

Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento

Gercina Ângela de Lima

Pós-doutorado pela Universidad Carlos III (UC3M) - Madrid, Espanha. Pós Doutorado pela Universidade de São Paulo (USP) - São Paulo, SP - Brasil. Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG). Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/3183050056105009>

E-mail: limagercina@gmail.com

Benildes Coura Moreira dos Santos Maculan

Doutora em Ciências da Informação pela Universidade Federal de Minas Gerais (UFMG) - Brasil. Professora da Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/5336218259257800>

E-mail: benildes@gmail.com

Submetido em: 10/07/2017. Aprovado em: 05/09/2017. Publicado em: 28/12/2017.

RESUMO

A elaboração de Sistemas de Organização do Conhecimento (SOCs) pressupõe o estabelecimento de relações semânticas entre seus conceitos, visando à criação de uma estrutura conceitual coerente. Diferentes SOC's possuem distintos níveis de representação relacional, que são dependentes do tipo de estrutura que cada instrumento requer para representar a base do conhecimento. Nota-se, entretanto, que existe um problema no entendimento sobre quais são os objetivos e propósitos de diferentes tipos de SOC's, principalmente sobre sua elaboração e uso. Este artigo pretende apresentar um estudo comparativo sobre a construção de SOC's do tipo taxonomia, sistema de classificação, tesouros e ontologias, a partir da apresentação de uma análise prática, com o intuito de demonstrar as características de suas estruturas conceituais, usos, potencialidades e limitações, verificando e demonstrando como as relações semânticas vão sendo incorporadas aos instrumentos.

Palavras-chave: Sistemas de Organização do Conhecimento. Relações semânticas. Estrutura conceitual.

Comparative study of semantic structures in different systems of knowledge organization

ABSTRACT

The elaboration of Knowledge Organization Systems (KOS) is based on the establishment of semantic relations between their concepts, aiming the creation of a coherent conceptual structure. Different KOS have different levels of relational representation, which are dependent on the type of structure that each tool requires to represent the knowledge base. It is observed, however, that there is a problem in the comprehension of what are the objectives and purposes of different types of KOS, especially their elaboration and use. This paper intends to present a comparative study on the construction of KOS about the type of taxonomy, classification system, thesauri and ontologies, from the presentation of a practical analysis. It has the purpose to demonstrate the characteristics of KOS conceptual structures, uses, potentialities and limitations, verifying and demonstrating how the semantic relationships are being incorporated into the tools.

Keywords: Knowledge Organization Systems. Semantic relationships. Conceptual structure.

Estudio comparativo de estructuras semánticas en diferentes sistemas de organización del conocimiento

RESUMEN

La elaboración de Sistemas de Organización del Conocimiento (SOCs) presupone el establecimiento de relaciones semánticas entre sus conceptos, buscando la creación de una estructura conceptual coherente. Diferentes SOCs poseen distintos niveles de representación relacional, que son dependientes del tipo de estructura que cada instrumento requiere para representar la base del conocimiento. Se observa, sin embargo, que existe un problema en el entendimiento sobre cuáles son los objetivos y propósitos de diferentes tipos de SOCs, principalmente sobre su elaboración y uso. Este artículo pretende presentar un estudio comparativo sobre la construcción de SOCs del tipo taxonomía, sistema de clasificación, tesauros y ontologías, a partir de la presentación de un análisis práctico, con el propósito de demostrar las características de sus estructuras conceptuales, usos, potencialidades y limitaciones, verificando y demostrando cómo las relaciones semánticas se van incorporando a los instrumentos.

Palabras clave: *Sistemas de Organización del Conocimiento. Relaciones semánticas. Estructura conceptual.*

INTRODUÇÃO

Sistemas de Organização do Conhecimento (SOCs), tais como taxonomias, sistemas de classificação, tesauros e ontologias, são instrumentos de representação do conhecimento; têm sido estudados na ciência da informação (CI) para organizar e recuperar informações, assim como para outros propósitos. Assim, na área de CI, entender como esses instrumentos são conceitualmente construídos auxilia a compreensão sobre o tema da organização do conhecimento.

Alguns estudiosos (SOERGEL, 1997; VICKERY, 1997; GILCHRIST, 2003; KLESS; SIMON; KAZMIERCZAK, 2012; KLESS et al., 2015; ALMEIDA, 2013) têm realizado estudos comparativos entre ontologias e outros tipos de SOCs, porém a maioria deles acontece em relação ao tesauro. Pressupõe-se que isso ocorra porque o tesauro, assim como a ontologia, pode ser utilizado como base de conhecimento de um domínio. Além disso, o tesauro parece possuir também semelhanças com as ontologias na forma em que as relações semânticas são estruturadas, em especial, no que diz respeito às relações *é-um* e *parte-de* nas ontologias e algumas relações hierárquicas (BT/NT) nos tesauros. Entretanto, há pouca literatura que compara ontologias e os outros tipos de instrumentos.

No âmbito da formação de alunos do curso de biblioteconomia, há disciplinas que trabalham com os diferentes tipos de SOCs. Contudo, na formação em pós-graduação para a área de ciência da informação, é usual ter-se alunos de outras diferentes áreas do conhecimento (computação, engenharia, sistemas, entre outros), que contam apenas com os fundamentos conceituais e com as definições existentes encontradas na literatura e em normas de construção dos instrumentos, para os quais não é simples o entendimento das características conceituais dos diferentes tipos de SOCs. Assim, há a necessidade de se fazer conhecer as características das estruturas conceituais, potencialidades e limitações desses instrumentos pelos profissionais da área.

Para suprir essa lacuna, foi desenhada a disciplina *Representação da Informação*, que vem sendo aplicada a diferentes turmas desde 2013. Ela tem como objetivo estudar a modelagem conceitual em diferentes SOCs, analisando a representação das estruturas conceituais, sobretudo quanto às relações semânticas estabelecidas, visando identificar elementos comparativos entre eles. Para viabilizar esse empreendimento, a proposta da disciplina é um trabalho prático de desenvolvimento de quatro diferentes instrumentos (taxonomia, sistema de classificação, tesauros e ontologias) de representação do conhecimento.

Partiu-se do pressuposto de que, por meio da prática, em conjunto com a fundamentação teórica, e à medida que os alunos fossem realizando as etapas da modelagem de cada instrumento, suas características ficariam mais evidentes.

Este artigo apresenta uma compilação de análises observadas a partir de uma amostra selecionada dentre os trabalhos apresentados pelos alunos, com exemplos específicos que permitiram uma comparação de construção das estruturas desses instrumentos. Demonstram-se, também, as características das estruturas conceituais e as relações semânticas que ocorrem em cada um dos quatro SOCs, assim como os procedimentos metodológicos empregados na construção deles.

Não é objetivo deste artigo aprofundar na descrição dos instrumentos, mas demonstrar o nível de representação relacional que acontece no momento da sua construção, apresentando uma análise comparativa quanto às suas semelhanças e diferenças, com apontamento para possibilidades de usos e limitações.

SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Para Vickery (2012, p. 203), os SOCs são “instrumentos complementares que ajudam o usuário a encontrar seu caminho no texto”, e, como ferramentas auxiliares, têm a finalidade de facilitar uma pesquisa genérica ou mais específica, assim como a escolha dos termos de pesquisa. Os princípios teóricos para a elaboração de SOCs são advindos das teorias da classificação, da teoria do conceito e da terminologia, que orientam os agrupamentos de conceitos pelas suas características, de acordo com seus graus de semelhanças e diferenças. As bases da classificação remontam a Aristóteles, em que um todo pode ser dividido em partes, sucessivamente, formando classes e subclasses. Essa lógica da divisão para representar o conhecimento parte do princípio classificatório de gênero e espécie para formar uma estrutura conceitual hierárquica para representar o conhecimento.

O núcleo principal desse processo é a escolha dos conceitos e seus relacionamentos e sob quais aspectos se quer representar determinado domínio. Pode-se dizer, então, que os conceitos são a base para a construção da estrutura conceitual, e são as relações entre eles que os mantêm reunidos dentro de um determinado domínio e princípio de representação. Para Green (2008), os relacionamentos são o coração da organização do conhecimento:

Tentamos localizar informações relacionadas para a necessidade de um usuário; tentamos localizar documentos contendo palavras que possam ser usadas para transportar a informação procurada; e/ou recuperamos documentos indexados por um descritor de assunto relevante ou classificado em uma categoria de assunto relevante. Usamos relacionamentos para navegar entre assuntos. Ajudamos a localizar documentos que juntos dizem mais do que a soma de suas partes (GREEN, 2008, p. 158, tradução nossa).

Existem vários tipos de relações semânticas (ISO 25964-1, 2011), porém, neste trabalho, vamos considerar os três tipos básicos utilizados na estruturação de um SOC: (1) de equivalência, que ocorrem entre termos sinônimos (total; parcial, etc.); (2) hierárquicas, com níveis de superordenação e subordinação, formando renques¹ e cadeias²; (3) associativas, entre termos que mantêm algum tipo de ligação que não de superordenação e subordinação.

A expressividade semântica dos relacionamentos e a forma de representá-los variam segundo o sistema a ser construído, porém, os quatro sistemas (taxonomia, sistemas de classificação, tesauro e ontologia), descritos na próxima seção neste artigo contemplam a noção de hierarquia entre conceitos.

¹ O renque é constituído de conceitos subordinados a um mesmo conceito, ou seja, conceitos coordenados; são conceitos ‘irmãos’. O renque é, portanto, uma série horizontal de conceitos. Tutorial Elaboração de tesauro documentário. Disponível em: <<http://www.conexaorio.com/bitit/tesauro/>>. Acesso em: 10 ago. 2017.

² Como as cadeias são séries verticais de conceitos, é necessário estabelecer a ordem em que cada conceito deverá estar em relação aos outros conceitos. Tutorial Elaboração de tesauro documentário. Disponível em: <<http://www.conexaorio.com/bitit/tesauro/>>. Acesso em: 10 ago. 2017.

ESTRUTURA E RELAÇÕES NOS SOCS

Nesta seção, serão descritas as características estruturais dos SOCs, na seguinte ordem: (1) taxonomia, (2) sistemas de classificação, (3) tesouro e (4) ontologia.

ESTRUTURA E RELAÇÕES EM TAXONOMIAS

A norma ISO 25964-2 (2013) sugere que a unidade básica de uma taxonomia seja denominada “categoria”, porém, na literatura, ela é também designada como “rótulo” (MARTINEZ et al., 2004, p. 106; MOREIRO GONZÁLEZ, 2011, p. 52), nome que será adotado neste trabalho. A taxonomia é composta de um conjunto de termos arranjados em hierarquias representando o domínio modelado, que organiza conceitos segundo as suas semelhanças e diferenças. Entre suas funções, estão o mapeamento do conhecimento de um domínio e o estabelecimento de um rótulo para as informações disponibilizadas, sendo um mecanismo de acesso em ambiente digital.

Uma taxonomia: (a) possui uma lista estruturada de termos, que representam conceitos; (b) apresenta apenas relações hierárquicas (gênero/espécie e todo-parce); (c) não oferece definições dos termos; (d) permite navegação (*navigation*) e busca (*browsing*) por tópicos (rótulos) e, assim, a organização e a recuperação de informação (*searching*) em bibliotecas digitais, portais institucionais, corporativos e menus de e-commerce, seja na Web seja em intranets; (e) comporta agregação de metadados e dados; (f) proporciona um modelo conceitual e mapa conceitual do domínio modelado; (g) não inclui códigos (notações), que são mais usuais em sistemas de classificação (CAMPOS; GOMES, 2007; MACULAN et al., 2009; MACULAN; LIMA; PENIDO, 2011). Além disso, ela pode ser customizada para refletir a terminologia, a cultura, as metas e os objetivos de uma organização, para a gestão organizacional (ISO 25964-2, 2013).

Lambe (2007) ainda ressalta que a taxonomia hierárquica é mais usual, mas podem ser criados outros tipos de estruturas: a poli-hierárquica ou a facetada. A última segue os mesmos

princípios utilizados na construção de sistemas de classificação facetados e de tesouros, a partir da Teoria da Classificação Facetada desenvolvida por Ranganathan (1967). Contudo, para atender ao escopo deste estudo, será considerado o tipo de taxonomia mais usual, o hierárquico.

Uma taxonomia é um instrumento pré-coordenado, pois estabelece a formação dos vários rótulos (assuntos simples ou compostos) no momento de sua construção. Logo, cada rótulo pode ser representado por um conceito único, por exemplo: “Direitos humanos”; pela combinação de conceitos: “Governo, cidadãos e direitos”; uma frase ou expressão: “Enfrentando câncer”, “Busca por ensaios clínicos?” ou “Cuide-se no sol”, de forma que seja mais intuitivo, comunicativo ou que antecipe a demanda do usuário (ISO 25964-2, 2013, p. 61). A norma ainda aponta que os rótulos podem não ser um assunto dentro de uma categoria: (a) público destinado: “Crianças e jovens”; (b) tempo ou local: “Arquivo”; (c) formato do documento: “comunicado de imprensa”; (d) uma tarefa: “Apresentação de reivindicação”, e que haja o uso concomitante desses tipos de rótulos.

No âmbito da ciência da informação, Vickery (1975, p. 10) definiu a taxonomia como um instrumento utilizado para “organizar todas as entidades de um universo em uma simples hierarquia: uma grande árvore na qual cada entidade pertence apenas a uma única classe, cada espécie a um único gênero, e assim por diante”. Ademais, Martinez et al. (2004) afirmam que, em um sentido amplo, a taxonomia cria ordem e rótulos que auxiliam localizar a informação relevante, e, de forma mais específica, “é o ordenamento e rotulação de metadados, que permite organizar sistematicamente a informação primária” (MARTINEZ et al., 2004, p. 106). Assim, uma taxonomia pode atender a objetivos diversos e em qualquer domínio e contexto de uso.

De modo geral, as relações empregadas na modelagem de uma taxonomia são as relações hierárquicas. Segundo Campos e Gomes (2007), a taxonomia de tipo canônica vai estabelecer relações de gênero-espécie, em unidades sistemáticas. Para

as autoras, nas taxonomias de domínio e nas de processos e tarefas gerenciais, aplica-se o “princípio classificatório policotômico [...] [que] possibilita uma organização que representa classes de conceitos com um princípio de divisão (coisas e seus tipos, processos e seus tipos), não priorizando nenhum dos aspectos” (*idem*, p. 5). Assim, as autoras afirmam que resultará em uma simples hierarquia de especificação, representando diversos aspectos do conhecimento modelado (períodos temporais, fenômenos, objetos, fatos, tarefas, atividades).

ESTRUTURA E RELAÇÕES EM SISTEMAS DE CLASSIFICAÇÃO BIBLIOGRÁFICA

Os sistemas de classificação são construídos a partir de signos estruturados e normalizados, podendo ser gerais, quando buscam cobrir todos os campos do conhecimento, ou especializados, quando representam um assunto específico. Ademais, os sistemas de classificação podem ser analíticos, quando sistematizam fenômenos físicos, viabilizando seu entendimento, ou podem ser documentais (de assuntos ou bibliográficos) e implicam a classificação de documentos, em qualquer tipo de suporte e mídia, visando facilitar, principalmente, a localização do item em um dado acervo.

A estrutura dos sistemas de classificação usualmente é constituída por: (1) um esquema de classificação, agrupando elementos por suas semelhanças; (2) uma notação da classificação, que representa os cabeçalhos; e (3) um índice, que facilita a busca do usuário. Segundo Guinchat e Menou (1994), os sistemas de classificação possuem cabeçalhos que permitem representar os conceitos e objetos de um campo determinado do conhecimento, ordenados de forma sistemática, em função de um ou de vários critérios materiais e intelectuais. Portanto, todo sistema de classificação possui uma divisão por classes, que é dividida numa subclasse e, assim, sucessivamente, até chegar ao mais específico (TRISTÃO; FACHIN; ALARCON, 2004). Essa sequência, que respeita a ordem natural das coisas, é composta de classes subordinadas, classes superordenadas e classes coordenadas. A divisão

por classes se baseia em: (1) princípio de divisão: critérios preestabelecidos, obedecendo a uma característica específica de divisão; (2) classes simples e classes compostas: a classe simples identifica uma única característica, na qual não cabem mais subdivisões, e a classe composta identifica mais de uma característica, na qual podem caber mais subdivisões.

Existem diversificados sistemas de classificação de assuntos (ou classificações bibliográficas), dentre os quais se citam: (1) Classificação Decimal de Dewey, de 1876; (2) Classificação Expansiva, de 1891; (3) Classificação da Biblioteca do Congresso, de 1902; (4) Classificação Decimal Universal, de 1905; (5) Classificação de Assuntos, de 1906; (6) Classificação dos Dois Pontos, de 1933; (7) Classificação Bibliográfica, de 1935; e (8) Classificação Internacional, de 1961.

A Classificação dos Dois Pontos, criada por Ranganathan, é um sistema de classificação por facetas, bastante difundido na atualidade. A Teoria da Classificação Facetada criada por ele e a técnica da análise facetada permitem a análise dos assuntos de um documento em facetas, privilegiando os seus diferentes aspectos, podendo-se estabelecer relações entre esses aspectos. Essa característica imprime maior flexibilidade ao sistema de classificação, possibilitando receber novos conceitos e, assim, adequar-se às constantes mudanças e evoluções do conhecimento. De modo geral, entretanto, as principais relações empregadas na modelagem de um sistema de classificação são as mesmas relações hierárquicas estabelecidas em uma taxonomia, já descritas na seção anterior.

ESTRUTURA E RELAÇÕES EM TESAUSOS

Um tesouro é composto de um conjunto de termos descritores (preferidos e não preferidos) relacionados semântica e genericamente entre si, permitindo diferentes tipos de organização (alfabética, por campos semânticos, entre outros). Os termos descritores devem ser padronizados quanto à sua forma e significado, compondo-se de palavras ou grupos de palavras. É um sistema pós-coordenado,

pois adota termos descritores que designam conceitos simples, que podem ser combinados, no processo de indexação, ou combinados pelo usuário no momento da busca. Ressalta-se, porém, que o tesauro pode ter maior ou menor nível de pré-coordenação na formação de descritores.

O tesauro é criado a partir de um conjunto de regras e princípios de contextualização, tendo a função de representar, de forma abreviada, o conhecimento de um domínio, com níveis diferenciados de controle da terminologia e padronização, para dar maior consistência à indexação. Em geral, irá atuar na representação de recursos informacionais, na análise e síntese de textos, assim como na formulação de estratégias de buscas por informações. Não é comum a disponibilização de tesouros aos usuários para a formulação da busca; contudo, esse tipo de uso facilita a recuperação, uma vez que orienta o usuário na escolha dos termos.

Os tesouros possuem três elementos básicos: (1) um léxico; (2) uma rede paradigmática: indica relações essenciais e estáveis entre descritores, numa rede lógico-semântica; (3) uma rede sintagmática: indica relações contingentes entre descritores, válidas apenas em determinado contexto de uso, através de regras de sintaxe. A elaboração de tesouros deve prever a definição de cada um dos termos descritores preferidos e, também, o elemento “Nota de Escopo”, que fornecem orientação sobre seu emprego, significado de um descritor e a relação com outros termos. Os tesouros podem ser apresentados por uma visualização gráfica, em estrutura alfabética sistemática, que permite observar as relações entre os descritores que compõem o grupo (*cluster*) semântico do conceito em questão.

Na elaboração de um tesauro terminológico (ou conceitual), a sua abrangência conceitual será definida pela natureza do serviço, pelo seu objetivo e propósito, levando em consideração o domínio (garantia literária), as especificidades das necessidades informacionais do usuário (garantia do usuário), assim como do seu comportamento de busca e o contexto (garantia da instituição) no qual o serviço é oferecido.

Os fundamentos de construção possuem suas bases na Teoria da Classificação Facetada, Teoria do Conceito e Terminologia.

As relações semânticas estabelecidas em um tesauro são, basicamente: (a) hierárquicas (gênero-espécie; todo-parte); (b) de equivalência (sinônimos); (c) associativas (ligações semânticas). De forma geral, esses relacionamentos são explicitados a partir de símbolos tais como: (1) USE: indica sinonímia e que há outro termo preferencial; (2) UF (usado para): indica sinonímia e que é o termo preferencial; (3) BT: indica a classe mais geral do termo; (4) NT: indica termo específico; (5) RT: indica termo em relações associativas diferentes.

ESTRUTURA E RELAÇÕES EM ONTOLOGIAS

A ontologia é estudada em diversas áreas do conhecimento e tem sua origem na filosofia; e a definição de ontologia como uma ciência se deve a Aristóteles (1978). Na ciência da informação e na ciência da computação, a ontologia é estudada como instrumento de representação do conhecimento, em geral, como um artefato computacional. Nesse contexto, uma ontologia descreve formalmente termos e relações que podem existir entre eles dentro de determinado domínio, inteligível pelo computador.

Os estudos sobre a ontologia ganharam mais destaque no contexto da Web Semântica, por possibilitar o entendimento da linguagem natural, integração de fontes de informação, interoperabilidade semântica e por servir como base de conhecimento em várias aplicações. Um ponto comum existente em todas as ontologias é a dimensão semântica, que é inerente à sua estrutura, sempre com a inclusão de um vocabulário, com as especificações dos termos, possibilitando a identificação das categorias fundamentais do domínio em questão, como esses termos relacionam entre si, e quais relacionamentos serão possíveis de serem criados (GRUNINGER et al., 2008).

Vickery (1997, p. 284) foi o primeiro autor na área de ciência da informação a estudar sobre ontologia; ele a define como “uma formalização sistemática

de definições de conceitos, relacionamentos e regras que capturam o conteúdo semântico de um domínio em um formato legível por máquina.” Já Smith (2003) define ontologia como:

um dicionário de termos formulados em uma sintaxe canônica e com definições comumente aceitas, concebidas para produzir um lexical ou uma estrutura taxonômica de representação do conhecimento que pode ser compartilhada pela comunidades de diferentes sistemas de informações (SMITH, 2003, p. 6, tradução nossa).

Assim, pode-se afirmar que a ontologia é uma lista de conceitos ou entidades dentro de um domínio específico, os quais podem ser estruturados de forma hierárquica, por meio de relações semânticas explicitadas formalmente, em meio informatizado. A estrutura da ontologia nem sempre é a mesma, e, para Gruber (1992), os componentes básicos são: (1) classes, as quais são organizadas em estrutura conceitual hierárquica; (2) relações, que representam a relação entre os conceitos; (3) axiomas, que representam sentenças verdadeiras; (4) instâncias, que representam os dados. Todos esses componentes devem possuir uma representação explicitamente formal para que a ontologia seja processável por máquina.

Existem várias relações que podem ser representadas em uma ontologia, sendo as mais comuns: a relação classe-classe (a relação “é um”), a relação instância-classe (“é instância de”) e a relação instância-instância (a relação “é parte de”). As ontologias são consideradas linguagens mais semanticamente ricas, em comparação com os outros tipos de SOCs, porque todas as relações entre os termos podem ser explicitadas, inclusive de equivalências e as associativas. Além disso, não existem limites para as variações das relações; assim podem ser feitas deduções e inferências que não são possíveis de serem realizadas nos outros SOCs.

Entre as tecnologias utilizadas na construção de ontologias, estão as linguagens *Resource Description Framework* (RDF), *Resource Description Framework Schema* (RDF-S) e *Web Ontology Language* (OWL). A RDF descreve metadados permitindo criar triplas

para fazer o relacionamento sujeito-predicado-objeto, permitindo representar a semântica contida neles. Já o RDF-S define as *tags* especificando as classes, propriedades e relacionamentos que podem ser utilizados nas triplas, que incluem as *tags* e a sua estrutura hierárquica (taxonomia). Já a OWL foi construída a partir da RDF-S e da DARPA *Agent Markup Language* (DAML) + *Ontology Inference Layer* ou *Ontology Interchange Language* (OIL), sendo considerada a linguagem de ontologias na Web Semântica. A OWL é uma linguagem computacional lógica, que possibilita melhor interoperabilidade entre os conteúdos, satisfazendo os formalismos exigidos para a representação dos recursos de informação. Segundo Isotani e Bittencort (2015), as principais características da OWL são a descrição explícita e formal de conceitos num dado domínio (classes), exibindo suas propriedades, características e atributos, e a possibilidade de expressar o raciocínio lógico da descrição para a Web semântica.

DESCRIÇÃO DOS PROCEDIMENTOS E RESULTADOS

Este estudo se caracteriza como descritivo, que permitiu estabelecer relações de dependência entre variáveis (características dos quatro instrumentos: taxonomia, sistema de classificação, tesouros e ontologias), que serviram de base para as análises sobre as semelhanças e diferenças entre elas. Assim, o estudo tem caráter comparativo, de natureza qualitativa, demonstrando o nível semântico da representação dos relacionamentos entre termos nas estruturas dos quatro instrumentos.

O estudo teve por base as discussões e análises realizadas no âmbito da disciplina Representação da Informação, oferecida no Programa de Pós-Graduação em Ciência da Informação (PPGCI), da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG). Para o trabalho prático de construção dos quatro instrumentos, foram selecionados, aleatoriamente, cinco documentos acadêmicos do tipo teses e dissertações, defendidos no PPGCI/ECI/UFMG,

para a elaboração de instrumentos de representação do conhecimento, a partir do assunto contido nesses documentos. Definiu-se o público-alvo como sendo alunos de graduação e professores.

A metodologia para os processos classificatórios empregados na construção dos instrumentos foi respaldada na Teoria do Conceito e na Teoria da Classificação Facetada. A técnica da análise facetada foi aplicada utilizando-se os princípios dos três planos propostos por Ranganathan: (1) plano das ideias: (a) análise de assunto e identificação das características que constituem aquele assunto; (b) decisão sobre as classes básicas (mais abstratas) para agrupamento dos assuntos em facetas e subfacetas, realizado a partir dos princípios das cinco categorias Personalidade, Matéria, Energia, Espaço, e Tempo (PMEST) de Ranganathan; (c) subdivisão e agrupamento dos assuntos conforme suas características comuns, separando os de características diferentes; (d) arranjo dos elementos dentro dos grupos e subgrupos; (2) plano verbal: utilizado para determinar o(s) termo(s) que designam cada conceito; (3) plano notacional: utilizado somente na construção do sistema de classificação, criando-se notações simbólicas (com números e/ou letras), de acordo com as regras estabelecidas.

A amostra para a comparação apresentada neste artigo foi um dos trabalhos entre os cinco criados no âmbito da disciplina já mencionada. A escolha da amostra foi intencional, pois se buscou selecionar o trabalho que melhor pudesse exemplificar em sua estrutura as características dos quatro instrumentos construídos.

Em síntese, a metodologia incluiu: (1) seleção do documento a ser analisado; (2) captura de conhecimento: análise no documento para identificação dos assuntos representativos do conteúdo; (3) organização dos assuntos com a subdivisão e agrupamento, alinhados com cada tipo de instrumento; (4) identificação das relações semânticas, em conformidade com a necessidade de cada instrumento; (5) apresentação da estrutura dos instrumentos. A seguir, apresentam-se os procedimentos de construção de cada instrumento.

PERCURSO METODOLÓGICO E ELABORAÇÃO DOS INSTRUMENTOS

1. Seleção do trabalho: para avaliação dos quatro SOCs construídos:

SILVA, A.R. *Estudo dos princípios de categorização na biblioteconomia e ciência da informação*: Ranganathan: entre a teoria clássica e a abordagem cognitiva contemporânea. 2010. 195 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/ECID-8ELHMA>>. Acesso em: 10 ago. 2017.

2. Descrição da construção dos Sistemas de Organização do Conhecimento-SOC

2.1. Construção da taxonomia

2.1.1. (2.1.1) Tipo de taxonomia: de domínio, expressando os assuntos do documento.

2.1.2. Captura de conhecimento: levantamento dos assuntos do documento, com a leitura e a marcação dos termos considerados significativos do conteúdo do documento e o contexto de uso. Resultados da taxonomia: foram levantados 99 termos.

2.1.3. Normalização dos rótulos pré-coordenados: a escolha dos rótulos obedeceu aos seguintes princípios: (a) comunicabilidade: linguagem do usuário (garantia do usuário); (b) utilidade: representativo do documento; (c) estimulação: termos intuitivos e fáceis de entender; (d) compatibilidade: termos alinhados à área de conhecimento do documento.

2.1.4. Criação da estrutura classificatória: mapeamento dos assuntos do documento a partir dos princípios das categorias PMEST e, depois, as

classes e subclasses foram organizadas a partir do plano das ideias e do plano verbal. No primeiro, foram criadas cadeias (hierárquicas genéricas: tipo de; hierarquias partitivas: parte de e todo de) e renques. Para a ordem dos conceitos dentro das cadeias, utilizaram-se os cânones: extensão decrescente (crescente em relação à intensão) e modulação (sequência dos elos). Para os renques, utilizaram-se os cânones: exaustividade, exclusividade (elementos mutuamente exclusivos), sequência útil (do geral para o particular) e sequência consistente (criar classes semelhantes de forma homogênea). A ordenação dos elementos das classes e subclasses foi feita em ordem alfabética. A validação pelo especialista foi realizada pelas autoras deste artigo.

2.2. Construção do sistema de classificação

2.2.1. Captura de conhecimento: foi iniciada a partir da taxonomia já elaborada, conforme os princípios de criação de um sistema classificatório.

2.2.2. Determinação das classes básicas (maior abstração): feita a partir da análise da taxonomia criada, resultando em quatro classes básicas: (a) Entidades compreendem os princípios, objetos imateriais, objetos materiais; (b) Atividades compreendem as operações, estados, processos; (c) Propriedades compreendem as quantidades, qualidades, relações; (d) Dimensões compreendem tempo, posição, espaço.

2.2.3. Criação da notação: inclusão apenas de números ordinais, em ordem crescente, separados por ponto, conforme a endentação. Não houve elaboração de índice, pois não fazia parte do escopo da proposta do trabalho.

2.2.4. Sintaxe para formação de notação: uso do PMEST na ordem de citação.

2.3. Construção do tesauro

2.3.1. Captura de conhecimento: também foi iniciada a partir da taxonomia já elaborada conforme os princípios de criação de um sistema classificatório.

2.3.2. Determinação das classes básicas: utilizaram-se as mesmas quatro classes básicas da taxonomia criada: (a) Entidades; (b) Atividades; (c) Propriedades; (d) Dimensões.

2.3.3. Glossário: foram elaboradas definições para os termos descritores.

2.3.4. Estabelecimento de relações: foram identificadas as relações hierárquicas da taxonomia (gênero-espécie, todo-parte) que foram representadas no tesauro pelas relações *Broader Term* (BT) e *Narrow Term* (NT). Depois, pelo teor das definições, foram estabelecidas as relações associativas utilizando-se o símbolo *Related Term* (RT). Em seguida, com nova análise do documento e consulta a dicionários da área, foram estabelecidas as relações de equivalência, utilizando-se os símbolos *Use For* (UF).

2.3.5. Forma de apresentação: alfabética, juntamente com a parte sistemática.

2.4. Construção da Ontologia

2.4.1. Captura e conhecimento: foi iniciada a partir da taxonomia e do tesauro já elaborados, conforme os princípios de criação de um sistema classificatório.

2.4.2. Determinação das classes de conceitos: utilizou-se a linguagem OWL, por meio da ferramenta Protégé³.

³ A Ferramenta Protégé é um dos softwares mais referenciados e utilizados na construção de ontologias. Foi desenvolvida na Stanford Center for Biomedical Informatics Research na Escola de Medicina da Universidade de Stanford, EUA.

2.4.3. Criação das entidades e estabelecimento de relações: ampliação das relações possíveis entre as classes do domínio, além das relações entre termos definidas pelo tesouro, sendo “use” e “use para”; assim como as propriedades “Termo Geral”, “Termo Específico” e “Termo Relacionado”, foram acrescentadas as relações “requires” e “isRequiredBy”, “isPartOf” e “hasPart”, “isBasedOn” e “isBasesFor”, “references” e “isReferencedBy”. Essas relações são definidas no padrão SCORM, que consiste em um modelo de referência para a construção e publicação de objetos de aprendizagem na web, por meio de sistemas de gerenciamento de conteúdo para educação a distância (NEVES, D.E; BRANDÃO, W. C.; ISHITANI, L., 2016).

2.4.4. Esquematização dos relacionamentos: as classes de conceitos e relações definidas nos itens 2.4.2 e 2.4.3 foram formalmente descritas por meio das propriedades e das relações estabelecidas na linguagem OWL.

2.4.5. Forma de apresentação: uma vez formalizada e descrita, a ontologia pôde ser visualizada na ferramenta *Protegé*, a partir do arquivo interoperável no formato OWL/RDF. Assim, foi possível gerar uma visualização gráfica navegável assim como realizar consultas sobre a ontologia por meio da linguagem SPARQL (*Protocol and RDF Query Language*).

3. Escolha da Classe: amostra para a análise e demonstração das relações semânticas: foi selecionada a Classe: Cognição.

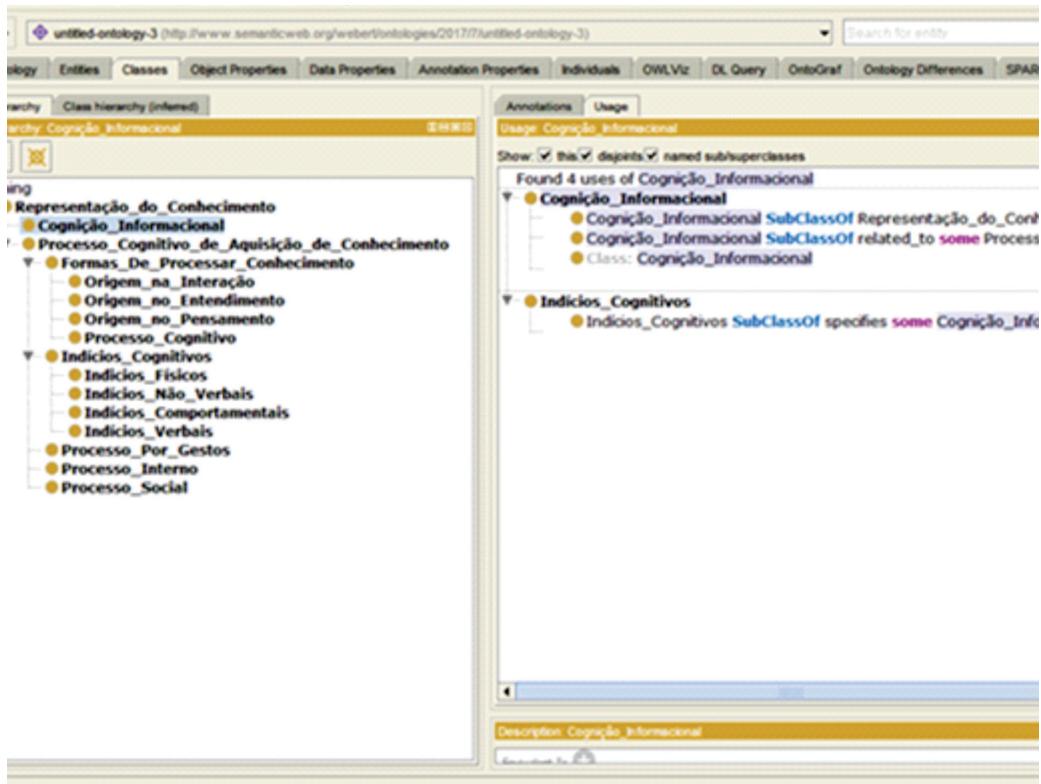
Após elaboração dos SOCs, as classes apresentaram-se nas seguintes estruturas:

Figura 1 – Estruturas da taxonomia, sistema de classificação e tesouro

<p>Taxonomia: Processo cognitivo de aquisição de conhecimento Processo social Processo interno Processo por gestos Indícios cognitivos Indícios verbais Indícios comportamentais Indícios não-verbais Indícios físicos Formas de processar conhecimento Origem no pensamento Origem no entendimento Origem na interação</p>	<p>Sistema de Classificação 1 ATIVIDADES 1.1 Ciências Cognitivas 1.1.1 Teorias cognitivas 1.1.2 Processos Cognitivos 1.1.2.1 Acordo cognitivo; Processo social 1.1.2.2 Enfoque cognitivo; Processo interno; Disposição para o aprendizado 1.1.2.3 Gesto categorial; Linguagem corporal; Expressividade não-verbal; Movimentos de cabeça, olhos, mãos e posição do corpo; Expressões faciais 1.1.2.4 Indícios cognitivos 1.1.2.4.1 Indícios verbais; Aderência de conceitos; Assimilação; Aprendizagem de conceitos 1.1.2.4.2 Indícios comportamentais; Critério para atribuição; Mudança cognitiva 1.1.2.4.3 Estrutura cognitiva; Padrões de ações física e mental; Estrutura do conhecimento; Conflitos; Reestruturação; Estágios do desenvolvimento; Conhecimento novo; Assimilação e acomodação 1.1.2.4.4 Limites entre categorias; Conhecimento prévio 1.1.2.4.5 Indícios físicos; Estereótipos 1.1.2.4.5.1 Função socio-cognitiva 1.1.2.4.5.2 Função socioafetiva 1.1.2.5 Componentes cognitivos 1.1.2.5.1 Origem no pensamento: Memorização 1.1.2.5.2 Origem no entendimento: Percepção de relações; Aprendizagem 1.1.2.5.3 Origem na interação: Abordagem de Vigotsky 1.1.2.5.3.1 Afetiva 1.1.2.5.3.2 Moral; Juízos 1.1.2.5.4 Comportamento reprodutivo</p>	<p>Tesouro COGNIÇÃO INFORMACIONAL NE É um tipo de representação do conhecimento que envolve elementos como a percepção, o acordo cognitivo, o gesto categorial, o pensamento e os indícios cognitivos. A noção de conhecimento é primária e advém das Ciências Cognitivas, de diferentes teorias cognitivas, se estruturando no paradigma cognitivo, sobretudo com Piaget e Vigotsky. TG Representação do conhecimento TE Gesto categorial Indícios cognitivos Pensamento Percepção TR Ciências cognitivas Paradigma cognitivo Piaget Teoria cognitiva Vigotsky</p>
---	--	---

Fonte: Elaborado pelas autoras (2017).

Figura 2 – Estrutura da ontologia



Fonte: Elaborado pelas autoras (2017).

A taxonomia foi elaborada pensando-se mais em um *menu* de busca em um *site* sobre o processo cognitivo de conhecimento. Já tesouro, como a intenção é indexar documentos, a modelagem está mais voltada para a cognição informacional, pois o documento é da área da CI. O sistema de classificação foi modelado de modo mais geral (enciclopédico) com o intuito de classificar assuntos em âmbito mais amplo, para organização de recursos informacionais (acervos). A ontologia foi modelada a partir dos interesses dos usuários e o que eles esperam encontrar dentro da lógica da base do conhecimento deste domínio, no contexto da Web Semântica.

A estrutura de uma taxonomia é bastante similar à estrutura de um sistema de classificação, a não ser pelo fato de que não é da natureza de uma taxonomia possuir notações. Tal como ocorre nos sistemas de classificação, também nas taxonomias acontece a repetição de termos. Mas, ainda que os termos sejam reproduzidos, eles têm uma dimensão

conceitual diferente. Ressalta-se que, com isso, uma característica tanto das taxonomias quanto dos sistemas de classificação é que o rótulo ou cabeçalho, fora da estrutura, não pode identificar o significado de forma unívoca.

Nas taxonomias e nos sistemas de classificação, não há o controle de equivalências, como existe nos tesouros. Contudo, com certeza, os rótulos e cabeçalhos poderiam ser substituídos por algum termo equivalente (sinônimos, quase-sinônimos, variações linguísticas), o que pode ocasionar alguma frustração no usuário, caso o termo escolhido não seja intuitivo para ele. Uma das vantagens do tesouro em relação à taxonomia e aos sistemas de classificação, é a possibilidade de expressar o conjunto completo de relações associativas entre conceitos e não apenas as relações hierárquicas que compõem esses instrumentos.

Ainda que bastante completas, as relações semânticas explicitadas nos tesouros não são capazes de alcançar o nível descritivo e de formalidade

encontrado em ontologias consistentes, uma vez que nestas as relações estabelecidas conduzem à possibilidade de inferência pelo sistema. Nas ontologias, os conceitos concretos, que representam as propriedades de domínio e seus relacionamentos, são distinguidos dos conceitos abstratos, que são representados como características e atributos. No tesouro, na taxonomia e no sistema de classificação, esse tipo de especificação não é requerido e não pode ser determinado. Assim, observa-se que a ontologia possibilita uma representação formal, com base em lógica, o que permite a este artefato fazer inferência sobre o domínio modelado, criando novos conhecimentos.

Os sistemas de organização do conhecimento possuem características específicas para sua aplicação em determinados contextos, e alguns exigem normas e metodologias diferenciadas de construção e atualização de acordo com suas finalidades e aplicações específicas.

CONSIDERAÇÕES FINAIS

Este artigo apresenta resultados do trabalho prático realizado com os alunos da disciplina Representação da Informação. Esse trabalho partiu do pressuposto de que, por meio da prática, em conjunto com a fundamentação teórica, e à medida que os alunos fossem realizando as etapas da modelagem de cada instrumento, suas características ficariam mais evidentes, e seus objetivos e propósitos, mais claros. Conforme relato dos alunos da disciplina, à medida que eles foram realizando as etapas da modelagem, agrupando os termos, a partir de suas relações semânticas, eles puderam distinguir com maior clareza as características, os objetivos e os propósitos de cada um dos instrumentos que estavam sendo construídos.

Ressalta-se que a integração entre os alunos de diferentes áreas do conhecimento, que geralmente matriculam-se nesta disciplina, com formação em biblioteconomia, em ciências da computação, engenharia, sistemas de informação, letras e ciências cognitivas, entre outras, auxilia na elaboração

dos instrumentos, pois se cria uma interação de conhecimentos que facilita a realização de todas as etapas dos procedimentos metodológicos na criação dos instrumentos, desde a modelagem até a implementação da ontologia.

A proposta de utilizar a prática como forma de complementação, ou em conjunto com os fundamentos conceituais encontrados na literatura, e com as recomendações das normas de construção de tesouros e ontologias, a ISO 25964-1 (2011) e a 25964-2 (2013), foram eficientes, pois se considera que o objetivo proposto foi atingido, já que facilitou o entendimento dos alunos quanto às semelhanças e diferenças existentes nestes sistemas de organização do conhecimento, conforme proposta do conteúdo programático da disciplina Representação da Informação.

REFERÊNCIAS

- ALMEIDA, M.B. Revisiting ontologies: a necessary clarification. *Journal of the American Society for Information Science and Technology*, v.64, n.8, p.1682-1693. 2013. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.22861/full>>. Acesso em: 11 ago. 2017.
- ARISTÓTELES. *Metafísica*: livro alfa. São Paulo: Abril Cultural, 1978. (Coleção Os Pensadores).
- CAMPOS, M.L.A.; GOMES, H.E. Taxonomia e classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2007, Salvador, Bahia. *Anais...* Salvador: UFBA. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf>>. Acesso em: 11 ago. 2017.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v. 59, n. 1, p. 7-18, 2003. DOI: 10.1108/00220410310457984.
- GREEN, R. Relationship in Knowledge Organization. *Knowledge Organization*, v.35, n.2/3, p.150-159, 2008. Disponível em: <https://cluster10-files.instructure.com/files/25425605/download?download_fid=1&verifier=T2jUGuHtAb6PBobt8Vajx9w4n3Fa1FBwKM2r0vd9>. Acesso em: 11 ago. 2017.
- GRUBER, T. What is an Ontology?, 1992. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 11 ago. 2017.

- GRUNINGER, M. Ontology, taxonomy, folksonomy: understanding the distinctions. *Applied Ontology*, n. 3 p. 191–200. 2008. Disponível em: <<https://mor.nlm.nih.gov/pubs/pdf/2008-ao-mg.pdf>>. Acesso em: 11 ago. 2017.
- GUINCHAT, C.; MENO, M. *Introdução geral às ciências e técnicas da informação e documentação*. Brasília, DF: IBICT, 1994.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964-1: thesauri and interoperability with other vocabularies*. Part 1: thesauri for information retrieval. Geneve: International Standard Organization, 2011.
- INTERNATIONAL STANDARD ORGANIZATION. *ISO 25964-2: thesauri and interoperability with other vocabularies*. Part 2: interoperability with other vocabularies. Geneve: International Standard Organization, 2013.
- ISOTANI, S.; BITTENCORTT, I.I. *Dados abertos conectados*. São Paulo, SP: Novatec, 2015. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 11 ago. 2017.
- KLESS, D. et al. Thesaurus and Ontology Structure: Formal and Pragmatic Differences and Similarities. *Journal of The Association for Information Science and Technology*, v. 66. n. 7, p. 1348–1366, 2015. Disponível em : <<http://onlinelibrary.wiley.com/doi/10.1002/asi.23268/abstract>>. Acesso em: 11 ago. 2017.
- KLESS, D.; KAZMIERCZAK, E.; MILTON, S. K. Relationships and Relata in Ontologies and Thesauri: Differences and Similarities. *Applied Ontology*, v. 7, n. 4, nov., 2012, p. 401–428. Disponível em: <<http://content.iospress.com/articles/applied-ontology/ao118>>. Acesso em: 11 ago. 2017.
- LAMBE, P. *Organising knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford, England: Chandos, 2007.
- MACULAN, B.C.M.S. et al. Taxonomia, folksonomia, acessibilidade e usabilidade: proposta de interseção na área de organização do conhecimento, com foco na recuperação de informação. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO - SECIN, 3., Londrina. *Anais...* Londrina: UEL, 2009. p. 1-26. Disponível em: <<http://eprints.rclis.org/23854/>>. Acesso em: 11 ago. 2017.
- MACULAN, B.C.M.S.; LIMA, G.A.B.O.; PENIDO, P. Taxonomia facetada como interface para facilitar o acesso à informação em bibliotecas digitais. *Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis*, v. 16, n. 1, p. 234-249, jan./jun. 2011. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/746>>. Acesso em: 11 ago. 2017.
- MARTÍNEZ, A. et al. Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios web argentinos. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 106-111, 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a11v33n2.pdf>>. Acesso em: 11 ago. 2017.
- MOREIRO GONZÁLEZ, J.A. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFA, 2011.
- NEVES, D.E; BRANDÃO, W.C.; ISHITANI, L. Metodologia para recomendação e agregação de Objetos de Aprendizagem no padrão SCORM. RBIE, v.24, n.1, 2016. Disponível em: <<http://www.wladimirbrandao.com/docs/publications/2016@RBIE.pdf>>. Acesso em: 11 ago. 2017.
- RANGANATHAN, S.R. *Phelogema library classification*. Bombay, Ásia: Publishing House, 1967, 540 p.
- SMITH, B. Ontology: Philosophical and computational. In: FLORIDI, L. (Ed.). *The Blackwell Guide to the Philosophy of Computing and Information*. New Jersey: Blackwell Publishing.
- SOERGEL, D. *Functions of a thesaurus: classification: ontological knowledge base*. [S.l.]: College of Library and Information Services, University of Maryland, 1997. Disponível em: <<http://www.umiacs.umd.edu/~oard/teaching/878/spring99/readings/functclass.pdf>>. Acesso em: 11 ago. 2017.
- TRISTÃO, A.M.D.; FACHIN, G.R.B.; ALARCON, O.E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 161-171, ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a17v33n2.pdf>>. Acesso em: 11 ago. 2017.
- VICKERY, B.C. *Classification and indexing in science*. 3. ed. Londres: Butterworths, 1975.
- _____. On knowledge organisation. In: FACETS OF KNOWLEDGE ORGANIZATIONS, 2., 2011. *Proceeding of the ISKO UK Second Biennial Conference*, Londres: Emerald, 2012. 416 p.
- _____. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997. DOI: 10.1177/016555159702300402.
- ZENG, M.L. Knowledge organization systems (KOS). *Knowledge Organization*, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.