

O USO INTEGRADO DO TESAURO E ONTOLOGIA EM MECANISMOS DE BUSCA PARA A ORGANIZAÇÃO DO CONHECIMENTO

THE INTEGRATED USE OF THE THESAURUS AND ONTOLOGY IN SEARCH MECHANISMS FOR THE ORGANIZATION OF KNOWLEDGE

Ana Carolina Rodrigues do Prado^a
Vânia Mara Alves Lima^b

RESUMO

Objetivo: descrever como o tesauro e a ontologia podem atuar em mecanismos de busca para organização do conhecimento por meio do uso integrado de seus elementos estruturais. **Metodologia:** foi desenvolvida uma pesquisa de caráter exploratório baseada em revisão bibliográfica na Ciência da Informação para descrever o tesauro e a ontologia enquanto sistemas de organização do conhecimento para posterior discussão de sua integração na representação do conhecimento. **Resultados:** por meio dos levantamentos realizados foram identificadas ações vinculadas ao uso integrado do tesauro e ontologia na representação do conhecimento em mecanismos de busca para fins de desambiguação, refinamento de relações conceituais e reuso de estruturas. **Conclusões:** a integração de elementos estruturais do tesauro e da ontologia viabiliza a qualificação da organização do conhecimento em mecanismos de busca, uma vez que o tesauro adiciona trabalho terminológico à ontologia com o uso dos termos descritores e a ontologia viabiliza o refinamento das relações conceituais trazendo efetividade à representação do conhecimento.

Descritores: Organização do conhecimento. Representação do conhecimento. Conceito. Sistemas de organização do conhecimento. Ontologia. Tesauro.

1 INTRODUÇÃO

Da mesma forma que podemos reconhecer a crescente necessidade de tecnologias que assegurem a autonomia do usuário na internet, cujo contexto é permeado de um número massivo de dados, há um esforço para a concepção

^a Mestre em Ciência da Informação pela Universidade de São Paulo (USP), São Paulo, Brasil. E-mail: ana.carolina.prado@usp.br

^b Doutora em Ciências da Comunicação pela Universidade de São Paulo (USP). Docente do Departamento de Informação e Cultura da Escola de Comunicações e Artes da Universidade de São Paulo (ECA-USP), São Paulo, Brasil. E-mail: vama@usp.br

de mecanismos de busca que atuem na recuperação de conteúdo específico para cada tipo de público. No campo da Ciência da Informação (CI), esta discussão está atrelada ao estudo dos contextos de produção de conhecimento e dos recursos para torná-lo acessível. E mais especificamente, é a disciplina Organização do Conhecimento (OC), na CI, que se ocupa do processo homônimo, de organização do conhecimento, o qual está embasado na análise de conceitos e objetivando que a representação da informação, isto é, a informação registrada em um sistema de informação, seja recuperada pelo usuário deste sistema. Por isso, ao longo do tempo, foram concebidas as classificações universais e as classificações facetadas, que partem das “unidades de conhecimento”, os conceitos (DAHLBERG, 2011, p. 69) para a recuperação dos itens informacionais (livros, documentos ou conteúdo na internet) de acordo com o interesse do usuário. Assim, a OC implementa a organização do conhecimento por meio de processos denominados representação do conhecimento, representação da informação e recuperação da informação. Portanto, os conceitos, ou seja, o plano abstrato das ideias e a formação de classes é o ponto de partida da organização do conhecimento. A sua aplicação, no entanto, depende dos elementos que surgem desses estudos, determinando como será o tratamento de objetos, assuntos e variedade de documentos (DAHLBERG, 2014).

No âmbito da OC, os denominados Sistemas de Organização do Conhecimento (SOC) são instrumentos responsáveis por realizar tarefas relacionadas à organização do conhecimento, mais especificamente na representação do conhecimento, ao identificar os conceitos, as relações entre esses conceitos e os seus termos descritores. Os termos, identificados por Soergel (2009, p. 19) como as “designações de conceitos”, vão constituir então um SOC. Os SOC variam tanto em sua tipologia, quanto na forma como realizam a representação do conhecimento, o que consiste na exposição dos termos preferidos para os conceitos, assim como das relações entre eles (BRÄSCHER, 2014). Entre estes sistemas, temos o tesauro e a ontologia.

O tesauro é definido pela ISO 25964-1 (2011) como um vocabulário controlado, onde os conceitos são representados por termos que demonstram

as relações entre os conceitos de um domínio. Trata-se de um SOC estruturado hierarquicamente, que permite a visualização das relações gênero-espécie ou parte-todo, além de apresentar as relações associativas e as indicações de sinônimos em uma determinada área. Sendo um caso especial de esquema de classificação facetada (TRISTÃO; FACHIN; ALARCON, 2004, p. 167), a organização lógica e hierárquica do tesauro parte da noção de faceta de Ranganathan – privilegiando determinados pontos de vista (CINTRA *et al.*, 2002, p. 41).

Já a ontologia é um SOC considerado de grande complexidade na representação do conhecimento, pois traduz para a linguagem computacional as relações associativas entre conceitos e realiza inferências ou julgamentos a partir de regras e enunciados (ZENG, 2008; VICKERY, 2008; PIETERSE; KOURIE, 2014). A ontologia viabiliza a organização do conhecimento a partir de seus elementos estruturais, apontados por Gruber (1993) como: classes (organização dos conceitos de um domínio); relações; axiomas (regras que restringem a interpretação dos termos) e instâncias (nomes para os dados).

Nos mecanismos de buscas da internet, que precisam operar com base na organização dos conceitos imersos nos domínios de conhecimento, o uso integrado do tesauro e da ontologia pode ser desenvolvido a partir do uso de seus elementos estruturais distintos, visando assim a adequação da organização do conhecimento a uma determinada demanda. Ambos têm a sua estrutura básica fundamentada nos conceitos de um domínio de conhecimento, realizando a representação do conhecimento de forma diversa: enquanto o tesauro apresenta verbetes e etiquetas para representar os conceitos por meio de termos descritores, e age prescrevendo a maneira como o usuário deve empregar esses termos, a ontologia auxilia no diálogo entre agentes inteligentes e mecanismos a partir do uso de axiomas ou regras (que geram inferências) e as classes acabam formando uma taxonomia. A sua atenção, portanto, está voltada para a explicitação dos enunciados a respeito dos conceitos, sendo as ligações conceituais dispostas formalmente com base em linguagem computacional, preocupada com o fluxo destes dados.

Considerando então que é necessário utilizar instrumentos para acessar

ao conhecimento, que muitas vezes está disperso devido a fluidez de informações na internet, este trabalho tem como objetivo geral descrever como os elementos estruturais do tesauro e ontologia podem interagir na organização do conhecimento, mais especificamente na representação do conhecimento em mecanismos de busca da internet, por meio de seu elemento essencial em comum, que é o conceito, sendo este designado (representado) pelo termo. Assim, cabe identificar quais são as propostas de uso integrado do tesauro e ontologia para a organização do conhecimento, procurando analisar como estes SOC podem compartilhar elementos como maneira qualificar a organização do conhecimento em mecanismos de busca na internet.

2 PROCEDIMENTO METODOLÓGICOS

Este estudo tem como alicerce uma revisão de literatura no âmbito da Ciência da Informação, abrangendo os procedimentos metodológicos da pesquisa bibliográfica e exploratória com abordagem qualitativa. Primeiro, foi realizada a busca por artigos científicos, capítulos de livros e trabalhos publicados em bases de dados nacionais e internacionais, entre 2015 e 2020, acerca dos termos: organização do conhecimento; conceito; teoria do conceito; sistemas de organização do conhecimento; ontologias e tesouros. Além das fontes de informação encontradas neste primeiro levantamento, foram reveladas referências que sustentavam a sua argumentação.

Na pesquisa exploratória, buscou-se descrever os Sistemas de Organização do Conhecimento (SOC), e, especialmente os elementos estruturais essenciais dos tesouros e ontologias (os conceitos e seus termos descritores), evidenciando a Teoria do Conceito enquanto um dos alicerces na concepção destes instrumentos para a organização do conhecimento e mais especificamente, para a representação do conhecimento partindo do conceito.

Em relação à forma de abordagem, foi utilizada a qualitativa, pois o estudo objetiva discutir como dois instrumentos, que atuam de forma integrada na organização do conhecimento, partem do universo abstrato dos domínios permeados por conceitos, possibilitando a recuperação da informação em mecanismos de busca na web. Dessa forma, além de textos que analisassem

tanto o tesauro, quanto a ontologia, o foco do estudo estava em encontrar artigos e livros que tratassem de sua atuação enquanto SOC, além da interação de seus elementos estruturais para fins específicos da organização do conhecimento. Por isso, não se objetivou recorrer quantitativamente a experiências de interação entre tesauro e ontologia, mas sim, discutir quais atividades poderiam ser apontadas como exemplos de integração, focando na necessidade de apontar como isto ocorre a partir do elemento essencial dos SOC, que é o conceito (sendo este designado por termos descritores).

No levantamento bibliográfico, as buscas foram realizadas em bases como a *Scientific Electronic Library Online* (SciELO); *Scencedirect*; o Banco de dados bibliográficos da Universidade de São Paulo (DEDALUS); Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); Portal de periódicos CAPES/MEC e *Library and Information Science Abstracts* (LISA). O período escolhido para coleta de textos de 2015 a 2020 tinha como objetivo trazer uma revisão de literatura recente para apresentar a estrutura do tesauro e da ontologia enquanto SOC, assim como as abordagens acerca do papel do conceito na organização do conhecimento, principalmente na representação do conhecimento. A partir das referências iniciais, identificamos em Dahlberg, com a Teoria do Conceito, um dos alicerces para concepção destes instrumentos a partir da análise sistematizada do conceito.

Dessa maneira, o desenvolvimento da pesquisa foi dividido em três etapas: a primeira, consistindo na identificação e fichamento para reconhecimento dos objetos que devem ser analisados para chegar à discussão da temática do estudo, ou seja: integração entre tesauro e ontologia no processo de organização do conhecimento. Os termos requisitados eram então: organização do conhecimento; representação do conhecimento; organização da informação; Sistemas de Organização do Conhecimento; tesauro; ontologia e Teoria do Conceito.

Após isto, na segunda etapa, houve a análise dos textos e a sua leitura integral, visando a identificação dos aspectos relevantes para a organização dos dados para fins de discussão e definição de resultados. Por último, a terceira etapa resultou da revisão de literatura desejada, sendo realizada a discussão

acerca da atuação de tesauro e ontologia de forma integrada na organização do conhecimento partindo de seu elemento estrutural em comum, o conceito.

A discussão foi realizada tendo como ponto principal o reconhecimento dos elementos estruturais destes SOC, largamente discutidos nas etapas anteriores, buscando trazer a Teoria do Conceito como fundamento teórico que viabiliza a qualificação da organização do conhecimento resultante da interação entre tesauro e ontologia por meio de uma representação do conhecimento mais detalhada.

Foram selecionados 30 trabalhos por abordarem os elementos teóricos para estudar o termo organização do conhecimento, quando foi revelado o papel essencial do conceito no processo de organização do conhecimento e atuação dos Sistemas de Organização do Conhecimento (SOC), tendo como embasamento principal a Teoria do Conceito de Dahlberg (1978, 2009, 2011, 2014), Pieterse e Kourie (2014) e Soergel (2009). A estruturação do tesauro e seu funcionamento enquanto SOC teve contribuições principais de Tristão; Facchin e Alarcon (2004); Clarke (2019); Cintra *et al.* (2002) e ISO 25964-1 (2011).

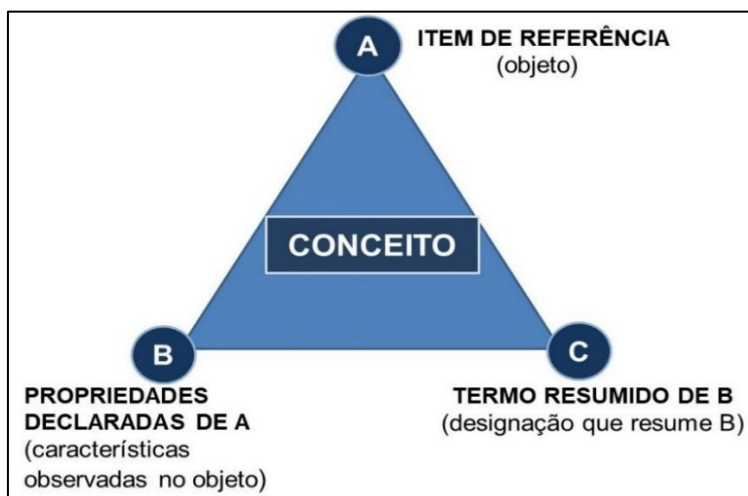
Posteriormente, destacaram-se como núcleo de referência principal os trabalhos de Gruber (1993); Jasper e Uschold (1999) e Ramalho (2015), que proporcionaram maior conhecimento a respeito da abordagem da ontologia enquanto SOC, descrevendo também sua estrutura e atuação na representação do conhecimento. Por fim, foi realizada a discussão a respeito do uso integrado do tesauro e ontologia na organização do conhecimento. Autores como Almeida, Mendonça e Aganette (2013) e Maculan e Aganette (2017) auxiliaram na elucidação de como as diferenças e semelhanças entre estes SOC possibilitam a sua integração na organização do conhecimento, sendo o conceito elemento essencial para recuperação de itens. Os exemplos de interação entre elementos estruturais do tesauro e ontologia foram coletados especialmente nos textos de Constâncio (2017) e Maculan (2015).

3 A ORGANIZAÇÃO DO CONHECIMENTO (OC)

O processo de organização do conhecimento parte da ideia de que

organizar é conceber um arranjo de elementos em uma só estrutura. Já conhecimento é a representação de algo, falso ou verdadeiro, que está sendo afirmado (SOERGEL, 2009, p. 12-13). A atribuição do termo organização do conhecimento como um campo de estudo específico tem início do século XX com os trabalhos de Charles A. Cutter, WC Berwick Sayers e Ernest Cushington Richardson. Em 1929, Bliss (1870-1955) adota o termo ao lançar *The Organization of Knowledge and the System of the Sciences* (HJØRLAND, 2016a, p. 475), cuja tradução literal pode ser “A Organização do Conhecimento e os Sistemas das Ciências”. Na década de 1970, Dahlberg usa o termo alemão *Wissensordnung* (que significa *ordenação do conhecimento*) para designar a organização conceitual e sistemática do conhecimento, declarando ser necessária uma teoria para estabelecer critérios para a análise do elemento essencial do conhecimento, o conceito ou “unidade de conhecimento” (DAHLBERG, 2011, p. 69). Assim, a Teoria do Conceito, de 1978, revisitada em textos de 1978, 2009, 2011, 2014, defende o conceito como elemento essencial para a formação dos campos de conhecimento (DAHLBERG, 2009). Com a Teoria do Conceito, Dahlberg promove a complementaridade entre a análise do conceito (DAHLBERG, 1978, 2009) e eclosão de uma área de pesquisa autônoma chamada Organização do Conhecimento (OC) (DAHLBERG, 2011, 2014). E é união destes objetos de estudo que gera a preocupação com a criação de critérios e embasamento teórico para organizar o conhecimento, assim como para estruturar os Sistemas de Organização do Conhecimento (SOC) a partir da análise de seu elemento essencial, o conceito. A Teoria do conceito, que visa sistematizar a análise do conceito, traz a visualização de um triângulo cujos vértices “A”, “B” e “C” trazem aspectos para reconhecermos o conceito, conforme demonstra a Figura 1:

Figura 1 – O triângulo conceitual



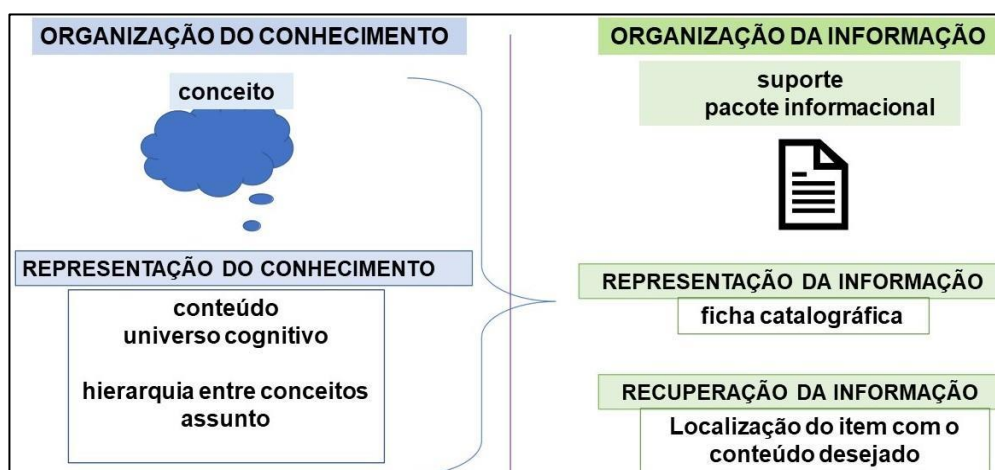
Fonte: Adaptação de Prado (2020, p. 23) a partir de Dahlberg (2009, p. 171).

De acordo com a Figura 1, o vértice “A” é o objeto ou assunto observado, “B” trata das “propriedades declaradas de A” e o vértice “C” é o “termo resumido das propriedades declaradas de A”, ou seja, para que cheguemos ao conceito, observamos algo, fazemos afirmações sobre as suas características e utilizamos um nome para resumir nossas declarações a respeito do objeto. Assim identificamos a definição do conceito, que, ainda para Dahlberg (2009), é a forma mais diminuta de chegarmos às declarações sobre as características observadas em um objeto. E, reconhecendo assim as características essenciais, visualizamos a relação hierárquica entre os conceitos em um determinado domínio (DAHLBERG, 2014, p. 86), interferindo na forma como acervos são organizados e recuperados.

Complementar à organização do conhecimento na CI, a organização da informação é o processo que busca registrar o conhecimento por meio das “características físicas e de conteúdo” dos itens que o armazenam – ela atua então no resgate de um determinado registro informacional, quando descrevemos as características do suporte que carrega o conhecimento (MAIMONE; TÁLAMO, 2015, p. 63). Ou seja, realizamos a representação da informação, que pode ser entendida como: “(...) um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico” (BRÄSCHER; CAFÉ, 2008, p. 5). Parte-se então do aspecto abstrato para o

prático, a partir da ação de representar o conteúdo e o pacote informacional em sua individualidade, trazendo à tona os materiais ou conteúdo solicitado pelo usuário. A Figura 2 demonstra como a organização do conhecimento e da informação são complementares:

Figura 2 – Organização do conhecimento e organização da informação



Fonte: Adaptação a partir de Prado (2020, p. 38).

De acordo com o exposto na Figura 2, a representação do conhecimento, baseada no nível conceitual e abstrato do material, viabiliza a recuperação da informação, sendo esta também consequente da representação realizada no nível do suporte, descrevendo o item físico ou digital. Dessa maneira, organizar o conhecimento tem relação com atividades como estabelecimento de hierarquia entre conceitos; classificação; atribuição de assunto a documentos e condensação de conteúdo (ou resumo). Em uma relação complementar à organização do conhecimento, a organização da informação está concentrada no suporte quando é recuperado determinado conteúdo, em sua especificidade, que foi solicitado pelo usuário.

4 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Instrumentos que realizam a organização do conhecimento, mais especificamente a representação do conhecimento, os Sistemas de Organização do Conhecimento (SOC) são formados por conceitos, termos ou “designações de conceitos” e relações entre conceitos (SOERGEL, 2009, p. 19). Estes instrumentos

buscam traduzir, por meio de uma linguagem controlada, a linguagem natural, permitindo que os conceitos imersos nas áreas de conhecimento sejam designados por termos. A padronização terminológica define então quais são os termos que se adequam aos conceitos de um dado domínio, o que repercute na recuperação da informação. Essa padronização também auxilia na definição dos conceitos e das relações entre conceitos, viabilizando o compartilhamento entre usuários de uma comunidade (BRÄSCHER, 2014).

Quando tratamos dos conceitos, para que possamos organizar o conhecimento, é na Teoria do Conceito, de Ingetraut Dahlberg (1978), que encontramos embasamento para a representação do conhecimento, processo que se inicia ao identificarmos a natureza dos objetos observados. Isso porque, o ser humano consegue com o enunciado gerar conexão com o mundo, diferenciando assim uma coisa da outra, a partir de critérios gerais e específicos para agrupar objetos ou assuntos. Assim, a natureza geral é importante para classificarmos, uma vez que as classes apresentam objetos com características gerais em comum, possuindo um elemento essencial. Como exemplo, temos o reino animal, onde estão inseridos todos os tipos de animais. Também é partindo do agrupamento que criamos sistemas conceituais baseados nas características dos conceitos de um determinado campo de estudo.

As relações entre conceitos podem ser divididas em: lógicas; hierárquicas; partitivas; de oposição e funcionais (DAHLBERG, 1978, p. 104-105). Por meio dos tipos de relações entre os conceitos, assim como da definição dos conceitos, melhor estruturamos um SOC. Portanto, ao criar um instrumento para organizar o conhecimento sobre planetas, sabemos que os planetas do sistema solar recebem nomes distintos. Porém, quando descobrimos novos planetas, precisamos definir o que é planeta, sendo esse o ponto de partida para distingui-los dos outros corpos celestes (DAHLBERG, 1978, p. 106).

4.1 TESAURO

O termo *thesaurus* foi usado pela primeira vez, em 1852, no dicionário de Peter Mark Roget *Thesaurus of English words and phrases* (TORRES; ALMEIDA, 2015, p. 6-7). Considerado um caso especial de esquema de

classificação facetada (TRISTÃO; FACHIN; ALARCON, 2004, p. 167), trata-se de um Sistema de Organização do Conhecimento (SOC) cuja organização lógica e hierárquica parte da noção facetada do ideal de Ranganathan – privilegiando o ponto de vista (CINTRA *et al.*, 2002, p. 41).

O tesouro acaba se adequando aos objetivos de um grupo de especialistas, lidando com termos específicos para representar o conteúdo de uma área de conhecimento. A ISO 25964-1 (2011) o define como vocabulário controlado, onde os conceitos são representados por termos. E, estes, por sua vez, são apresentados de maneira que as relações conceituais fiquem claras.

Como é estruturado hierarquicamente, o tesouro possibilita a representação de relações entre conceitos, que podem ser de equivalência, hierárquicas ou associativas e identificadas por meio de etiquetas ou *tags*, conforme exposto no Quadro 1:

Quadro 1 – Termos e etiquetas no tesouro

RELAÇÕES REPRESENTADAS POR TERMOS			
	EQUIVALÊNCIA	HIERÁRQUICA	ASSOCIATIVA
Etiquetas	✓ (Use) (Termo preferencial)	❖ Genérica Animais (TG) Aves (TE)	❖ Causa/efeito Fogo/ incêndio (TA)
	✓ Usado para (UP) (Termo não preferencial)	❖ Partitiva Árvore (TG) Folhas (TE)	❖ Processo/agente Colar/ fita adesiva (TA)
		❖ Instância Mar (TG) Mar Vermelho (TE)	❖ Processo/contra-agente Incêndio/ água (TA)

Fonte: Adaptado a partir de Clarke (2019), ANSI/NISO Z39.19 (2005) e ISO 25964-1 (2011).

Ainda de acordo com o Quadro 1, percebemos o uso de etiquetas distintas para especificar qual a relação que os termos representam. Por exemplo, nas relações hierárquicas, as etiquetas “TG” e “TE” manifestam os conceitos genéricos e específicos, de parte-todo e de instância. As relações associativas entre conceitos, são descritas por meio da etiqueta “TA” e mostram as relações

de causa e efeito, processo e agente ou processo e contra-agente. Dessa forma, a padronização de etiquetas visa trazer clareza às ligações entre conceitos, facilitando a consulta do usuário, permitindo que sistemas distintos possam dialogar. O Tesauro Brasileiro de Ciência da Informação, publicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), por exemplo, traz os termos descritores dos conceitos e suas etiquetas por meio de verbetes. Abaixo, na Figura 3, trazemos um exemplo de verbete para o termo “base de dados”:

Figura 3 - Verbetes para termo “bases de dados” do Tesauro Brasileiro de Ciência da Informação

bases de dados	
ING:	databases (UF data banks, data files, data warehouses, files)
ESP:	bases de datos (UP bancos de información)
TG	documentos por estrutura representação da informação
TE	bases de dados bibliográficos bases de dados científicos bases de dados de imagens bases de dados de textos completos bases de dados em linha bases de dados factuais bases de dados numéricos bases de dados referenciais
TR	dados gestão de bases de dados metadados modelos de bases de dados produtores de bases de dados serviços de indexação e resumo serviços de informação sistemas baseados em conhecimento
CAT:	2.1.1 Representação da informação 5.3.1 Bases de dados e extração da informação

Fonte: Adaptado a partir de Pinheiro e Ferrez (2014).

Conforme exposto na Figura 3, “ING”, “ESP” e “CAT” indicam a tradução do termo “bases de dados” para inglês e espanhol, assim como a categoria a qual o conceito pertence é “Representação da informação” e “Bases de dados e extração da informação”. Também são apresentados os Termos Gerais “documentos por estrutura” e “representação da informação”, assim como termos específicos e relacionados, que são, respectivamente, “bases de dados” “bibliográficos e metadados”.

A lista de termos ou terminologia de domínio, que também pode ser

entendida como um tipo de SOC, sugere as designações para os conceitos na pesquisa (SOERGEL, 2009). E no caso do tesauro, trata-se de elemento prévio para sua estruturação, sendo a Terminologia, enquanto disciplina, responsável por fornecer diretrizes e princípios para que seja realizado um trabalho terminológico de uma área de conhecimento específica (CABRÉ, 1995, p. 2), promovendo assim uma organização do sistema conceitual.

4.2 ONTOLOGIA

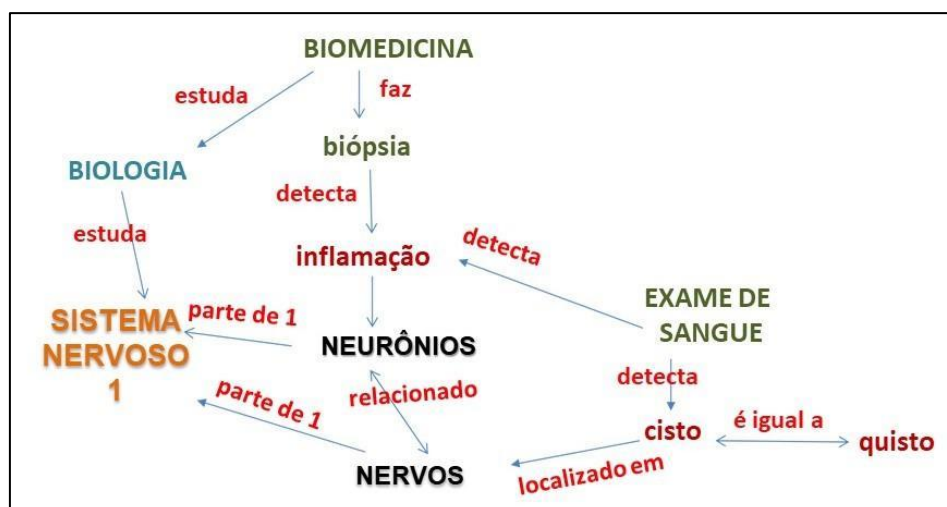
Na Filosofia, o termo ontologia remete ao ramo da Metafísica focada em designar as categorias das entidades e seus relacionamentos (LOWE, 2007). Quando inserida no ideal de Web Semântica, a ontologia é tratada como um instrumento que viabiliza o diálogo entre agentes inteligentes (PIETERSE; KOURIE, 2014). Na Ciência da Informação, partimos da ideia de que se trata de um Sistema de Organização do Conhecimento (SOC) de grande complexidade na representação do conhecimento, conseguindo explicitar as relações associativas entre conceitos e realizar inferências (ZENG, 2008; VICKERY, 2008; PIETERSE; KOURIE, 2014). Assim, reconhecendo que há abordagens do termo em áreas distintas, no âmbito deste estudo, a ontologia é compreendida como uma estrutura formada essencialmente por conceitos e termos, assim como por relações conceituais que permitem descrever os domínios de conhecimento, realizando assim a representação do conhecimento. Optamos também pela descrição de seus elementos estruturais definidos por Gruber (1993): classes (organização dos conceitos de um domínio); relações; axiomas (regras que restringem a interpretação dos termos) e instâncias (nomes para os dados). Assim, na visão da CI, podemos inferir que a ontologia acaba especificando os conceitos, os relacionamentos e outras distinções para conceber a estrutura de um domínio. Se a relação entre os conceitos e seus domínios de conhecimento nos transmite a tarefa de atribuir significados e repercute na descrição de recursos informacionais, visualizamos igualmente uma aplicação semântica da ontologia, com o adicional de fazer inferências automáticas para que os dados possam fluir entre computadores (RAMALHO, 2015). Percebemos assim que a ontologia é capaz de simular os julgamentos

que um ser humano faria por meio das relações entre os conceitos:

Uma ontologia pode assumir vários formatos, mas necessariamente deve incluir um vocabulário de termos e alguma especificação de seu significado. Esta deve abranger definições e uma indicação de como os conceitos estão inter-relacionados, o que resulta na estruturação do domínio e nas restrições de possíveis interpretações de seus termos (JASPER; USCHOLD, 1999, p. 25).

Neste SOC, a representação do conhecimento acontece por meio da exposição de relações conceituais, de acordo com as características dos conceitos, sendo divididas em relações conceituais de: equivalência (sinônimos), hierárquicas, partitivas, de oposição e funcionais – estas como reflexos da interação entre as características de conceitos distintos (DAHLBERG, 1978, p. 104-105). A ontologia também apresenta as relações lógicas entre conceitos apresentadas por Dahlberg (1978), sendo estas divididas em identidade; implicação; intersecção; disjunção e negação, além da relação funcional (referente a processos). Por isso, sua concepção envolve inicialmente a análise dos conceitos, determinando assim o uso e posicionamento do conceito em determinado domínio. A partir disto, que um determinado termo de busca pode gerar a conexão em rede, quando a relação semântica passa a ser identificada. É esta condição que possibilita o diálogo entre sites distintos para a recuperação da informação (BRÄSCHER, 2014). A Figura 4 demonstra como podem ser esboçadas as relações conceituais na taxonomia de uma ontologia do domínio da Biomedicina:

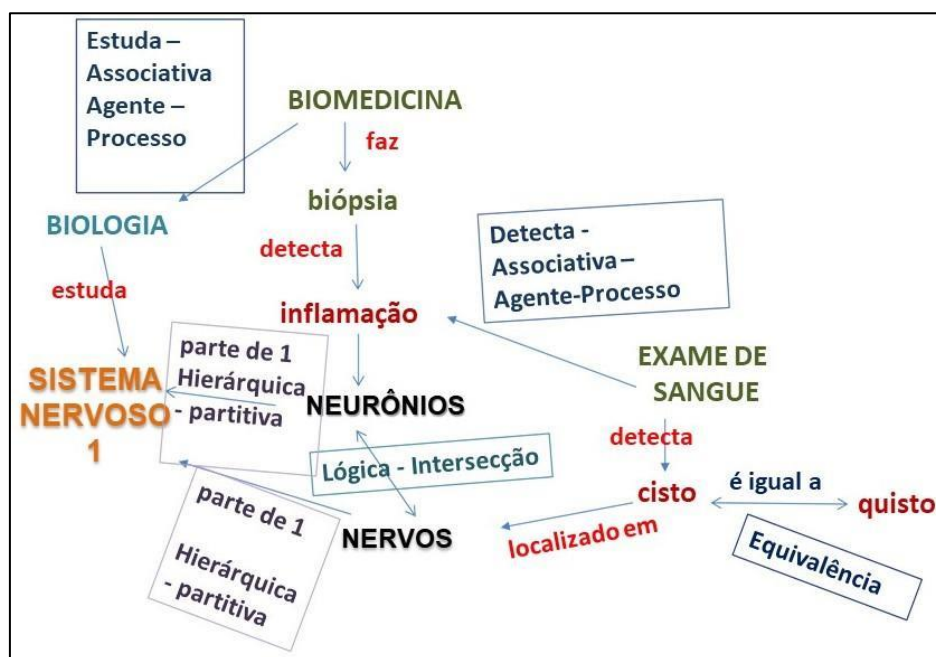
Figura 4 – Relações entre conceitos em ontologia para Biomedicina



Fonte: Adaptado a partir de Felon (2008) e Prado (2020, p. 81).

Na Figura 4 é possível observar alguns termos usados no domínio da Biomedicina. Descrevendo uma parte diminuta da área, a taxonomia faz parte de uma das etapas iniciais para a estruturação de uma ontologia para representação do conhecimento. Percebe-se que há um encadeamento que une os termos partindo de algum tipo de relação ou conexão, por exemplo, “cisto” é um termo sinônimo de “quisto”, representando uma relação de equivalência conceitual. A Figura 5 apresenta os tipos de relações conceituais da Figura 4, a partir, sobretudo, da perspectiva da Teoria do Conceito, de Dahlberg (1978):

Figura 5 – Visualização dos tipos de relações entre conceitos



Fonte: Elaborado a partir de Dahlberg (1978), Gruber (1993) e Bräscher (2014), Fenelon (2008) e Prado (2020, p. 92).

Conforme a Figura 5, percebemos como os conceitos (representados por termos), de uma taxonomia prévia para estruturação da ontologia, podem ser conectados com base nas relações, gerando declarações e inferências. A partir da especificação de relações, o agente faz julgamentos para recuperar itens ou conteúdos de interesse. Como no caso das relações associativas, a partir das quais são criadas declarações que conectam os conceitos por meio de verbos (“detecta” e “estuda”, por exemplo). Ao tornar essas relações legíveis para agentes inteligentes, a ontologia torna explícita as relações associativas – causa/efeito; agente/processo; processo/ferramenta (BRÄSCHER, 2014, p.

177), – o que repercute na ligação de conteúdos de fontes distintas. No Quadro 2, especificamos as relações apresentadas na Figura 5 acrescidas de exemplos:

Quadro 2 – Relações conceituais representadas nas ontologias

RELAÇÃO ENTRE CONCEITOS	Tipo	Exemplo
EQUIVALÊNCIA	sinônimo	<i>cisto - quisto</i>
HIERÁRQUICA	genérica	<i>biologia – biologia celular</i>
	todo- parte (partitiva)	<i>sistema nervoso – neurônios</i>
	instância	<i>biomédica Luana Flores</i>
ASSOCIATIVA	causa-efeito	<i>corte (causa) – sangramento</i>
	processo - agente (funcional)	<i>a) biologia (estuda)– sistema nervoso b) exame de sangue (detecta) – cisto</i>
	processo-contragente	<i>ressecação – cisto</i>
LÓGICA	intersecção (alguma característica em comum)	<i>neurônios – nervos</i>

Fonte: Elaborado pela autora a partir de Dahlberg (1978), Gruber (1993) e Bräscher (2014).

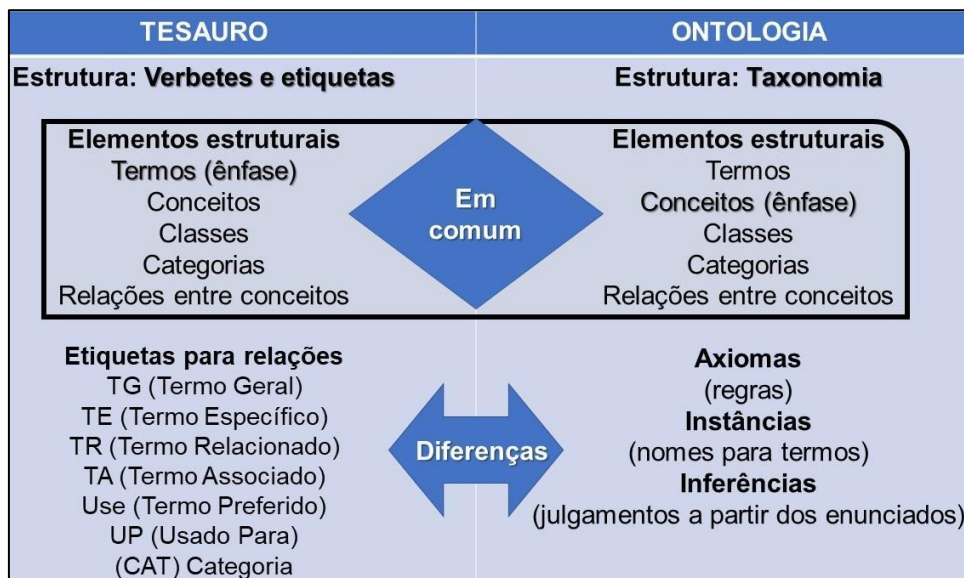
Dessa maneira, de acordo com o Quadro 2, podemos visualizar os exemplos para tipos de relações de equivalência, hierárquicas, associativas e lógicas entre conceitos no domínio da Biomedicina. Ao descrevermos as relações conceituais de forma complexa, há inserção de regras (axiomas) e inferências (julgamentos) com base na elaboração de enunciados, permitindo a criação de uma rede conectada entre itens ou conteúdo para a recuperação da informação em mecanismos de busca.

5 ONTOLOGIA E TESAURO: USO INTEGRADO NA ORGANIZAÇÃO DO CONHECIMENTO EM MECANISMOS DE BUSCA

Considerando as estruturas distintas do tesauro e da ontologia, podemos presumir que o seu uso integrado pode ser baseado na união de elementos estruturais para a representação do conhecimento partindo do conceito e das relações entre conceitos: de hierarquia, de equivalência e de associação. A representação é realizada com ênfases distintas: no tesauro, são descritas no trabalho terminológico partindo de uma perspectiva de uma dada comunidade

usuária; na ontologia, estão focadas nos conceitos de forma abrangente, não considerando necessariamente o domínio de sua aplicação. A Figura 6 relaciona as semelhanças e diferenças em suas estruturas:

Figura 6 - Semelhanças e diferenças na representação do conhecimento



Fonte: Elaborado pelas autoras a partir de Gruber (1993), ANSI/NISO Z39.19 (2005), ISO 25964-1 (2011) e Prado (2020).

De acordo com a Figura 6, o tesauro é organizado em verbetes e etiquetas. A ontologia faz uso de axiomas para criar enunciados a respeito dos termos, assim como instancia os termos, tornando-os particulares. Dessa maneira, os elementos estruturais básicos em comum, que são os termos, conceitos e relações entre conceitos, são acrescidos de componentes distintos para que ambos atuem na representação do conhecimento.

Considerando suas diferenças e similaridades estruturais, é possível questionar a possibilidade de uma integração baseada em seus elementos para responder às inconstâncias da linguagem natural manifestadas em mecanismos de busca, como: homonímia (um termo usado para representar conceitos diferentes) e sinonímia (dois termos usados para representar um conceito apenas). Na prática, isto refletiria na extensão dos resultados com base nas relações entre conceitos com tradução para a linguagem computacional, o que impactaria nos resultados, ou seja, na recuperação da informação.

Para Clarke (2019), as hipóteses de uso integrado de SOC estão ainda em construção, enquanto padrões e regras surgem, sendo ainda recente essa

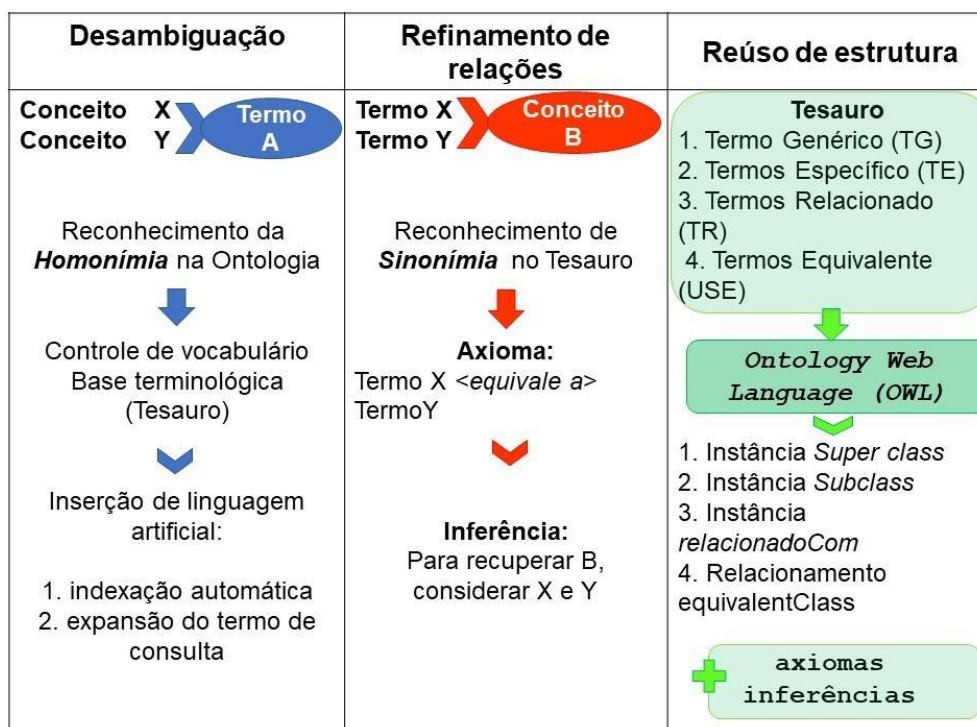
discussão. Questiona-se também se o tesouro será usado em nichos específicos ou como um instrumento integrado em rede. Neste caso, seu diálogo com a ontologia seria uma condição para agir em ambiente web, baseado na combinação de suas estruturas (MACULAN; AGANETTE, 2017).

Entende-se que, por assumir uma estrutura em taxonomia, a ontologia não possui limitações para demonstrar relações, já que estas estão amparadas em axiomas e declarações. E, pelo fato de o tesouro atribuir às distintas relações entre conceitos, muitas vezes, a mesma etiqueta “TR”, ao utilizarmos a integração de elementos, surgiria uma “ontologia específica” considerando quais conceitos e conjuntos de relações conceituais são apropriados para determinado domínio, traduzindo-os para a linguagem computacional (HJØRLAND, 2016b, p. 151). A representação do conhecimento estaria de acordo com a importância de cada tipo de relação entre conceitos para cada área, por exemplo, a relação “local/área” importa no contexto da geografia e relação de “causa/efeito”, na medicina, por exemplo. Por isso, o uso de um tesouro com “padrão uniforme para todos os campos” possivelmente não seria tão útil quanto a composição de uma ferramenta semântica de domínio específico com auxílio da ontologia (HJØRLAND, 2016b, p. 152). De acordo com Almeida, Mendonça e Aganette (2013), a atividade de restrição da linguagem natural é possível a partir do uso das etiquetas do tesouro para uma ontologia mais específica.

Assim como, quando a homonímia fosse verificada na ontologia, para eliminação de ambiguidade ou desambiguação, a base terminológica do tesouro iria definir quais são os termos preferidos e relacionados, principalmente com base na garantia literária, permitindo o uso mapeado de termos distintos para o mesmo conceito, gerando a recuperação da informação (MACULAN, 2015). Percebe-se então que a união de elementos destes SOC traz embasamento terminológico para criação de regras e inferências, viabilizando a indexação automática e ampliação de termos de busca realizada pela ontologia, que possui a faculdade de incorporar estes dados à linguagem computacional (SOERGEL, 2009). O reuso do tesouro como modelo para a elaboração de ontologia – com todas as relações entre conceitos levantadas a partir do embasamento terminológico e etiquetas – também é apontado pela literatura (CONSTÂNCIO, 2017).

Como brevemente exposto, observa-se que, no tesauro, o conceito é revelado pelo emprego do termo, sendo a faceta ou emprego contextual determinado sobretudo nas relações associativas. Já na ontologia, o uso do conceito é revelado com axiomas e inferências. A Figura 7 resume as atividades e elementos estruturais utilizados na integração entre tesauro e ontologia para a representação do conhecimento:

Figura 7 – A interação entre tesauro e ontologia



Fonte: Prado (2020, p. 114).

A partir da Figura 7 percebemos que a integração de componentes do tesauro e ontologia é fundamental para a qualificação da organização do conhecimento em mecanismos de busca. Também foi possível notarmos que as formas de integrar elementos respondem a demandas e contextos específicos no processo de representação do conhecimento.

6 CONSIDERAÇÕES FINAIS

A organização do conhecimento enquanto processo transita entre aspectos abstratos e pragmáticos inerentes que permeiam os domínios de conhecimento. Os Sistemas de Organização do Conhecimento, por sua vez, são

instrumentos que organizam o conhecimento, mais especificamente, a partir da representação do conhecimento. Este processo, na prática, diz respeito a reconhecer os conceitos, a estrutura hierárquica do sistema de conceitos de uma determinada área, assim como, as relações entre conceitos presentes – atividade que recebe contribuições de estudos relacionados à análise conceitual, como a Teoria do Conceito.

Especialmente na web, os mecanismos de buscas têm como base SOC como as ontologias e tesouros, que realizam a representação do conhecimento partindo dos conceitos, relações entre conceitos e termos descritores. Tratam-se de instrumentos que eliminam ambiguidades e sinônimos da linguagem natural, estabelecendo as relações hierárquicas e associativas entre conceitos. Por meio dessa estrutura prévia, os agentes inteligentes identificam os conceitos para navegação online e pesquisa em mecanismos de busca, promovendo a recuperação da informação. Dessa maneira, a análise do conceito permite a criação de SOC para o contexto da internet – onde a circulação da informação ocorre de maneira mais dinâmica e é preciso alcançar a sua fluidez por meio da troca constante entre linguagem natural, linguagem documentária e linguagem computacional.

Integrando os elementos do tesauro e da ontologia para fundamentar os mecanismos de busca podemos realizar atividades como desambiguação, refinamento de relações entre conceitos e reutilização de estrutura. Por isso, deve-se identificar quando um elemento pode ser útil em mecanismos de busca, viabilizando assim a organização do conhecimento e conseqüentemente, trazendo ao usuário, de acordo com sua demanda e contexto de uso, um SOC mais completo.

REFERÊNCIAS

ALMEIDA, M. B.; MENDONÇA, F. M.; AGANETTE, E. C. F. Interfaces entre ontologias e conceitos seminais da Ciência da informação: em busca de avanços na organização do conhecimento. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 14., 2013. **Anais** [...]. Florianópolis: UFSC, 2013.

ANSI/NISO Z39. 19-2005: R2010. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. 2005.

BLISS, H. E. **Organization of Knowledge and the System of the Sciences**. Harry Holt and Company, New York, 1929.

BRÄSCHER, M. Semantic relations in knowledge organization systems. **Ko Knowledge Organization**, [S. l.], v. 41, n. 2, p. 175-180, 2014.

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou do conhecimento. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008. **Anais [...]**. São Paulo: USP, 2008.

CABRÉ CASTELLVÍ, M. T. La terminologia hoy: concepciones, tendencias y aplicaciones. **Ciência da Informação**, Brasília, v. 24, n. 3, p. 289-298, set./dez. 1995.

CINTRA, A. M. M.; TÁLAMO, M. de F. G. M.; LARA, M. L. G. de; KOBASHI, N. Y. **Para entender as linguagens documentárias**. São Paulo: Polis, 2002.

CLARKE, S. G. D. The Information Retrieval Thesaurus. **Knowledge Organization**, [S. l.], v. 46, n. 6, p. 439-459, 2019.

CONSTÂNCIO, A. S. **Ontologia para um motor de busca semântica para recuperação jurisprudencial no Brasil**. Dissertação (Mestrado). Universidade Federal do Paraná, Curitiba, 2017.

DAHLBERG, I. Teoria do conceito. **Ciência da informação**, [S. l.], v. 7, n. 2, 1978.

DAHLBERG, I. **Concepts and terms—ISKO's major challenge**. 2009.

DAHLBERG, I. Brief Communication: How to improve ISKO's standing: Ten desiderata for knowledge organization. **Ko Knowledge Organization**, [S. l.], v. 38, n. 1, p. 68-74, 2011.

DAHLBERG, I. What is Knowledge Organization?. **Knowledge Organization**, [S. l.], v. 41, n. 1, p. 85-91, 2014.

FENELON, S. **Glossário de termos médicos e exames de imagem**. 2008. Disponível em: <http://imaginologia.com.br>. Acesso em: 22 out. 2021.

GRUBER, T. A translation approach to portable ontology specifications. **Knowledge acquisition**, [S. l.], v. 5, n. 2, p. 199-220, 1993.

HJØRLAND, B. Knowledge organization (KO). **Knowledge Organization**, [S. l.], v. 43, n. 6, p. 475-484, 2016a.

HJØRLAND, B. Does the traditional thesaurus have a place in modern information retrieval? **Ko Knowledge Organization**, [S. l.], v. 43, n. 3, p. 145-159, 2016b.

INTERNACIONAL STANDARD ORGANIZATION. **ISO 25964-1: 2011**: Information and documentation - Thesauri and interoperability with other vocabularies – part 1 – Thesauri for information retrieval. Geneva: ISO, 2011.TNI

LOWE, E. J. **The Four-Category Ontology**: A Metaphysical Foundation for Natural Science. New York, NY: Oxford University Press, 2007.

MACULAN, B. C. M. dos S. **Estudo e aplicação de metodologia para reengenharia de tesouro**: remodelagem do THESAGRO. 2015. 345 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MACULAN, B. C. M. dos S.; AGANETTE E. C. Desambiguação de relações em tesouros e o seu reuso em ontologias. **Ciência da Informação**, Brasília-DF, v. 46, n. 1, 28 dez. 2017.

MAIMONE, G. D.; TÁLAMO, M. de F. G. M. Contribuições da organização e representação do conhecimento para a epistemologia do campo da Ciência da Informação. **Revista Ciencias de la Documentación**, Santiago, v. 1, n. 3, 2015. Disponível em:
<http://www3.eca.usp.br/sites/default/files/form/biblioteca/acervo/producao-academica/002731481.pdf>. Acesso em: 20 abr. 2020.

PIETERSE, V.; KOURIE, D. G. Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle. **Ko Knowledge Organization**, [S. l.], v. 41, n. 3, p. 217-229, 2014.

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.

PRADO, A. C. R. do. **A interação entre o tesouro e a ontologia na organização do conhecimento**. 2020. Dissertação (Mestrado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2020. doi:10.11606/D.27.2020.tde-27032021-114627. Acesso em: 2022-10-17.

RAMALHO, R. A. S. Ontologias e Knowledge Organization System (SKOS): aproximações e diferenças. In. GUIMARÃES, José Augusto CHAVES; DODEBEI, Vera (Org.). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil, FUNDEPE, 2015. p. 100-107.

SOERGEL, D. Digital libraries and knowledge organization. *In*: **Semantic digital libraries**. Springer, Berlin, Heidelberg, 2009. p. 9-39.

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, v. 33, n. 2, p. 161-171, ago. 2004. Disponível em: <https://doi.org/10.1590/S0100-19652004000200017>. Acesso em: 28 jul. 2020.

TORRES, S.; ALMEIDA, M.. Classificação: uma operação inerente às linguagens documentárias. **DataGramaZero - Revista da Informação**, [S. l.], v. 16, p. 1-13, 2015.

VICKERY, B. C. A note on knowledge organization. 2008. Disponível em: http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/Vickery_a_note_on_knowledge_organisation. Acesso em: 20 out. 2021.

USCHOLD, M.; JASPER, R. **A framework for understanding and classifying ontology applications**, 1999. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.6456>. Acesso em: 20 abr. 2020.

ZENG, M. L. Knowledge organization systems (KOS). **Ko Knowledge Organization**, [S. l.], v. 35, n. 2-3, p. 160-182, 2008.

THE INTEGRATED USE OF THE THESAURUS AND ONTOLOGY IN SEARCH MECHANISMS FOR KNOWLEDGE ORGANIZATION

ABSTRACT

Objective: to describe how the structural elements of the thesaurus and ontology can act together in the organization of knowledge in search engines as knowledge organization systems. **Methodology:** an exploratory research was developed based on a bibliographic review in the field of Information Science to analyze how thesaurus and ontology work in the organization of knowledge as knowledge organization systems for further discussion of their interaction in the representation of knowledge in search engines. **Results:** through the surveys carried out, actions linked to the interaction of the thesaurus and ontology in the representation of the knowledge in search engines for the purpose of disambiguation, refinement of conceptual relationships and reuse of structures. **Conclusions:** the interaction between the thesaurus and ontology enables the qualification of the knowledge organization, since the thesaurus adds previous terminological work. o the ontology structure with the use of descriptor terms and ontology enables the refinement of conceptual relationships in the thesaurus, bringing effectiveness to the representation of knowledge.

Descriptors: Knowledge organization. Knowledge representation. Concept. Knowledge organization systems. Ontology. Thesaurus.

EL USO INTEGRADO DEL TESAURO Y LA ONTOLOGÍA EN MECANISMOS DE BÚSQUEDA PARA LA ORGANIZACIÓN DEL CONOCIMIENTO

RESUMEN

Objetivo: describir cómo los elementos estructurales del tesauro y la ontología pueden actuar en conjunto en la organización del conocimiento en los motores de búsqueda como sistemas de organización del conocimiento. **Metodología:** se desarrolló una investigación exploratoria basada en una revisión bibliográfica en el campo de las Ciencias de la Información para analizar cómo funciona el tesauro y la ontología trabajan en la organización del conocimiento como sistemas de organización del conocimiento para profundizar en la discusión de su interacción en la representación del conocimiento en los motores de búsqueda. **Resultados:** a través de las encuestas realizadas, acciones vinculadas a la interacción del tesauro y la ontología en la representación del conocimiento en motores de búsqueda con el propósito de desambiguación, refinamiento de relaciones conceptuales y reutilización de estructuras. **Conclusiones:** la interacción entre el tesauro y la ontología posibilita la cualificación de la organización del conocimiento, ya que el tesauro agrega trabajo terminológico previo. o la estructura de la ontología con el uso de términos descriptores y ontología permite el refinamiento de las relaciones conceptuales en el tesauro, trayendo eficacia a la representación del conocimiento.

Descriptores: Organización del conocimiento. Representación del conocimiento. Concepto. sistemas de organización del conocimiento. Ontología. Tesauro.

Recebido em: 29.03.2022

Aceito em: 10.01.2024