

Desambiguação de relações em tesouros e o seu reúso em ontologias

Benildes Coura Moreira dos Santos Maculan

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

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

E-mail: benildes@gmail.com

Elisângela Cristina Aganette

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

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

E-mail: elisangelaaganette@hotmail.com

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

RESUMO

Um tesouro é composto de conceitos organizados por uma rede paradigmática (relações essenciais e estáveis entre termos e conceitos), e uma rede sintagmática (relações específicas em um contexto de uso). Tesouros e ontologias são sistemas de organização do conhecimento (SOCs) capazes de representar a base de conhecimento de um domínio, a partir de regras e critérios de estruturação. A base terminológica de tesouros pode ser utilizada na construção de ontologias. Porém, nem sempre as relações estabelecidas na estrutura dos tesouros, por terem uma semântica ainda considerada fraca, podem ser reutilizadas nas ontologias. Isso ocorre porque as ontologias exigem a explicitação das relações atribuídas entre objetos, propriedades, classes e indivíduos, e também a declaração de axiomas. Em 2015, trabalho apresentou o refinamento semântico de relações na estrutura de um tesouro de agropecuária. Tendo como parâmetros (1) o resultado do trabalho mencionado, (2) a literatura sobre a conversão de tesouros em ontologias e (3) as recomendações para a construção de tesouros e ontologias, este artigo explora a questão sobre o reúso do conhecimento estruturado na estrutura de tesouros na construção de ontologias. Os resultados apontam que os relacionamentos semanticamente refinados na estrutura do tesouro podem ser mais bem aproveitados na criação de ontologias, mas que não há uma correlação perfeita. Este estudo contribui com a ciência da informação, uma vez que traz insumos para maior entendimento sobre as semelhanças e diferenças estruturais entre tesouros e ontologias para o reúso de conhecimento.

Palavras-chave: Tesouro. Ontologias. Reúso de conhecimento.

An approach to deal with ambiguity of relationships in thesauri for reuse in ontologies

ABSTRACT

A thesaurus is composed of concepts organized by a paradigm network (essential and stable relations between terms and concepts), and a syntagmatic network (specific relations in a context of use). Thesauri and ontologies are knowledge organization systems (SOCs) that are capable of representing the knowledge base of a domain, based on rules and structuring criteria. The terminological basis of thesauri can be used in the construction of ontologies. However, not always the relations established in the structure of thesauri, because they have a semantic still considered weak, can be reused in the ontology. This is because ontologies require explicit relations between objects, properties, classes and individuals, as well as the declaration of axioms. In 2015, a work presents the semantic refinement of relationships in the structure of a thesaurus of the Agricultural area. (1) the results of the aforementioned work, (2) the literature on the conversion of thesauri into ontologies and (3) recommendations for the construction of thesauri and ontologies, this article explores the question about the reuse of structured knowledge in the structure of thesauri in the construction of ontologies. The results point out that the semantically refined relationships in the thesaurus structure can be better used in the creation of ontologies, but that there is no perfect correlation. In Information Science, this study contributes because it brings inputs to a greater understanding about the similarities and structural differences between thesauri and ontologies for reuse of knowledge.

Keywords: *Thesaurus. Ontologies. Reuse of knowledge.*

Desambiguación de relaciones en tesauros y su reutilización en ontologías

Resumen: Un tesoro se compone de conceptos organizados por una red paradigmática (relaciones esenciales y estables entre términos y conceptos), y una red sintagmática (relaciones específicas en un contexto de uso). Tesauros y ontologías son sistemas de organización del conocimiento (SOC) que son capaces de representar la base de conocimiento de un dominio, a partir de reglas y criterios de estructuración. Se puede utilizar la base terminológica de tesauros en la construcción de ontologías. Pero no siempre las relaciones establecidas en la estructura de los tesauros, por tener una semántica aún considerada débil, pueden ser reutilizadas en las ontologías. Esto ocurre porque las ontologías exigen la explicitación de las relaciones asignadas entre objetos, propiedades, clases e individuos, y también la declaración de axiomas. En 2015, trabajo presentó el refinamiento semántico de relaciones en la estructura de un tesoro del área de la Agropecuaria. Teniendo como parámetros (1) el resultado del trabajo, (2) la literatura sobre la conversión tesoro en ontologías y (3) las recomendaciones para la construcción de tesauros y ontologías, este artículo explora la cuestión de la reutilización de conocimiento estructurado en la estructura de tesauros en la construcción de ontologías. Los resultados apuntan que las relaciones semánticamente refinadas en la estructura del tesoro pueden ser más bien aprovechadas en la creación de ontologías, sin embargo, que no hay una correlación perfecta. Este estudio contribuye con la Ciencia de la Información, una vez que trae insumos para una mejor comprensión de las semejanzas y diferencias estructurales entre tesauros y ontologías para reúso de conocimiento.

Palabras-clave: *Tesoro. Ontologías. Reúso de conocimiento.*

INTRODUÇÃO

Os tesouros, um tipo de sistema de organização do conhecimento (SOC), são desenvolvidos para representar o conhecimento de um domínio, a partir de um conjunto de termos descritores, preferidos e não preferidos, inter-relacionados em um sistema conceitual (estrutura do tesouro), fazendo o controle da terminologia em níveis diferenciados de controle e padronização. No âmbito da ciência da informação (CI), as ontologias são também um tipo de SOC, desenvolvidas como um modelo conceitual e base de conhecimento de um domínio, para serem interpretáveis pela máquina. A sua criação, em geral, é realizada utilizando-se linguagem formal de representação de dados capazes de proporcionar inferências a partir das relações atribuídas entre seus diferentes elementos.

Na literatura sobre tesouros, é usual a recomendação para que a terminologia de tesouros seja utilizada na construção de outros tesouros de mesmo domínio, com propósitos diferentes, e a mesma recomendação é encontrada para a criação de ontologias. Porém, não há indicação para reuso das relações dos tesouros. Em pesquisa realizada por Maculan (2015), a autora observa que na literatura internacional há substancial reuso de conhecimento estruturado de tesouros na criação de ontologias, inclusive das relações, fato que não se repete no Brasil. Para Boccato, Ramalho e Fujita (2008, p. 207), essa reutilização é um “fato [que se] justifica [...], pois eles [os tesouros] podem servir como substratos teóricos para a construção de ontologias”. Considera-se que o conhecimento estruturado nos tesouros está pautado nas bases teóricas da biblioteconomia, CI e terminologia, sendo, a princípio, uma fonte confiável para ser reutilizada na construção de outros instrumentos. Contudo, o problema que se apresenta é se essa fonte, sobretudo no que concerne às relações semânticas estabelecidas entre termos e conceitos, efetivamente contribui para a criação de ontologias. Nesse sentido, este artigo parte do pressuposto de que, no Brasil, há pouca discussão sobre o reuso da estrutura de tesouros para o desenvolvimento de ontologias, em especial das relações semânticas. Para

analisar essa questão, utilizam-se como parâmetros a literatura sobre o tema, as normas de construção e os resultados do refinamento semântico de relações em tesouros de Maculan (2015).

O artigo está organizado da seguinte maneira: a metodologia descreve os procedimentos adotados; a seção sobre ambiguidade aponta os problemas que afetam a elaboração de proposições em tesouros e ontologias; a seção dos tesouros apresenta as relações de sua estrutura conceitual; a seção das ontologias proporciona as bases de sua construção; o estudo de caso oferece a proposta de refinamento de relações em um tesouro; a seção de reuso de conhecimento estruturado dos tesouros na criação de ontologias apresenta os resultados e as análises; por fim, têm-se as considerações finais, seguidas da lista de referências utilizadas no artigo.

METODOLOGIA

Este artigo apresenta resultados de um estudo exploratório-descritivo que visou ao maior entendimento sobre a temática do reuso de conhecimento estruturado de tesouros para a construção de ontologias, em especial no que se refere às relações estabelecidas. Este estudo se justifica, uma vez que é tema ainda pouco explorado na literatura nacional. Primeiramente, buscou-se na literatura melhor visão do problema e, em seguida, delimitou-se este estudo, com os procedimentos: (1) estabelecer a proposta de estudo; (2) formular o problema estudado; (3) mapear a literatura sobre o tema; (4) estabelecer os parâmetros para a análise; (5) apresentar a análise dos resultados. Nesse sentido, estabeleceu-se a proposta de estudo, a partir do pressuposto da carência de estudos nacionais sobre o reuso de conhecimento estruturado de tesouros na construção de ontologias. Em seguida, realizou-se um mapeamento na literatura sobre o tema, ao se constatar a escassez de discussão sobre essa temática em âmbito nacional. Segundo Maculan (2015), cerca de 15 trabalhos trataram a respeito da conversão de tesouros em ontologias na literatura internacional até julho de 2013, e, no Brasil, há somente Campos et al. (2008), que trabalharam com a conversão do tesouro do folclore brasileiro em uma ontologia de domínio

e discutiui o reuso das relações de tesouros na criação de ontologias. A busca bibliográfica realizada neste estudo confirmou o pressuposto aqui estipulado.

Os parâmetros para as análises tiveram por base a literatura sobre conversão de tesouros em ontologias, recomendações para o desenvolvimento de tesouros e ontologias, em especial, da norma ISO 25964, Parte 1 e 2 (2011; 2013), e os resultados de Maculan (2015). A autora trabalhou com o modelo de Soergel et al. (2004) para a reengenharia de tesouros que determina a explicitação dos relacionamentos na estrutura dos tesouros, tornando-os com maior expressividade (refinamento) semântica. Com base nisso, este estudo buscou responder às questões: (1) a explicitação das relações de equivalência nos tesouros pode contribuir para a criação de equivalências na estrutura de ontologias?; (2) a explicitação das relações hierárquicas (gênero-espécie, todo-parte e instância) pode contribuir para a criação de hierarquias na estrutura de ontologias?; (3) a explicitação das relações associativas nos tesouros pode contribuir para a criação de relações não hierárquicas na estrutura de ontologias?; (4) que função pode ter a estrutura dos papéis *Qualia* na construção de ontologias?

AMBIGUIDADE, DESAMBIGUAÇÃO E PROPOSIÇÕES

A ambiguidade é o emprego de palavras, termos ou expressões que permitem mais de uma interpretação. No contexto linguístico, a ambiguidade pode ser estrutural (ou gramatical) e lexical (ou polissêmica). Na primeira, os problemas são relativos à construção de sentenças, tais como a posição de complementos, adjuntos ou redução de orações (GOMES, 2009). Na segunda, ocorre a identidade de duas lexias que são “formas perfeitamente iguais que se distinguem semanticamente (um significante para dois significados, no plano do conteúdo)” (ZAVAGLIA, 2003a, p. 83).

A ambiguidade estrutural pode ocorrer na elaboração de textos para as definições de termos que representam conceitos em tesouros e ontologias, pois os enunciados não podem carregar diferentes

possibilidades discursivas. Nos tesouros, grande parcela da comunicação da estrutura do domínio está expressa nas definições e nas notas de escopo, que estabelecem uma finalidade e intenção única (clara, precisa, coerente e unívoca) para cada conceito. Nesse sentido, elaborar esses enunciados exige refletir a intenção de comunicação pretendida no escopo do instrumento criado.

A ambiguidade lexical pode acontecer de diferentes maneiras, porém, neste artigo, serão tratados aqueles termos que mais interessam na construção de tesouros: sinonímia, polissemia e homonímia. A sinonímia ocorre quando há um ou mais termos (palavras ou expressões) com sentido igual ou semelhante. Nos tesouros, esse fenômeno é controlado ao se estabelecer uma relação de equivalência entre um descritor preferido e outro(s) descritores não preferido(s). Essas relações informam que somente um descritor é autorizado para uso, indicando: origem linguística diferente: poliglota e multilíngue; nomes populares e nomes científicos: pinguins e enfenisciformes; substantivos comuns e nomes comerciais: fotocópia e xerox; variantes para conceitos novos: deletar e remover; termos reduzidos e termos antiquados: rádio e radiocomunicação; grafias variantes: Rumênia e Romênia; variações linguísticas: termos originais de culturas diferentes – mandioca e aipim; abreviaturas e nomes por extenso: PVC e polivinil clorido; forma fatorada e não fatorada de um termo: carvão + minério e minério de carvão.

A ambiguidade da polissemia é um fenômeno natural das línguas em geral e ocorre pela associação de diferentes sentidos a um único termo (palavra, expressão ou imagem), e pode ser considerada contrária à sinonímia. Ela pode ocorrer a partir de quatro formas básicas: (1) mudança de aplicação: mesma palavra difere conforme o contexto – indexação (economia), indexação (biblioteconomia); (2) especialização: mesma palavra pode ter diferentes sentidos especializados, mas apenas um deles será aplicado – análise de assunto, análise conceitual; (3) linguagem figurada: uma palavra usada em sentidos figurados, sem

perder o significado original – uso de metáfora e metonímia; (4) influência estrangeira: empréstimo entre idiomas diferentes.

Por fim, a ambiguidade da homonímia pode acontecer de (1) forma gráfica, quando termos (palavras ou expressões) possuem fonias iguais e sentidos diferentes, tal como em, por exemplo, acento (sinal gráfico) e assento (cadeira); e de (2) forma fônica, quando possuem identidade igual (grafia) e sentidos diferentes em quatro circunstâncias (ZAVAGLIA, 2003b): (a) homonímia semântica, mesma classe gramatical, com sentidos diferentes – cabo (militar) e cabo (da enxada); (b) homonímia categorial, quando possuem sentido e classes gramaticais distintas – conto (verbo) e conto (substantivo); (c) homonímia etimológica: étimo (morfema base para a formação de uma palavra, por derivação ou composição) e sentidos diferentes – manga (fruto: do idioma malaiala) e manga (do vestuário: do latim *manica*); (d) homonímia heterófona: quando têm sentido e fônica distintas – acerto (substantivo) e acerto (verbo).

Na organização conceitual da estrutura de tesouros e de ontologias, os termos estão correlacionados por tipos diferentes de relações. Segundo Zavaglia (2003a, p. 84), “o levantamento e a identificação das relações léxico-semânticas” são igualmente necessários no processamento da linguagem natural (PNL), porque são “recursos linguísticos para implementação computacional” (idem), sendo importante tratar o fenômeno da homonímia para a interpretação da máquina. Assim, as redes de significação podem ser estabelecidas por proposições: uma sentença declarativa sobre algo, que pode ser declarado e representado por termos, palavras ou símbolos, que permite determinar se o seu conteúdo possui um valor lógico verdadeiro ou falso (GOMES, 2009).

Na filosofia, a proposição é uma entidade abstrata que exprime, em uma frase declarativa, aquilo que tem valor (validade) de verdadeiro ou falso (MURCHO, 2003). De acordo com o autor, o conjunto de proposições (premissas; primitivas) utilizadas para provar uma coisa é denominado

“argumento”. Almeida (2014) afirma que fundamentos filosóficos vêm sendo utilizados para dar condição ontológica de proposição para o conteúdo de um documento. O autor argumenta que, para isso ocorrer, uma análise deveria seguir “o seguinte princípio: ‘a informação transportada pela sentença S é uma proposição apropriadamente associada a S’ (FOX, 1983, p. 84)” (ALMEIDA, 2014, p. 250). Nesse sentido, a informação indexada seria o conteúdo proposicional do documento.

De modo geral, entretanto, sendo a linguagem o meio do homem se expressar sobre as verdades, a ambiguidade está sempre presente e, com ela, a inconsistência. Assim, a construção de tesouros e de ontologias busca por sua eliminação. Nesses instrumentos, as relações semânticas são representadas por proposições verbais (conexão entre termos e conceitos), determinando fatos, ações, lugares, períodos temporais, modos, causas, efeitos, destinos, procedências, circunstâncias, qualidades do objeto, entre outros, que explicitam o sentido da ligação. Portanto, conhecer o comportamento dos verbos é importante na construção de tesouros e ontologias.

TESAUROS E SUAS RELAÇÕES SEMÂNTICAS

Um tesouro é um SOC constituído por um conjunto de termos descritores, preferidos e não preferidos, que formam um sistema de conceitos inter-relacionados, sendo desenvolvidos com base em conceitos (ISO 25964-1, 2011). Tem como função fazer o controle do vocabulário de um dado domínio de especialidade, em diferentes níveis de controle, sendo utilizado no processo de indexação, na etapa de tradução da linguagem natural utilizada nos documentos para a linguagem adotada pelo sistema (UNESCO, 1971). Pode também ter a função de recuperação da informação, pois, quando acoplado na saída de dados, permite a conversão da expressão de busca, em linguagem natural, para a linguagem controlada que foi utilizada na entrada de dados no sistema (CAVALCANTI, 1978). Com isso, é ainda possível ampliar (termos mais gerais)

ou restringir (termos mais específicos) a busca, tornando a recuperação mais relevante à demanda do usuário.

Um tesouro é constituído por quatro elementos básicos: (1) um léxico; (2) uma estrutura gramatical, que representa o padrão de relações existentes entre conceitos e termos; (3) uma rede paradigmática (relações essenciais e estáveis); (4) uma rede sintagmática (relações válidas no contexto de uso) (SVENONIUS, 2000). A construção dessa estrutura leva em consideração as bases da Teoria do Conceito e da Terminologia para o levantamento do vocabulário (conceitos e termos que os designam) e para a elaboração da definição, com o registro das relações semânticas. Também considera a Teoria da Classificação Facetada (TCF), que respalda a análise analítico-sintética dos assuntos e a formação de classes gerais, subclasses e facetas, e regras de validação das escolhas realizadas.

Dentre as normas e diretrizes para a construção de tesouros, destaca-se a norma internacional mais atual, a ISO 25964-1 (2011), com as linhas gerais para o estabelecimento de relações entre termos e conceitos, assim como as formas de desambiguação que são recorrentemente discutidas na literatura. O sistema conceitual de um tesouro abarca relações de equivalência, hierárquicas e associativas, conectadas por relações binárias, que podem ter propriedades tais como reflexivas e irreflexivas, simétricas e assimétricas ou transitivas e intransitivas.

As relações de equivalência ocorrem no nível do campo semântico do termo, pois, também na linguagem de especialidade, diferentes termos podem representar conceitos idênticos, semelhantes ou com sentidos sobrepostos. Dodebei (2002, p. 91) afirma que a relação de equivalência ocorre quando “um conceito pode ser representado por vários símbolos distintos ou quando se quer reduzir, por questões pragmáticas, os níveis de implicação conceitual”, sendo uma relação recíproca. De forma resumida, as equivalências abarcam os tipos: (1) sinonímia (total e parcial); (2) correspondência em outro idioma; (3) variações ortográficas; (4) variações linguísticas; (5) erro de soletração; (6) forma completa do termo; (7)

abreviatura; (8) siglas; (9) forma singular e plural e (10) nome popular e nome científico (MACULAN, 2015; ISO 25964-1, 2011).

Quanto às relações hierárquicas, ocorrem no nível conceitual e são de três tipos: (1) gênero-espécie: incluem relacionamentos reflexivos, transitivos, assimétricos e taxonômicos (semelhanças entre táxons), que conectam uma classe (gênero: grupo de seres ou objetos não individualizados) e seus membros (espécie: ser individualizado ou objeto específico). Nessas relações, incidem os fenômenos de hiperonímia (hiperônimo; superordenação), e de hiponímia (hipônimo; subordinação), que indicam uma inclusão semântica de maior especificidade (FERREIRA, 2009); (2) todo-parte: incluem relacionamentos transitivos e assimétricos, nos quais o todo é um termo mais abrangente (superordenado) que abarca outro conceito, de forma lógica e inerente, que é uma parte em posição subordinada, que, em geral, pode incidir em objetos (segmentação) ou na divisão de uma área do conhecimento (DAHLBERG, 1979). A norma ISO 25964-1 (2011) determina um número limitado de situações para a representação dessa relação: (a) sistemas e organismos do corpo; (b) localização geográfica; (c) disciplinas e campos de discurso; (d) estruturas sociais. A diretriz estabelece que outros tipos de relação todo-parte, ainda que possíveis, não devem ser incluídos nos tesouros, sobretudo quando a tal parte puder ser, também, parte de outro todo; e (3) de instância: incluem classes básicas (mais gerais) de elementos, tais como objetos, eventos, propriedades ou fenômenos, que representam um conjunto desses elementos, que, em geral, é representado por um substantivo comum, e as instâncias, por nomes próprios. Por exemplo, “Regiões Montanhosas” (superordenado: substantivo) e suas instâncias, “Alpes” e “Himalaias” (subordinado: nomes próprios) (ISO 25964-1, 2011).

Também as poli-hierarquias podem ocorrer em tesouros em circunstâncias tais como em relações: (a) de gênero-espécie: instrumento musical, órgão, que é subordinado dos superordenados instrumentos de teclado e de vento; (b) de todo-parte: bioquímica, que é subordinado aos superordenados biologia e química; (c) mistas: crânio, que é subordinado

aos superordenados osso (gênero-espécie) e cabeça (todo-parte).

As relações associativas, ou não hierárquicas, acontecem no nível conceitual e são as ligações entre conceitos de uma mesma categoria ou não, que implica um elo com propriedade simétrica. Soergel (1974) observa que, ainda que não seja a regra, as relações associativas podem ser assimétricas, unidirecionais, pois nem sempre há a necessidade de se indicar reciprocidade. Em especial, isso ocorre em tesouros multilíngues, pois em comunidades com diferentes culturas e/ou idiomas que compartilham um mesmo tesouro, nem sempre um conceito é adequado para todas as comunidades assistidas (ISO 25964-1, 2011). Svenonius (2000) afirma que, se na definição de um conceito está implícito outro conceito não ligado por relações hierárquicas, então haverá uma relação associativa entre eles; por exemplo, se para definir o conceito “ave” emprega-se o conceito “ornitologia”, é apropriado estabelecer uma relação associativa entre eles. Também é recomendado criar relações associativas entre conceitos com significados sobrepostos dentro do escopo coberto por um tesouro, visando orientar o usuário; por exemplo, entre os conceitos “navio” e “barco” que podem representar conceitos diferentes em dado contexto. Ressalta-se

que, se os dois conceitos estiverem em coordenação em uma mesma classe, a relação associativa não deverá ser estabelecida.

O quadro 1 apresenta situações nas quais a relação associativa deve ser criada, porém esse conjunto de situações não necessariamente limita-se a esses exemplos, sendo preciso mapear o domínio para determinar o conjunto necessário e suficiente ao escopo (ISO 25964-1, 2011).

Nos tesouros, as diferentes relações (equivalência, hierárquicas e associativas), em geral, são identificadas na sua estrutura pelos símbolos apresentados no quadro 2.

Como pode ser observado, apesar de todo o trabalho intelectual para o mapeamento e estabelecimento das diferentes relações (equivalência, hierárquicas e associativas) que ocorrem entre termos e conceitos, os símbolos utilizados na sua representação ainda carecem de um refinamento mais específico, ainda que haja a recomendação para que esse aprimoramento possa ser realizado. Um trabalho sobre refinamento de relações em tesouros é apresentado na seção do estudo de caso.

Quadro 1– Situações para estabelecimento de relações associativas em tesouros

SITUAÇÃO	EXEMPLO
1) Uma disciplina ou campo de estudo e os objetivos ou fenômenos estudados	Estética e Beleza
2) Uma operação ou processo e seu agente ou instrumento	Controle de temperatura e Termostato
3) Uma ação e o resultado da ação	Tecelagem e Tecido
4) Uma ação e o seu recipiente ou alvo	Colheita e Produtos agrícolas
5) Objetos e materiais e suas propriedades	Venenos e Toxicidade
6) Artefatos e suas partes (se não se qualifica na relação hierárquica todo/parte)	Instrumento ótico e Lentes
7) Conceitos ligados por uma dependência causal	Duelo e Morte
8) Um objeto ou processo e seu contra-agente	Plantas e Herbicidas
9) Um conceito e sua unidade de medida	Corrente elétrica e Amperes
10) Um conceito representado por um termo composto e o conceito representado pelo substantivo foco	Barcos para amarrar e Barcos
11) Um organismo ou substância derivada ou produzida de outra	Latão e Cobre

Fonte: Elaborado pelas autoras, com base na norma ISO 25964-1 (2011).

Quadro 2 – Descrição dos símbolos utilizados em tesouros

SÍMBOLO DAS RELAÇÕES	DESCRIÇÃO
Equivalência (sem especificação) USE; UF / USE; UP	USE, apontando o descritor preferencial UF (<i>used for</i> ; UP: usado para), apontando o descritor não preferencial
Equivalência (com especificação) USE; UF / USE; UP FT; AB / TC; AB *Outros	USE, apontando o descritor preferencial; e UF (<i>used for</i> ; UP: usado para), apontando o descritor não preferencial FT (<i>full term</i> ; TC: termo completo), apontando o descritor do nome completo; e AB (<i>abbreviation</i> ; abreviatura), apontando o descritor do nome abreviado ou sigla *Recomendação para subdividir as relações de equivalência para refinamento da semântica das ligações estabelecidas entre conceitos
Hierárquica (sem especificação) BT; TG NT; TE	BT (<i>broader term</i> ; TG: termo geral), apontando o descritor da classe mais geral NT (<i>narrower term</i> ; TE: termo específico), apontando o descritor mais específico
Hierárquica (com especificação) BTG; NTG / TGG; TEG BTP; NTG / TGP; TEP BTI; NTI / TGI; TEI	BTG (<i>broader term (generic)</i>); TGG: termo geral de gênero); BTP (<i>broader term (partitive)</i>); TGP: termo geral partitivo); BTI (<i>broader term (instantial)</i>); TGI: termo geral de instância), apontando o descritor da classe mais geral NTG (<i>narrower term (generic)</i>); TEG: termo específico de gênero); NTP (<i>narrower term (partitive)</i>); TEP: termo específico partitivo); NTI (<i>narrower term (instantial)</i>); TEI: termo específico de instância), apontando o descritor mais específico
Associativa RT; TR; TA	RT (<i>related term</i> ; TR: termo relacionado; TA: termo associado), apontando para o descritor sugerido para ser levado em consideração
Associativa (com especificação) RT; TR; TA *CAUSE/EFFECT; CAUSA/ EFEITO	RT (<i>related term</i> ; TR: termo relacionado; TA: termo associado), apontando para o descritor sugerido para ser levado em consideração *Recomendação para subdividir as relações específicas para refinamento da semântica das ligações estabelecidas entre conceitos

Fonte: Elaborado pelas autoras, com base na norma ISO 25964-1 (2011).

ONTOLOGIAS E AS RELAÇÕES SEMÂNTICAS

No papel de um SOC e no âmbito da CI, as ontologias são criadas a partir de representações de conhecimento para serem interpretáveis pela máquina como artefatos capazes de proporcionar inferências a partir das relações atribuídas entre seus diferentes elementos. É um modelo conceitual e base de conhecimento, podendo representar relacionamentos complexos entre objetos, por meio de uma modelagem formal de dados. Diferentes autores têm investigado sobre ontologias, tanto em relação a questões teóricas quanto a aplicações (GRUBER, 1992; GUARINO, 1995; 1998; VICKERY, 1997; SOERGEL, 1997; BORST, 1997; NOY; HAFNER, 1997; WAND; STOREY; WEBER, 1999; HOVY, 2002; GILCHRIST, 2003; ALMEIDA, 2006; 2014). Na CI, Vickery (1997) define ontologia como

um esquema (ou sequência formal de eventos), que, de alguma forma, pode envolver o uso de categorias semânticas, de conceitos significantes em um domínio particular (que pode ser amplo, como o universo de conhecimento), juntamente com uma definição ou uma nota de escopo para cada conceito, assim como mecanismos para a exibição das relações entre conceitos (VICKERY, 1997, p. 284).

O autor aponta, na definição, uma analogia entre os tesouros e as ontologias, embora deixe claro que há diferenças, em especial, nos propósitos e funções previstas para cada instrumento. De modo geral, uma ontologia é formada por: (a) classes ou conceitos: uma taxonomia; (b) relações: representadas pela interação entre termos e classes; (c) atributos e propriedades (*slots*): descritas pelas características dos conceitos e instâncias; (d) instâncias ou indivíduos: elementos que pertencem a uma classe e que possuem propriedades; (e) axiomas: representam sentenças verdadeiras; (f) regras para permitir inferências (NOY; MCGUINNESS, 2001).

Desde a concepção filosófica, o objeto material da ontologia é a formalidade (a forma) do “Ser”, que determina o “que é” (*quid*), e, na CI, a busca pela formalidade continua sendo priorizada (BORCHERT, 2006), mas a partir de níveis diferenciados (ALMEIDA; BAX 2003): (a) altamente informais: em linguagem natural; (b) semi-informais: em linguagem natural, de forma restrita; (c) semiformais: em linguagem artificial e definidas de maneira formal; (d) rigorosamente formais: com semântica formal, teoremas e provas. Apesar desses distintos níveis de formalidade, uma ontologia controla a ambiguidade, sinonímia, hierarquias, associações, com relacionamentos complexos, regras e axiomas.

Almeida e Bax (2003) destacam que não há consenso sobre a classificação das tipologias de ontologias. Dentre os tipos discutidos na CI, destacam-se (GUARINO, 1998; ALMEIDA; BAX, 2003): (1) ontologias de representação; (2) ontologias genéricas; (3) ontologias genéricas de domínio (*core ontologies*); (4) ontologias de domínio; (5) ontologia de aplicação; (6) ontologias de tarefas. A construção de ontologias é desafiante devido ao elevado consumo de tempo requerido em seu desenvolvimento e em sua implantação, em especial quando exige a integração de diferentes aplicações (SCHIESSL; BRÄSCHER, 2011). Os autores asseguram que ela é uma aliada em ambiente Web, pois “possibilita o trabalho cooperativo entre computadores e pessoas” (idem, p. 304). Entretanto, as conclusões de um estudo realizado por Silva, Souza e Almeida (2008, p. 74) demonstraram problemas com a “falta de um padrão para construção de ontologias e na falta de explicações sistemáticas de como, onde e sob quais limites podem ser utilizadas as abordagens teóricas dentro do processo de elaboração”. Esse fato tem tornado o desenvolvimento de ontologias, nas discussões científicas promovidas na área da CI, suscetível a algumas críticas (AGANETTE, 2015).

Quanto às relações semânticas nas ontologias, elas são estabelecidas por meio de uma organização taxonômica de conceitos, com hierarquias de generalidade e de especialidade que compõem o

conceito de heranças entre as distintas entidades (CARVALHEIRA, 2007). Pode-se afirmar que as relações estabelecidas em sua estrutura acontecem tal qual nos tesouros, já especificadas na seção anterior, porém, com maior complexidade de representação e condicionadas ao tipo e ao propósito do instrumento. Nas ontologias aplicadas de orientação aristotélicas, por exemplo, na *Basic Formal Ontology* (BFO), o instrumento trata somente de universais e particulares, com um conjunto limitado de relações ontológicas, que estão especificadas na *Relation Ontology* (RO) (ALMEIDA; MENDONÇA; AGANETTE, 2013), que visa permitir a interoperabilidade entre diferentes aplicações na biomedicina.

De modo geral, as linguagens para construção de ontologias proporcionam primitivas (regras de formação) para a determinação de classes, subclasses e decomposições: disjuntiva, quando “os conceitos são subclasses de um conceito comum e onde nenhum membro de uma subclasse é membro de outra classe”; exaustiva de subclasse, quando “todos os membros da classe ou conceito comum são membros de alguma de suas subclasses”; e de negação (não é uma subclasse de), quando “um conceito não é uma especialização de um outro conceito (CARVALHEIRA, 2007, p. 15-16). Assim, as relações entre classes e indivíduos podem ser especificadas a partir de combinações lógicas de interseção, união e complemento.

Dentre as linguagens que vêm sendo utilizadas para a implementação de ontologias, destacam-se: *Web Ontology Language* (OWL; Lite, DL, Full), *Knowledge Interchange Format* (KIF), *Resource Description Framework* (RDF e RDFS), *Ontology Inference Layer* (OIL), *DARPA Agent Markup Language + OIL* (DAML+OIL). Em geral, a linguagem utilizada é a OWL (desenvolvida para Web); é considerada de maior expressividade, capaz de declarar fatos, estruturar classes (hierarquias de subsunções; tipologia e instanciações) e propriedades (hierarquias de subpropriedades) em RDF e RDFS, além de possibilitar declarações de equivalência para classes e propriedades, que podem ser determinadas iguais ou desiguais, a depender da intencionalidade delas.

Por meio da OWL, é possível estabelecer as seguintes declarações taxonômicas (classe: generalidade e subclasse: especialidade), de instanciação (instâncias), de propriedade (tipo de dado; objeto; hierarquia de propriedade) e de relações.

A ISO 25964-2 (2013) é a primeira norma internacional que traz regras específicas para a construção de ontologias e estabelecimento de relações. As normas se restringem às ontologias formais de domínio, adotando a terminologia da linguagem OWL, mas determinam que, em geral, as ontologias devem ser expressas em lógica de primeira ordem, e que podem ser criadas com distintos níveis de especificidade, propósito e para diversos domínios.

Para o estabelecimento de classes, a norma recomenda que sejam feitas declarações explícitas de restrições de propriedade e critérios para que um indivíduo possa ser membro de uma classe. Ademais, pode-se definir as relações de uma classe com outras classes, usando conectores lógicos (proposições de relação) ou qualquer outro tipo de restrição necessária.

Por exemplo, “uma classe pode ser definida como sendo uma interseção entre duas classes (denotando todos os indivíduos que pertencem às duas classes) ou como complemento de outra classe (denotando os indivíduos que não pertencem àquela classe)” (ISO 25964-2, 2013, p. 73).

As hierarquias de classe são criadas por estabelecimento de subclasses que vão obedecer aos axiomas declarados para cada classe, pois as subclasses (filho) irão herdar (princípio da herança) as características e propriedades (atributos; relações entre membros da classe e com outras classes) da classe (pai). Todas as relações entre as classes e subclasses são transitivas e as propriedades podem possuir restrições declarativas. Os indivíduos (instâncias) são os objetos de um domínio em particular, sobre os quais são feitas declarações, tais como: é uma pessoa, é um livro, tem determinado sintoma, é um processo ou evento, entre outros. Ainda que as ontologias, em si, não necessitem do uso de linguagem natural, todas as classes, subclasses, propriedades e indivíduos são identificados por termos para que possam ser interpretáveis por humanos (ISO 25964-2, 2013).

Quadro 3 – Declarações utilizando linguagem OWL

Item	Declarações
País; Pessoa	Classe (generalidade)
Estudante	Subclasse de Pessoa (especialidade)
Brasil e Portugal	Instância da classe País
Brasil e Portugal	Disjunção (decomposição disjuntiva)
Nacionalidade	Propriedade de Pessoa (domínio) e País (imagem ou <i>range</i>)
Nacionalidade	Propriedade inversa igual a <tem_cidadão>
Idade	Propriedade funcional
Idade	Propriedade de tipo de dado onde Pessoa é o domínio e o range é um número inteiro
João	Instância da classe Brasileiro e que tem na propriedade Idade um valor 35
Brasileiro	Conjunto de membros da classe Pessoa
Multinacional	Conjunto de membros da classe Pessoa, que possui pelo menos dois valores da propriedade Nacionalidade

Fonte: Elaborado pelas autoras com base em Carvalheira (2007, p. 20).

REFINAMENTO DE RELAÇÕES EM TESAuros: UM ESTUDO DE CASO

Este estudo de caso apresenta resultados da pesquisa de Maculan (2015), que aplicou a proposta de reengenharia de tesauros (SOERGEL et al., 2004; LAUSER et al., 2006), cujos procedimentos visam ao refinamento semântico das relações na estrutura conceitual de um tesouro a partir de três etapas: (1) definição da estrutura do tesouro; (2) coleta de terminologia; (3) aplicação e edição da reengenharia no tesouro. Os procedimentos (1) e (2) seguiram as diretrizes de construção de tesauros da norma ISO 25964-1 (2011), que estão descritas em Maculan (2015). Neste artigo, para atender ao escopo, será apresentado o procedimento (3) do refinamento das relações.

A reformulação foi realizada no tesouro Thesagro, do domínio da Agropecuária, no recorte temático da Intensificação Agropecuária (Intagro), a partir da concepção teórica de Boserup (1965). Originalmente, a estrutura conceitual do Thesagro possui as relações de equivalência, hierárquicas e associativas, e a representação delas é feita pelos símbolos na língua inglesa: *USE* e *Used For* (UF), *Broader Term* (BT) e *Narrower Term* (NT), e *Related Term* (RT). O refinamento semântico das relações levou em consideração os fundamentos da Teoria do Conceito, Terminologia e Teoria da Classificação Facetada, contando com a validação de especialista do domínio a cada estágio do trabalho, seguindo três fases:

(1) Estabelecimento das classes básicas e conjunto de conceitos: o domínio temático foi modelado de forma híbrida, com os métodos (1) dedutivo: análise macro para obter uma visão geral do domínio temático, que resultou em um conjunto de nove classes básicas que representam o campo semântico – agricultura extensiva, agricultura intensiva, material e métodos, ambiente, agronomia, território e paisagem, socioeconômica, espaço e tempo, instituições; (2) indutivo: obter diversas visões micro (particulares) do domínio temático, a partir de dados coletados (*corpus*) na literatura do domínio Agropecuário, nacional e internacional.

Dessa análise, resultaram cerca de 600 termos para serem estruturados dentro das nove classes básicas, tendo sido feito um recorte amostral composto de 30 conceitos representativos das classes básicas, e, assim, do domínio temático.

(2) Estabelecimento da estrutura conceitual: foram modelados os grupos (*clusters*) semânticos da amostra, com o estabelecimento de relações semânticas (equivalência, hierárquica e associativa), realizado a partir das definições (delimitação de significado) para cada conceito. As definições foram elaboradas tendo como insumos os excertos definitórios recolhidos na literatura (glossários, dicionários, bases de dados da agropecuária, textos técnicos e científicos). Como resultado, foram obtidos cerca de 600 relacionamentos.

(3) Refinamento dos relacionamentos: a explicitação das relações foi realizada utilizando-se 44 diferentes tipos, incluindo os seus inversos (por exemplo, <influencia> e a sua inversa <é_influenciado_por>), sendo criadas 13 novas relações e 31 relações provenientes do *Agrontology*¹, do tesouro Agrovoc (*Multilingual Agricultural Thesaurus*).

Também para auxiliar na construção do sistema conceitual do tesouro, foi aplicada a estrutura dos papéis *Qualia*, desenvolvida por Pustejovsky (1995), para a identificação da semântica dos verbos, pois toda relação é explicitada a partir de uma expressão verbal (proposição). A estrutura *Qualia* possui os papéis: (a) Formal; (b) Constitutivo; (c) Tético; (d) Agentivo. Essa identificação deu sentido aos verbos das 44 relações e facilitou a determinação das equivalências, dos gênero-espécies (hiperonímia/hiponímia) e dos todo-partes (holonímia/meronímia).

A identificação dos papéis *Qualia* auxiliou a associação de propriedades às relações, facilitando o refinamento semântico das expressões verbais utilizadas nas relações, permitindo apreender os seus significados específicos, sobretudo das relações hierárquicas e associativas.

¹Disponível em: <<http://aims.fao.org/sites/default/files/uploads/file/aos/agrontology/index.htm>>.

Quadro 4 – Conjunto relações com refinamento semântico

NOME DA RELAÇÃO	PAPEL QUALIA	TIPO RELAÇÃO
1. afeta (TR)	AGENTIVO	ASSOCIATIVA
2. causa (TR)	AGENTIVO	ASSOCIATIVA
3. éFeitoAPartirDe (TR)	AGENTIVO	ASSOCIATIVA
4. influencia (TR)	AGENTIVO	ASSOCIATIVA
5. seDesenvolveEm (TR)	AGENTIVO	ASSOCIATIVA
6. éInsumoPara (TR)	CONSTITUTIVO	ASSOCIATIVA
7. inclui (TR)	CONSTITUTIVO	ASSOCIATIVA
8. fazUsoDe (TR)	TÉLICO	ASSOCIATIVA
9. temPrática (TR)	TÉLICO	ASSOCIATIVA
10. temSinonímiaAproximada (UP EQ-APR)	FORMAL	EQUIVALÊNCIA
11. temSinônimo (UP EQ-SIN)	FORMAL	EQUIVALÊNCIA
12. temTraduçãoEN	FORMAL	EQUIVALÊNCIA
13. éProdutoDe (TGG)	AGENTIVO	GENÉRICA
14. termoEspecificoGênero (TEG)	FORMAL	GENÉRICA
15. termoGenéricoGênero (TGG)	FORMAL	GENÉRICA
16. termoEspecificoInstância (TEI)	CONSTITUTIVO	INSTÂNCIA
17. éComponenteDe (TGP)	CONSTITUTIVO	PARTITIVA
18. éParteDe (TGP)	CONSTITUTIVO	PARTITIVA
19. temComponente (TEP)	CONSTITUTIVO	PARTITIVA
20. temParte (TEP)	CONSTITUTIVO	PARTITIVA

Fonte: Maculan (2015, p. 222-223).

O REÚSO DE RELAÇÕES DE TESAUROS NAS ONTOLOGIAS: ANÁLISES

Primeiramente, destaca-se uma diferença básica entre tesouros e ontologias. Um tesouro é uma linguagem documentária e, mais especificamente, uma linguagem de indexação, que tem como primordial função ser instrumento de tradução da linguagem natural utilizada nos documentos para o vocabulário controlado adotado por um sistema de recuperação da informação (SRI). Já as ontologias são criadas como um modelo conceitual de mundo, alinhadas a sistemas formais de representação, a partir de diferentes acepções. Nesse sentido, haverá ontologias de alto nível, nível médio e de domínio, por exemplo. Segundo Almeida, Mendonça e Aganette (2013), seria possível uma ontologia ser utilizada como uma linguagem documentária, no sentido de restringir a linguagem natural, “caso esteja referindo-se ao uso de linguagens de representação adotadas para especificar ontologias [...] implementada[s] em

um sistema de recuperação da informação baseado na Web” (idem, p. 13). Porém, esse não é o papel típico de uma ontologia.

No mapeamento de um domínio, tanto para criação de tesouros quanto de ontologias, as classes de maior abstração são oriundas da linguagem natural. Como os tesouros almejam uma restrição do vocabulário utilizado em um domínio em particular, em que, pelo menos, as garantias da literatura e do usuário são levadas em consideração, o nível de representação ainda ficará bem próximo da linguagem natural. Por sua vez, nas ontologias, o estabelecimento das classes mais gerais (categorias) se inicia no nível da linguagem natural e é elevado ao nível da semântica formal (ALMEIDA; MENDONÇA; AGANETTE, 2013), no qual o sentido de referencialidade é fundamental, pois são declaradas asserções sobre objetos (indivíduos, fatos, eventos, propriedades, entre outros), que são externos à própria língua.

Assim, o “significado é entendido como uma relação entre a linguagem, por um lado, e, por outro, aquilo sobre o qual a linguagem fala” (MÖLLER; VIOTTI, 2003, p. 138). Dessa maneira, as ontologias exigem relações explícitas, em um contexto de linguagem não natural, como a lógica, para que sejam passíveis de interpretação pelas máquinas e não apenas dispostos à abstração pelo ser humano (AGANETTE, 2015).

Na perspectiva das considerações anteriores, este artigo se atém às discussões relativas às possibilidades de reúso das relações semânticas (equivalência, hierárquicas e associativas), estabelecidas na estrutura de um tesouro, para a construção de ontologias.

Como já mencionado, as relações de equivalência acontecem, na estrutura dos tesouros, entre um descritor preferido e um não preferido, pois manter o elo “permitirá que a busca por qualquer um deles recupere informações referentes a todos os outros que estão conectados entre si” (MACULAN, 2015, p. 146). Assim, ela é operacionalizada no nível da significância (campo semântico) do termo, que é o conjunto dos significados que o termo possui dentro de um contexto. Nas ontologias, Schulz et al. (2009) afirmam que a relação de equivalência acontece no nível da referencialidade, pois trata-se de uma representação descritiva formal, no nível referencial, e apresentam uma relação de equivalência entre dois conceitos e a explicam

Hepatite \equiv Doença_Inflamatória \sqcap \exists tem_localização.
Fígado

O operador de equivalência \equiv nesta fórmula nos diz que: (I) cada ocorrência específica de hepatite é uma ocorrência de doença inflamatória localizada em algum fígado, e também (ii) que todas as ocorrências de doença inflamatória localizadas em algum fígado são ocorrências de hepatite. Assim, em qualquer situação, o termo à esquerda pode ser substituído pela expressão à direita, sem qualquer perda de significado (SCHULZ et al., 2009, p. 37-38).

Como pode ser observado, essa relação não ocorre da mesma maneira que acontece nos tesouros. Na declaração formal de equivalência, não há qualquer conhecimento implícito e ela será sempre verdadeira em toda e qualquer situação.

De maneira semelhante, ocorre equivalência entre classes, propriedades e indivíduos

EquivalentClasses <ontA:Person ontB:Human>;
EquivalentObjectProperties <ontA:hasChild
ontB:child>; EquivalentDataProperties <ontA:hasAge
ontB:age> (ISOTANI; BITTENCOURT, 2014, s.p.).

Esse exemplo foi declarado usando linguagem OWL e os autores esclarecem que esse tipo de equivalência é feito, em geral, para a integração e o reúso de ontologias (ontA e ontB).

Nos tesouros, usualmente, as relações hierárquicas são representadas por *Broader Term* (BT) e *Narrower Term* (NT), incluindo as relações gênero-espécie, todo-parte e de instâncias, com semântica ainda considerada restrita. Porém, no trabalho de Maculan (2015), cada um dos tipos de relação foi explicitado, aumentando a expressividade semântica das relações, como descrito na seção do estudo de caso. Com isso, todas as relações hierárquicas, refinadas, poderiam ser reutilizadas na modelagem de uma ontologia no mesmo domínio, ainda que seus elementos e partes devam ser representados com maior complexidade. Cada entidade (classe, subclasse, propriedade, indivíduo) na ontologia é tratada de maneira isolada, situação que não ocorre na construção de tesouros em geral, mas aconteceu no trabalho de Maculan (2015). Cuidado também deve ser tomado para a eliminação de poli-hierarquias nas ontologias, pois não são permitidas, atendendo ao princípio da não herança dupla (KLESS et al., em 2012; SANAA et al., 2013). Ademais, ressalta-se o já mencionado sobre os rótulos das classes de maior abstração, assim como das suas subclasses, que devem ser levadas para o nível da semântica formal e não ser em linguagem natural.

As relações associativas, em geral, são representadas nos tesouros por *Related Term* (RT), podendo incluir uma série de diferentes tipos de associações entre conceitos, o que é considerado com semântica ainda mais restrita do que nas relações hierárquicas. Do ponto de vista da recuperação de informações, esse elemento do tesouro permite a expansão de buscas pelo usuário. Entretanto, como já mencionado, a recuperação de documentos não

é um papel típico das ontologias, ainda que haja, segundo Schulz et al. (2009), alguns trabalhos nesse sentido (MÖLLER et al., 2004; GIUNCHIGLIA; DUTTA; MALTESE, 2013). Como no trabalho de Maculan (2015) as relações associativas foram explicitadas, considera-se que elas podem ser insumos para o mapeamento de relações para a criação de ontologias, sobretudo as de domínio.

Schulz et al. (2009) alertam, entretanto, para o fato de que, nas ontologias formais, não seria possível, utilizando o mesmo exemplo da entidade hepatite ($\langle \text{Hepatite} \equiv \text{Doença_Inflamatória} \sqcap \exists \text{tem_localização.Fígado} \rangle$), apresentado anteriormente, afirmar que, algumas vezes, a hepatite vem acompanhada de um sintoma, a febre ($\langle \text{Inflamação} \sqcap \exists \text{tem_localização.Fígado} \sqcap \exists \text{normalmente_tem_sintoma.Febre} \rangle$). Para os autores, a máquina não conseguiria interpretar, “por exemplo, que a hepatite tem o sintoma febre na maioria dos casos (mas não em todos)” (idem, p. 38), pois, quando se atribui uma afirmativa para uma entidade, há de se admitir que o sintoma da febre irá ocorrer em todos os casos. Não podendo ser assim, as ontologias não podem utilizar esse tipo de associação, coisa que é possível nos tesouros. Portanto, ao reutilizar conhecimento estruturado dos tesouros na criação de ontologias, esse tipo de situação sempre deverá ser analisado, mesmo que as relações estejam explicitadas, conforme o trabalho de Maculan (2015). Assim, nas ontologias, é preciso verificar a existência de interpretações ambíguas, por meio de declarações definitórias e axiomas expressos em linguagem formal.

Outro ponto que se considera interessante na construção de hierarquias e relações associativas em ontologias, proposto por Maculan (2015), é a determinação da estrutura dos quatro papéis *Qualia*: Constitutivo, Formal, Télico e Agentivo (PUSTEJOVSKY, 1995; ZAVAGLIA, 2003a; 2003b) para as diferentes entidades. Na construção de ontologias, os papéis *Qualia* podem ser utilizados para tratar as extensões do significado do termo (ou conceito) e das proposições verbais (relações semânticas), caracterizando-os (léxicos e verbos)

em um nível semântico, pois é interessante para a organização e representação de conhecimento e, também, para a interpretação desse significado (lexical e verbal) por sistemas computacionais.

Segundo Zavaglia (2003b, p. 2), a estrutura *Qualia* “é capaz de suprir o vocabulário básico para expressar aspectos diferentes do significado lexical (*wordmeaning*)”, apresentando o exemplo para a unidade lexical “pudim”:

Quadro 5 – Significados de uma unidade léxica a partir de papéis *Qualia*

Pudim	
<i>Qualia</i> =	CONST = ingredientes
	FORMAL = substância
	TELIC = comer
	AGENT = fazer

Fonte: Zavaglia (2003b, p. 3).

Pelo exemplo, notam-se as dimensões de significado que a unidade lexical “pudim” possui, expressas pelos quatro papéis *Qualia*, que se referem às relações envolvidas no contexto em que a unidade lexical é utilizada. A autora aponta que esse tipo de especificação de papéis foi utilizado para a desambiguação semântica entre unidades lexicais homônimas (ZAVAGLIA, 2003a), pois a caracterização das informações nos quatro tipos de papéis *Qualia* permite interpretar, também pela máquina, em qual contexto a unidade lexical está inserida.

Zavaglia (2003a; 2003b) também investigou sobre os tipos de relações semânticas a partir da determinação dos seus papéis dentro da estrutura *Qualia*, conforme exemplos no quadro 6.

É possível notar que a atribuição de papéis *Qualia* às proposições verbais (quadro 6) permite determinar o valor semântico das diferentes expressões verbais, auxiliando na desambiguação semântica das relações. Zavaglia (2003a) assegura que determinar os papéis *Qualia* às relações semânticas dá maior expressividade a elas, sendo possível estabelecer um padrão, principalmente, para relações de hiperonímia (todo) e hiponímia (partes), identificado pelo papel *Qualia* Constitutivo.

Quadro 6 – Relações Semânticas conforme a estrutura *Qualia*

PAPEL DA ESTRUTURA QUALIA	TIPOS DE RELAÇÕES SEMÂNTICAS
FORMAL	<é_um>; <é_um_sinônimo>; <é_um_antônimo>
CONSTITUTIVO	<é_um_membro_de>; <contém>; <quantifica>; <vive_em>; <atividade_constitutiva>; <está_em>; <tem_como_cor>; <tem_como_membro>; <feito_de>; <produzido_por>; <é_parte_de>; <propriedade_de>; <medido_por>
TÉLICO	<é_uma_atividade_de>; <objeto_da_atividade>; <é_a_habilidade_de>; <usado_para>; <usado_por>; <destinado_a>; <usado_contra>
AGENTIVO	<experiência_agentiva>; <resultado_de>; <origem>; <derivado_de>; <produzido_por>

Fonte: Zavaglia (2003b, p.3; 2003a, p.92-93).

CONSIDERAÇÕES FINAIS

Este estudo teve como proposta explorar a questão do reuso, na construção de ontologias, de conhecimento estruturado existente nos tesouros, em especial, das relações semânticas estabelecidas entre seus termos e conceitos. As análises das possíveis contribuições foram realizadas tendo como parâmetros a literatura sobre o tema, a norma ISO 25964, Parte 1 e 2 (2011; 2013) e os resultados apresentados por Maculan (2015). As apreciações realizadas neste artigo não são cabais, mas reflexões iniciais das autoras, e visam fornecer insumos para maior entendimento sobre as convergências e as divergências entre tesouros e ontologias, assunto de interesse nos estudos em representação do conhecimento na ciência da informação.

A partir das análises realizadas, considerou-se que as relações estabelecidas na estrutura de tesouros, quando explicitadas como no trabalho de Maculan (2015), em certo ponto podem contribuir com a determinação de relações de equivalência na construção de ontologias, desde que tudo que estiver relacionado com as entidades em equivalência puder ser, necessariamente, verdadeiro, em qualquer ocasião. Assim, percebeu-se que há similaridades na atribuição de relacionamentos no refinamento semântico descrito no trabalho de Maculan (2015) e na atribuição de relacionamentos na ontologia, porém, não uma correlação perfeita.

Ressalta-se que, como a ontologia estabelece hierarquias a partir da lógica descritiva, é possível determinar particularidades sobre uma propriedade

geral dos indivíduos e fazer generalizações que valem para um indivíduo. Isso é possível porque, na ontologia, o que ocorre são sentenças declarativas (verdadeira ou falsa) para o domínio, que sejam argumentos dedutivamente válidos (se as premissas são verdadeiras, a conclusão também é verdadeira). Porém, esse tipo de inferência, a partir da representação estabelecida na estrutura dos tesouros, não é possível que ocorra. Nos tesouros, as hierarquias podem ser construídas de maneira mais informal, podendo incluir, em uma mesma classe, por exemplo, diferentes indicadores de facetas (rótulo nodal), que não têm a mesma função de descritores preferidos, mas que compõem a parte sistemática do tesouro, indicando a base lógica do tipo de subdivisão que está sendo representado. Esse tipo de elemento pode dar pistas para a criação de ontologias, contudo, as ontologias não podem ser empregadas de forma indiscriminada.

Considerando as premissas de Guarino (1998), por exemplo, um dos componentes de uma ontologia são os conceitos (classes), representados por entidades concretas, reais ou fictícias, descrevendo funções, tarefas ou processos de raciocínio (GÓMEZ-PÉREZ, 1999). Para o autor, nessa abordagem (não realista) o compromisso primordial nem sempre está com a realidade, mas com a aplicação em si (necessidades e contexto). Assim, a contribuição da estrutura do tesouro para reuso na criação de ontologias também será dependente da abordagem adotada pelo modelador.

Verificou-se, ainda, que as diferenças nos níveis de granularidade representacional (subdivisão de entidades; relações hierárquicas) do conhecimento no tesouro, para ser reutilizadas na criação de ontologias, irão depender do tipo de aplicação, nível de formalidade e propósito, podendo exigir maior detalhamento e especificação. Nas ontologias, a granularidade não é um atributo de indivíduos, mas do próprio sistema, e se refere ao número de objetos representados nele.

Outro ponto que se considerou importante é a atribuição de papéis a partir da estrutura dos papéis *Qualia*, tanto na construção de tesouros quanto no desenvolvimento de ontologias. Se o tesouro for criado com refinamento das relações semânticas em sua estrutura, assim como com a atribuição de papéis semânticos, tal como ocorreu no trabalho de Zavaglia (2003a; 2003b) e de Maculan (2015), é esperado que as ontologias possam fazer, de maneira mais proveitosa, o reúso das proposições utilizadas nas relações semânticas que foram estabelecidas.

REFERÊNCIAS

AGANETTE, E.C. *Representação do conhecimento biomédico: uma investigação sobre a teoria da terminologia e a teoria da ontologia aplicada no domínio do sangue humano*. 2015.

238 f. Tese (Doutorado em Ciência da Informação)- Programa de Pós-Graduação em Ciência da Informação, Escola de Ciência da Informação, Belo Horizonte, ECI/UFMG, 2015.

ALMEIDA, M.B. *Um modelo baseado em ontologias para representação da memória organizacional*. 2006. 316 f. Tese (Doutorado em Ciência da Informação)- Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

_____. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 242-258, jul./set. 2014.

_____. BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, v. 32, n. 3, p. 7-20, set./dez. 2003.

_____. MENDONÇA, F. M.; AGANETTE, E. C. 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., 29 de outubro a 1 de novembro de 2013, Florianópolis. *Anais...* Florianópolis: ENANCIB, 2013.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. *Ibersid*, v. 2, p. 199-209, 2008.

BORCHERT, D. M. (Ed.). *Encyclopedia of Philosophy*. 2nd ed. New York: Thomson/Gale, Macmillan Reference, 2006. Disponível em: <<http://www.lambsound.com/Downloads/philosophy/Volume%2001.pdf>>. Acesso em: 22 jul. 2017.

BORST, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. 1997. 227 f. Tese (Doutorado)- Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.

BOSERUP, E. *The conditions of agricultural growth: the economics of agrarian change under population pressure*. London: Allen e Unwin, 1965.

CAMPOS, M. L. M. et al. O uso de tesouro como base terminológica para a elaboração de ontologias de domínio: uma experiência com o domínio do Folclore e Cultura Popular. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. *Anais...* São Paulo: USP, 2008.

CARVALHEIRA, L. C. C. *Método semi-automático de construção de ontologias parciais de domínio com base em textos*. 2007. 143 f. Dissertação (Mestrado em Engenharia de Sistemas Digitais)- Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Computação e Sistemas Digitais, São Paulo, USP, 2007.

- CAVALCANTI, C.R. *Indexação e tesouro: metodologia e técnicas*. Brasília: ABDE, 1978.
- DAHLBERG, I. Teoria da classificação, ontem e hoje. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1972, Rio de Janeiro. *Anais...* Brasília: IBICT/ABDE, 1979. P. 352-370, V. 1.
- _____. Teoria do conceito. *Ciência da Informação*, v. 7, n. 2, p. 101-07, 1978.
- DODEBEI, V. L. D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto, 2002.
- FERREIRA, A. B. H. *Novo dicionário Aurélio da Língua Portuguesa*. 4. ed. Curitiba: Positivo, 2009.
- FOX, C. J. *Information and misinformation: an investigation of the notions of information, misinformation, informing, and misinforming*. Westport: Greenwood, 1983.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v, 59, n. 1, p. 7-18, 2003.
- GIUNCHIGLIA, F.; DUTTA, B.; MALTESE, V. *From knowledge organization to knowledge representation*. Trento, Italy: Università di Trento, 2013.
- GOMES, R. M. *Desambiguação de sentido de palavras dirigida por técnicas de agrupamento sob o enfoque da mineração de textos*. 2009. 118 f. Dissertação (Mestrado em Engenharia Elétrica)- Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.
- GÓMEZ-PÉREZ, A. G. Ontological engineering: a state of the art: expert update: knowledge based systems and applied Artificial Intelligence. *British Computer Society*, v. 3-2, p. 33-43, 1999.
- GRUBER, T. R. *What is an ontology?*. 1992. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 22 jul. 2013.
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, v. 43, n. 5/6, p. 625-640, 1995.
- GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGIES IN INFORMATION SYSTEMS, 1., 1998, Trento, Italy, *Proceedings...* Trento, Italy: FOIS, IOS Press, 1998. P. 1-13.
- HOVY, E. Comparing sets of semantic relations in ontologies. In: GREEN, R.; BEAN, C. A.; MYAENG, S. H. (Ed.). *The semantics of relationships: an interdisciplinary perspective*. Norwell, MA, USA: Springer, 2002. Cap. 6, p. 91-110. V. 3.
- INTERNATIONAL STANDARD ORGANIZATION - ISO. *ISO 25964-1: thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval*. Genève, 2011.
- _____. *ISO 25964-2: thesauri and interoperability with other vocabularies: part 2: interoperability with other vocabularies*. Genève, 2013.
- ISOTANI, S.; BITTENCOURT, I. I. *Dados abertos conectados*. São Paulo: Centro de Estudos sobre Tecnologia Web, 2014.
- KLESS, D. et al. A Method of re-engineering a thesaurus into an ontology. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS; INTERNATIONAL CONFERENCE – FOIS, 7., 2012, Amsterdam. *Proceedings...* Amsterdam: IOS Press, 2012. P. 133-146.
- LAUSER, B. et al. From Agrovoc to the agricultural ontology service: concept server an OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. *Proceedings...* México: DCMI, 2006.
- MACULAN, B. C. M. S. *Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do Thesagro*. 2015. 345 f. Tese (Doutorado em Ciência da Informação)- Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, ECI/UFMG, 2015.
- MÖLLER, A. L. P.; VIOTTI, E. C. *Introdução à linguística II: princípios de análise*. São Paulo: Contexto, 2003.
- MÖLLER, H. et al. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *Plos Biology*, v. 2, n. 11, p. 1984-1998, 2004.
- MURCHO, D. *O lugar da lógica na Filosofia*. Lisboa: Plátano Edições Técnicas, 2003.
- NOY, N. F.; HAFNER, C. D. The state of the art in ontology design: a survey and comparative review. *AI Magazine*, v. 36, n. 3, p. 53-74, 1997.
- _____. MCGUINNESS, D. L. *Ontology development 101: a guide to creating your first ontology*. Stanford, CA: Stanford University, 2000.
- ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA - UNESCO. *Guidelines for the establishment and development of monolingual thesauri for information retrieval*. Paris, 1971.
- PUSTEJOVSKY, J. *The generative lexicon*. Cambridge, MA: MIT Press, 1995.
- SANAA, M. et al. A methodological approach for converting thesaurus to domain ontology: application to tourism. *International Journal of Engineering and Innovative Technology (IJEIT)*, v. 3, n. 5, Nov. 2013.
- SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. *Ciência da Informação*, v. 40, n. 2, p. 301-311, maio/ago. 2011.
- SCHULZ, S. et al. Strengths and limitations of formal ontologies in the biomedical domain. *Reciis: Electronic Journal of Communication, Information and Innovation in Health*, v. 3, n. 1, p. 31-45, 2009.

- SILVA, D. L.; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, n. 3, p. 60-75, set./dez. 2008.
- SOERGEL, D. *Indexing languages and thesauri: construction and maintenance*. Los Angeles, CA: Wiley, 1974.
- _____. *Functions of a thesaurus / classification / ontological knowledge base*. Maryland: College of Library and Information Services, University of Maryland, 1997.
- _____. et al. Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, v. 4, n. 4, 2004.
- SVENONIUS, E. *The intellectual foundations of information organization*. Cambridge: The MIT Press, 2000.
- VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997.
- WAND, Y.; STOREY, V. C.; WEBER, R. An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions*, v. 24, n. 4, p. 494-528, Dec. 1999.
- ZAVAGLIA, C. A homonímia no português: tratamento semântico segundo a estrutura *Qualia* de pustejovsky com vistas a implementações computacionais. *Alfa*, v. 47, n. 2, p. 77-99, 2003a.
- _____. Base de conhecimento léxico-ontológico para o Português do Brasil: uma proposta de modelo. In: WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA, 1., 12 de outubro de 2003, São Carlos – SP. *Anais...* São Carlos/SP: USP, 2003b. P. 1-9.