

Diferenças entre tesouros e ontologias

Rodrigo de Sales

**Mestre em Ciência da Informação - UFSC.
Bacharel em Biblioteconomia - UFSC. Editor da
revista *Texto Digital*; Membro do Núcleo de
Pesquisa em Informática, Literatura e Lingüística
(NUPILL/UFSC).**

Lígia Café¹

**Doutora em Lingüística - Université Laval, U.L.,
Canadá. Professora do Departamento de Ciência
da Informação e Programa de Pós-Graduação em
Ciência da Informação da Universidade Federal de
Santa Catarina.**

Tesouros e ontologias são modelos de representação do conhecimento baseados no controle terminológico de domínios específicos. Este artigo relata parte de uma pesquisa de Mestrado em Ciência da Informação desenvolvida na Universidade Federal de Santa Catarina e foca as diferenças existentes entre aqueles dois modelos de representação. Foi empregado, no plano metodológico, o Método de Análise de Conteúdo. No plano teórico, a pesquisa foi subsidiada pela Teoria Comunicativa da Terminologia (TCT). A constatação de que ambos os modelos apresentam diferenças no tocante aos termos, conceitos e objetivos permite um entendimento mais profundo desses instrumentos, tão imprescindíveis à comunicação técnica e científica.

Palavras-Chave: *Tesouro; Ontologia; Representação do conhecimento; Teoria comunicativa da terminologia.*

Differences between thesauri and ontologies

Thesauri and ontologies are models of knowledge representation based on the terminological control of specific domains. This article describes a research project

1 Artigo baseado em Dissertação de Mestrado desenvolvida no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, com auxílio financeiro da CAPES.

developed at the Universidade Federal de Santa Catarina as part of a Masters degree in Information Science and it focusses on the current differences between these two models of representation. The Content Analysis Method was used as the methodology, and the theory was supported by the Communicative Terminology Theory. The discovery that both models exhibit differences regarding terms, concepts and objectives allows a deeper understanding of these tools which are considered as vital to technical and scientific communications.

Keywords: *Thesaurus; Ontology; Knowledge representation; Communicative theory of terminology*

Recebido em 14.04.2008 Aceito em 12.03.2009

1 Introdução

A comunicação científica e profissional se manifesta por meio de linguagens especializadas, codificadas e decodificadas por especialistas de uma determinada área do conhecimento ou de uma determinada área profissional. Essas linguagens especializadas, também chamadas de linguagens de especialidade, são constituídas por terminologias próprias, que por sua vez são responsáveis pela transmissão do conhecimento especializado. As terminologias também pertencem à linguagem natural (linguagem como um todo, usada para a comunicação e expressão humanas) utilizada por um sujeito-falante comum, pois, antes de um indivíduo ser um especialista em determinado assunto ele é um sujeito-falante natural. Melhor dizendo, a terminologia pertence a uma linguagem especializada, que por sua vez pertence à linguagem natural. Com isso, constata-se que uma língua de especialidade é uma sublinguagem pertencente à linguagem natural. Tal fato leva à consideração de que a comunicação técnico-científica nada mais é (de uma maneira geral) do que uma comunicação natural ocorrida dentro dos campos especializados.

Uma particularidade das linguagens utilizadas por especialistas é a impreterível precisão da informação comunicada. Para que tal precisão seja consolidada são necessários instrumentos que zelem pela não-ambigüidade do discurso especializado. As linguagens documentárias são instrumentos que visam a representação da informação com o objetivo de recuperar conteúdos informacionais em ambientes específicos.

É importante um esclarecimento a respeito de qual o sentido empregado para o termo 'representação do conhecimento' neste trabalho. Recorrendo à etimologia da palavra latina *repraesentare* (representar), têm-se como significados: "fazer presente", "apresentar novamente", "trazer à memória". Nesse sentido, é comum encontrar na literatura corrente o uso do termo 'representação do conhecimento' para nomear o processo de tornar presente, por meio de termos, o conteúdo de uma

informação. Porém, cabe frisar que o processo de tornar presente conteúdos informacionais é uma atividade que diz respeito à 'representação da informação', que por sua vez se instrumentaliza por meio de modelos de representação do conhecimento. Melhor dizendo, a representação do conhecimento é um processo mental (campo das idéias) responsável pela organização do conhecimento. Desta organização surgem os sistemas de representação do conhecimento, *Knowledge Organization Systems (KOS)*, também chamados de modelos de representação do conhecimento ou linguagens documentárias, que servem de ferramentas para a atividade de representação da informação. Com isso, afirma-se que os modelos de representação do conhecimento, como os tesouros e as ontologias, que representam uma parte do mundo das idéias (representação do conhecimento), são aplicados nas atividades de representação da informação.

Cabe ainda esclarecer que o termo representação do conhecimento é empregado distintamente pelas áreas da Ciência da Informação (CI) e Ciência da Computação (CC). Para a CI, o referido termo é utilizado, como exposto acima, para designar a organização mental dos conceitos, pois, para esta ciência, o conhecimento e sua organização dizem respeito aos processos desenvolvidos e compreendidos na mente humana. Já para a CC, a representação do conhecimento diz respeito, segundo Branchman e Levesque (2004), à área da Inteligência Artificial (IA), dedicada ao estudo das formas de representar e manipular simbolicamente o conhecimento de maneira automática, por meio de raciocínio de programas informáticos. De uma maneira geral, a representação do conhecimento para a CI reside no campo das idéias e é articulado pelos processos mentais, enquanto que para a CC reside no campo informático e é articulado pelos programas de raciocínio lógico que alimentam as bases de conhecimento.

No âmbito terminológico da representação do conhecimento, os tesouros e as ontologias são modelos principais abordados pela Ciência da Informação. O tesouro é uma linguagem documentária caracterizada pela complexidade existente no relacionamento entre os termos que comunicam o conhecimento especializado. A ontologia é um modelo de representação do conhecimento, por vezes empregado como linguagem documentária, que, a exemplo do tesouro, é utilizada para representar e recuperar informação por meio de estruturas conceituais. As ontologias possibilitam compartilhar uma visão de determinado campo de conhecimento, compartilhar uma forma de pensar de determinado assunto, proporcionando um mapa semântico e uma estrutura conceitual de um domínio específico por meio de um vocabulário comum.

O presente artigo relata parte de uma pesquisa de Mestrado, desenvolvida no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, que consiste, sobretudo, na identificação e na discussão de aspectos que aproximam e distanciam os fundamentos dos tesouros aos das ontologias. Neste artigo, somente as diferenças entre eles são abordadas. O intuito é ajudar a suprir uma carência da literatura correspondente, já que são raros os estudos que lançam mão de rigor teórico e metodológico para comparar esses dois

modelos de representação, fato que inibe uma definição de suas diferenças fundamentais.

Na esfera teórica, a pesquisa foi subsidiada pelos fundamentos da Teoria Comunicativa da Terminologia (TCT), criada por Maria Teresa Cabré da Universitat Pompeu Fabra (Barcelona). No quadro metodológico, a Análise de Conteúdo, definida por Laurence Bardin, conduziu todo o processo de coleta e tratamento das informações, bem como as interpretações dos resultados obtidos.

O texto está estruturado com algumas definições significativas correspondentes aos tesouros e às ontologias, com uma apresentação em linhas gerais das idéias da TCT, e com a descrição metodológica da pesquisa, além, obviamente, das afirmações conclusivas a respeito dos resultados encontrados na investigação.

2 Definições sobre tesouros e ontologias

Tesouros são vocabulários controlados formados por termos-descretores semanticamente relacionados, e atuam como instrumentos de controle terminológico. Os tesouros podem estar estruturados hierarquicamente (gênero-espécie e todo-parte) e associativamente (aproximação semântica), e são utilizados principalmente para indexar e recuperar informações por meio de seu conteúdo. Cabré (1993) afirma que os tesouros são 'recopilações' de termos relacionados semanticamente, que servem como ferramenta para organizar e recuperar informação.

Segundo as diretrizes para a construção de tesouros descritas no Padrão Norte-Americano,

a thesaurus is a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators that are employed reciprocally² (AMERICAN, 2003, p.?).

O Padrão ressalta ainda que os tesouros não são utilizados somente pelos especialistas da informação, no momento da indexação, mas também por usuários da informação, no momento da busca de documentos. Essa afirmação é enaltecida por Moreira (2003), que, além de concordar que o tesouro é o elo entre a linguagem utilizada pelos indexadores e pelos usuários, afirma que os termos e as relações dos termos contidos nos tesouros fazem deles instrumentos essenciais para que ambos (indexador e usuário) busquem o melhor termo (ou termos) em um sistema de informação.

² Um tesouro é um vocabulário controlado organizado em uma ordem preestabelecida e estruturado de modo a que os relacionamentos de equivalência, de homografia, de hierarquia, e de associação entre termos sejam indicados claramente e identificados por indicadores de relacionamento padronizados empregados reciprocamente (Tradução nossa).

As ontologias são aparatos desenvolvidos pela engenharia computacional, e consistem em especificações formais que visam descrever estruturas conceituais de domínios específicos. Para Gruber (1993a), ontologias são esquemas conceituais em sistemas de bancos de dados. Um esquema conceitual fornece uma descrição lógica de dados compartilhados, permitindo programas de aplicação e interoperabilidade entre os bancos de dados. Uma ontologia define o vocabulário usado para compor expressões complexas. O objetivo da ontologia é viabilizar um acordo no uso do vocabulário compartilhado de uma maneira coerente e consistente. A definição mais difundida para ontologia no âmbito da representação do conhecimento é a de Gruber (1993b), na qual o autor afirma que uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. No entendimento do autor, 'conceitualização' se refere a um modelo abstrato dos fenômenos no mundo, identificando os conceitos relevantes daqueles fenômenos. O termo 'formal' se refere ao fato de que a ontologia deve ser legível por máquina; 'explícito' se justifica porque os tipos de conceitos usados e suas restrições de uso são definidos explicitamente; e 'compartilhado' reflete que a ontologia deve capturar o conhecimento consensual aceito pelas comunidades.

Guarino e Giaretta (1995) afirmam ser problemático o fato de Gruber (1993b) considerar que a conceitualização reside no nível das relações extensionais, descrevendo um estado particular das coisas. Para eles, a conceitualização reside no nível das relações intensionais, e descreve os vários estados das coisas. Guarino e Giaretta (1995) ressaltam que ontologia é uma teoria lógica que fornece um relato explícito e parcial de uma conceitualização, e afirmam também que ontologia pode ser considerada sinônimo de conceitualização, ou seja, uma estrutura semântica intensional que codifica as regras implícitas, legitimando uma estrutura de uma parte da realidade. Guarino (1998) expõe que o termo 'ontologia' denota o resultado da atividade de análise conceitual, que modela um domínio, realizada por meio de metodologias padrão.

Ding e Foo (2001) afirmam ser ontologia uma estrutura de termos que possibilita o compartilhamento de informações de determinado domínio do conhecimento, sendo que domínio pode também ser entendido como uma tarefa específica.

Segundo Noy e McGuinness (2005), uma ontologia é uma descrição explícita e formal de: a) conceitos em um domínio de discurso; b) propriedades de cada conceito, descrevendo as características e atributos do conceito; e c) restrições sobre as propriedades.

Em artigo que visa esclarecer questões referentes à construção e ao compartilhamento de ontologias, Sowa (1999) afirma que

Ontology is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D . The types in the ontology represent the *predicates*, *word senses*, or *concept and relation types* of the language L when used to discuss topics in the domain D ³ (SOWA, 1999, p. 1-2).

As definições aqui apresentadas afirmam essencialmente que a ontologia proporciona um vocabulário formal e comum baseado em uma estrutura de conceitos específicos de um dado domínio.

As diferenças entre tesouros e ontologias não ficam muito evidentes com uma simples observação em suas definições. Por isso, buscaram-se, na Teoria Comunicativa da Terminologia, parâmetros para investigar mais detidamente o que há de comum entre esses instrumentos.

3 Teoria Comunicativa da Terminologia (TCT)

Dentre as teorias modernas da Terminologia que ancoram os estudos mais recentes de organização e representação do conhecimento, merecem destaque, sem pormenorizar as diferenças de abordagens, a Teoria Geral da Terminologia (TGT) de Eugen Wüster, a Teoria da Socioterminologia de François Gaudin e a Teoria Comunicativa da Terminologia (TCT) de Maria Teresa Cabré.

Definida no final da década de 1990, a TCT é uma teoria descritiva de base lingüística e perspectiva funcionalista, focada no caráter comunicativo do termo. Cabré (1999) definiu uma teoria generalizada, levando em consideração que a Terminologia é interdisciplinar (integrando aspectos da Lingüística, das Ciências Cognitivas e das Ciências Sociais) e transdisciplinar (atua em todas as disciplinas). Segundo a autora, a TCT não considera os termos como unidades isoladas que constituem seu próprio sistema, mas sim, considera-os como unidades que se incorporam no léxico de um falante, no momento em que este adquire o *know how* de especialista por meio da aprendizagem do conhecimento especializado.

Bem como toda teoria terminológica, a teoria de Cabré direciona sua luz às implicações que dizem respeito ao termo e ao conceito. Mas a lente concebida pela TCT permite visionar o termo de uma maneira diferente. Dando ênfase ao exame da estrutura e do funcionamento terminológico, e levando em conta o aspecto da variação, o termo é visto na teoria de Cabré como uma unidade denominativo-conceitual, como uma unidade de conhecimento. Assim, com base em orientações epistemológicas voltadas à dimensão comunicativa das línguas naturais, a TCT visualiza o termo como uma unidade de conhecimento, composta por uma forma e um

³ Ontologia é um catálogo dos tipos de coisas que se admite existir em um domínio do interesse D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D . Os tipos na ontologia representam os predicados, os sentidos da palavra [termo], ou os tipos de conceito e relação da língua L quando usados para discutir tópicos no domínio D (Tradução nossa).

conteúdo, sendo a forma a unidade lexical que denomina o conceito (conteúdo).

Com base nos fundamentos da TCT, foram extraídos os seguintes elementos de observação para a análise: a) o termo, considerando seu caráter de unidade de conhecimento pertencente à linguagem natural e as distintas funções deste no contexto discursivo. Considerando também sua característica pragmática inserida no discurso e sua simultaneidade quanto à forma e ao conteúdo; b) o conceito e seus diferentes tipos de relações, formadores da estrutura conceitual; e c) os objetivos (teóricos e práticos) atribuídos aos modelos em questão.

4 Metodologia

A pesquisa aqui relatada é qualitativa do ponto de vista da abordagem do problema, pois está pautada em análises e interpretações de conteúdos. Na ótica dos procedimentos técnicos, trata-se de uma pesquisa documental, que empregou técnicas da Análise de Conteúdo para o levantamento, tratamento e análise das informações.

Conforme a determinação do método de Análise de Conteúdo, definido por Bardin (2003), a análise foi dividida em três fases: a) Pré-análise; b) Exploração do material; e c) Resultados e interpretações;

4.1 Pré-análise

O conjunto documental da investigação foi constituído por relatórios de pesquisas da área Ciência da Informação e Ciência da Computação, localizados nas bases de dados da [Library and Information Science Abstracts](#) (LISA), da *Wilson Library Literature and Information Science Full Text*, e da Biblioteca Digital de Teses e Dissertações do IBICT (INSTITUTO..., 2007). Os documentos não disponíveis *on-line* foram solicitados pela comutação bibliográfica por meio do sistema COMUT. O período de abrangência foi de 1998 a 2007, e os idiomas foram delimitados em português, inglês e espanhol.

O Corpus de Análise foi construído seguindo as regras de exaustividade, representatividade, homogeneidade e pertinência, postuladas por Bardin (2003). Foi composto por 34 documentos, sendo 33 artigos técnico-científicos e uma Tese. Dos 34 documentos selecionados para o Corpus de Análise, 17 correspondem ao tema 'tesauro', sendo 16 artigos e a tese, 16 artigos são relativos ao tema 'ontologia', e um artigo cobre ambos os temas.

Acatando as orientações concebidas pela Análise de Conteúdo, foram determinadas as seguintes categorias de análise:

1. Termo: registra uma síntese contendo a perspectiva apresentada no documento com relação aos termos do modelo de representação do conhecimento em questão.
2. Conceito: registra uma síntese contendo a perspectiva apresentada no documento com relação aos conceitos e

estruturas conceituais do modelo de representação do conhecimento em questão.

3. Objetivo: registra uma síntese contendo a abordagem apresentada no documento com relação aos objetivos (teóricos e práticos) do modelo de representação do conhecimento em questão.

O registro dos elementos correspondentes às categorias focou os seguintes parâmetros baseados na TCT: a) Categoria Termo – as funções das unidades terminológicas, a relação forma-conteúdo do termo e os níveis de relacionamento entre os termos; b) Categoria Conceito – a relação com a designação do termo e a relação entre os próprios conceitos e; c) Categoria Objetivo – finalidades das linguagens documentárias em questão.

4.2 Exploração do Material

A exploração do material foi realizada por meio da técnica de fichamento de textos e auxiliada por uma base de dados criada no *Microsoft® Access 2003*. A leitura de cada documento foi devidamente fichada, com base nas categorias descritas acima, e registrada na referida base de dados. Para cada texto analisado foi elaborada uma ficha contendo registros relativos às unidades de registro Referência, Tema, Resumo e Observações, e às categorias Termo, Conceito e Objetivo. A FIG. 1 mostra a ficha de registro de um dos textos analisados.

REFERÊNCIA	CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. <i>Perspectivas em Ciência da Informação</i> , Belo Horizonte, v.11, n.3, p. 348-353, set./dez. 2006. Disponível em < http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&layout=abstract >. Acessado em: 23.mai.2007.
	TESAURO <input checked="" type="checkbox"/> ONTOLOGIA <input type="checkbox"/>
RESUMO	Faz uso da Teoria do Conceito e da Teoria de Classificação Facetada para apresentar uma metodologia para elaboração de tesouros fundamentada nas questões que envolvem o conceito e as categorias. Não considera a expressão tesouro terminológico como apropriada, pois afirma que o que se evidencia é o conteúdo conceitual de uma etiqueta linguística e não a palavra ou a expressão verbal. Assim, as bases para a elaboração do tesouro conceitual se encontram fundamentalmente nos princípios que enfatizam o processo de conceitualização e em sua ordenação sistemática.
TERMO	Signo verbal que designa um objeto/referente. Em nível de abstração: signo verbal que denota o conceito. Unidade lexical que designa um objeto e um conceito. À análise do objeto/referente se dá a partir de um determinado domínio. O contexto do tesouro é o domínio e não o discurso. Tesouro conceitual é um tesouro com base em conceitos: seu nome indica que cada termo denota um conceito, ou seja, uma unidade de conhecimento. Função: denominar o conceito. Relação com o conceito: designação de um conceito. Relação com o próprio termo: depende das características do conceito.
CONCEITO	O ponto de partida para estabelecer as relações conceituais e determinar a forma verbal mais adequada para representá-lo. O conceito é unidade de conhecimento, como propõe Dahlberg (1978), incluindo a definição como elemento essencial para a fixação do conceito. O tesouro conceitual reúne dois princípios fundamentais para sua elaboração: o conceito, como unidade de representação, e o uso de categorias, como base para organização de sistemas de conceitos. Relação com o termo: conceitualiza-lo, como o significado de uma palavra. Relação com o próprio conceito: se dá de acordo com as características do conceito (os predicados de um objeto/referente).
OBJETIVOS	Serve para auxiliar a indexação e a recuperação em um sistema de recuperação de informação.
OBSERVAÇÕES	Teorias utilizadas: Teoria do Conceito, Teoria da Classificação Facetada. Defende que primeiro se determina o conceito e depois o termo.

FIGURA1 – Interface da Base de Dados Criada para o Fichamento

Com o auxílio da base *Microsoft*® *Access*, após o fichamento do material analisado, foi possível gerar relatórios de acordo com a necessidade do analista. Foram gerados relatórios contendo a descrição de todas as informações registradas no processo de coleta de informações: Referência, Tema, Resumo, Termo, Conceito, Objetivo e Observações. E também um relatório contendo a descrição apenas das informações referentes às categorias de análise Termo, Conceito e Objetivo, com o fim de uma análise mais detida dos pontos de observação central.

4.3 Tratamento dos resultados e interpretações

Para direcionar a análise, foram determinadas variáveis de inferência. Essas variáveis foram escolhidas de acordo com a percepção, ocorrida durante todos os processos anteriores, sobretudo durante a leitura e o fichamento dos textos, de elementos recorrentes na grande maioria dos documentos analisados. Tais elementos foram identificados e considerados como variáveis de inferência inseridas nas categorias da análise, a saber: a) Categoria Termo – definição/função, tipos, relação entre termos e relação com os conceitos; b) Categoria Conceito –

definição/função, organização dos conceitos, relação entre conceitos e relação com os termos; e c) Categoria Objetivo – teórico (relativo à terminologia em geral) e prático (relativo aos sistemas de informação).

Com o intuito de facilitar a comparação dos aspectos que caracterizam os tesouros e as ontologias, foram elaborados um Quadro A, que descreve o conteúdo concernente aos documentos relativos aos tesouros, e um Quadro B, que descreve o conteúdo referente aos documentos relativos às ontologias. Ambos os quadros foram orientados pelas categorias de análise e variáveis de inferência. O quadro abaixo (QUADRO 1) ilustra o modelo dos quadros A e B da pesquisa, que não são expostos neste artigo devido sua grande extensão.

QUADRO 1- Modelo dos Quadros Utilizados para Análise de Conteúdo

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	CONTEÚDO
Termo	Definição/função	
	Tipos de termos	
	Relação entre termos	
	Relação com o conceito	
Conceito	Definição/função	
	Organização dos conceitos	
	Relação entre conceitos	
	Relação com o termo	
Objetivo	Teórico (relacionado com a terminologia em geral)	
	Prático (relacionado com os sistemas de informação)	

Fonte: Elaborado pelos autores.

Com as características dos tesouros e das ontologias devidamente registradas, o passo seguinte foi o cruzamento dos conteúdos dos Quadros A e B, com o fim de identificar as diferenças existentes entre ambos os modelos de representação do conhecimento.

4.3.1 As diferenças

No que diz respeito ao termo, as diferenças encontradas nos textos residem predominantemente na função exercida por eles. Os textos relativos às ontologias focam mais as funções desempenhadas pelos termos, sem se (pre)ocuparem em defini-los, ficando apenas notória a idéia de que um termo é uma etiqueta que se refere a um conceito. A literatura relativa aos tesouros atribui aos termos a função de evitar ou diminuir a flexibilidade da linguagem e descrever um conceito de maneira

unívoca em um sistema de informação, ao passo que a literatura relativa às ontologias atribui aos termos a função de definir formalmente coisas em um domínio de interesse e viabilizar a consulta a um sistema de informação fazendo uso de conceitos pré-estabelecidos por especialistas. As funções dos termos atribuídas pelos textos referentes ao tesouro são funções de caráter terminológico e conceitual, ao passo que os outros textos atribuem aos termos funções mais práticas em ambientes de aplicação especializados.

Na parte do corpus que representa os tesouros, foram encontrados nove tipos de termos não identificados nos textos correspondentes às ontologias: Termo Simples, Termo Composto, Termo Equivalente, Termo Preferido, Termo Proibido, Termo Relacionado, Termos Polissêmicos, Identificador e Termo Qualificado. No que diz respeito às ontologias, foi identificada apenas uma classificação de termo que se distancia dos tipos encontrados nos tesouros, a saber: a Entidade (termo que mostra uma substância).

Nos documentos correspondentes às ontologias são definidas apenas duas informações a respeito da relação entre os termos. A primeira afirma que a relação semântica está diretamente ligada à apresentação sintática em um discurso; a segunda expõe que o relacionamento entre os termos é realizado pelos especialistas, ou seja, dá-se por meio do consenso em um determinado domínio. Já a literatura referente aos tesouros apresenta três tipos de relações entre os termos (diferentes das ontologias): a) Equivalência – quando um termo apresenta uma relação de sinonímia com outro, e, neste caso, o termo adotado pelo tesouro (termo preferido), também conhecido como descritor, é determinado na elaboração do tesouro lançando mão da sigla UP (Usado Para). O termo preterido é marcado pela sigla USE (que o remete para o descritor correspondente); b) Associativo – apresenta relação semântica não hierárquica; c) Nota Explicativa - orientação que elucida o emprego de determinado termo, fornecendo informações como a definição do termo e sua relação com outros termos.

Com relação ao conceito, a literatura voltada aos tesouros o considera como o conjunto formado pelas características de um objeto, que por sua vez são sintetizadas por um termo, definição pautada na Teoria do Conceito. Para os estudos de ontologias, os conceitos são unidades de um vocabulário especializado que representam classes, entidades, atributos e processos. É possível identificar que, no âmbito dos tesouros, o conceito é abordado sob uma ótica mais teórica (abstrata), como uma unidade representante de um objeto. No âmbito das ontologias, embora o conceito também seja uma unidade representante de um objeto, o conceito é tratado sob uma ótica mais aplicada. Essa diferença de visões reflete a diferença de abordagens das áreas de conhecimento que cobrem os estudos aqui analisados. A área da Ciência da Informação, que predominantemente cobre os estudos de tesouros, objetiva uma investigação de cunho mais reflexivo, mapeando todo um campo teórico-conceitual, em busca de embasamentos e entendimentos teóricos e metodológicos referentes a aplicações passadas e futuras. A

área da Ciência da Computação, responsável pela maioria dos estudos de ontologias, não negligenciando as reflexões teóricas, mas sim priorizando a construção dos aparatos informáticos, centra suas investigações no desenvolvimento e na aplicação de seus produtos (nesse caso as ontologias). Isso explica a diferente maneira com que os artigos relativos aos tesouros abordam o assunto, se comparado com os textos relativos às ontologias. Com base na análise do corpus, ao imaginar um ciclo de desenvolvimento científico para este caso, é possível visualizar a pesquisa da Ciência da Informação alimentando e sendo alimentada pela pesquisa da Ciência da Computação, e vice-versa. Obviamente, ambas as pesquisas, além de se complementarem, perpassam por outras áreas (como a Lingüística, por exemplo) para fortalecerem este ciclo.

Os textos relativos ao tesouro não fazem uso da expressão 'organização de conceitos', ao passo que os textos relativos às ontologias, nesse aspecto, distinguem os conceitos concretos principais, que são aqueles que apresentam propriedades do domínio, bem como seus relacionamentos, dos conceitos abstratos, que são as características.

Os tesouros apresentam dois tipos de relacionamento entre conceitos que não foram identificados na literatura referente às ontologias: o relacionamento ontológico, que diz respeito à proximidade situacional dos elementos na realidade (contigüidade dos conceitos no espaço), e que é considerado a relação entre conceito e realidade; e o relacionamento de equivalência, que ocorre quando um conceito é representado por mais de uma forma. Enquanto os textos relativos às ontologias relatam que interligações entre conceitos mais refinados e conceitos mais periféricos formam as relações adicionais, os textos voltados aos tesouros apresentam relacionamentos dos tipos: descendência, instrumental, causa e efeito, benefício, prejuízo, material, aparência, processo e estado. Finalizando a questão dos tipos de relações possíveis entre os conceitos contidos em um tesouro, afirma-se que essa relação é determinada pelo uso que o domínio faz dos conceitos, além das características próprias do respectivo domínio. Por outro lado, na literatura das ontologias é mencionada a seguinte informação a respeito do relacionamento entre conceitos atrelados ao domínio: as relações conceituais, que se dão no nível intensional, são definidas em um espaço do domínio, e podem ser representadas em grupos de mundos possíveis (conjunto de coisas, estados e relações de coisas que são convencionalmente determinados como possíveis, mas que estão sob a égide de um conjunto de regras também determinado). Tal informação pode não caracterizar uma diferença, mas evidencia uma função que transcende as possibilidades de relacionamento conceitual dos tesouros.

Outra particularidade do relacionamento entre conceitos, possível somente nas ontologias, é a relação concebida por meio de axiomas, os quais definem a interpretação pretendida. Embora os textos referentes aos tesouros enumerem uma quantidade maior de tipos de relacionamentos entre conceitos, a flexibilidade do relacionamento por meio de axiomas, viabilizada pelo formalismo informático das ontologias,

proporciona maior dinamicidade no tangente ao relacionamento conceitual.

Ao que se refere aos objetivos teóricos, a análise de conteúdo identificou como meta dos tesouros auxiliar a inter-relação entre linguagem natural e linguagem artificial, fornecendo um sistema de símbolos lingüísticos para agrupar e relacionar informações de uma temática. Do lado das ontologias, foram identificados como objetivos fornecer um mapa semântico aos campos individuais e o relacionamento entre os campos, servindo como uma ferramenta que crie uma estrutura lógica, uma filosofia, uma classificação em um domínio. Enquanto os tesouros almejam orientar qual o termo mais adequado para representar um conceito, as ontologias visam esclarecer o significado pretendido de um vocabulário por meio de axiomas.

Com isso, fica evidente que, embora ambos tenham (em teoria) o objetivo de servir como uma ferramenta de referência para a representação de assuntos especializados, os tesouros estão voltados ao elo que une a linguagem do usuário com a linguagem utilizada pelos sistemas de informação, preocupando-se em conceder um sistema simbólico que esclareça a relação entre os termos e os conceitos. Já as ontologias transcendem esta meta de padronizar a linguagem utilizada na indexação e na recuperação da informação, propondo ser um mapa semântico, uma estrutura formal para um dado domínio, ou até mesmo servir como ferramenta capital para a elaboração de bases de conhecimento. Para Noy e McGuinness (2005), a linha que define onde termina uma ontologia e onde começa uma base de conhecimento é muito tênue. Segundo as autoras, uma ontologia, somada a um conjunto de instâncias individuais de classes, formam uma base de conhecimento. Na área da Ciência da Computação, o termo 'base de conhecimento' é definido como uma ferramenta de suporte para o atendimento a dúvidas, usada para definir bases de dados ou conhecimento acumulado sobre um determinado assunto. Pode ser utilizada na solução de problemas, por meio do uso de ferramentas de Inteligência Artificial (IA). De uma maneira geral, base de conhecimento pode ser considerada uma coleção de informação que opera com base em ontologias.

Partindo para o âmbito das aplicações dos modelos de representação do conhecimento, fica evidente que os objetivos dos tesouros são a padronização e a normalização terminológica das atividades de indexação e recuperação nos sistemas informacionais. Já as ontologias, devido ao seu formalismo informático, vão em busca de uma estrutura de conceitos com alto nível de dinamicidade no que diz respeito aos modelos de representação do conhecimento. Enquanto os tesouros pretendem servir como pontes que ligam as necessidades de informação aos sistemas de recuperação da informação, as ontologias pretendem ajudar a responder perguntas em um corpo de informação, não apenas relacionando os conceitos aos termos e os definindo, mas, também, esclarecendo-os e contextualizando-os em uma classificação, baseando-se nas disciplinas, nas línguas e nas culturas. Enquanto os tesouros se voltam à atividade de indexação baseada em linguagem natural, as

ontologias servem como uma espécie de dicionário que é usado tanto por humano quanto por base de conhecimento (máquina) para processar linguagem natural. As ontologias não visam à 'tradução' de linguagens naturais para linguagens especializadas e vice-versa, mas, sim, atuam no próprio processamento dessas linguagens. O uso da palavra 'dicionário' torna evidente que as ontologias vão além de propor uma estrutura conceitual por meio do relacionamento controlado de termos, pois os dicionários têm como característica apresentar definições de palavras. Os tesouros almejam ser um vocabulário oficial para a indexação e recuperação de documentos, deixando explícita sua função de controle terminológico para as respectivas atividades, ao passo que as ontologias visam a um entendimento comum e compartilhado de um determinado domínio, deixando claro que uma de suas funções é possibilitar que bases de conhecimento respondam perguntas solicitadas. Assim como os tesouros estão voltados para a normalização terminológica de um sistema de informação, as ontologias estão voltadas para a 'identificação e definição' dos 'conceitos relevantes' que caracterizam um domínio.

Nota-se que, assim como a literatura aponta, como objetivo dos tesouros, propor um conjunto estruturado de termos sob a base de um sistema de conceitos aptos a organizar conteúdos, auxiliando a representação desse conteúdo e evitando as ambigüidades lingüísticas, aponta também, como objetivo das ontologias, possibilitar por meio de aplicações lógicas a construção de modelos computacionais para um determinado domínio de aplicação. Embora isto não evidencie uma oposição direta com características próprias dos tesouros, denota mais uma vez que os recursos informáticos possibilitam que os objetivos das ontologias vão além daqueles almejados pelos tesouros.

Devido a esse fato, de as ontologias serem criadas e desenvolvidas no meio informático, são inúmeros os objetivos atribuídos a elas que transbordam a esfera de atuação dos tesouros. Dentre eles: a) interoperabilizar sistemas; b) proporcionar a interoperabilidade dos serviços ofertados na *web*, potencializando os níveis de serviços ofertados; ou seja, sustentar a *web* semântica; c) propiciar benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação); d) fornecer base conceitual para projetos de pesquisa e execução; e) melhorar a comunicação e o aprendizado, assistindo a escritores e leitores, fornecendo estruturas conceituais que sustentem o ensino; f) possibilitar serviços baseados em operacionalizações semânticas; g) ligar parte do conhecimento humano aos processamentos computacionais; h) descrever a semântica de um domínio de modo que seja compreensível por homens e máquinas; i) melhorar a consistência e o reuso da informação e o compartilhamento do conhecimento; j) recuperação da informação e raciocínio automático de um dado domínio; k) capturar significados e relações entre os significados; l) servir como um catálogo dos tipos de coisas que existem em um domínio D, da perspectiva de uma pessoa que usa uma língua L, com a finalidade de falar sobre D; m) representar os predicados, os sentidos das palavras ou os tipos de conceitos e relações da língua L, quando usados para discutir tópicos no

domínio D; n) na *web* semântica, serve para fornecer uma conceitualização parcial de um dado domínio de forma compartilhada entre usuários comuns e que seja formalmente definida por uma linguagem processada por máquinas; o) oferecer meios concisos e sistemáticos para definir a semântica dos recursos *web* e; p) proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação.

Fica subentendido que as ontologias transcendem a questão da simples representação de conteúdos documentais, mais frequentemente atribuída aos tesouros, para assumir um papel de ferramenta elementar para os sistemas de informação automatizados, para as bases de conhecimento e os serviços ofertados pela *web*, sobretudo no tocante à *web* semântica. Isso não significa que ontologias não cumpram o papel de representar conteúdos de informação, pelo contrário, significa que a representação da informação proporcionada por uma ontologia vai além das tradicionais descrições simbólicas de conteúdos de documentos.

As diferenças descritas acima, bem como as características que, embora não apresentem uma distinção direta, demonstram peculiaridades dos modelos, tornam evidente que o distanciamento entre tesouros e ontologias é mais bem esclarecido na esfera das aplicações, haja vista que as diferenças de ordem conceitual, muitas vezes, correspondem a pontos de vista. Claro está o fato de que a literatura que trata dos tesouros fortalece seus argumentos na questão do controle terminológico necessário para as atividades de indexação e recuperação de informação, e no elo entre o usuário e a informação. Por outro lado, o foco argumentativo da literatura das ontologias está na concepção de uma estrutura conceitual formal e compartilhada, que conceda uma visão de parte de mundos especializados.

5 Conclusões

A pesquisa relatada parcialmente neste artigo foi fruto de um processo de Análise de Conteúdo. Portanto, os resultados aqui alcançados são fruto de inferências extraídas dos documentos analisados, ou seja, são os resultados de uma interpretação controlada por variáveis julgadas como relevantes para este estudo. Torna-se, portanto, inevitável o esclarecimento de que esses resultados pertencem a uma investigação que deu vazão à subjetividade do analista. No entanto, essa subjetividade não significa uma falta de rigor científico quanto à análise do conteúdo dos documentos, mas, sim, representa que o objetivo aqui alcançado teve uma interferência 'controlada' do sujeito em relação ao objeto observado. Isso leva à consideração de que as características que distanciam tesouros e ontologias, identificadas neste estudo, não são últimas e acabadas, mas, sim, características significativas que possibilitam afirmações importantes ao estudo da diferença entre tesouros e ontologias.

A quantidade de características diferentes entre os modelos de representação do conhecimento por ora analisados reflete algo que já

poderia ter sido previsto pelo fato de ambos pertencerem genuinamente a áreas de conhecimento distintas (tesouro – disciplinas da Ciência da Informação; ontologia – disciplinas da Ciência da Computação). Porém, o emprego de ambos os modelos como linguagens documentárias para o universo da informação os tornam passíveis de análises como esta, que os colocam como modelos afins localizados em espaços similares.

Embora ambos os modelos de representação do conhecimento sejam utilizados em situações similares e com funções por vezes concomitantes, seria uma redução afirmar que ontologias são linguagens documentárias que visam o controle terminológico nas atividades de indexação e recuperação da informação, bem como seria, também, uma incoerência afirmar que tesouros são especificações comuns e compartilhadas de uma conceitualização. A simples afirmação de que ontologias são linguagens documentárias já se evidencia como uma incoerência. Embora as ontologias possam cumprir papéis desempenhados pelas linguagens documentárias, já foi enfatizado neste trabalho que linguagens documentárias são instrumentos que auxiliam o processo de classificação, indexação e recuperação de documentos por assunto, uma definição muito aquém das possibilidades de execução de uma ontologia.

As diferenças apontadas neste estudo não são indicadores de vantagens e desvantagens de um modelo em relação ao outro, mas sim índices que apontam para a evolução dos modelos de representação do conhecimento que, ao longo da história, caminharam da classificação filosófica e bibliográfica de assuntos para a construção de uma estrutura conceitual, uma estrutura de pensamento comum. Servidas pelos aparatos informáticos, e toda a potencialidade que eles concedem, as ontologias são hoje o exemplo de sofisticação no que tange à representação do conhecimento.

Avançando um pouco mais o assunto, e trazendo para esta reflexão as idéias funcionalistas da Teoria Comunicativa da Terminologia (TCT), é possível ousar a seguinte afirmação: a ontologia é a potencial concretização dos postulados da TCT. Tal afirmação se constrói com base no seguinte raciocínio: se a TCT persegue uma Terminologia baseada em uma linguagem natural (embora contida de especificidade) e real (efetivamente usada nos ambientes especializados), e se as ontologias concebem uma especificação comum e compartilhada de uma conceitualização - o que permite a especialistas compartilharem do mesmo vocabulário -, tem-se que as ontologias são capazes de formalizar (não no sentido de padronizar, mas sim no sentido de legitimar uma estrutura) uma terminologia efetivamente utilizada e compartilhada entre especialistas. Obviamente que a ousadia de tal afirmação surge mais como uma 'hipótese' a ser investigada do que uma 'tese' a ser defendida. Porém, sua presença neste trabalho se justifica no fato de ser uma reflexão resultante da análise que buscou estudar os modelos de representação do conhecimento com base na visão comunicativa da terminologia.

A contribuição principal da presente investigação está na identificação, devidamente comentada, das diferenças existentes entre os

tesouros e as ontologias, assunto pouco tratado na literatura corrente. Tal contribuição traz ao universo da Ciência da Informação, dedicado, dentre outros esforços, ao fluxo informacional nos ambientes científicos e profissionais, alicerces para melhor compreender essas ferramentas de representação do conhecimento que, em se tratando de controle terminológico, são as principais da área da Ciência da Informação.

Referências

AMERICAN NATIONAL STANDARDS INSTITUTE. *Guidelines for the Construction, format and management of monolingual thesauri*. Bethesda, USA: American National Standards Institute, 2003. (ANSI Z39.19-2003). Disponível em: <<http://www.niso.org/standards/index.html>>. Acesso em: 30 ago. 2005.

BARDIN, L. *L'analyse du contenu*. 7ème. Paris: PUF, 2003. 296 p. (Le Psychologue, 69).

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBCT. *Biblioteca Digital de Teses e Dissertações*. Disponível em: <<http://bdtd.ibict.br/>>. Acesso em: 25 abr. 2007.

BRANCHMAN, R. J; LEVESQUE, H. J. *Knowledge representation and reasoning*. San Francisco: Morgan Kaufmann, 2004. 381p. CABRÉ, M. T. *La terminología: representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, 1999. 369 p.

_____. *La terminologia: teoría, metodología, aplicaciones*. Traducción castellana de Carles Tebé. Barcelona: Editorial Antártida/ Empúries, 1993. 526 p.

DING, Y.; FOO, S. A review of ontology generation. In: ONTOLOGY RESEARCH AND DEVELOPMENT. Parte 1. 2001. Disponível em: <http://homepage.uibk.ac.at/~c703205/download/01jis01_final_revision.pdf>. Acesso em: 13 mar. 2006.

GUARINO, N. Formal ontology and information systems. In: FOIS'98, 1998, Trento, Italy. *Proceedings...* Trento, Italy. Disponível em: <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Acesso em: 05 fev. 2007.

GUARINO, N.; GIARETTA, P. *Ontologies and knowledge bases: towards a terminological clarification*. 1995. Disponível em: <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Acesso em: 12 jan. 2007.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, n. 2, p.199-220, 1993a. Disponível em: <<http://tomgruber.org/writing/ontologia-kaj-1993.pdf>>. Acessado em: 07.fev.2007.

_____. *Toward principles for the design of ontologies used for knowledge sharing*. 1993b. Disponível em: <<http://tomgruber.org/writing/onto-design.pdf>>. Acesso em: 13 jan. 2007.

LISA: Library and Information Science Abstracts. Disponível em <<http://www.csa.com/factsheets/lisa-set-c.php>>. Acesso em: 25.abr.2007.

MOREIRA, A. *Tesouros e ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético*. Belo Horizonte, 2003. 150 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2003. Disponível em <[http://opus.grude.ufmg.br/opus/opusanexos.nsf/4d078acf4b397b3f83256e86004d9d55/915f0db8ceb5bb3583256fb0006a1d5e/\\$FILE/mestrado%20-%20Alexandra%20Moreira.pdf](http://opus.grude.ufmg.br/opus/opusanexos.nsf/4d078acf4b397b3f83256e86004d9d55/915f0db8ceb5bb3583256fb0006a1d5e/$FILE/mestrado%20-%20Alexandra%20Moreira.pdf)>. Acesso em: 16 abr. 2006.

NOY, N. F.; McGUINNESS, D. L. *Desarrollo de ontologías – 101: guía para crear tu primera ontología*. 2005. 29 p. Disponível em <http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf>. Acesso em: 06 nov. 2007.

SOWA, J. F. Building, sharing and merging ontologies. Tutorial. [S. 1. : s. n.], 1999. Disponível em: <<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>>. Acesso em: 16 jan. 007.

WILSON Library literature and Information Science full text. Disponível em <<http://www.ovid.com/site/catalog/DataBase/203.jsp?top=2&mid=3&bottom=7&subsection=10>>. Acesso em: 26 abr. 2007.