
A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais

The contribution of thesauri in the construction of ontologies as tools for the organization and retrieval of information in digital environments

**Vera Regina Casari BOCCATO (1), Rogério Aparecido Sá RAMALHO (2)
y Mariângela Spotti Lopes FUJITA (3)**

(1) Universidade Federal de São Carlos, UFSCar, Departamento de Ciência de Informação, Rodovia Washington Luiz, Km 235, Campus Universitário, CEP 13565905, São Carlos, SP, Brasil, Caixa-Postal: 676, vbocato@ufscar.br. (2) Faculdades de Dracena-Unifadra, Rua Bahia, 322, MetrÓpole, CEP 17900-000, Dracena, SP, Brasil, rogerio@ramalho.eti.br (3) Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Departamento de Ciência de Informação, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP 17525900, Marília, SP, Brasil, Caixa-Postal: 421, fujita@marilia.unesp.br

Resumen

A partir da interdisciplinaridade existente entre la Ciencia de la Información y la Ciencia de la Computación, y con el objetivo de demostrar la interrelación existente entre los tesouros y las ontologías, se discuten los principios básicos que orientan el desarrollo de las ontologías y los principales aspectos teóricos y prácticos que afectan a la construcción de tesouros. Se concluye que los tesouros pueden contribuir significativamente al desarrollo de las ontologías, constituyendo sustratos teóricos por medio del reaprovechamiento de sus estructuras lógico-semánticas.

Palavras-chave: Tesouros. Ontologías. Organización de la información. Recuperación de la información. Ambiente digital.

1. Introdução

O tratamento documentário teve um avanço significativo nos últimos anos em decorrência da intensificação da utilização das Tecnologias da Informação e Comunicação (TICs), favorecendo a automatização dos processos de desenvolvimento e administração de bibliotecas e unidades de informação, possibilitando a diminuição dos custos de manutenção, agilizando processos operacionais e viabilizando melhores formas de acesso às informações de interesse dos usuários.

Observa-se, também, uma crescente demanda quanto ao desenvolvimento de projetos e estudos relacionados à organização de informações em ambientes digitais, tendo como

Abstract

From the interdisciplinary connections between the Information Science and Computer Science; and, in order to demonstrate the interrelation between thesauri and ontologies, this research comprises a discussion of the literature about the basic principles that guide the development of ontologies and of the main theoretical aspects identified in the construction of thesauri that can be used to develop computer ontologies. It was concluded that thesauri can contribute to the development of ontologies both with theoretical approaches and by (re)using their hierarchical, equivalence and associative relations.

Keywords: Thesauri. Ontology. Information organization. Information retrieval. Digital environment.

objetivos comuns possibilitar melhorias qualitativas nos processos de organização e recuperação de recursos informacionais, garantindo respostas mais consistentes em relação às solicitações de buscas realizadas pelos usuários.

Dentro desse cenário, uma nova gama de possibilidades vem sendo incorporada aos processos de gestão de recursos informacionais e, em meio de tal desenvolvimento, o termo ontologia tem sido cada vez mais utilizado para designar uma nova e instigante categoria de instrumentos de organização e representação de informações, de modo que é possível identificar aproximações entre as novas abordagens computacionais e os instrumentos tradicionais de representação e organização de informações, já utilizados no âm-

bito da área de Ciência da Informação, evidenciando uma forte tendência de reaproveitamento de tais instrumentos para a elaboração de ontologias.

Na concepção das linguagens documentárias o tesouro é o instrumento que possui maior familiaridade e relacionamento com as ontologias por serem linguagens de estruturas combinatórias, de caráter especializado, constituídos por termos providos de suas relações semânticas que possibilitam a representação temática do conteúdo de um documento, bem como sua posterior recuperação.

Considerando-se a interdisciplinaridade existente entre a Ciência da Informação e a Ciência da Computação e com o objetivo de demonstrar a inter-relação existente entre tesouros e ontologias, realiza-se a discussão dos princípios básicos que norteiam o desenvolvimento de ontologias e dos principais aspectos teóricos e metodológicos identificados na construção de tesouros. Espera-se que os resultados dessa discussão possam ser utilizados na elaboração de ontologias computacionais, favorecendo um reaproveitamento dos aportes teóricos já sedimentados na área de Ciência da Informação para o desenvolvimento de novas abordagens tecnológicas que favoreçam uma melhor organização e recuperação das informações disponíveis em ambientes digitais.

2. Os tesouros como instrumento de organização e recuperação da informação documentária

A concepção inicial dos tesouros, no contexto de vocabulários controlados sistematizados, é atribuída a Peter Mark Roget com a elaboração, em 1852, de um dicionário analógico intitulado *Thesaurus of English Words and Phrases (Roget's Thesaurus)*. As palavras foram organizadas pelo campo das idéias que expressavam, apresentadas em seis categorias conceituais, a saber: relações abstratas, espaço, matéria, intelecto, volição e afeições, tendo sido subdivididas em classes e, sucessivamente, por seções, subseções e conceitos isolados (Dodebei, 2002).

Observa-se, nesse momento, um prenúncio dos tesouros na era moderna, como instrumentos de representação e recuperação da informação documentária, nessa proposta do *Roget's Thesaurus*, desenvolvido no campo da língua inglesa, como fonte de pesquisa de um conceito, sem necessariamente se conhecer o termo que o designa, com o intuito de

auxiliar na formulação das idéias expressas pelos autores.

A partir desse cenário, e decorridos mais de cem anos, assistiu-se a um processo evolutivo de teorias e métodos para construção de tesouros, caracterizado pelas correntes norte-americana e européia.

A corrente norte-americana revitalizou os princípios de construção apresentados no *Roget's Thesaurus* e partiu dos cabeçalhos de assuntos para as representações documentárias por meio de unitermos, sistema desenvolvido em 1951 por Mortimer Taubé, demarcando a adoção de um novo modelo de sistema pré-coordenado para um sistema pós-coordenado. A corrente européia se destacou, principalmente, com a criação do *Classification Research Group* – CRG, com a finalidade primeira no desenvolvimento de pesquisas voltadas para a classificação facetada, de Ranganathan e seguidas por Vickery, Foskett, Aitchison, Farradane, Austin, Mills, entre outros, influenciando uma geração de pesquisadores e profissionais no âmbito da representação e recuperação da informação.

A “Teoria Analítica do Conceito voltada para o Referente” de Ingetraut Dahlberg, apresenta-se também como fundamento importante na construção de tesouros conceituais (Campos; Gomes, 2006). Dahlberg (1978) define conceito como sendo uma unidade de conhecimento que compreende afirmações verdadeiras sobre um referente representado numa forma verbal. Toda afirmação (termo) sobre um referente (objeto, ação) possui um elemento do respectivo conceito, que se identifica como característica (propriedade) desse conceito. Os conceitos são constituídos por inúmeras características, nas quais características iguais suscitam relacionamentos entre conceitos, que analisados e interpretados permitem a sistematização em categorias e subcategorias.

Notoriamente, dentro desse cenário, tanto a Teoria da Classificação Facetada, como a “Teoria Analítica do Conceito voltada para o Referente”, tornam-se contribuintes na construção de tesouros conceituais, fornecendo subsídios na compreensão do que seria conceito, bem como no estabelecimento e na sistematização de termos e de suas relações lógico-semânticas.

A norma internacional ANSI/NISO Z39.19-2005 apresenta diretrizes para a elaboração, formato e manutenção de tesouros, complementadas neste estudo por elementos norteadores definidos por Gomes e Campos (2008), na concepção de uma metodologia de construção de tesouros, a saber:

- planejamento: delimitação do domínio; classificação (método de categorização); levantamento das fontes de informação; forma de apresentação (alfabética/ sistemática); período de atualização; seleção do software de construção; divulgação e manutenção do tesouro;
- levantamento dos termos: escolha e forma de apresentação (termos simples e/ou compostos);
- organização dos termos/conceitos (aplicação de fichas terminológicas);
- apresentação final (alfabética/sistemática: definição da forma gráfica aplicada a ordem sistemática; impresso e/ou eletrônico);
- interoperabilidade: definição de necessidades para a ocorrência da interoperabilidade entre o sistema e o tesouro (referentes ao processo de indexação e busca da informação);
- definição de critérios para testes e avaliação do tesouro.

Ressalta-se que o processo de construção de tesouros requer conhecimentos teórico-metodológicos da equipe envolvida (profissionais bibliotecários, analistas de sistemas, especialistas de áreas, entre outros), além de infra-estrutura material e financeira para o desenvolvimento desse instrumento de organização e recuperação da informação documentária.

Isso posto, e de acordo com a ANSI/NISO Z39.19-2005, os tesouros são vocabulários controlados organizados em uma ordem conhecida e estruturada de modo que os vários relacionamentos entre os termos sejam identificados e exibidos claramente por meio de indicadores de relações normalizadas.

Nessa perspectiva, pode-se definir tesouros como sendo linguagens de estruturas combinatórias e pós-coordenadas, constituídas de termos – unidades lingüísticas provenientes da linguagem de especialidade, e da linguagem natural – denominados de descritores, providos de relações sintático-semânticas, referentes a domínios científicos especializados, possibilitando a representação temática do conteúdo de um documento, bem como a recuperação da informação.

Lancaster (2002) apresenta dois objetivos para a utilização de tesouros em sistemas de recuperação de informação: o primeiro é o de facilitar a representação consistente dos assuntos por parte dos indexadores e usuários

que o recuperam, evitando a dispersão dos elementos relacionados. Isso se consegue com o controle (agrupamento) dos sinônimos e quase-sinônimos, com a distinção dos homógrafos e com a reunião dos termos que possuem uma relação mais próxima entre si. O segundo objetivo é o de facilitar a realização de uma busca ampla sobre um assunto atrelando os termos com relações sintagmáticas ou paradigmáticas.

As relações sintagmáticas opõem-se às relações paradigmáticas referindo-se ao eixo horizontal de relações de sentido entre os signos lingüísticos da cadeia falada, verbal ou escrita enquanto que as relações paradigmáticas ocorrem no eixo vertical das relações virtuais entre os signos lingüísticos comutáveis, isto é, por associação (Tálamo, 1997).

Os tesouros se compõem de uma base léxica (descritores e não-descritores) definida e descrita sob a denominação de nota de escopo ou nota de aplicação e estruturada em relações hierárquicas (termos genéricos e específicos), não-hierárquicas (associativas – termos relacionados) e de equivalência (não-descritores – sinônimos ou quase-sinônimos).

Segundo Fujita (1998, p. , 109),

[...] as relações hierárquicas são elaboradas a partir de um termo que denomina uma categoria ou classes de termos. [...]. Isto significa que os termos de um tesouro são classificados segundo uma ordem hierárquica existente com a finalidade de oferecer uma visão geral do assunto.

A relação hierárquica corresponde a níveis de superordenação, representando uma categoria, classe ou o todo, e em nível de subordinação, se referindo a seus membros ou partes. Essas podem ser estabelecidas por relações genéricas, isto é, de gênero/espécie ou coisa/tipo, pelas relações partitivas, isto é, todo/parte e pelas relações de instância (Cintra et al., 2002). As relações não-hierárquicas ou associativas, “[...] indicam a ligação entre os termos que estão em campos semânticos distintos, porém próximos”. (Zavitoski, 2001, p. 40). Boccato (2005, p. 59) afirma que:

essas relações podem ocorrer em duas situações: dentro de uma hierarquia, isto é, os termos podem estar no mesmo nível de relacionamento (termos/elementos coordenados) ou podem se apresentar numa relação de dependência entre si. Neste último caso, essa dependência pode acontecer em níveis seqüenciais como: causa/efeito, coordenação, atividades complementares, entre outros(1).

As relações de equivalência correspondem à relação entre o termo preferido (descriptor) e o não-preferido (não-descriptor) “onde dois ou mais termos são considerados, para fins de indexação,

como referentes ao mesmo conceito". (Austin; Dale, 1993, p. 42). O não-descritor pode ser um sinônimo (termos com o mesmo significado) em relação ao termo principal (descritor) e um quase-sinônimo – termos geralmente possuidores de significados diferentes na linguagem natural, mas que, para fins de indexação, são considerados como sinônimos.

A estrutura de um tesauro baseia-se de maneira lógico-semântica. O conjunto das noções de um determinado domínio se apresenta na vertical (relações hierárquicas), as quais se agregam às unidades informacionais que se relacionam horizontalmente (relações não-hierárquicas). Assim, as relações hierárquicas se apresentam como relações lógicas entre os termos, e as não-hierárquicas se associam semanticamente. (Zavitoski, 2001). Segundo Lancaster (2002), a estrutura de um tesauro deverá consistir de duas partes: uma lista alfabética de termos e uma organização sistemática desses termos. Essas partes poderão ser separadas ou totalmente integradas. Quanto ao critério de coordenação (combinação), a linguagem se apresentará com termos pós-coordenados.

Atualmente verifica-se uma tendência de desenvolvimento de tesouros para serem utilizados a partir de ambientes digitais. Nesta perspectiva Arano (2005) apresenta uma compilação de alguns dos reflexos do ambiente digital na estrutura dos tesouros, pontuando quatro itens:

- o enriquecimento da funcionalidade da estrutura dos tesouros a partir da hipertextualidade, possibilitando o estabelecimento de hiperlinks entre os elementos estruturais e as diferentes partes do tesauro.
- a redução dos custos de atualização e manutenção. Devido à crescente informatização dos processos de construção de tesouros e o progressivo abandono do suporte papel para a publicação destas ferramentas, viabilizando uma redução dos custos.
- a integração do usuário nos processos de criação, gestão e otimização dos tesouros, por meio de testes de viabilidade e uso de técnicas ajustadas ao usuário. Isso permite elaborar ferramentas que levem em conta os requerimentos dos usuários, e descartar sua construção como simples estruturas teóricas.
- a possibilidade de aplicar medidas de reutilização e interoperabilidade no momento de planejar e construir os tesouros. Possibilitando assim o aproveitamento e enriqueci-

mento da informação conceitual e lingüística que é gerada para outros recursos.

Todavia, como ressalta Pedraza-Jiménez et al. (2007), a Web é um cenário informacional heterogêneo completamente distinto do que se presencia no universo das disciplinas vinculadas no âmbito da documentação. Deste modo, por mais que os tesouros sejam aprimorados a partir das necessidades dos ambientes digitais, torna-se evidente que novos instrumentos tecnológicos concebidos inicialmente em contextos computacionais apresentem contribuições mais significativas para a organização e recuperação de informações em ambientes digitais, destacando-se as ontologias.

3. As ontologias como instrumento de organização e recuperação da informação em ambientes digitais

A palavra ontologia tem origem no grego *ontos* (ser) e *logos* (palavra), e apesar do estudo do *ontos* originar-se com Aristóteles e Platão, a utilização do termo Ontologia para designar um ramo da Filosofia é muito mais recente, tendo sido introduzido na transição da Idade Média para a Idade Moderna, na escolástica, por volta dos séculos XVII e XVIII. Segundo Welty e Guarino (2001), o termo foi cunhado na área de Filosofia em 1613 por Rudolf Goclenius e aparentemente de forma independente por Jacob Lorhard.

Nos últimos anos, verifica-se uma demanda cada vez maior de pesquisas relacionadas à utilização de ontologias no âmbito da representação da informação. A origem da apropriação de tal termo remete ao início da década de 1990, quando pesquisas na área de Ciência da Computação, mais especificamente na subárea de Inteligência Artificial (IA), impulsionaram o desenvolvimento de projetos relacionados à criação e organização de bases de conhecimento computacionais.

Guarino (1998, tradução nossa), apresenta uma interessante distinção quanto à utilização do termo ontologia para designar instrumentos de representação da informação, e o sentido que tal termo abarca no âmbito filosófico:

No sentido do filosófico, nós podemos referir a uma ontologia como um sistema particular de categorias que consideram uma certa visão do mundo. Como tal, este sistema não depende de uma linguagem particular: ontologia de Aristóteles é sempre a mesma, independentemente da linguagem utilizada para descrevê-la. Por outro lado, em seu uso mais predominante na IA, uma ontologia é referida como um artefato de engenharia, constituído por um vocabulário específico utilizado para descrever uma certa realidade e um conjunto de pressupostos explícitos re-

lacionados com o significado pretendido para as palavras do vocabulário.

No âmbito da representação da informação a definição mais freqüentemente identificada na literatura é a de Gruber (1993, p. 1, tradução nossa), que define uma ontologia como: “uma especificação explícita de uma conceitualização”, considerando que o termo ‘explícita’ significa que um objeto de nível simbólico deve ser expresso formalmente e de maneira clara, e uma ‘conceitualização’ será composta por objetos, conceitos e as relações existentes em um determinado domínio.

Apesar de tal definição ser a mais referenciada na literatura relacionada às ontologias, no contexto da representação da informação, Guarino e Giaretta (1995) apontam problemas quanto à noção de conceitualização utilizada por Gruber (1993), devido ao fato de considerar uma conceitualização como um conjunto de relações extensíveis (*extensional relations*) descrevendo um estado particular das coisas, enquanto que comumente a noção que se tem de conceitualização é de algo intencional, como uma grade que deve ser preenchida com os vários estados das coisas.

Guarino (1998, p. 5, tradução nossa), preocupando-se em estender a definição apresentada por Gruber (1993), define ontologia como: “uma teoria lógica para relacionar o significado pretendido de um vocabulário formal, seu comprometimento com uma conceitualização particular do mundo”.

Assim, torna-se conveniente estabelecer uma definição formal de conceitualização, que pode ser apresentada a partir de uma tripla ordenada $C = \langle D, E, \rangle$, onde “D” representa um domínio, “E” o conjunto máximo de estados relacionais desse domínio e “ ” o conjunto de relações conceituais sob o espaço de domínio $\langle D, E \rangle$. De acordo com tais considerações, uma ontologia é uma teoria lógica cujo modelo restringe uma conceitualização particular, sem especificar exatamente qual, ou em outras palavras, pode-se definir como uma caracterização axiomática do significado de um vocabulário lógico, a qual tem o compromisso apenas com a consistência em um determinado domínio, e não com a completude.

Ramalho (2006, p. 97), define uma ontologia como:

Um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de infe-

rências automáticas por programas computacionais.

Tal definição destaca que as ontologias quando utilizadas em sistemas de recuperação não possuem “propriedades transcendentais”, sendo utilizadas para favorecer uma melhor representação das informações, descrevendo formalmente os conteúdos informacionais a partir de linguagens que possam ser processadas computacionalmente, permitindo a realização de buscas mais consistentes e a realização de inferências automáticas, dentro de um domínio previamente determinado e restrito.

Assim como os tesouros, as ontologias também são compostas a partir de um conjunto de termos estruturados hierarquicamente, formando, desta maneira, uma estrutura de classes e subclasses. As classes de uma ontologia representam as entidades do “mundo real”, que são agrupadas e categorizadas de acordo com suas similaridades, levando-se em consideração um domínio concreto. Tais entidades podem representar coisas físicas ou conceituais, desde objetos inanimados até teorias científicas ou correntes teóricas, de modo que uma classe pode possuir diversas subclasses, que podem organizar as entidades a partir de características mais específicas.

As entidades descritas em uma ontologia são denominadas como instâncias e consistem de representações de objetos ou indivíduos pertencentes a uma classe, ou subclasse. Cada instância possui determinadas propriedades, ou atributos, que são representadas a partir de pares atributo/valor que descrevem as características mais relevantes de acordo com a classe a qual pertencem, o domínio representado e o enfoque da ontologia.

Segundo Ramalho (2006), as relações são um dos principais componentes que formam a estrutura básica das ontologias, pois são estes que possibilitam descrever formalmente os relacionamentos existentes entre as instâncias, classes, subclasses e suas respectivas propriedades, a partir de restrições lógicas que possam ser processadas de forma automatizada, permitindo inclusive relacionamentos baseados na Lógica de Segunda Ordem.

Complementando essa definição, as exposições de Jasper e Uschold (1999, p. 2) destacam a necessidade de explicação e relacionamento entre os termos, aproximando-se das estruturas de construção de linguagens documentárias, onde:

Uma ontologia pode possuir uma variedade de formas, mas necessariamente incluirá um vocabulário de termos, e alguma especificação de seus significados. Isto inclui definições e uma indicação de como conceitos estão inter-relacionados o que impõe

uma estrutura no domínio e restringe as possíveis interpretações dos termos.

As finalidades das ontologias, segundo Guzman Luna et al. (2006) podem ser assim definidas:

- compartilhar a compreensão da estrutura da informação entre pessoas ou agentes construtores de softwares;
- permitir a reutilização do conhecimento pertencente a um domínio;
- permitir a representação explícita dos conceitos de um domínio;
- separar o conhecimento de um domínio, do conhecimento que se pode denominar de operacional;
- possibilitar a análise de um domínio.

Nesse contexto, conforme apresenta Sales (2006, p. 49),

É importante destacar aqui que, na área de Ciência da Informação, os estudos não se dão no âmbito do desenvolvimento de Ontologias e sim no âmbito da construção de linguagens documentárias verbais e notacionais. Acredita-se, porém, que possam também ser aplicáveis às ontologias, já que estas, assim como outros tipos de linguagens de indexação, podem ser consideradas como instrumento de representação e recuperação de informação.

Em sua pesquisa Sales (2006), apresentada também dois critérios de classificação de ontologias: 1) por sua natureza, representadas pelas ontologias de tarefa e ontologias de domínio; e 2) por seu grau de formalismo, sendo as ontologias altamente informais, ontologias semi-informais, ontologias semi-formais, ontologias formais e ontologias rigorosamente formais a seguir (Figura 1).

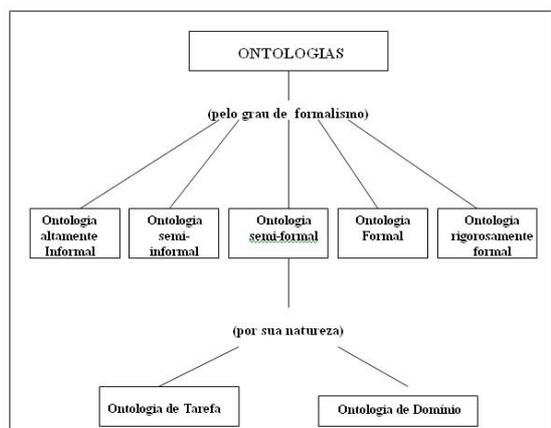


Figura 1. Classificação das ontologias
[Fonte: Sales, 2006, p. 30].

Nessa perspectiva, observa-se atualmente inúmeras pesquisas relacionadas ao desenvolvimento de ontologias de tarefa e de domínio relacionadas com a área de Ciência da Informação, apontadas como linhas de desenvolvimento capazes de fornecer instrumentos que permitam otimizar os processos de recuperação de informações no ambiente Web, podendo-se destacar os projetos: JeromeDL, MarcOnt, SWED e POIS.

Conforme Kruk et al. (2005) apresentam no artigo sugestivamente intitulado como JeromeDL - *Reconnecting Digital Libraries and the Semantic Web*, o projeto JeromeDL(2) consiste de uma biblioteca digital de código aberto baseada nas principais tecnologias presentes no projeto Web Semântica, permitindo a descrição de recursos informacionais a partir da linguagem computacional RDF e a realização de buscas semânticas baseadas em ontologias.

Conforme afirmam Synak e Kruk (2005), o projeto MarcOnt(3) possui como objetivo criar uma ontologia capaz de tornar-se um padrão de representação de informações para bibliotecas digitais, possibilitando a descrição dos aspectos semânticos dos conteúdos e favorecendo a integração de bibliotecas.

Quanto ao SWED (4) (*The Semantic Web Environmental Directory*), é parte do projeto da SWAD-Europe (5) (*Semantic Web Advanced Development – Europe*), e consiste em um repositório temático sobre informações ambientais, que tem como objetivo facilitar a disseminação de informações de organizações e projetos ambientais desenvolvidas no Reino Unido.

Segundo Almeida et al. (2005), o projeto POIS (*Portuguese Ontology in Information Science*) está sendo desenvolvido a partir de uma iniciativa interinstitucional, entre a Escola de Ciência da Informação da Universidade Federal de Minas Gerais – Brasil (ECI-UFMG) e o curso de Ciência da Informação da Pontifícia Universidade Católica de Minas Gerais – Brasil (CI-PUCMINAS), tendo como objetivo a construção de uma ontologia para Ciência da Informação, em português, para uso da comunidade acadêmica e para divulgação da área de pesquisa.

Entre os projetos propostos pela SWAD-Europe, pode-se destacar também o desenvolvimento da SWAD-E *Thesaurus Activity* (6) uma iniciativa que tem como objetivo apresentar tecnologias que possibilitem expressar, de maneira formal, a estrutura básica e o conteúdo de tesouros e vocabulários controlados, favorecendo a construção de ontologias e sua utilização em *Knowledge Organization Systems* (KOS), Sistemas de Organização do Conhecimento.

4. A contribuição dos tesouros na construção de ontologias

A classificação é um processo fundamental da natureza humana. O homem, em todas as suas atividades diárias, está acostumado a classificar as coisas. Tesouros e ontologias são artefatos que objetivam estabelecer um vocabulário compartilhado que permita classificar e representar informações. (Breitman, 2005).

Garcia Jiménez (2004) destaca que apesar das similaridades existentes entre tesouros e ontologias, ambos os formatos de representação possuem conseqüências teóricas e práticas distintas, e dessa maneira, é difícil afirmar que os tesouros podem ser entendidos como tipos de ontologias, não apenas por suas diferentes trajetórias históricas, como também por suas diferentes vinculações operativas e teóricas.

A construção de tesouros dispõe de princípios provenientes de campos científicos como a Terminologia e de diretrizes estabelecidas por normas internacionais, norteadores para a sua elaboração, possibilitando a qualidade e a precisão na determinação dos termos/descriptores representativos de conceitos de um determinado domínio científico.

Quanto às ontologias, sua origem remete a subárea de Inteligência Artificial, a partir de projetos técnicos relacionados à criação e organização de bases de conhecimento computacionais, de modo que seu desenvolvimento baseia-se principalmente em linguagens computacionais e não existem diretrizes ou normas de desenvolvimento consolidadas.

Conforme apresentado anteriormente e de acordo com a ANSI/NISO Z39.19-2005, os tesouros podem ser representados de forma sistemática por meio de uma apresentação estruturada de conceitos ou, de forma alfabética, representando, para cada descritor, todas as suas relações com os outros descritores. Na Figura 2, a seguir, são apresentados tipos de relações presentes em Tesouros.

Apesar dos tesouros permitirem a construção de uma estrutura flexível de relacionamentos entre conceitos, que estejam interligados em uma estrutura lógico-semântica, possuem também certas limitações quanto às possibilidades de exploração de determinados conteúdos informacionais, uma vez que não têm suficientemente expressividade que permita representar relações semânticas mais ricas entre os conceitos.

Relação	Código utilizado	Descrição
Relação Hierárquica	TG/TE	Descritores Gerais e Específicos da hierarquia
Relação Hierárquica Genérica	TGG / TEG	Descritores Gerais e Específicos da hierarquia, segundo a relação entre uma classe ou Gênero e seus membros ou espécies
Relação Hierárquica Partitiva	TGP/TEP	Descritores Gerais e Específicos da hierarquia, segundo a relação Partitiva: todo/parte.
Relação Hierárquica de Instância	TGI/TEI	Descritores Gerais e Específicos, segundo a relação existente entre uma categoria geral de coisas ou eventos, expressas por um nome comum, e uma Instância individual dessa categoria.
Relação Associativa	TR	Descritores aparentados, segundo relações simétricas de causa/efeito; coordenação, atividades complementares, entre outras.
Relação de Equivalência	UP/USE	Relação ao nível da mesma sinonímia lingüística: descritor e o não-descritor, possuidores do mesmo conceito.

Figura 2 – Tipos de relações estabelecidas nos Tesouros (Fonte: Elaborado pelos autores).

Conforme destaca Ramalho (2006), as ontologias possibilitam descrever níveis de relacionamentos mais avançados e que podem ser processados de forma automatizada por computadores, permitindo ao usuário rotular os relacionamentos.

Assim, com o intuito de exemplificar os tipos de relacionamentos existentes em uma ontologia e que não podem ser expressos por meio de um tesouro, pode-se considerar o diagrama que representa parte de uma ontologia de um domínio acadêmico, conforme apresentado na figura 3.

Considerando-se a figura 3, é possível identificar alguns exemplos de situações que requerem a utilização de ontologias, pois apresentam necessidades que não podem ser satisfeitas completamente a partir de tesouros, apesar de suas relações não-hierárquicas (associativas), tais como:

- relações transitivas entre conceitos que estejam em hierarquias distintas: verifica-se que apesar dos conceitos “Universidade” e “Professor” pertencerem a diferentes categorias, e não possuírem um relacionamento direto pré-estabelecido possuem um relacionamento indireto na ontologia, que não poderia ser explorado rigorosamente a partir de tesouros;

- relações inversas: em uma ontologia qualquer um dos conceitos pode representar um ponto de entrada, de modo que a partir das relações inversas observa-se que se um Professor “e especialista” em um determinado Assunto, automaticamente é possível inferir que determinado assunto “e_estudado_por” um Professor;
- inferências complexas relacionando vários conceitos: supondo-se a necessidade de responder questionamentos do tipo: que unidade contém especialistas em um assunto específico “X”? Verifica-se que partir da ontologia apresentada na figura 3, seria possível um sistema computacional inferir que, se uma determinada Unidade “U” contém um Departamento “D”, o Departamento “D” possui um Professor “P”, e o Professor “P” é especialista no assunto “X”, logo a unidade “U” contém ao menos um especialista no assunto “X”.

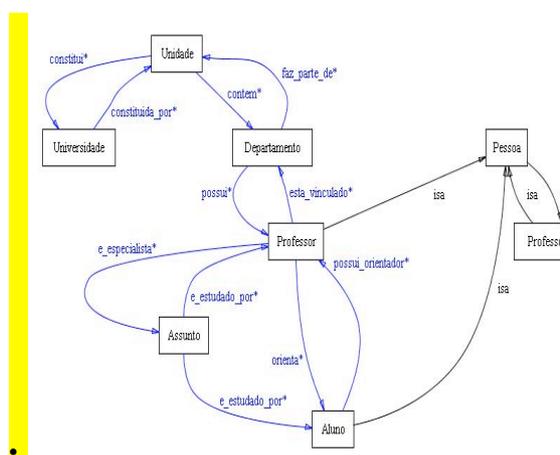


Figura 3. Diagrama de parte de ontologia de um domínio acadêmico

Nesta perspectiva, Noy e McGuinness (2001) propuseram uma metodologia simplificada para a criação de ontologias denominada de Método 101 (7), na qual as autoras utilizaram de suas experiências no desenvolvimento de ontologias para sistematizar resumidamente o processo de construção de ontologias, a partir dos seguintes passos:

- determinar o domínio e escopo da ontologia;
- considerar o reuso de outras ontologias;
- enumerar os termos importantes da ontologia;
- definir classes e a hierarquia de classes;
- definir as propriedades das classes;

- definir os valores das propriedades;
- criar instâncias.

No entanto, deve-se ressaltar que o processo de desenvolvimento de ontologias não é uma tarefa linear, pois muitas interações, reajustes e refinamentos são necessários até se atingir o objetivo esperado. Conforme afirma Breitman (2005), não existe uma maneira correta de se modelar um domínio, sempre existem várias possibilidades, devendo-se levar em consideração o tipo de aplicação desejada, possíveis extensões e o enfoque utilizado. Esta mesma autora ilustrou o método 101, proposto Noy e McGuinness (2001), conforme apresentado na Figura 4:

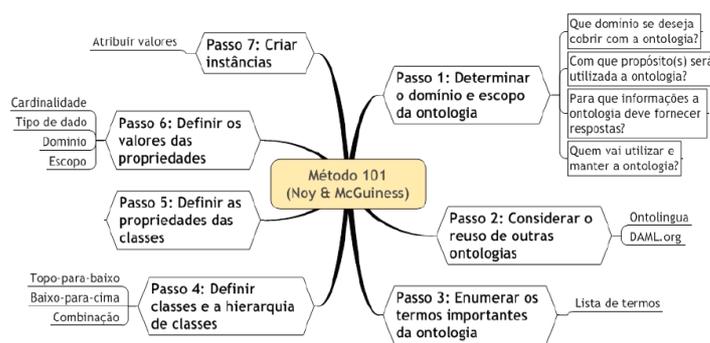


Figura 4. Método 101 para o desenvolvimento de ontologias (Fonte: Breitman, 2005, p. 76).

Levando-se em consideração este método para a construção de ontologias, pode-se verificar que alguns dos passos/processos mencionados muito se aproximam daqueles utilizados para a elaboração de tesouros, sendo que muitas das dificuldades são identificadas nos passos iniciais, que abordam aspectos como: determinação do conhecimento parcial de um domínio; definição do escopo/objetivo; dificuldades na definição dos conceitos; e necessidade de verificação da consistência dos termos utilizados.

Conforme apresentado anteriormente e ratificado por Campos (2004), a área de Ciência da Informação possui bases teóricas bem fundamentadas ligadas à representação de sistemas de conceitos, como a teoria da classificação facetada de S. R. Ranganathan e a teoria do conceito de I. Dahlberg, tal embasamento teórico já sedimentado há décadas favorece a uma melhor determinação dos conceitos e suas relações.

Segundo Pedraza-Jiménez et al. (2007, p. 574),

as ontologias podem se considerar linguagens documentárias com distintos níveis de estrutura, porém a diferença do tesouro tradicional está na sua elaboração a partir de uma sintaxe compreensível para os computadores.

Nesse contexto, há de se destacar a grande tradição dos tesouros frente ao estado ainda incipiente das ontologias, de modo que se pode afirmar que os tesouros há décadas têm evoluído a partir dos aportes teóricos da área de Ciência da Informação. Tal fato justifica as vantagens de se desenvolver ontologias a partir de tesouros, pois eles podem servir como substratos teóricos para a construção de ontologias.

Além disso, analisando o método 101, proposto Noy e McGuinness (2001) para a construção de ontologias, observa-se que um tesouro já estruturado eliminaria a necessidade de uma série de etapas iniciais, de modo que os termos presentes no tesouro seriam mantidos, buscando-se apenas explorar mais detalhadamente os níveis de relacionamentos que as ontologias são capazes de descrever e os tesouros não.

Comprovando as contribuições que os tesouros podem propiciar na construção de ontologias Guzman Luna et al. (2006) apresentaram em seu trabalho uma proposta de migração de um tesouro documentário para uma ontologia no contexto da Web Semântica (8), visando à recuperação da informação.

O trabalho de conversão do tesouro do *Centro de Naciones Unidas sobre Asentamientos Humanos* para uma ontologia especializada contou com a utilização da linguagem SKOS-Core, permitindo a compatibilização de 1.263 descritores, partindo-se das seguintes modificações:

- representar cada termo em um único conceito;
- converter a hierarquia existente em uma hierarquia de conceitos agrupados por categorias de conceitos, partindo-se do geral para o específico ou do específico para o geral;
- relacionar os conceitos, adicionando-se os termos associativos e de equivalência.

Essas modificações permitiram uma adaptação do tesouro tradicional para as propriedades do SKOS-Core. Para a implementação da ontologia foi utilizado o software Protégé – versão 3.1, e a linguagem de representação OWL.

O trabalho foi concluído, temporariamente, com a conversão de 444 termos dos 1263 constituintes do tesouro documentário, apresentando a aplicação de tecnologias de Web Semântica para a compatibilização de tesau-

ros em ontologias, por meio da utilização do SKOS Core.

5. Considerações Finais

Este estudo possibilitou a sistematização de referenciais teóricos que proporcionaram elencar os elementos constituintes dos tesouros, norteadores para a modelagem das relações existentes entre as instâncias, classes, subclasses, e suas respectivas propriedades em ontologias, a saber: relações hierárquicas, relações associativas, relações de equivalência e notas de escopo.

A partir da análise da literatura apresentada verificou-se que tanto os tesouros quanto as ontologias atuam como instrumentos de organização e recuperação da informação, buscando favorecer a identificação das informações que realmente interessam ao usuário em um contexto adequado.

Verificou-se que, no âmbito das linguagens documentárias, o tesouro é um instrumento que possui maior proximidade com as ontologias, pois ambos são constituídos a partir de linguagens de estruturas combinatórias, operando em níveis semelhantes, isto é, epistemológicos, porém, devem ser observadas as divergências de seus propósitos iniciais: o primeiro, com a finalidade de representar conteúdos documentários e atuando como um veículo de comunicação entre o usuário e o sistema de informação e o segundo, de descrever formalmente recursos informacionais disponíveis em ambientes digitais, a partir da utilização de linguagens lógicas que possibilitem a realização de inferências automáticas por programas computacionais.

Assim, conclui-se que os tesouros podem contribuir significativamente para o desenvolvimento de ontologias, constituindo-se como substratos teóricos, por meio do reaproveitamento de suas estruturas lógico-semânticas.

Deste modo, tais ontologias, desenvolvidas a partir de tesouros, podem ser denominadas como “ontologias referenciais”, pois têm como principal objetivo possibilitar a descrição dos recursos informacionais, permitindo que sistemas computacionais possam criar referências automáticas a tais recursos, a partir de inferências automatizadas, facilitando assim sua posterior recuperação.

Notas

- (1) Níveis sequenciais: Atributiva; Disciplina ou campo de estudo/objetos ou fenômenos estudados; Processo ou operação/seu agente ou instrumento; Influência; Matéria-prima/produto; Coisa/aplicação; Ação/resultado da ação; Causalidade ou causa/conseqüência; Efeito/causa; Dependência causal; Atividade/agente; Atividade/propriedade; Ativi-

dades complementares; Opostos; Ação/seu paciente; Coisa ou atividade/suas propriedades ou agentes; Coisa/seu contra-agente; Atividade/produto; Pessoas ou coisas/suas origens; Associação implícita; Expressões sincategoremáticas/substantivos incluídos; Interfacetada, entre outras (Lima; Boccato et al., 1996, p. 180).

(2) <http://www.jeromedl.org>

(3) <http://www.marcont.org>

(4) <http://www.swed.org.uk>

(5) <http://www.w3.org/2001/sw/Europe/>

(6) <http://www.w3.org/2001/sw/Europe/reports/thes>

(7) A etiqueta 101 em inglês é o código utilizado para a primeira de uma série de disciplinas em uma universidade.

(8) "Web Semântica é uma extensão da Web atual, onde a informação possui um significado claro e bem definido, possibilitando uma melhor interação entre computadores e pessoas". (Berners-Lee; Hender; Lassila; 2001).

Referências

- Almeida, M. B.; et al. (2005). Uma iniciativa interinstitucional para construção de ontologia sobre Ciência da Informação: visão geral do projeto P. O.I.S. // *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*. ISSN 1518-2924. 19:1º sem. (2005). http://www.encontros-bibli.ufsc.br/Edicao_19/4_Almeida.pdf. (2006-03-04).
- ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO, 2005. ISBN 1-880124-65-3. <http://www.niso.org/standards/resources/Z39-19-2005.pdf> (2008-05-17).
- Arano, S. (2005). Los tesauros y las ontologías en la Biblioteconomía y La Documentación. // *Hipertext.net*. ISSN 1695-5498. 3 (May. 2005). <http://www.hipertext.net>. (2008-02-10).
- Austin, D.; Dale, P. (1993) Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngües. Tradução de Bianca Amaro de Melo. Brasília: IBICT, 1993. ISBN 85-7013-022-8.
- Berners-Lee, T.; Hender, J.; Lassila, O. (2001). The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. // *Scientific American*. ISSN 0037-7333. (May 2001). <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&pageNumber=2&catID=2>. (2007-06-05).
- Boccato, V. R. C. (2005). Avaliação de linguagem documentária em Fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal. Marília: Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005. Dissertação de Mestrado.
- Breitman, K. (2005). Web semântica: a internet do Futuro. Rio de Janeiro: LTC, 2005. ISBN 85-21614-66-7.
- Campos, M. L. A. (2004). Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*. ISSN 0100-1965. 33:1 (jan./abr. 2004) 22-32. <http://www.scielo.br>. (2008-04-28).
- Campos, M. L. A.; Gomes, H. G. (2006). Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. // *Perspectivas em Ciência da Informação*. ISSN 1413-9936. 11:3 (Set./Dez. 2006) 348-359. <http://www.scielo.br/pdf/pci/v11n3/a05v11n3.pdf>. (2008-04-30).
- Cintra, A. M. M.; et al. (2002). Para entender as linguagens documentárias. 2. ed. rev. e ampl. São Paulo: Polis, 2002. ISBN 85-72280-12-X.
- Dahlberg, I. (1978). A referent-oriented, analytical concept theory for INTERCONCEPT. // *International Classification*. ISSN 0340-0050. 5:3 (1978) 142-150.
- Dodebei, V. L. D. (2002). Tesouro: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro; Interciência, 2002. ISBN 85-7193-064-3.
- Fujita, M. S. L. (1998). A estrutura de categoria do tesouro: modelos de elaboração. // *Cadernos da F.F.C.* ISSN 1415-868X. 7:1/2 (1998) 107-120.
- García Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. // *Anales de Documentación*. ISSN 1697-7904. 7 (2004) 79-95. <http://www.um.es/fccd/anales/ad07/ad0706.pdf> (2008-01-28).
- Gomes, H. E.; Campos, M. L. de A. (coords) (2008). Elaboração de tesouro documentário: tutorial. <http://www.conexaorio.com/bitl/> (2008-03-31).
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. Denter: Kluwer Academic Publishers, 1993.
- Guarino, N. (1998). Some ontological principles for designing upper level lexical resources. // *International Conference on Language Resources and Evolution - LREC 98'*, 1., 1998, Granada. Proceedings. Granada: [s.n], 1998. <http://www.loa-cnr.it/Papers/LREC98.pdf> (2007-09-03).
- Guarino, N.; Giaretta, P. (1995). Ontologies and knowledge bases: towards a terminological clarification. // Mars, N. J. I. (ed). Towards very large knowledge bases: knowledge building and knowledge sharing. Amsterdam: IOS Press, 1995. ISBN 978-9051992175. 25-32.
- Guzman Luna; Torres Pardo; Lopez Garcia (2006). Desarrollo de una ontología en el contexto de la web semántica a partir de un tesouro documental tradicional. // *Revista Interamericana de Bibliotecología*. ISSN 0120-0976. 29:2 (Jul./Dic. 2006) 79-95.
- Jasper R., Uschold, M. A. (1999). Framework for understanding and classifying ontology applications. // *International Joint Conference on Artificial Intelligence - IJCAI-99*, 1999, Stockholm. Proceedings... Stockholm: Professional Book Center, 1999. ISBN 0938075934.
- Kruk, S. R. et al. (2005). JeromeDL reconnecting digital libraries and the semantic Web. http://www.marcont.org/marcont/pdf/www2005_jeromedl.pdf. (2005-10-10).
- Lancaster, W. F. (2002). El control del vocabulario en la recuperación de información. Traducción: Alejandro de la Cueva Martin. València: Universitat de València, 2002. ISBN 84-37054-44-3.
- Lima, V. M. A.; Boccato, V. R. C. et al. (1996). Atualização da lista de assuntos USP: compatibilização de linguagens documentárias. // *Ciência da Informação*. ISSN 0100-1965. 25:2 (maio/ago. 1996) 177-181.
- Noy, N. F.; McGuinness, D. L. (2001). Ontology development 101: a guide to creating your first ontology. Stanford: Stanford University, 2001.
- Pedraza-Jiménez, R.; Codina, L.; Rovira, C.. (2007). Web semántica y ontologías en el procesamiento de la información documental. // *El profesional de la información*. ISSN 1386-6710. 16:6 (Nov./Dic. 2007) 569-578.
- Ramalho, R. A. S. (2006). Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito

- da Ciência da Informação. 2006. Marília: Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2006. Dissertação de Mestrado.
- Sales, L. F. (2006). Ontologias de domínio: estudo das relações conceituais e sua aplicação. 2006. Niterói: IBICT-UFF, 2006. Dissertação de Mestrado.
- Synark, M.; Kruk, S. R. (2005). MarcOnt initiative the ontology for the librarian world. // European Semantic Web Conference ESWC, 2., 2005, Heraklion, Grécia. Proceedings... Heraklion: 2005. ISBN 978-3540261247.
- Tálamo, M. de F. G. M. (1997). Linguagem documentária. São Paulo: APB, 1997.
- Welty, C., Guarino, N. (2001). Supporting ontological analysis of taxonomic relationships. // Data and Knowledge Engineering. ISSN 0169-023X. 39:1 (2001) 51-74.
- Zavitoski, M. T. (2001). Exploração do uso do tesauro como instrumento de recuperação da informação. São Paulo: Escola de Comunicações e Artes, Universidade de São Paulo, 2001. Dissertação de Mestrado.