>>. Acesso em: 15 fev. 2008.
- MILES, A.; BRICKLEY, D. **SKOS Core Guide**: W3C Working Draft, 2 November 2005a. Disponível em: <<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>>. Acesso em: 20 nov. 2012.
- MILES, A.; BRICKLEY, D. **SKOS Core Vocabulary**: W3C Working Draft, 2 November 2005b. Disponível em: <<http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/pt>>. Acesso em: 20 nov. 2012.
- MOREIRA, M. A. **Mapas conceituais e a aprendizagem significativa**. Porto Alegre: UFRGS, 2012. Disponível em: <<http://www.if.ufrgs.br/~moreira/mapasport.pdf>>. Acesso em: 23 dez. 2012.
- MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2011. p. 1-128.
- OCLC. **Dewey summaries as Linked Data**. Dublin, Ohio: OCLC, 2009. Disponível em: <<http://www.oclc.org/dewey/webservices/default.htm>>. Acesso em: 20 nov. 2012.
- RDF WORKING GROUP. **Resource Description Framework**. W3C, 2004. Disponível em: <<http://www.w3.org/RDF/>>. Acesso em: 10 out. 2011.
- UDC Consortium. **UDC Summary Linked Data**. 2011. Disponível em: <<http://udcdata.info/>>. Acesso em: 20 nov. 2012.

W3C. **Semantic Web**. W3C, 2010. Disponível em: <<http://www.w3.org/standards/semanticweb/>>. Acesso em: 1 set. 2010.

W3C BRASIL. **Missão do W3C**. 2012. Disponível em: <<http://www.w3c.br/Sobre/MissaoW3C>>. Acesso em: 10 nov. 2012.

W3C. **Vocabularies**. 2012. Disponível em: <<http://www.w3.org/standards/semanticweb/ontology>>. Acesso em: 10 nov. 2012.

Auxílio recebido:

A pesquisa coordenada por M.E.CATARINO contou com o apoio financeiro do MCT/CNPq, Edital Universal 2010, processo 474215/2010-5.