

**RODRIGO DE SALES**

**TESAUROS E ONTOLOGIAS SOB A LUZ DA  
TEORIA COMUNICATIVA DA TERMINOLOGIA**

**Florianópolis, 2008.**

**RODRIGO DE SALES**

**TESAUROS E ONTOLOGIAS SOB A LUZ DA  
TEORIA COMUNICATIVA DA TERMINOLOGIA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre

Orientadora: Dra. Lígia Maria Arruda Café

**Florianópolis, 2008.**

**RODRIGO DE SALES**

**TESAUROS E ONTOLOGIAS SOB A LUZ DA  
TEORIA COMUNICATIVA DA TERMINOLOGIA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do Centro de Ciências da Educação da Universidade Federal de Santa Catarina em cumprimento ao requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

*Pesquisa desenvolvida com o apoio financeiro da  
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES*

**APROVADO PELA COMISSÃO EXAMINADORA  
EM FLORIANÓPOLIS, 15 DE ABRIL DE 2008**

**Dra. Miriam Figueiredo Vieira da Cunha  
Coordenadora do Curso**

**Dra. Lígia Maria Arruda Café – PGCIN/UFSC  
Orientadora**

**Dr. José Augusto Chaves Guimarães – PPGCI/UNESP  
Examinador**

**Dra. Miriam Figueiredo Vieira da Cunha – PGCIN/UFSC  
Examinadora**

**Dra. Edna Lúcia da Silva – PGCIN/UFSC  
Suplente**

S163 Sales, Rodrigo de.  
Tesouros e ontologias sob a luz da Teoria Comunicativa da Terminologia / Rodrigo de Sales. — 2008.  
164 f.

Dissertação (Mestrado em Ciência da Informação) — PGCIN/  
UFSC, Florianópolis, 2008.

1. Linguagem Documentária. 2. Tesouro. 3. Ontologia. 4. Teoria Comunicativa da Terminologia. I. Título

CDD – 025.49

CDU – 025.43

*Você não pode provar uma definição.  
O que você pode fazer, é mostrar que ela faz sentido.*

**A. Einstein**

**Dedicado a  
Valdir, Edméya,  
Fernanda e Cristiano.**

## AGRADECIMENTOS

À Família pelos 28 anos de história

À Ana (Pimpa) por estar sempre perto

À Professora Lígia pela orientação e pela luz

Aos professores José Augusto, Miriam e Edna pela avaliação deste trabalho

Aos professores do PGCIN por minha formação

Ao NUPILL/UFSC pelo espaço e pelo aprendizado

À CAPES pelo auxílio financeiro

**LISTA DE ILUSTRAÇÕES**

<b>Figura 1 – Base de Dados Criada para o Fichamento</b>	<b>111</b>
<b>Quadro 1 – Elementos que Caracterizam os Tesouros</b>	<b>114</b>
<b>Quadro 2 – Elementos que Caracterizam as Ontologias</b>	<b>118</b>
<b>Quadro 3 – Inferências Relativas às Semelhanças da Categoria Termo</b>	<b>124</b>
<b>Quadro 4 – Inferências Relativas às Semelhanças da Categoria Conceito</b>	<b>122</b>
<b>Quadro 5 – Inferências Relativas às Semelhanças da Categoria Objetivo</b>	<b>126</b>
<b>Quadro 6 – Inferências Relativas às Diferenças da Categoria Termo</b>	<b>128</b>
<b>Quadro 7 – Inferências Relativas às Diferenças da Categoria Conceito</b>	<b>131</b>
<b>Quadro 8 – Inferências Relativas às Diferenças da Categoria Objetivo</b>	<b>135</b>

SALES, Rodrigo de. **Tesauros e ontologias sob a luz da Teoria Comunicativa da Terminologia**. 2008. 164f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal de Santa Catarina, Florianópolis, 2008.

## RESUMO

O acelerado desenvolvimento das tecnologias de informação e comunicação promove um avanço vertiginoso na veiculação do conhecimento. Tal fato incide diretamente na comunicação entre especialistas que, por meio de linguagens de especialidade, laçam mão de terminologias próprias e específicas para transmitir conteúdos informacionais das mais diversas áreas científicas e tecnológicas. Linguagens Documentárias são modelos de representação do conhecimento que, servindo como instrumentos de controle terminológico, auxiliam o processo de indexação e recuperação de documentos por assunto. O tesauro é uma linguagem documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos que comunicam o conhecimento especializado. A ontologia é um modelo de representação do conhecimento que, a exemplo do tesauro, é utilizada para representar e recuperar informação por meio de uma estrutura conceitual (no caso da ontologia o meio de ação é o informático). A presente pesquisa consiste sobretudo na identificação, por meio de uma pesquisa documental, de aspectos que aproximam e distanciam os fundamentos do tesauro e o da ontologia. Para tanto, foi empregado, no plano metodológico, o Método de Análise de Conteúdo de Laurence Bardin. No plano teórico, a pesquisa está fundamentada pela Teoria Comunicativa da Terminologia (TCT) de Maria Teresa Cabré, que zela pelo caráter comunicativo no ambiente científico e técnico, baseando-se no uso social da linguagem efetivamente utilizada por especialistas. Para a constituição do corpus de análise foram utilizadas as 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 (BDTD). A análise dos resultados e interpretações proporcionaram significativas contribuições aos estudos dos modelos de representação do conhecimento, visto que inúmeras características semelhantes e diferentes entre tesauros e ontologias foram identificadas e comentadas. Constatou-se que ambos os instrumentos se aproximam no que diz respeito ao esclarecimento dos aspectos relativos aos termos e suas estruturas conceituais. Suas diferenças se acentuam na esfera das aplicações, pois os recursos informáticos que suportam as ontologias as concedem objetivos que vão além daqueles atribuídos aos tesauros.

**Palavras-chave:** Tesauro. Ontologia. Linguagem Documentária. Teoria Comunicativa da Terminologia. Método de Análise de Conteúdo

SALES, Rodrigo de. **Tesauros e ontologias sob a luz da Teoria Comunicativa da Terminologia**. 2008. 164f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação. Universidade Federal de Santa Catarina, Florianópolis, 2008.

### ABSTRACT

The accelerated development of information and communication technologies fosters a vertiginous growth in the transmission of knowledge. Fact which come across communication among specialists, who by means of specialized languages, make use of their own particular terminologies in order to transmit informational content belonging to several scientific and technological areas. Documentary languages are models of knowledge representation. They can be employed as tools for terminology control and assistance in the process of indexing and retrieving documents by subject. Thesaurus is a documentary language characterized by the existing specificity and complexity found in the relationship among terms which communicate the specialized knowledge. Ontology is a knowledge representation model (taking for example the thesaurus) used for representing and retrieving information by means of a concept structure (in the case of ontology, its means of action is done through information technology). This piece of research consists specially of identifying, by means of a documental research, aspects which approximates and distantiate thesaurus's basis. The methodology used for this piece of research is that of Laurence Bardin's Content Analysis Method. The theoretical framework is affiliated to Maria Teresa Cabré's Communicative Theory of Terminology (CTT). This framework manage the communicative feature in the scientific and technical environment based on the social use of language effectively used by specialists. For the constitution of the corpus of analysis it was employed the Library and Information Science Abstracts' (LISA) database, Wilson Library Literature and Information Science Full text, and the *Biblioteca Digital de Teses e Dissertações* of IBICT (BDTD). Analysis and interpretation of the output provided meaningful contributions to the studies of knowledge representation models. It was possible to observe a large range of similar and different characteristics between thesauri and ontologies that were identified and commented. It was brought forth that both instruments get approached in what is referred to the clarifying of aspects related to terms and their conceptual structures. Their differences get highlighted in their application sphere while the informatic resources which supports ontologies concedes them objectives that goes beyond those attributed to thesauri.

**Keywords:** Thesaurus. Ontology. Documentary Language. Communicative Theory of Terminology. Content Analysis Method

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	12
<b>1.1</b>	<b>Justificativa</b>	17
<b>1.2</b>	<b>Objetivos</b>	19
<i>1.2.1</i>	<i>Objetivo Geral</i>	20
<i>1.2.2</i>	<i>Objetivos Específicos</i>	20
<b>1.3</b>	<b>Estrutura do Texto</b>	21
<b>2</b>	<b>MARCO CONCEITUAL</b>	24
<b>2.1</b>	<b>Linguagem Documentária: Breve Histórico</b>	24
<b>2.2</b>	<b>Linguagem Documentária: Panorama</b>	34
<b>2.3</b>	<b>Tesouro</b>	45
<b>2.4</b>	<b>Ontologia</b>	59
<b>3</b>	<b>MARCO TEÓRICO</b>	77
<b>3.1</b>	<b>Alguns Subsídios Teóricos que Fundamentam a Reflexão sobre as Linguagens Documentárias</b>	77
<b>3.2</b>	<b>Teoria Comunicativa da Terminologia (TCT)</b>	87
<b>4</b>	<b>PROCEDIMENTOS METODOLÓGICOS: Método de Análise de Conteúdo</b>	99
<b>4.1</b>	<b>Pré-Análise</b>	100
<i>4.1.1</i>	<i>Construção do Corpus de Análise</i>	100
<i>4.1.2</i>	<i>Hipóteses, Índices e Indicadores</i>	106
<i>4.1.3</i>	<i>Categorização</i>	107
<b>4.2</b>	<b>Exploração do Material</b>	110
<b>4.3</b>	<b>Tratamento dos Resultados Obtidos e Inferências</b>	112
<i>4.3.1</i>	<i>As Características Semelhantes</i>	121
<i>4.3.2</i>	<i>As Características Diferentes</i>	127
<b>5</b>	<b>CONCLUSÕES</b>	142
	<b>REFERÊNCIAS</b>	148
	<b>Referência do Corpus de Análise</b>	156
	<b>APÊNDICES</b>	
	<b>Apêndice A – Relatório 1</b>	
	<b>Apêndice B – Relatório 2</b>	

## 1 INTRODUÇÃO

A comunicação técnico-científica, assim como a comunicação geral, se manifesta por meio da linguagem. No entanto, enquanto a comunicação geral faz uso de uma língua geral falada por indivíduos que compartilham do mesmo idioma, respeitando assim o conjunto lexical e as regras gramaticais desse idioma, a comunicação científica e profissional faz uso de uma linguagem especializada falada por especialistas de uma determinada área do conhecimento ou de uma determinada área profissional. O léxico da língua geral é o conjunto de palavras conhecidas e dominadas por um indivíduo falante da respectiva língua (idioma), ao passo que o léxico da língua de especialidade é o conjunto de termos conhecidos e dominados por um indivíduo especialista em determinado domínio de conhecimento. A especificidade das linguagens especializadas restringe seu uso aos conhecedores de determinado conjunto de códigos muitas vezes não conhecido por qualquer sujeito falante, pois as linguagens de especialidade são aquelas criadas e comunicadas entre membros possuidores de um conhecimento específico. As linguagens de especialidade possuem terminologias próprias (terminologias no sentido de conjuntos de termos específicos de um determinado domínio), que nem sempre são de conhecimento geral. A Matemática, por exemplo, possui uma linguagem de especialidade própria e, conseqüentemente, uma terminologia própria. É relevante ressaltar que uma terminologia se desenvolve consensualmente em acordo com os membros de determinada comunidade especializada. É da necessidade da comunicação especializada, usada para adquirir e transferir o conhecimento, que surgem as terminologias e as línguas de especialidades.

Embora a terminologia faça uso de termos específicos em vez de palavras de âmbito geral, fato que lhe dá um caráter determinador de sentidos discursivos, ela também pertence à linguagem natural (linguagem como um todo, utilizada para a comunicação e expressão humanas) usada por um sujeito-falante, pois, antes de um indivíduo ser um especialista em determinado assunto ele é um sujeito-falante natural. A terminologia pertence a uma linguagem especializada, que por sua vez pertence à linguagem natural. Com isso, não é descabido afirmar que uma língua de especialidade é uma sublinguagem pertencente à linguagem natural. Tal fato se leva a considerar que a comunicação técnico-científica nada mais é 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 exemplos de instrumentos que visam o controle dessa precisão terminológica no âmbito da representação do conhecimento e da recuperação da informação. O termo recuperação da informação foi cunhado em 1951 por Calvin Mooers, que englobou em sua definição os aspectos intelectuais concernentes à descrição e busca de informações, além do aparato técnico utilizado para sua operacionalização. Para Saracevic (1996), a recuperação da informação foi (e continua sendo) a principal solução para o primeiro grande ‘problema’ da Ciência da Informação, a “explosão informacional”, que, anunciada em 1945 por Vannevar Bush, foi tema central das discussões voltadas ao universo informacional no Pós-Segunda Guerra. Porém, embora o termo ciência da informação tenha sido cunhado nesse período (Pós-Guerras) na Conferência da *Royal Society* em Londres (em 1948), e difundido nas

reuniões do *Georgia Institute of Technology*, 1961 e 1962, a essência da Ciência da Informação já estava presente em princípios do século XVII. Ao entender que a essência dessa ciência está no acesso à informação, tem-se que considerar o fato de que o fluxo livre da informação e sua distribuição equitativa começaram a ser arquitetados pelas academias de ciência (Itália, Inglaterra, França e Alemanha) no século XVII, que davam ao povo a possibilidade de saber a respeito das descobertas científicas. Nos séculos seguintes a história relata importantes acontecimentos que convergiram com essa idéia de conhecimento distribuído, como é o caso da elaboração da enciclopédia de ciências e artes de d'Alemberte e Diderot (publicada entre 1750 e 1772), que visava uma rede de distribuição do saber, e as idéias preconizadas por Paul Otlet e Henry la Fontaine, que apontavam para o pretendido 'mundialismo' do saber, culminando no Instituto Internacional de Bibliografia (1895). Isso sem falar da almejada biblioteca internacional (conjunto de bibliotecas conectadas mundialmente) vislumbrada por Otlet e la Fontaine. É possível assertar que os pensamentos indicativos de uma ciência voltada à informação já estava sendo traçado nesse período anterior às Guerras Mundiais, e que, se tornaram tecnologicamente viáveis no período Pós-Guerras. Sendo assim, torna-se mais prudente a afirmação de que Vannevar Bush e seus sucessores foram pioneiros na tecnologia da informação. Barreto (2007) diria que os acontecimentos do período anterior às Guerras Mundiais pertencem à sociedade do conhecimento, tendo sido a sociedade da informação desenhada posteriormente com base nas tecnologias de comunicação e informação surgidas no período pós 1945.

Saracevic (1996), revisando e reformulando conceitos apresentados nas décadas de 1960 e 1970, define a Ciência da Informação como “um campo dedicado às questões científicas e à prática profissional voltadas para o problema da efetiva comunicação do

conhecimento e de seus registros entre os seres humanos” (SARACEVIC, 1996, p. 47). A Ciência da Informação, que se ocupa em estudar toda a amplitude e implicações acerca da informação e sua comunicação, é uma das áreas do conhecimento que centra seus esforços teóricos e práticos na precisão da transferência da informação especializada. Mais especificamente, é na atividade de análise documentária que a Ciência da Informação estuda meios de representar o conhecimento de modo que ele seja facilmente armazenado e recuperado. A atribuição principal da análise documentária é tratar o conteúdo da informação visando à eficiência de sua recuperação. Para tanto, é aconselhável que ela adote uma linguagem documentária.

Para prosseguir, é necessário esclarecer que, embora a língua portuguesa determine que os adjetivos derivados de substantivos terminados em ‘nto’ sejam constituídos pelo sufixo ‘al’, fato que determinaria como correto o uso da expressão ‘linguagem documental’ em vez de ‘linguagem documentária’, o emprego da segunda expressão é justificado pela aceitação (convenção) da comunidade da ciência da informação. Este é um exemplo claro de um termo especializado pertencente a uma linguagem de especialidade. Chama-se atenção também para outra expressão comumente empregada pela comunidade da Ciência da Informação, que é a ‘representação do conhecimento’. De maneira simples e de acordo com a etimologia da palavra ‘representação’, que quer dizer apresentar novamente, este trabalho entende a ‘representação do conhecimento’ como o ato de rerepresentar o conhecimento por meios variados, ou seja, rerepresentar o conteúdo de uma informação de maneira concisa, econômica e codificada de maneiras diversas.

As linguagens documentárias são instrumentos de controle terminológico utilizados no processo de indexação e recuperação da informação. O tesauro é uma linguagem documentária caracterizada pela especificidade e 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 tesauro, é utilizada para representar e recuperar informação por meio do conteúdo documental (no caso da ontologia o meio de ação é o informático). Na literatura, não é tarefa fácil encontrar estudos que analisem a fundo as diferenças e semelhanças entre esses dois modelos de representação do conhecimento que de uma maneira geral perseguem uma meta em comum, minimizar os obstáculos da representação e da recuperação da informação e, conseqüentemente, reduzir os ruídos e interferências conceituais da comunicação técnico-científica. Além do mais, 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, pois ambos os modelos de representação do conhecimento são utilizados em situações semelhantes e com funções por vezes concomitantes.

Este estudo propõe uma análise teórico-comparativa desses dois modelos de representação do conhecimento de modo a identificar quais as semelhanças e diferenças existentes entre ambos. Para tanto, a questão que norteia esta pesquisa é: **QUAIS AS SEMELHANÇAS E DIFERENÇAS ENCONTRADAS NA LITERATURA REFERENTE AOS TESAuros E ÀS ONTOLOGIAS NA PERSPECTIVA DA TEORIA COMUNICATIVA DA TERMINOLOGIA?**

Para fundamentar teoricamente a referida análise é empregada a Teoria Comunicativa da Terminologia (TCT), por considerá-la a teoria terminológica que melhor investiga o uso efetivo das terminologias como um processo inserido em uma comunicação especializada real. A TCT foca seus fundamentos calcada no fato de que o processo de comunicação entre especialistas se dá por meio de uma linguagem real e natural.

Como procedimento metodológico para a respectiva análise, que se trata de uma pesquisa documental, adota-se a Análise de Conteúdo definida por Laurence Bardin para a coleta e tratamento dos dados. A importância tanto da TCT quanto da Análise de Conteúdo como alicerces teórico e metodológico para o presente estudo será devidamente explicada no decorrer deste trabalho.

### **1.1 Justificativa**

A imensurável quantidade e variedade de informações produzidas e publicadas na atualidade, em decorrência do exponencial desenvolvimento ocorrido nas tecnologias de informação e comunicação, acarreta, dentre outras coisas, em dificuldades e obstáculos para as eficientes representação e recuperação da informação. Tal fato compromete significativamente a comunicação do conhecimento veiculada pela informação. No âmbito do conhecimento técnico-científico, no qual a precisão e a pertinência da informação comunicada são imprescindíveis, minimizar os obstáculos da representação e recuperação da informação é fator capital. As linguagens documentárias, ou modelos de representação do conhecimento, são instrumentos que procuram reduzir ruídos na representação e recuperação do conhecimento especializado;

conseqüentemente, perseguem a eficiência da comunicação entre os especialistas. Esta comunicação entre especialistas, como afirmado anteriormente, é realizada por meio de uma linguagem de especialidade, que por sua vez produz uma determinada terminologia. Um estudo detalhado sobre tesouros e ontologias, instrumentos de controle terminológico, é um esforço comprometido com o desenvolvimento do conhecimento técnico-científico, visto que este conhecimento se movimenta por meio da comunicação e da terminologia técnica e científica.

Para fundamentar uma pesquisa dessa natureza é necessário recorrer a teorias cujo objetivo de análise diga respeito aos instrumentos de controle terminológico, sobretudo, teorias que esclareçam as implicações do ‘termo’. Dentre as teorias da Terminologia moderna, a Teoria Comunicativa da Terminologia (TCT) é a que foca atenção no caráter comunicativo do termo, abordando-o como uma unidade pertencente à linguagem natural, fato que proporciona uma perspectiva embasada em uma linguagem efetivamente usada pelos especialistas. A perspectiva da TCT contempla a comunicação ‘real’ no âmbito científico e profissional.

O interesse em realizar um estudo desta natureza surgiu ao participar da equipe do Núcleo de Pesquisa em Informática, Literatura e Lingüística da Universidade Federal de Santa Catarina (NUPILL/UFSC), o qual desenvolve e mantém um banco de dados e uma biblioteca digital na área de Literatura. Durante os trabalhos realizados no NUPILL/UFSC foi identificado que a maior carência dos respectivos banco de dados e biblioteca digital é a impossibilidade de realizar uma consulta ao acervo por meio do assunto das obras. Esta carência foi reclamada pelos próprios usuários do banco de dados e da biblioteca digital. Tal fato motivou a equipe do NUPILL/UFSC à elaboração

de uma linguagem documentária em meio digital que viabilizasse a indexação e, conseqüentemente, a recuperação das informações por assunto. A primeira dificuldade encontrada na empresa de elaborar uma linguagem documentária para o banco de dados e biblioteca digital do NUPILL/UFSC foi identificar qual seria o modelo mais adequado, tesouro ou ontologia. Para saber qual modelo de linguagem documentária empregar nas diversas situações é necessário conhecer detalhadamente os elementos fundamentais de cada uma, e foi da necessidade de saber qual a linguagem documentária mais apropriada à necessidade do NUPILL/UFSC que nasceu a idéia desta pesquisa.

O fato de serem poucos os estudos preocupados em analisar os elementos que aproximam e distanciam o tesouro da ontologia é o motivo propulsor do presente estudo. Se as semelhanças e as diferenças entre tesouros e ontologias não são teoricamente explicadas e esclarecidas, torna-se problemático empregá-los adequadamente no processo de comunicação técnico-científica. A realização desta pesquisa brotou do desejo de buscar elucidações teóricas para uma melhor concepção da prática de representação e recuperação do conhecimento.

## **1.2 Objetivos**

Perseguindo meios que viabilizassem encontrar respostas à pergunta-chave deste estudo, a lembrar: *quais as semelhanças e diferenças encontradas na literatura referente aos tesouros e as ontologias na perspectiva da Teoria Comunicativa da Terminologia?*, foi determinado um objetivo principal, que por sua vez exige objetivos específicos, para tornar factível o resultado desta investigação.

### *1.2.1 Objetivo Geral*

- Investigar as semelhanças e diferenças teóricas encontradas na literatura sobre tesouros e ontologias sob a luz da Teoria Comunicativa da Terminologia (TCT)

### *1.2.2 Objetivos Específicos*

- Identificar nos fundamentos da TCT os critérios adequados para uma análise teórica comparativa entre tesouros e ontologias;
- Localizar na produção bibliográfica da área da Ciência da Informação e Ciência da Computação documentos de conteúdos teóricos sobre as temáticas, identificando os elementos bibliográficos fundamentais (autoria, título, ano de publicação, local de publicação, etc);
- Caracterizar as abordagens dadas aos temas, considerando aspectos conceituais e teóricos;
- Analisar as semelhanças e diferenças encontradas na literatura usando como parâmetro a TCT.

Acredita-se que, identificar na TCT os critérios norteadores desta análise e, debruçar-se na produção bibliográfica da Ciência da Informação e Ciência da Computação (áreas que predominantemente articulam as idéias correspondentes aos tesouros e às ontologias) para realizar um estudo das características que aproximam e distanciam ambos os modelos, são escopos que permitem alcançar respostas à pergunta por ora ansiada.

### 1.3 Estrutura do Texto

A presente dissertação está constituída, para além desta introdução (Capítulo 1), por mais cinco capítulos, sendo o Capítulo 2 destinado à fundamentação conceitual, o Capítulo 3 à fundamentação teórica, o quarto Capítulo focado na descrição dos métodos investigativos adotados, incluindo aí todos os processos de construção e análise do corpus e tratamento dos resultados obtidos, o quinto Capítulo dedicado às conclusões e o sexto às referências bibliográficas que fundamentaram este estudo.

A construção dos pilares que sustentam conceitualmente e teoricamente esta pesquisa é constituída por elementos extraídos fundamentalmente da literatura das áreas da Ciência da Informação, Ciência da Computação, Lingüística e Terminologia. A revisão bibliográfica é formada pelos capítulos 2 e 3 (Marco Conceitual e Marco Teórico, respectivamente). O marco conceitual apresenta quatro seções que expõem a) um breve histórico da evolução das linguagens documentárias, b) um panorama dos conceitos definidos às linguagens documentárias, c) a conceitualização dos tesouros, e d) a conceitualização das ontologias. O marco teórico está constituído por duas partes que apresentam respectivamente a) um panorama das teorias que vêm sendo utilizadas para amparar estudos relativos às linguagens documentárias e b) a descrição e a reflexão dos postulados da Teoria Comunicativa da Terminologia.

O quadro metodológico, descrito no Capítulo 4, relata detalhadamente todos os procedimentos efetuados durante a pesquisa. Está constituído pela descrição de todas as etapas sugeridas pela Análise de Conteúdo e adotadas para esta investigação. As etapas

e os procedimentos metodológicos relatados vão desde o planejamento, o qual define variáveis e categorias de análise, passando pelo relato de todo o processo de exploração do material, até chegar à apresentação e interpretação dos resultados da pesquisa. Nas conclusões, como é de praxe, são tecidas algumas reflexões e ousadas algumas afirmações com relação aos resultados obtidos com o intuito de expor o que de significativo foi apreendido com esta pesquisa.

Cabe finalizar este texto introdutório com um esclarecimento tocante à formalidade da dissertação por ora apresentada. O fato de este trabalho ter sido escrito na forma impessoal não quer de forma alguma simular um afastamento do sujeito/analista com o objeto/observado, forjando uma eventual neutralidade, tão corriqueira no discurso científico; trata-se apenas do estilo do autor. Essa perspectiva, que não acredita nem na neutralidade da Língua nem na neutralidade da Ciência, torna-se evidente na escolha tanto da teoria quanto da metodologia empregadas para regerem este estudo. A Teoria Comunicativa da Terminologia (TCT) defende a tese de uma terminologia pertencente a uma linguagem de especialidade que constitui uma sub-linguagem natural em constante funcionamento, sendo regida não pela padronização sintática mas sim pela função pragmática do processo de comunicação. Para a TCT o escopo principal de qualquer língua de especialidade é a comunicação. Tal perspectiva em nada converge com a afirmação de que a língua pode ser uma manifestação neutra. O mesmo pode ser dito com relação à Análise de Conteúdo em contraposição à neutralidade da Ciência. Tal metodologia tem seu argumento fortalecido no espaço cedido à interpretação do analista, inferências de análise.

O emprego de uma teoria com visão funcionalista que foca atenção na pragmática da linguagem comunicada e de uma metodologia que dá vazão às inferências do pesquisador acusa de forma clara que o texto escrito de maneira impessoal se trata de uma formalidade e não de uma neutralidade. Tornando a enfatizar a descrença na neutralidade da Língua e da Ciência, lança-se mão de um estilo formal para relatar uma investigação sob um olhar científico, porém humano.

## 2 MARCO CONCEITUAL

O marco conceitual deste estudo está dividido em três partes que apresentam uma revisão de conceitos e definições relativos às linguagens documentárias (incluindo aqui um breve histórico de sua evolução), aos tesouros e às ontologias respectivamente. Para isso, a investigação se debruçou sobre a literatura da Ciência da Informação (Biblioteconomia e Documentação) e da Ciência da Computação (Inteligência Artificial), sendo que a primeira área norteia o entendimento a respeito das linguagens documentárias, sobretudo dos tesouros, e a segunda esclarece a respeito das ontologias.

### 2.1 Linguagem Documentária: Breve Histórico

A origem das linguagens de indexação, melhor dizendo, as primeiras iniciativas em se criar instrumentos capazes de ajudar no armazenamento e recuperação da informação podem ser encontradas antes da era cristã, na Biblioteca de Alexandria, mais especificamente no Catálogo de Calimachus, embora os termos indexação e recuperação da informação se encontrem no referido período a milênios de distância de sua real utilização. Segundo Piedade (1977, p. 62), “entre os anos de 260 e 240 a.C., o chefe da Biblioteca de Alexandria, Calimachus, poeta e sábio grego, considerado o primeiro bibliotecário que se tem notícia, publicou um catálogo, o *Pinakes*, no qual os livros aparecem escritos segundo o tipo de escritores”. Posteriormente, merecem destaque os inventários de manuscritos organizados pelas bibliotecas medievais. Mas o primeiro exemplo seguro de indexação e análise documentária são as bibliografias surgidas no final do século XV. Porém, Cavalcanti (1978) lembra que os primeiros índices de manuscritos datam da Grécia Antiga, a saber, *De libris propriis liber*, de Cláudio

Galeno, século II d.C., e, *Notitia de seipso et de libris suis*, lista dos escritos de Beda, séculos VII e VIII.

Em meados do século XVI, o filólogo Conrad Gesner apresentou um catálogo, denominado *Bibliotheca Universalis*, constituído de 15.000 obras manuscritas e impressas em latim, grego e hebraico, segundo afirma Cavalcanti (1978). O catálogo incluía obras de aproximadamente 3.000 autores. Posteriormente, Gesner publicou um segundo volume codificado de acordo com um sistema de classificação por ele idealizado.

Em 1852, em Londres, Peter Mark Roget publicou um dicionário analógico denominado *Thesaurus of English Words and Phrases*, o que explica a origem da palavra “tesauro” para o universo da Documentação. Segundo Gomes (1990), Roget criou o “Thesaurus” realizando uma analogia contrária à apresentada pelos dicionários, pois, enquanto estes relacionam em ordem alfabética as palavras e seus respectivos significados, aquele estava organizado por ordem de significados e tinha a função de atribuir uma palavra a cada idéia já relacionada.

É imprudente discorrer a respeito da evolução das linguagens documentárias sem antes perpassar pela evolução histórica dos sistemas de classificação, que são o que podem ser chamados de precursores das linguagens documentárias. Esclarecendo que não é o objetivo deste trabalho pormenorizar os sistemas de classificação do conhecimento, segue-se uma breve menção de alguns dos principais sistemas de classificação, com o intuito apenas de contextualizar cronologicamente o percurso dos modelos de representação do conhecimento.

Os sistemas de classificação podem ser tipificados sob três aspectos: segundo a característica, segundo a finalidade e segundo o campo de conhecimento que abrange. Quanto à característica, a classificação pode ser: a) natural, quando provém de uma aplicação natural regida pelo objeto ou; b) artificial, quando provém de uma aplicação de característica mutável. Quanto à finalidade, uma classificação pode ser: a) filosófica, como aquelas criadas pelos filósofos com a finalidade de definir, esquematizar e hierarquizar o conhecimento, focadas na ordem das ciências e das coisas ou; b) bibliográfica, destinada a servir de base à organização de documentos de modo a serem facilmente localizados. No que diz respeito ao campo de conhecimento que abrangem, as classificações podem ser: a) gerais (enciclopédicas), que podem ter finalidades filosóficas ou bibliográficas ou; b) especializadas, que podem ser científicas ou bibliográficas.

As lógicas de classificação existem desde quando o homem se preocupou em estabelecer algum tipo de ordem ao conhecimento como um todo. Seja com a finalidade de organizar as ciências ou de organizar uma coleção de biblioteca, os sistemas de classificação foram sendo desenvolvidos pelos sábios, pensadores, livreiros e pelos responsáveis e proprietários de acervos bibliográficos. Para se ter uma idéia da longevidade deste tipo de fazer intelectual, Piedade (1977) afirma que as primeiras pseudoclassificações que se tem registro datam de 669 e 626 a.C.. A autora se refere à divisão em dois grandes grupos, Ciência da Terra e Ciência do Céu, para a organização dos tabletes de argila da biblioteca de Assurbanipal, rei da Assíria. Com o passar dos tempos, muitos foram os instrumentos criados com a finalidade de organizar e classificar o conhecimento existente, seja o conhecimento existente no mundo seja o

conhecimento existente em uma biblioteca. Os já citados catálogo de Calimachus (260 e 240 a.C), índices de manuscritos da Grécia Antiga, catálogo de Gesner (século XVI) e o dicionário analógico de Roget (1852), são exemplos deste esforço intelectual.

Para os sistemas de classificação, sobretudo para o Ocidente, teve significativa contribuição o aparecimento da classificação de Francis Bacon no século XVII. O filósofo britânico Francis Bacon (1561-1626) traçou uma classificação filosófica baseada nas faculdades mentais do homem divididas respectivamente em Memória (História), Imaginação (Poesia) e Razão (Filosofia). Na segunda metade do século XIX, em meio a uma mudança de paradigma das classificações do conhecimento para bibliotecas, pois as classificações deixam de focar apenas a macro organização das coleções e passam a focar o assunto tratado nas coleções, o norte americano William Torrey Harris, empenhado em desenvolver uma classificação para a biblioteca da Escola Pública de St. Louis (EUA), inverte a lógica de classificação de Bacon. A inversão preconizada por Harris em 1870, a qual definia uma classificação organizada respectivamente em Ciências, Artes e História, foi a fundamentação da primeira classificação bibliográfica de grande importância mundial, a Classificação Decimal de Dewey (CDD).

A classificação desenvolvida em 1876 pelo norte americano Melvin Dewey (1851-1931) é uma das mais utilizadas em bibliotecas do mundo inteiro, e está dividida em dez classes principais, as quais são identificadas por notações numéricas: 000. Generalidades, 100. Filosofia e Psicologia, 200. Religião, 300. Ciências Sociais, 400. Linguagem, 500. Ciências Naturais, 600. Tecnologias (Ciências Aplicadas), 700. Artes, 800. Literatura, 900. Geografia e História. A principal novidade da CDD foi a notação

decimal que, dentre outras coisas, possibilitou a classificação de assuntos mais específicos. Por meio da notação numérica se tornou possível especificar conhecimentos sob uma perspectiva hierárquica, por exemplo:

**500** Ciências Naturais

**540** Química

**541** Química física e teórica

**542** Técnicas, Equipamentos e Materiais

**543** Química analítica

Ainda na segunda metade do século XIX, que é considerado por muitos como o período incipiente das chamadas linguagens documentárias, outro bibliotecário norte americano, Charles Ammi Cutter (1837-1903), famoso por sua classificação de autoria, durante os anos de 1891 e 1893 desenvolveu uma classificação bibliográfica denominada *Expansive Classification*, que também foi inspirada pela inversão da classificação de Bacon. A Classificação Expansiva de Cutter, que representava assuntos por meio de letras, não atingiu alcance internacional devido à complexidade de sua aplicação. Seu sistema consistia em sete sistemas classificatórios que deveriam ser aplicados gradativamente do mais simples ao mais complexo conforme o aumento do acervo bibliográfico. Ou seja, conforme a coleção de uma biblioteca fosse evoluindo, a biblioteca teria que passar por um processo de reclassificação. É seguro afirmar que a maior contribuição da Classificação Expansiva de Charles Cutter foi ter servido de pilar para a classificação bibliográfica da Biblioteca do Congresso Americano, a *Library of Congress* (LC).

A classificação da Biblioteca do Congresso Americano foi desenvolvida em 1901 especialmente para atender as necessidades organizacionais da própria biblioteca, portanto, estava estruturada para dar conta de um único acervo, desconsiderando

aspectos e assuntos externos à sua coleção. Obviamente que seu esquema enumerativo, suas tabelas auxiliares e suas divisões de classes e seções principais, bem como suas subdivisões alfanuméricas, estavam restritos ao uso interno da Biblioteca do Congresso.

Em 1905, surge outra classificação bibliográfica de grande importância mundial, a Classificação Decimal Universal (CDU). Criada pelos belgas, o advogado Paul Otlet (1868-1944) e o professor Henri La Fontaine (1853-1843), a CDU, originada da classificação de Dewey, é, segundo Piedade (1977), um sistema hierárquico, com base filosófica, mas que devido ao emprego de sinais gráficos, esboça uma tentativa de classificação em facetas, que surge conscientemente apenas com a Classificação de Dois Pontos de Ranganathan. A maior articulação veiculada por meio de dispositivos sintagmáticos para traduzir linguagem natural por meio de notações, fez com que a CDU se tornasse o primeiro sistema de classificação a viabilizar a síntese de dois ou mais assuntos. Com suas divisões de classes principais e subdivisões derivadas da CDD, a CDU avança a Classificação preconizada por Dewey ao adotar em suas notações sistemas semióticos que cumprem funções distintas de relacionamento entre os assuntos. As notações da CDU podem ser formadas por números, letras, símbolos gregos, marcas de pontuação, ou ainda a combinação de todos.

Um ano depois da criação da CDU, o bibliotecário inglês James Duff Brown (1862-1914), considerado o Dewey da Inglaterra, por sua significativa produção bibliográfica na área de Biblioteconomia daquele país, desenvolveu em 1906 a *Subject Classification*. Seu sistema de classificação, que se tornou o sistema geral da Inglaterra na ocasião, era pautado na lógica de que todo conhecimento deriva de um dos quatro princípios fundamentais: Matéria e Força, Vida, Razão e Registro. O britânico que

introduziu a ação de consultas livres às estantes dos acervos das bibliotecas públicas, era defensor da idéia de que a Matéria e a Força geram a Vida, que por sua vez realiza a ação Racional, que resulta nos Registros do conhecimento. Regido por esses princípios fundamentais, Brown definiu as classes de conhecimentos principais de sua classificação e as ordenou por ordem de notações alfabéticas. Além disso, eram empregados algarismos aritméticos, sinais gráficos, números de 0 a 999, ou ainda a combinação de todos. Sua classificação contava ainda com tabelas que descreviam categorias que auxiliavam e representação mais detalhada de assuntos mais específicos.

Em 1933, o universo da classificação do conhecimento se transforma definitivamente com a consolidação da era da Classificação Facetada, liderada pelo bibliotecário indiano Shiyali Ramamrita Ranganathan (1892-1972). Porém, o marco inicial da idéia de facetar coordenando conceitos por meio de termos de indexação pode ser encontrada anteriormente, em 1911, no sistema (*Systematic Indexing*) definido pelo bibliotecário norte-americano Julius Kaiser (1868-1927). O sistema proposto por Kaiser se pautava no pensamento de que todo assunto é composto por duas categorias básicas: o *concreto*, que consistia no objeto tratado, ou seja, tudo que existe a respeito do próprio assunto; e o *processo*, que consistia no modificador do concreto, melhor dizendo, toda ação que incidia sobre o assunto era um processo, ou processos. Kaiser denominou a combinação entre o concreto e o processo de *statement* (enunciado), e o termo principal, ou termo de entrada, era o concreto, nunca o(s) processo(s), o(s) qual(is) servia(m) apenas para especificar o concreto. O sistema de Kaiser considerava ainda uma terceira categoria de termos: o *lugar*, local em que o concreto está localizado. É possível ponderar que as categorias de termos *concreto*, *processo* e *lugar*, definidas por Kaiser,

tenham influenciado Ranganathan na criação das categorias *personalidade*, *energia* e *espaço* respectivamente.

Duas décadas mais tarde surge a *Colon Classification* (ou Classificação de Dois Pontos), definida por Shiyali Ramamrita Ranganathan, que era baseada no agrupamento de assuntos que apresentam as mesmas características e os mesmos tipos de relacionamentos. Ranganathan amplia a visão existente até então relativa aos sistemas de classificação mais difundidos que se pautavam na relação de conhecimento mais geral para conhecimento mais específico. Segundo o bibliotecário indiano, as subdivisões do conhecimento podem se originar por meio de diversas relações, entre elas, gênero-espécie, todo-parte, propriedade-possuidor, ação-paciente ou agente etc.

As facetas definidas por Ranganathan correspondem ao total das subdivisões resultantes da aplicação de uma única característica. Ou seja, são praticamente o mesmo que categorias de elementos formados por características iguais. Por exemplo: na área da Ciência da Informação no Brasil, dentre as várias facetas possíveis, pode-se definir as seguintes:

### **Ciência da Informação**

#### **A Faceta de Periódicos Científicos**

- a Encontros Bibli
- b DataGramZero
- c Transinformação

#### **B Faceta de Programas de Pós-Graduação**

- a PGCIN/UFSC
- b PPGCI/UNESP
- c PPGCI/UNB

Todas as facetas são consideradas por Ranganathan como manifestações de uma das cinco categorias fundamentais: Personalidade (objeto estudado), Matéria (materiais e substâncias), Energia (ações), Espaço (local) e Tempo (época). A *Colon Classification*, que é constituída de classes principais, subdivisões de classes, notações alfanuméricas e simbólicas, tabelas auxiliares de subdivisões (geográficas, políticas, fisiográficas, por língua), e símbolos para diversos tipos de relacionamento entre assuntos, não se difundiu para o mundo prático e pouco se tem notícia de sua efetiva utilização. Porém, a maior contribuição da Classificação Facetada desenvolvida por Ranganathan está na esfera teórica, pois suas idéias de classificação vêm sendo sistematicamente utilizadas para a elaboração de linguagens documentárias como tesouros, além de serem notoriamente empregadas como fundamentação teórica para estudos e aplicações relativas à representação do conhecimento e recuperação da informação.

Sistemas de indexação influenciados pela *Colon Classification* foram desenvolvidos por pesquisadores do *Classification Research Group*, por Derek Austin e por Neelamegahn e Gopinath. Fundado em Londres, 1952, por B. C. Vickery, o *Classification Research Group* (CRG) é um grupo formado por pesquisadores, bibliotecários e profissionais envolvidos com a discussão da classificação e organização do conhecimento. Os membros do CRG avançaram o desenvolvimento da teoria preconizada por Ranganathan transformando e ampliando as categorias de classificação de assunto. Segundo Piedade (1983), o CRG estabeleceu doze categorias possíveis: tipos de produto final, partes, todo, materiais, constituintes, processos, propriedades, operações, agentes, espaço (lugar), tempo e formas de apresentação.

Derek Austin foi o idealizador do sistema de indexação PRECIS (*Preserved Context Index System*), que consistia em um sistema pré-coordenada de indexação alfabética. Segundo Fujita (2003), Austin desenvolveu o sistema PRECIS em 1974 para a *British National Bibliograph* (BNB), com o fim de criar um instrumento cujo funcionamento estava pautado em estruturas semântica e sintática e esquemas de operadores de funções. Para Straioto (2001), o sistema de indexação PRECIS é significativamente influenciado pelo princípio da classificação facetada, sobretudo no que diz respeito aos esquemas de operadores de funções e às relações semânticas genérica (gênero/espécie) e hierárquica (todo/parte). Tal influência é justificada pelo fato de que Derek Austin estava diretamente ligado às pesquisas do *Classification Research Group*.

Fujita (2003), faz referência também ao sistema POPSI (*Postulated-based Permuted Subject Indexing Language*) idealizado por Neelamegahn e Gopinath em 1975. Para a autora, o POPSI “é um sistema inteiramente baseado em princípios classificatórios e que utiliza cabeçalhos de classificação como termos de entrada na produção dos índices cuja padronização é derivada das categorias da classificação de dois pontos de Ranganathan” (FUJITA, 2003, p.75)

Os instrumentos mencionados nos parágrafos precedentes são alguns exemplos de linguagens formais e artificialmente criadas para auxiliar o processo de tratamento da informação responsável pela descrição de conteúdos documentais. Com o avançar dos tempos, a descrição de conteúdos informacionais ganhou força ao se tornar atividade primordial no processo de recuperação da informação. Doravante essas linguagens, que têm por objetivo maior controlar o uso de uma determinada terminologia, seja no

processo de indexação seja no processo de recuperação da informação, serão chamadas de Linguagens Documentárias.

## **2.2 Linguagem Documentária: Panorama**

Linguagens Documentárias (LDs), também conhecidas como modelos de representação do conhecimento, são linguagens artificialmente construídas e constituídas de sistemas simbólicos que visam descrever sinteticamente conteúdos documentais, e são utilizadas nos sistemas documentários para indexação, armazenamento e recuperação da informação. O léxico utilizado pela linguagem natural é composto pelo vocabulário de uma determinada língua, ou seja, a relação de todas as palavras faladas por essa mesma língua, enquanto o léxico adotado pela linguagem documentária “é a relação de termos e símbolos empregados para a identificação temática nos sistemas de informação” (CAVALCANTI, 1978, p. 21). Para a autora, linguagem documentária “é o conjunto de regras, símbolos e termos previamente estabelecidos para indicação de assuntos constantes dos documentos” (op. cit.). Faz-se necessário distinguir linguagem natural (LN) e linguagem documentária (LD). Para Cintra et alii (2002), tanto a LN quanto a LD são sistemas simbólicos instituídos que visam facilitar a comunicação. Porém, a primeira é composta por palavras e seus diversos significados, ao passo que a segunda é composta por palavras que assumem o papel de termos, ou seja, restringem seus significados de acordo com um contexto ou de acordo com uma área específica de conhecimento. É possível complementar essa definição com a afirmação de Cabré (1999) de que unidades lexicais não são previamente nem palavras e nem termos, pois é o contexto situacional de aquisição da unidade lexical que determinará sua função de palavra ou de termo. Embora ambas

autoras concordem em afirmar que a linguagem natural faz uso de palavras, ao passo que a linguagem documentária faz uso de termos, Cabré (1999) reconhece a linguagem documentária como uma linguagem formal e artificial que tem como propósito controlar o uso de uma determinada terminologia, que por sua vez é uma sublinguagem da linguagem natural, pois a estrutura das terminologias respeita as regras gramaticais da linguagem natural.

Analisando sob uma ótica social, Cintra et alii (2002) afirmam que a função das linguagens documentárias vai além da simples tradução de conteúdos documentais; ela tem a função de possibilitar que os diferentes segmentos sociais tenham adequado acesso aos estoques de conhecimento. Sem as linguagens documentárias, não seria possível cumprir com o caráter público da informação, pois são elas que garantem o compartilhamento social entre estoques de conhecimento e usuários de informação. Numa visão também social, Lima (2004) reforça a afirmação descrita acima alegando que a representação e a recuperação eficientes da informação, possibilitadas pelas linguagens documentárias, potencializam a construção de novos conhecimentos, fato este que contribui para o desenvolvimento da sociedade, pois segundo a autora, os usuários de informações documentárias são produtores de conhecimento em potencial. Embora a vertente social exposta pelas duas autoras esteja em um âmbito de caráter público, diferente da vertente social voltada para o uso comunicativo de linguagens de especialidade apresentada por Cabré (1999), percebe-se uma convergente preocupação social do efetivo objetivo das linguagens documentárias, ou seja, o objetivo também social dos instrumentos terminológicos.

Lima (2004), em sua tese de Doutorado, faz uso do termo *informação documentária* para denominar o signo documentário que possibilita a articulação entre conceitos e termos de um domínio de conhecimento, melhor dizendo, a *informação documentária* é o signo que, auxiliado por uma linguagem documentária, representa o conteúdo informacional de um documento. O termo *informação documentária* foi anteriormente utilizado por Kobashi (1994) para identificar genericamente os índices e os resumos, que em sua pesquisa foram definidos como “representações condensadas do conteúdo informacional de documentos, que têm a função de facilitar a circulação da informação e de documentos nas várias esferas da atividade humana” (KOBASHI, 1994, p. 8). Para Lima (2004), a informação documentária é o resultado de um processo de codificação (realizado pelo indexador) e de um processo de decodificação (realizado pelo usuário que acessa a informação), e mais, são unidades que articulam significantes (descritores), significados (conceitos) e referentes (termos). Esta abordagem apresentada por Lima (2004), pode preliminarmente causar certa inquietação ao considerar o termo como o objeto/referente, fato que exige mais adiante (p. 43) voltar atenção para a referida abordagem.

No âmbito da era digital, Gomes e Campos (2004), Moreira, Alvarenga e Oliveira (2004) convergem em afirmar que a maioria dos trabalhos referentes à indexação e recuperação de informação realizados com recursos automáticos, que incluem evidentemente o uso de linguagens documentárias em meio digital, opera somente com base na frequência de ocorrências de palavras pertencentes aos documentos submetidos à análise, ou seja, abrange somente aspectos sintáticos e estatísticos, fator esse que desconsidera os aspectos semânticos dos conteúdos dos documentos, comprometendo a qualidade da indexação e permanecendo muito aquém

da indexação realizada por especialistas. Porém, os respectivos autores não mencionam trabalhos que vêm sendo realizados por pesquisadores da área da Ciência da Informação empenhados em desenvolver métodos de indexação automática considerando elementos semânticos contidos nas informações textuais. Estudos de Kuramoto (1996) e Moreiro, Marzal e Beltrán (2003) são exemplos de esforços voltados à qualidade da indexação e recuperação automáticas da informação tomando por base estruturas da linguagem natural, como sintagmas verbais e nominais, buscando identificar aspectos semânticos dos conteúdos informacionais. Souza (2005) avança esses estudos propondo uma nova metodologia para indexação automática utilizando sintagmas nominais. Evidentemente que não se trata de uma tendência de pesquisa notoriamente consolidada na área, pois os exemplos citados acima foram e estão sendo realizados no Brasil de maneira isolada, mas, não mencioná-los numa revisão de literatura seria colocá-los à margem daquilo que vem sendo estudado nessa temática.

Seguindo a preocupação de zelar pela eficiência da recuperação automática da informação, em um período em que a disseminação da informação conta com a potencialidade dos textos completos disponíveis no universo digital, Bräscher (2002), com uma visão funcionalista<sup>1</sup>, apresenta um estudo de desambiguação para os sistemas automáticos de tratamento da linguagem natural. Pautada na idéia de utilizar gráficos conceituais para solucionar ambigüidades que prejudicam a recuperação pertinente da informação, Bräscher (2002) afirma que, ao lançar mão de aspectos lingüísticos conectados as representações gráficas de conceitos, a aproximação entre o conteúdo da informação com a sentença de busca do usuário final se aplica em nível conceitual, que a autora denomina “estrutura profunda”. O oposto ocorre quando o estudo é realizado

---

<sup>1</sup> A perspectiva funcionalista, voltada a uma interação social das línguas naturais, visa especificar a função sintática, semântica e pragmática dos elementos lexicais dentro da expressão lingüística. A perspectiva funcionalista será explicada no Capítulo 3, reservado ao marco teórico.

exclusivamente no nível da forma, que está fadado à superficialidade da recuperação da informação. Evidenciando a insuficiência da recuperação ancorada somente na estatística e na sintaxe textuais, Bräscher (2002) é enfática ao afirmar que é possível, e necessário, inserir automaticamente solução para ambigüidades em sistemas de tratamento da linguagem natural.

Na questão da recuperação eficiente da informação pela normalização terminológica, Gomes e Campos (2004) ressaltam que os autores (sobretudo os cientistas) também podem contribuir para a precisão da comunicação da informação ao adotarem em seus textos termos normalizados de acordo com linguagens documentárias de sua área de domínio.

É relevante ressaltar que existem na literatura trabalhos como os de Mendonça (2000), Café (2003a), Bufrem (2005) e Pereira e Bufrem (2005), para citar alguns, que enfatizam ser necessário uma relação estreita entre Lingüística e Ciência da Informação para melhor entender e construir linguagens documentárias. Por meio de um estudo bibliométrico referente aos temas abordados nos artigos dos fascículos da *Revista Ciência da Informação*, no período de 1972 a 1998, Mendonça (2000) constatou que, em média, apenas 01 (um) artigo contendo como tema a interseção entre Lingüística e Ciência da Informação foi publicado por ano, durante os 26 anos investigados. Para a autora, que sugere a inserção da disciplina *Terminologia* nos currículos dos cursos de graduação e pós-graduação da área da Ciência da Informação, essa baixa produção literária (que aborda o respectivo tema) compromete a qualidade da organização do conhecimento e da recuperação da informação efetuadas por profissionais da área. Com base no estudo de Mendonça (2000), é seguro afirmar que essa baixa produção literária

afeta diretamente o processo de construção de linguagens documentárias, pois tanto a eficiência da representação do conhecimento quanto à da recuperação da informação estão diretamente ligadas à eficácia da linguagem documentária que será empregada para os respectivos processos. Essa afirmação leva à conclusão de que a Lingüística e a Terminologia devem fazer parte do processo de construção de linguagens documentárias. Café (2003a), ao dissertar a respeito das contribuições da Gramática Funcional na delimitação de segmentos descritores de informação, afirma que a relação entre a Lingüística e a Ciência da Informação é indispensável, e, ao lidar com o processamento automático de conteúdos de informação fazendo uso da linguagem natural, ou ao elaborar ferramentas de recuperação de informação, é possível evidenciar a importância da relação entre as respectivas áreas.

Convergindo com os pensamentos de Mendonça (2000) e Café (2003a), Pereira e Bufrem (2005) também afirmam ser imprescindível a relação entre Lingüística e Documentação para aprimorar a organização e a recuperação da informação (conseqüentemente aprimorar a construção de linguagens documentárias), e acrescentam, embasadas na literatura, que o diálogo deve ser composto pela Terminologia, pela Ciência da Informação e pela Ciência da Computação. As autoras afirmam que trabalhos como os de “Wanderley (1973) [...] Baranow (1983), Medeiros (1986), Miranda (1989), Boulanger (1995), Faulstich (1995 e 1995a) e Peixoto (1995)” (PEREIRA; BUFREM, 2005, p. 22-3) são enfoques diversificados que tratam o cientista da informação a partir de contribuições da Lingüística, da Terminologia e da Socioterminologia. Bufrem (2005) contempla ainda a literatura voltada à relação indispensável entre a Lingüística e a Documentação com uma resenha crítica do livro *El contenido de los documentos textuales: su análisis y representación mediante el*

*lenguaje natural*, de José Antonio Moreiro González, que segundo a autora trata-se de um “contexto teórico exaustivamente analisado e discutido, em que se realizam aproximações atuais relativas à análise documental” (BUFREM, 2005, p. 5).

A relação entre a Lingüística e a Ciência da Informação é assunto abordado desde as décadas de 1960 e 1970, por Jean-Claude Gardin e H. Borko, e 1990, por Nair Y. Kobashi e García Gutiérrez. Em sua Tese de Doutorado, Kobashi (1994), aborda a interface entre a Lingüística e a Análise Documentária a partir do final da década de 1960, quando Borko (1969) *apud* Kobashi (1994) enfatizava que o ponto de vista comunicacional deveria reger as observações a respeito da eficiência dos sistemas documentários, visto que, a interação entre usuário e documento está diretamente ligada à interação entre a linguagem do usuário e a linguagem do processo de indexação. Jean-Claude Gardin, visto por Kobashi (1994) como o precursor da Análise Documentária<sup>2</sup> e criador das bases teóricas e terminológicas da Documentação de orientação francesa, clamou por uma aproximação da Lingüística com a Documentação baseado em dois aspectos: primeiro, pelo vasto número de textos a serem tratados pela Análise Documentária; e segundo, pela necessidade de formalizar um modelo de linguagem documentária que desse conta da representação do conhecimento no ponto de vista semântico e sintático. Compartilhando das idéias de Gardin, García Gutiérrez (1990) fortalece o pensamento de uma relação Documentação/Lingüística ao eleger a Teoria Lingüística como base metodológica para os processos documentários, adentrando nas questões do discurso, do conhecimento, da análise, da tradução, da redução e da representação.

---

<sup>2</sup> Para Gardin (1973) Análise Documentária é entendida como a extração do significado do conteúdo do documento, no caso, documentos escritos. Sem fazer a usual distinção entre textos científicos e textos de âmbito geral, Gardin (1973) afirma ser pouco provável haver outra forma de entender as operações intelectuais envolvidas na análise de documentos científicos, que não seja por meio de um estudo de análise textual em geral.

Esclarecida a perspectiva da interação necessária entre a Lingüística e a Documentação/Ciência da Informação, parte-se agora para uma primeira descrição dos elementos constituintes das linguagens documentárias

### *Termos, conceitos e características*

Termos, conceitos e características são elementos comuns em toda linguagem documentária. Para Dahlberg (1978), filósofa alemã criadora da Teoria do Conceito, a soma total dos enunciados verdadeiros sobre um objeto forma o conceito desse objeto. Segundo a autora, os enunciados são compostos por elementos do conceito, esses elementos são as características do conceito, em outras palavras, características são os atributos predicáveis do objeto. Definições, que vão ao encontro das apresentadas por Dahlberg, são encontradas na ISO 1087-1 (2000), que afirma ser o conceito uma unidade de conhecimento criada pela combinação única de características. A mesma norma internacional define característica como abstração de uma propriedade de um objeto ou de um grupo de objetos.

Linguagens documentárias são formadas por um conjunto de termos, que são considerados por alguns como designações verbais; os termos são definidos pelas ISOs 704 (2000) e 1087-1 (2000) como sendo designações verbais de conceitos em um domínio específico. Afirmação semelhante é feita por Lara (2004a), para quem o termo é uma designação que corresponde a um conceito em uma linguagem de especialidade. O conceito é, para o Manual de Elaboração de Tesouros Monolíngües criado por Hagar Espanha Gomes em 1990, um construto mental que representa um objeto individual material ou imaterial. Pode-se complementar, pautado na TCT, que conceito é a

representação mental de um objeto, e que, somado a uma unidade lexical que o denomina, forma o “termo”. Os conceitos sempre estarão relacionados entre si, e o que determinará essas relações serão as características inerentes aos conceitos.

Em um estudo mais aprofundado, Lima (2004) apresenta uma série de distinções entre termos e descritores, pois não raramente se encontra na literatura um uso indiscriminado desses dois signos que segundo Lima se distinguem em vários aspectos. Embasada por Larivière (1989), Lima (2004) pontua as seguintes diferenças entre termos e descritores: a) enquanto os termos pertencem exclusivamente às linguagens de especialidade, os descritores podem pertencer às linguagens de especialidade e à linguagem comum<sup>3</sup>; b) os termos são geralmente extraídos dos documentos, ao passo que os descritores são geralmente signos conhecidos dos usuários e em hipótese alguma são criados; c) os termos remetem a um objeto da realidade e os descritores remetem a documentos que tratam dessa realidade; d) os termos visam a monorreferencialidade (fazer referência a uma só realidade) e os descritores são escolhidos (preferencialmente) entre outros descritores para representar uma determinada realidade.

Lima (2004) aponta também algumas semelhanças entre termos e descritores, a saber: ambos são unidades significantes que representam um conceito dentro de um domínio de conhecimento; e ambos são unidades da linguagem natural e não códigos artificiais.

---

<sup>3</sup> É provável que esse ‘pertencer também à linguagem comum’ diz respeito ao fato de que os descritores não são necessariamente extraídos dos documentos especializados, mas sim escolhidos de forma a aproximar ao máximo o vocabulário da indexação com o vocabulário do usuário, visando à recuperação da informação.

Lima (2004) lança mão das distinções descritas acima para explicar a respeito da construção das já mencionadas *informações documentárias*, que, para a autora, são signos capazes de articular conceito (significado), descritor (significante) e termo (objeto/referente). As distinções são utilizadas com o intuito de mostrar não apenas que termos e descritores são elementos distintos, mas principalmente esclarecer que no universo da Ciência da Informação um termo se torna um descritor com o fim de representar conteúdos documentais em um sistema de informação. Para a autora, o termo, ou conjunto de termos, equivale ao objeto/referente, ou seja, o termo faz referência a um objeto da realidade, ao passo que o descritor é um significante que, juntamente ao objeto e ao conceito do objeto, forma uma unidade (informação documentária) capaz de descrever conteúdos de informações. As distinções expostas acima não pretende de modo algum sugerir uma oposição ou estado de divergência entre termos e descritores (fato este impensável, já que descritores são termos preferencialmente selecionados), mas sim tornar claro que descritores são unidades terminológicas representantes de assuntos técnicos e científicos utilizados na codificação (indexação) e decodificação (busca de informação) de conteúdos documentais. Por serem unidades terminológicas, os descritores conservam os princípios de monorreferencialidade<sup>4</sup> e univocidade<sup>5</sup>. Sendo assim, permanece verdadeira a afirmação de que linguagens documentárias são constituídas por termos, ou melhor, termos-descritores.

---

<sup>4</sup> Um termo se refere a um único objeto. Para Krieger e Finatto (2004), os termos, que são por natureza signos lingüísticos monossêmicos, são monorreferências porque “veiculam apenas o significado específico de cada área, bem como estabelecem uma única referência com o mundo exterior, sempre na ótica da área em que a unidade lexical está inserida” (KRIEGER; FINATTO, 2004, p. 18). O princípio de monorreferencialidade visa dar conta dos casos de polissemia e homonímia, que nos tesouros são resolvidos com o emprego de qualificadores.

<sup>5</sup> O termo se refere a um único conceito. O princípio de univocidade visa dar conta dos casos de sinonímia, que nos tesouros são resolvidos por meio da relação de equivalência.

Se essas linguagens documentárias são constituídas por termos, e, são instrumentos de controle terminológico atuando em um domínio específico que faz uso de uma linguagem de especialidade, é imprescindível recorrer à Ciência do Termo (Terminologia) e à Ciência da Linguagem (Linguística) para fundamentar teoricamente estudos desta natureza.

Outro trabalho que apresenta uma maior preocupação em definir o que é o termo dentro de uma linguagem documentária, porém, menos detalhista em sua descrição, é o trabalho realizado por Tristão, Fachin e Alarcon (2004). Nele, as autoras alegam que o termo é o elemento que sintetiza o conceito, permitindo sua comunicação. Afirmam ainda que o termo denota um referente, ou seja, trata-se de um signo que representa um objeto. A tese de Lima (2004) converge com essa afirmação e complementa que esse objeto/referente é percebido e interpretado numa práxis, numa realidade social. Isso é melhor descrito nos fundamentos da TCT que são enfáticos ao afirmarem que o ‘termo’ é formado pelo conceito e pela(s) unidade(s) lexical(is) que o designa. Como exposto na descrição da TCT, um objeto é percebido por meio de um processo cognitivo e conceitualizado por meio de um processo de abstração, posteriormente passa por um processo onomasiológico que visa determinar uma (ou mais de uma) unidade lexical que irá designar o conceito do objeto. Conseqüentemente o termo é o conjunto-resultado de todos esses processos.

As linguagens documentárias que melhor representam as relações existentes entre os termos são os tesouros e as ontologias, embora essas últimas transbordem o escopo das linguagens documentárias como será apresentado nas seções seguintes.

### 2.3 Tesouro

A palavra tesouro vem do latim = thesauru e do grego = thesaurós e quer dizer tesouro, armazenamento, repositório. Em 1936, o Oxford English Dictionary definiu a palavra em inglês como sendo dicionário, enciclopédia e similares.

Para o universo da Documentação, o tesouro é uma linguagem documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos. O tesouro geralmente é temático, voltado para uma área específica de conhecimento, ou criado de forma multidisciplinar objetivando solucionar um problema em especial. Os tesouros funcionam como linguagens pós-coordenadas de indexação e podem atuar tanto no processo de indexação quanto no processo de recuperação, ou seja, seus termos são combinados ou correlacionados no momento em que o usuário de um sistema de informação realiza a busca. As LDs que operam nos sistemas pré-coordenados de indexação são aquelas cuja combinação de termos é feita no momento da indexação ou de sua elaboração. Nesse caso, devido à combinação prévia dos termos, têm-se como resultado extensos vocabulários controlados que podem apresentar limitações no processo de indexação de assuntos mais complexos. As listas de cabeçalho de assuntos são exemplos clássicos de linguagens documentárias pré-coordenadas. Constituem-se em listas de termos rigorosamente estruturadas que operam como um vocabulário controlado, e, os termos devem ser escolhidos em um léxico específico já existente. Como informa Cesarino e Pinto (1978), as listas de cabeçalho de assuntos são compostas por: um vocabulário, que é uma lista de termos aprovados para representar assuntos; uma gramática, que é rigorosa quanto ao controle de sinônimos, quase-sinônimos e homógrafos e; regras, que estabelecem o uso da linguagem, tais

como: a) forma de entrada direta ou invertida, b) formação de síntese, c) uso de singular e plural, d) uso de palavras abreviadas ou termos por extenso, e) termos traduzidos e, f) uso de pontuação. As listas de cabeçalhos de assunto, por serem sistemas pré-coordenados fechados e com características prescritivas, possibilitam maior precisão na recuperação de documentos. Porém, segundo Cesarino e Pinto (1978), não são suficientemente específicas para assuntos complexos que se transformam a cada dia.

Diferentemente das linguagens de indexação pré-coordenadas, que atuam sobretudo nos sistemas impressos, as linguagens de indexação pós-coordenadas foram potencializadas com o desenvolvimento de sistemas automatizados de informação. Nos sistemas que utilizam linguagens pós-coordenadas, os termos são combinados entre si no momento da busca da informação, preservando assim, segundo Lancaster (2004), a multidimensionalidade das relações entre os termos, e, atribuindo relevância proporcional a qualquer termo utilizado para a busca de informação. Com estas afirmações relativas às linguagens documentárias pré e pós-coordenadas, torna-se evidente uma vantagem que os tesouros apresentam em relação às demais linguagens documentárias que os antecederam.

Para Cavalcanti (1978),

Tesouro é a lista estruturada de termos associados empregada, por analistas de informação e indexadores, para descrever [o conteúdo de] um documento com a desejada especificidade, em nível de entrada, e para permitir aos pesquisadores a recuperação da informação que procura. (CAVALCANTI, 1978, p.27).

Gomes (1990) define tesouro como “linguagem documentária dinâmica que contém termos relacionados semântica e logicamente, cobrindo de modo compreensivo um domínio do conhecimento” (GOMES, 1990, p.16). A autora discorre sobre a função

e os tipos de tesouros afirmando que a função do mesmo é representar os conteúdos documentais e os conteúdos das buscas realizadas em uma unidade de informação. Quanto à tipologia dos tesouros, Gomes (1990) aponta os monolíngües e multilíngües (no que diz respeito ao idioma), os macrotesouros e microtesouros (no que diz respeito à especificidade de seus termos), e, com relação à abrangência temática, afirma que um tesouro pode estar voltado para a solução de um determinado problema (missão), ou dedicado a um assunto específico.

Do ponto de vista dos relacionamentos entre os termos, Chaumier (1982) apresenta dois tipos de tesouros, o tesouro temático e o tesouro facetado. O primeiro relaciona seus termos de acordo com o domínio de conhecimento (temas, disciplinas), e o segundo de acordo com categorias determinadas, como processos, materiais, equipamentos etc.

Merece destaque o trabalho realizado pela Equipe Hubert de Phalèse da Universidade Sorbonne Nouvelle (Paris III) que, coordenada pelo professor Henri Béhar, desenvolveu um tesouro para indexar por assunto as mais de 2.800 obras literárias de 750 autores franceses contidas atualmente no Banco de Dados de História Literária (BDHL) <http://www.cavi.univ-paris3.fr/phalese/bdhl/bdhl.htm>. Este é um exemplo de um tesouro elaborado para solucionar uma questão problemática. Segundo Bernard (1994), três foram os princípios norteadores da empreitada da Equipe Hubert de Phalèse: a necessidade de um banco de dados de história literária que indique qual o conteúdo das obras; o fato de que as pessoas que fazem grande uso de obras literárias, apesar da noção temática, nem sempre conseguem defini-la com precisão e; a constatação de que grande parte dos documentalistas e bibliotecários não possuem o

conhecimento necessário para indexar obras literárias. Todo o processo de investigação que acarretou na confecção do tesouro dividiu-se em três etapas principais, a saber: a fabricação de um protótipo, a realização propriamente dita e a avaliação. Esse é um trabalho de referência em se tratando de elaboração de tesouro para banco de dados de Literatura. É possível inferir que os motivos propulsores da construção do respectivo tesouro são basicamente os mesmos encontrados nas mais variadas unidades de informação: o desconhecimento temático de profissionais responsáveis pelo processo de indexação e a imprecisão das consultas realizadas pelos usuários dos acervos.

Discorrendo a respeito da relação existente entre a Terminologia e a Documentação, pois ambas se valem uma da outra para a representação do conhecimento por meio das unidades terminológicas, Cabré (1993) afirma que os tesouros são “recopilações” de termos relacionados semanticamente que servem como ferramenta para organizar e recuperar informação.

Vizcaya Alonso (1997) define tesouro como sendo “una lista estructurada de conceptos (descriptores) destinados a representar de forma unívoca el contenido semántico de los documentos y las necesidades de información” (VIZCAYA ALONSO, 1997, p. 144)<sup>6</sup>. Uma definição visivelmente próxima a essa é apresentada anteriormente por Slype (1991). Nela, o autor afirma que tesouro é

una lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar el usuario en la indización de los documentos y de las consultas (Slype, 1991, p. 23-24)<sup>7</sup>.

---

<sup>6</sup> Uma lista estruturada de conceitos (descriptores) destinados a representar de forma unívoca o conteúdo semântico dos documentos e as necessidades de informação

<sup>7</sup> Uma lista estruturada de conceitos destinada a representar de maneira unívoca o conteúdo dos documentos e a consulta dentro de um sistema documental determinado, e a ajudar o usuário na indexação dos documentos e nas consultas.

Laan (2002), ao tecer uma revisão conceitual crítica referente aos tesouros, expõe uma perspectiva distinta da apresentada por Slype (1991) e Vizcaya Alonso (1997), afirmando que:

Entendemos tesouros como a relação de termos de uma linguagem de especialidade, estruturada de modo a evidenciar as relações conceituais dessa área do conhecimento. Dessa forma estamos afirmando que os descritores serão melhor definidos se tratados na perspectiva de unidades lexicais especializadas utilizadas no discurso de uma determinada área do conhecimento (LAAN, 2002, p. 18).

Embora não hajam contraposições nas definições expostas acima, fica evidente que as definições apresentadas na década de 1990 por Slype (1991) e Vizcaya Alonso (1997), as quais convergem em utilizar a palavra ‘conceito’ em vez de ‘termo’, estão voltadas à aplicação dos tesouros nos sistemas de informação, ao passo que a definição de Laan (2002) aborda o tesouro como um instrumento inserido em uma linguagem de especialidade, composta por unidades lexicais especializadas.

O relacionamento entre os termos de um tesouro pode se dar por meio de relação de equivalência (termos em relação de sinonímia), relação hierárquica (gênero-espécie e todo-parte), relação não hierárquica (associativa), ou ainda por meio de uma nota explicativa que forneça informações a respeito do uso do termo. Chaumier (1982) apresenta cinco categorias de relações entre termos: relações de equivalências, relações hierárquicas, relações associativas, relação de definição e relações categoriais. Para o autor, a relação de equivalência, também conhecida como relação de substituição, relaciona descritores<sup>8</sup> com sinônimos e quase sinônimos, possibilitando, por meio de códigos remissivos, a tradução de termos equivalentes (sinônimos) em descritores.

---

<sup>8</sup> Para Chaumier (1982) os descritores são os termos autorizados à serem utilizados nos processos de indexação. Os descritores, que podem ser uma palavra ou uma expressão, têm a função de representar os conceitos e noções tratados nos conteúdos dos documentos.

Outra possibilidade desse tipo de relacionamento é o conhecimento do grupo semântico que um descritor pode cobrir. As relações hierárquicas servem, segundo Chaumier (1982), para explicitar a superioridade e a subordinação existente entre os conceitos expressos pelos descritores. Esse relacionamento pode ser do tipo genérico (gênero/espécie) e partitivo (todo/parte). No caso da relação de tipo genérica, o autor chama a atenção para a possibilidade de um termo estar subordinado a mais de um termo, o que caracteriza como uma relação poli-hierárquica. A relação associativa é descrita por Chaumier (1982) como sendo a responsável pela indicação das possíveis analogias existentes entre dois descritores. A relação de definição mencionada pelo autor equivale à nota explicativa, ou *Scope Note*, que serve para limitar o emprego de um descritor, explicar questões de forma ou possíveis acepções de um descritor. As relações categoriais são os agrupamentos de descritores por semelhança temática ou de facetas. Os temas são os domínios, como Documentação, Agricultura, Química Orgânica etc., e as facetas são as representações que diz respeito aos Processos, Materiais, Equipamentos etc.

Ao tratar de um tesauro eletrônico voltado para o universo do trabalho, Levacov *et alli* (2002) afirmam que as relações de termos nos tesauros impressos normalmente estão estruturadas da seguinte maneira: o termo genérico (TG) - *Broader Term* (BT) - com a sua descendência de termos específicos (TE) - *Narrower Term* (NT) - com as suas ligações com termos relacionados (TR) - *Related Term* (RT) - e com os termos cujo uso é recomendado - *Used* (USE) - ou termos que substituem outros - *Used for* (UF). Cada termo pode ainda ser classificado pela área temática - *Subject* (SG) - à qual ele pertence, pelo eventual delimitador - *Delimited by* (DG) - e por uma eventual nota

explicativa (NE) do conceito - *Scope Note* (SN) -, quando necessário. A título de ilustração, apresenta-se a seguinte organização estrutural dos tesouros:

**Biblioteca**

**TG1** Unidade de Informação

**TE1** Biblioteca Escolar

**TE2** Biblioteca Escolar Pública

**TE1** Biblioteca Universitária

**TR** Bibliotecário

O termo relacionado é proveniente da relação associativa. Além do relacionamento hierárquico baseado na relação gênero-espécie, outra ligação hierárquica possível entre os termos é a partitiva (todo-parte), como por exemplo as partes de uma árvore: raíz, tronco, galho, folha são 'partes' do 'todo' árvore. Em um tesouro as relações partitivas são representadas como termos relacionados. No exemplo exposto, o termo *Árvore* teria como termos relacionados *Raíz*, *Tronco*, *Galho*, *Folha*:

**Árvore**

**TR** Raíz

**TR** Tronco

**TR** Galho

**TR** Folha

Existem exceções quanto ao relacionamento hierárquico partitivo. Quando se trata de partes do corpo humano, partes de região geográfica, ou disciplinas de uma Ciência, o relacionamento pode ser considerado partitivo genérico, e os termos são precedidos da sigla TGP (termo genérico partitivo) e TEP (termo específico partitivo). Por exemplo, os termos *Biblioteconomia*, *Arquivologia* e *Museologia* podem ser termos específicos partitivos do termo *Ciência da Informação*, por se tratarem de disciplinas pertencentes à área da *Ciência da Informação*:

## **Ciência da Informação**

**TEP** Biblioteconomia

**TEP** Arquivologia

**TEP** Museologia

Com o exponencial desenvolvimento das tecnologias de comunicação e informação, surge uma tendência em se utilizar linguagens documentárias como tesouros para recuperar informações em meio digital. Dodebei (2002) alerta que, independentemente da adoção de *softwares* especializados para a elaboração de tesouros em meio digital, todo o planejamento conceitual e suas possíveis relações devem ser traçados por mentes humanas, pois ao *software* cabe somente a função de gerenciar todos os processos previamente determinados pelo humano, que é o único capaz de raciocinar quais são as aproximações e as relações úteis à linguagem que está sendo criada.

Segundo as diretrizes para a construção, formato e gestão de tesouros monolíngues descritas no padrão norte-americano Z39.19 do American National Standards Institute/ National Information Standards Organization – ANSI/NISO,

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 reciprocally. The primary purposes of a thesaurus are (a) to facilitate retrieval of documents and (b) to achieve consistency in the indexing of written or otherwise recorded documents and other items, mainly for postcoordinate information storage and retrieval systems<sup>9</sup>. (ANSI/NISO Z39.19, 2003)

---

<sup>9</sup> Um tesouro é um vocabulário controlado organizado em uma ordem preestabelecida e estruturado de modo 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. As finalidades primordiais de um tesouro são (a) facilitar a recuperação dos documentos e (b) alcançar a consistência na indexação dos documentos escritos ou registrados de outra forma e outros tipos, principalmente para sistemas de armazenamento e de recuperação de informação pós-coordenados. (Tradução nossa).

O ANSI/NISO Z39.19 (2003) ressalta que os tesauros 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 (consultentes), afirma que os termos e as relações dos termos contidos nos tesauros fazem esses instrumentos serem essenciais para ambos (indexador e usuário) buscarem o melhor termo (ou termos) em um sistema de informação. E acrescenta ser imprescindível destacar a linha de tesauros terminológicos, baseados na Teoria do Conceito de Dahlberg (década de 1970), que segundo a autora contribui para o melhor entendimento do conceito e do termo.

Para Tristão, Fachin e Alarcon (2004), tesauros são sistemas artificiais concebidos com o propósito principal de normalizar signos que permitam representar de forma eficiente o conteúdo dos documentos, proporcionando maior eficácia à recuperação da informação, manual ou automática.

Quanto à tipologia dos tesauros, a literatura predominantemente vai ao encontro da classificação concebida por Gomes (1990) que os categoriza da seguinte maneira: com base na língua (monolíngües e multilíngües), com base no grau de especificidade (macrotesauros e microtesauros), ou então com relação à cobertura do assunto. Neste último caso, os tesauros podem ser voltados para uma missão (problema) ou dedicados a um assunto.

Moreira, Alvarenga e Oliveira (2004) mencionam que o programa intergovernamental da UNISIST (United Nations International Scientific Information System), idealizado pela UNESCO, não pode deixar de ser lembrado ao abordar o tema tesouro. O programa define tesouro sob dois aspectos, o estrutural e o funcional. Para o aspecto estrutural, tesouro é um vocabulário controlado dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento (UNESCO *apud* MOREIRA; ALVARENGA; OLIVEIRA, 2004)<sup>10</sup>; e para o aspecto funcional é um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema restrito (*op. cit.*).

Embasadas por Lancaster (1987), Moreira, Alvarenga e Oliveira (2004) apresentam a origem e a linha evolutiva do tesouro. Duas vertentes fundamentam a evolução histórica dos tesouros. A primeira vertente é representada pelo Unitermo, sistema composto por um conjunto de fichas, em que cada ficha contém uma única palavra e os números dos documentos associados a esta palavra (*op.cit.*). Desenvolvido em 1951 por Mortimer Taube, o Unitermo é pautado na abordagem alfabética norte-americana. A segunda vertente, de ordem europeia, é influenciada pela Teoria da Classificação Facetada, pautada no estudo analítico-sintético de Ranganathan. Segundo Moreira, Alvarenga e Oliveira (2004), ambas vertentes se convergem no padrão ISO 2788 (Norma Internacional de Elaboração de Tesouros Monolíngues) e no padrão BS 5723 (British Standards), apesar da linha americana, influenciada pelo Unitermo, apresentar uma evolução mais pragmática, opondo-se à organização por categorias preconizada pela linha europeia, influenciada por sua vez pela classificação facetada.

---

<sup>10</sup> Documento disponível em html.

Com o desenvolvimento das tecnologias de informação e comunicação se torna possível a comunicação via rede de computadores entre bancos de dados e bibliotecas digitais, assim como surge a possibilidade da comunicação (interoperabilidade) entre tesouros de áreas afins. Para tanto, é necessário que os tesouros (ou qualquer linguagem documentária) sejam construídos com princípios baseados nas normas nacionais e internacionais para elaboração desses instrumentos. Focando a importância dessa padronização, Gomes e Campos (2004) realizaram uma análise da norma ISO 2788 com base na Teoria do Conceito de Dahlberg de 1978 e na Teoria da Terminologia de Wüster. Tal análise concluiu que a principal inconsistência do padrão ISO 2788 está na elaboração de diretrizes de construção de tesouros baseada inicialmente nas expressões verbais, quando o mais adequado seria partir de uma análise do conceito, como preconizam as teorias do Conceito e da Terminologia (Teoria Clássica). Para Gomes e Campos (2004), o tesouro não é apenas um instrumento fornecedor de termos de um determinado domínio, mas sim uma estrutura de conceitos, portanto deve ser construído seguindo princípios teóricos consistentes, abarcando pesquisas lingüísticas e de gestão de conhecimento.

Outro trabalho de revisão de normas de elaboração de tesouros é apresentado por Fayen (2004). Nele, é descrita uma revisão do padrão ANSI/NISO Z39.19 (Norma Americana para a Construção, Formato e Gestão de Tesouros Monolíngües), o qual foi elaborado em 1974 e revisado em duas ocasiões, 1980 e 1993, esta última fortemente pautada na Norma Internacional ISO 2788 e na Norma Britânica BS 8723. Segundo Fayen (2004), em 1998 foi realizada, com o intuito de aprimoramento do padrão Z39.19, uma oficina formada pelas American Psychological Association (APA),

American Society of Indexers (ASI), e Association for Library Collections and Technical Services (ALCTS). Essa oficina, que focou a operação de tesouros eletrônicos, identificou as principais limitações do padrão Z39.19 e criou um grupo para discutir melhorias do respectivo padrão. A norma revisada basicamente amplia a cobertura, que passa de documentos para conteúdos de objetos digitais, e amplia os tipos de vocabulários controlados, além de ampliar a interoperabilidade, possível no meio digital. Fayen (2004) é enfática ao dizer que o tesouro tem como vantagem sobre os demais vocabulários controlados a apresentação de termos como árvores hierarquicamente arranjadas.

A despeito dos padrões de elaboração de tesouros, nota-se na literatura que um número considerável de estudos relacionados à linguagem documentária, sobretudo aos tesouros, são embasados teoricamente pela Teoria do Conceito de Dahlberg e pela Teoria da Terminologia de Wüster. Segundo Tristão, Fachin e Alarcon (2004), para a compreensão do conceito e de suas relações é imprescindível adotar princípios da Teoria do Conceito, pois se trata de um método indicado para a fixação do conteúdo do conceito, conseqüentemente, determinante para o posicionamento desse dentro de um sistema de conceitos. Esse preceito é concordado por Gomes e Campos (2004), Moreira (2003), Moreira, Alvarenga e Oliveira (2004), Pereira e Bufrem (2005), entre outros. Pereira e Bufrem (2005), além de utilizarem as respectivas teorias do Conceito e da Terminologia, lançam mão da Teoria da Classificação Facetada de Ranganathan para analisarem princípios e métodos de organização e representação de conceitos para as linguagens documentárias.

Além de buscar fundamentos em teorias como a do Conceito, da Terminologia e da Classificação Facetada, e ater-se aos padrões internacionais de elaboração de tesouros, é necessário considerar a transformação que ocorre com o conceito de tesouro devido às mudanças de meios de atuação e implementação (mudança do meio impresso, para o meio digital e para os serviços ligados em redes). Rosa (1999), chamando a atenção para essas mudanças, destaca o processo de adaptação dos tesouros na *World Wide Web*. O autor enfatiza que a construção de tesouros automáticos deve melhor explorar os recursos da web, visando a elaboração de tesouros compartilhados integrados a sistemas de informação. Ao apresentar uma proposta para a construção de softwares de tesouros, Rosa (1999) afirma que os tesouros devem ser mais flexíveis, possibilitando que os usuários dos tesouros estabeleçam vocabulários (com base em estatísticas de uso e frequência) e definam relações entre os termos, além de escolherem a apresentação final do tesouro, ou seja, o usuário desenharia sua estratégia de busca de informação manipulando o próprio procedimento do tesouro, inserindo e apagando termos. Além de propor um software que possibilite que usuários auxiliem a construção de tesouros, Rosa (1999) ainda vislumbra a possibilidade de extrair termos da web.

No tangente aos tesouros que operam na web, Shiri e Revie (2000) apresentam os seguintes tipos de tesouros: 1) quanto à estrutura e formato: a) tesouro em formato de texto simples e estático; b) tesouro em formato HTML (*HyperText Markup Language*)<sup>11</sup>, ainda estático sem uso efetivo de *hiperlinks*; c) tesouro em formato dinâmico HTML, com uso de *hiperlinks*; d) tesouro com interface visual e gráfica e; e) tesouro em formato XML (*EXtensible Markup Language*)<sup>12</sup>; 2) quanto à funcionalidade

---

<sup>11</sup> Linguagem de marcação utilizada para produzir páginas na web.

<sup>12</sup> O XML é uma especificação técnica desenvolvida pela W3C (World Wide [Web](#) Consortium - entidade responsável pela definição da área gráfica da internet), para superar as limitações do HTML, que é o padrão das páginas da Web.

e usabilidade: a) tesouro autônomo, que não é parte de um sistema de informática e; b) tesouros totalmente integrados a bancos de dados e sistemas de recuperação de informação.

Outros artigos que abrangem os tesouros atrelados ao meio digital são os de Schmitz-Esser (1999), Gilchrist (2003), Schneider e Borlund (2004), Aitchison e Clarke (2004), Thomas (2004) e Nielsen (2004), para citar alguns. Schmitz-Esser (1999), ao descrever uma proposta para uma nova abordagem no planejamento e construção de tesouro com vistas na troca de informações entre tesouros multilíngues, lança mão da abordagem computacional para refletir questões de linguagem. Gilchrist (2003) expõe como os termos tesouro, taxonomia e ontologia vêm sendo utilizados na área de Ciência da Informação, práticas de inteligência artificial e trabalhos que focam a web semântica. Schneider e Borlund (2004) revêm os trabalhos que inspiraram a abordagem semi-automática, baseado na bibliometria, para a construção e manutenção de tesouros. Aitchison e Clarke (2004) abordam a construção de tesouros na era computacional, focando a questão do acesso à informação e a interoperabilidade. Thomas (2004) apresenta uma base lógica para auto-instruções de elaboração de tesouros focando na possibilidade de criar uma auto-aprendizagem de construção de tesouros. Nielsen (2004) reflete em uma revisão bibliográfica tanto as técnicas manuais quanto as automáticas para a elaboração de tesouros.

Com o exposto até então, é possível inferir que tesouro: a) é um vocabulário controlado formado por termos (descritores) semanticamente relacionados, b) é um instrumento de controle terminológico, c) pode estar estruturado hierarquicamente (gênero-espécie e todo-parte) e associativamente (aproximação semântica), e d) é

utilizado principalmente para indexar e para recuperar informação. Ademais, os estudos concernentes aos tesouros, sejam na esfera teórica sejam na esfera aplicada, estão concatenados com as transformações ocorridas em decorrência da mudança de meios de implementação dos tesouros.

Com efeito, apoiando-se na idéia de que o tesouro é uma linguagem documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos, não se pode perder de vista o fato de o mesmo ser um instrumento terminológico que visa auxiliar o processo de comunicação técnica e científica.

## **2.4 Ontologia**

Com o crescente desenvolvimento das tecnologias de informação e comunicação, a produção e divulgação do conhecimento são vertiginosamente potencializadas, o que, por sua vez, acarreta em dificuldades em recuperar pertinentemente o conhecimento produzido e disponível nos diversos espaços proporcionados por aquelas tecnologias. É neste contexto que surgem instrumentos que visam facilitar o processo de representação do conhecimento, organizando e estruturando a informação disponível no meio digital.

Ontologias são modelos de representação do conhecimento que, a exemplo dos tesouros, são utilizadas para representar e recuperar informação/conhecimento por meio de uma estrutura conceitual. Ao longo deste trabalho será possível constatar que as ontologias vão além dos demais instrumentos mencionados até então. Historicamente,

as ontologias (empregadas como modelo de representação do conhecimento e não como a definição adotada pela Filosofia, que será mencionada mais adiante) apresentam idade muito inferior em relação aos tesouros. A literatura relacionada às ontologias se encontra predominantemente em artigos de periódicos científicos das áreas da Ciência da Computação e da Informação. Não há ainda, pelo menos no Brasil, uma significativa produção livresca em torno dessa temática como ocorre com a literatura referente aos tesouros. Tal fato se deve não somente à recente história da ontologia como modelo de representação do conhecimento para a Informática, mas principalmente por se tratar de um instrumento que ainda busca consenso com relação às suas definições. É na Ciência da Computação que o termo vem sendo mais aceito e desenvolvido cientificamente, pois se trata de uma ferramenta cujo objetivo é representar e recuperar informação dentro do ambiente digital e de rede de computadores. Para bancos de dados e bibliotecas digitais e/ou *on-line*, o uso de ontologias se evidencia como relevante devido ao seu desenvolvimento estar atrelado às complexidades do universo computacional.

Para Gruber (1993a), ontologias podem ser consideradas como esquemas conceituais em sistemas da base de dados. Um esquema conceitual fornece uma descrição lógica de dados compartilhados, permitindo programas de aplicação e interoperabilidade entre as bases de dados. O autor enfatiza que uma ontologia tem a função de definir termos que representem o conhecimento. Uma ontologia define o vocabulário usado para compor expressões complexas. Gruber (1993b), ao traçar princípios para projetos de ontologias que visem o compartilhamento de conhecimento, enfatiza que ontologia é uma especificação explícita de uma conceitualização. E afirma ainda que pragmaticamente uma ontologia comum define o vocabulário das consultas realizadas no processo de busca de informação. O objetivo da ontologia é viabilizar um

comum 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), a qual o autor afirma que uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. Ao 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.

Esta definição foi enfaticamente questionada por Guarino e Giaretta (1995) que afirmam ser problemática a noção de conceitualização adotada por Gruber (1993b). Segundo os autores, o problema está no fato de Gruber considerar que a conceitualização reside no nível das relações extensionais<sup>13</sup> e que descreve um estado particular das coisas. Para eles, o mais adequado seria investir na noção de que a conceitualização reside no nível das relações intensionais e que descreve os vários estados das coisas.

Guarino e Giaretta (1995), em artigo que visa esclarecer terminologicamente os princípios das ontologias e das bases de conhecimento, enumeram uma série de possíveis interpretações conceituais a respeito das ontologias, a saber:

1. Ontology as a philosophical discipline;
2. Ontology as a an informal conceptual system;
3. Ontology as a formal semantic account;
4. Ontology as a specification of a “conceptualization”

---

<sup>13</sup> Para a ISO 1087-1 (2000) a extensão é a totalidade de objetos que correspondem a um conceito, e, a intensão é o conjunto de características que constituem um conceito.

[converge com a definição de Gruber (1993b)]; 5. Ontology as a representation of a conceptual system via a logical theory; 5.1 characterized by specific formal properties; 5.2 characterized only by its specific purposes; 6. Ontology as the vocabulary used by a logical theory; 7. Ontology as a (meta-level) specification of a logical theory<sup>14</sup> (GUARINO; GIARETTA, 1995, p.1).

Observa-se que são várias as interpretações terminológicas que correspondem à ontologia, pois são diversos conceitos utilizados em distintos contextos de conhecimento. Como conclusão do referido trabalho, Guarino e Giaretta (1995) afirmam 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 que confinam a estrutura de uma parte da realidade.

Ao analisar as ontologias relacionadas com os sistemas de informação, 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. Em muitos casos, as ontologias apresentam suas próprias peculiaridades metodológicas e de arquitetura. Metodologicamente, sua peculiaridade principal é a adoção de uma aproximação interdisciplinar, em que a Filosofia e a Linguística cumprem um papel fundamental em analisar a estrutura de uma dada realidade em um nível elevado de generalidade e, em formular um vocabulário desobstruído e rigoroso. Com relação à arquitetura, o maior interesse é o papel central que uma ontologia pode cumprir em um sistema de informação, conduzindo à perspectiva de sistemas de informação dirigidos por ontologias.

---

<sup>14</sup> 1. Ontologia como uma disciplina filosófica; 2. Ontologia como um sistema conceitual informal; 3. Ontologia como um sistema semântico formal; 4. Ontologia como uma especificação de uma "conceitualização" [converge com a definição de Gruber (1993b)]; 5. Ontologia como uma representação de um sistema conceitual por meio de uma teoria lógica; 5.1 caracterizado por propriedades formais específicas; 5.2 caracterizado somente por suas finalidades específicas; 6. Ontologia como o vocabulário usado por uma teoria lógica; 7. Ontologia como a especificação de uma teoria lógica. (Tradução nossa).

Com o intuito de distinguir, ou definir com maior precisão, tesauros e ontologias, Moreira, Alvarenga e Oliveira (2004), embasadas pela literatura da Ciência da Computação e da Ciência da Informação, comunicam um estudo de mestrado realizado por Moreira (2003), o qual visa responder a seguinte pergunta: “seriam tesauros e ontologias termos que denotam o mesmo instrumento ou denotariam objetos diferentes e, portanto, seu uso indistinto seria inadequado?” (MOREIRA; ALVARENGA; OLIVEIRA, 2004)<sup>15</sup>. Para responder a essa pergunta, Moreira (2003) fundamenta-se teoricamente na Teoria do Conceito de Dahlberg, por sua vez, influenciada pela classificação do conhecimento por meio das categorias idealizadas por Aristóteles. Dissertando sobre a influência da filosofia aristotélica na Ciência da Informação e na Inteligência Artificial – subárea da Ciência da Computação que mais tem fomentado estudos sobre ontologias – Moreira (2003) afirma que é na elaboração de linguagens documentárias que a classificação por meio das categorias do conhecimento de Aristóteles, se faz mais presente no âmbito da Ciência da Informação. A classificação por categorias de Aristóteles também exerce significativa influência na organização de conceitos realizada pela Inteligência Artificial, porém, segundo Moreira (2003), é a lógica a maior herança aristotélica presente nos estudos da Inteligência Artificial.

O termo ontologia tem origem grega, *ontos* (ser) e *logos* (palavra) e foi empregado primeiramente em estudos de Platão e Aristóteles. Porém, segundo Moreira (2003), foram os filósofos alemães dos séculos XVII e XVIII quem primeiro fizeram uso do termo para designar ontologia como um ramo da Filosofia. Os trabalhos de

---

<sup>15</sup> Documento disponibilizado no formato .html

Moreira (2003), de Almeida e Bax (2003) e Batres et alii (2005) evidenciam a diferença entre a ontologia da Filosofia e a ontologia da Inteligência Artificial. Para os autores, a ontologia é o ramo da Filosofia que estuda “o que existe” na natureza, ao passo que a ontologia da Ciência da Computação é estudada focando a representação do conhecimento. Moreira (2003) acrescenta que a ontologia na Filosofia se ocupa em pensar sobre o que existe no mundo, enquanto a Inteligência Artificial emprega seus esforços nos objetos virtuais, ou seja, naquilo que pode ser representado em um computador. Não é descabido afirmar que a área da Computação tomou de empréstimo o termo ontologia originário da Filosofia por se tratar, em ambas as áreas, de uma ação de descrição de objetos. Enquanto na Filosofia se descreve o ‘ente’ de uma maneira geral (estudo do ser), nas áreas que se ocupam da representação do conhecimento se descreve o ‘ente’ de um determinado domínio.

Jacob (2003), ao dissertar sobre ontologias na web semântica, afirma que a ontologia especifica conceitos relevantes de um domínio, bem como as propriedades desses conceitos e possíveis relacionamentos entre conceitos e propriedades. Isso faz com que as ontologias concebam suportes para processar recursos baseados em interpretação de significados de grandes conteúdos. A autora enfatiza que uma ontologia é um sistema representacional único que integra dentro de uma única estrutura características de abordagens tradicionais como classificações hierárquicas, tesouros facetados e vocabulários controlados.

Para Almeida e Bax (2003), ontologias são estruturas utilizadas para organizar informação com base em conceitos e seus relacionamentos. Segundo os autores

Uma ontologia é criada por especialistas e define as regras que regulam a combinação entre termos e relações em um domínio

do conhecimento. Os usuários formulam consultas usando conceitos definidos pela ontologia. O que se busca, em última instância, são melhorias nos processos de recuperação da informação. (ALMEIDA; BAX, 2003, p. 7).

Almejando proporcionar uma visão geral sobre o estado-da-arte no estudo de ontologias, Almeida e Bax (2003), além de fornecerem definições referentes à ontologia, apresentam diferentes tipos de ontologias e propostas para aplicação e construção de ontologias, como metodologias, ferramentas e linguagens. Pautados na literatura, os autores afirmam que ontologias têm como componentes básicos: classes (organizadas em uma taxonomia), relações (que representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados). Para eles as ontologias podem ser utilizadas em projetos de gestão do conhecimento, comércio eletrônico, processamento de linguagens naturais e recuperação da informação na *WEB*. Essas áreas de atuação das ontologias já haviam sido descritas por Guarino (1998), que foi mais além ao afirmar, ao final da década de 1990, que as ontologias já estavam sendo amplamente estudadas pela Inteligência Artificial, pela Linguística Computacional e pelas teorias de bases de dados, além de campos de pesquisas como: engenharia do conhecimento, engenharia da língua, informação modelar e integração da informação. Guarino (1998) prosseguiu em sua extensa lista de campos de aplicações de ontologias mencionando os seguintes: integração de empresa, medicina, mecânica projetada, padronização do conhecimento e de produto, sistema de informação geográfica, sistema de informação legal e sistema de informação biológica.

Missikoff, Velardi e Fabriani (2003) sintetizam os elementos constituintes de uma ontologia afirmando que necessariamente uma ontologia é formada por um

vocabulário de termos com seus significados e seus relacionamentos. Quanto a isso, Ding (2001) já havia afirmado que uma ontologia típica apresenta uma taxonomia que define classes e suas relações, e, um conjunto de regras de inferência operadoras de funções de raciocínio. É provável que devido a isso Missikoff, Velardi e Fabriani (2003) afirmam que por meio de um relacionamento semântico mais rico, a ontologia é capaz de representar um conhecimento conceitual. De forma sumária os autores afirmam que uma ontologia pode ser vista como um tesouro enriquecido que permite ir além do relacionamento entre termos. Cabe complementar (quase contrapor) a visão de Missikoff, Velardi e Fabriani (2003) com o que afirma Feitosa (2006)

Do ponto de vista da representação do conhecimento, uma ontologia não deve ser concebida apenas como um vocabulário informal, ou mesmo como uma linguagem de termos estruturados – como um tesouro, por exemplo -, mas requer uma possibilidade de interpretação algorítmica dos seus significados e, por conseguinte, uma representação em uma linguagem formal, cujo processamento dos significados pode ser realizado por máquinas. Dito de outro modo: uma ontologia requer a explicitação lógico-formal de significados e palavras, que devem ser expressos por meio de construtos matemáticos (FEITOSA, 2006, p. 73).

E mais, é possível acrescentar, com Denny (2002), que construir uma ontologia significa atrelar coisas diferentes a práticas diferentes, pois atua na descrição de algo, por isso é um avanço com relação aos vocabulários controlados, tesouros e taxonomias.

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. Uma ontologia juntamente a um conjunto de indivíduos de conceitos constitui uma base de conhecimento.

Em trabalho cujo objetivo é abordar o uso de ontologias para extração automática de informações administrativas de uma instituição pública, Batres *et alii* (2005) afirmam que ontologias são “sistemas conceituais utilizados como veículos promotores do compartilhamento e reutilização do conhecimento” (BATRES *et alli*, 2005, p. 76), e seguem afirmando que “são usadas na integração inteligente da informação, recuperação e extração de informação, sistemas de informação cooperativos, comércio eletrônico e gestão do conhecimento” (op. cit.). A definição acima converge com a apresentada por Ding e Foo (2001) que 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. Quatro anos antes das afirmações proferidas por Batres *et alii* (2005), Ding e Foo (2001) acrescentaram algumas informações fundamentais no que diz respeito à composição das ontologias as quais foram concordadas por Almeida e Bax (2003), conforme descrito em parágrafo anterior. Para Ding e Foo (2001) uma ontologia geralmente é composta por grupos de classes (conceitos), relações, funções, axiomas e instâncias. Cabe ressaltar que, a exemplo da maioria dos estudos sobre ontologias, as reflexões tecidas por Ding e Foo (2001) foram embasadas pela definição de Gruber (1993b).

Seguindo na questão dos elementos que formam uma ontologia, a literatura é bastante diversificada, porém, Feitosa (2006) descreve os elementos com base em autores que afirmam que, a despeito da diversidade de tipos e domínios de ontologias, alguns construtos são independentes dessa variação e aparecem como via de regra em qualquer ontologia. Baseado em Hyvonen (2002) *apud* Feitosa (2006): a) **classes** são as definições dos conceitos genéricos de um vocabulário; b) **relação de superclasse** é a

organização hierárquica das classes (superclasse quando genérica e subclasse quando específica); c) **propriedades de classe (slots)**, como o próprio nome diz, são as propriedades de uma classe (de um conceito), e podem ser hereditárias, ou seja, as propriedades de uma subclasse podem ser herdadas de sua superclasse, d) **características de propriedade (facetats)** podem ser associadas às propriedades para determinar ou restringir seu relacionamento com a classe; e) **indivíduos (instâncias)** são os objetos individuais sobre os quais trata a ontologia; e f) **axiomas e restrições**, ou, segundo o autor, regras de inferência, que podem fazer referência a termos, funções, predicados, operadores, quantificadores etc. Já Gómez-Pérez (1999) *apud* Feitosa (1996) define os construtos de uma ontologia da seguinte maneira: a) **classes** podem ser um objeto, uma tarefa, uma função, uma ação uma estratégia, um processo, pois podem ser tanto abstratas quanto concretas, tanto reais quanto fictícias; b) **relações** são os tipos de interações entre as classes; c) **funções** são relações especiais em que um elemento é único para os elementos precedentes; d) **axiomas** servem para definir sentenças verdadeiras; e e) **instâncias** representam objetos particulares.

Por ora, é seguro afirmar que o uso de ontologias potencializa a extração automática de informações, pois se trata de um sistema conceitual composto por termos e suas relações que, usufruindo os recursos informáticos, possibilita o compartilhamento e a reutilização do conhecimento. No que se refere à reutilização do conhecimento, Ding e Foo (2002) dedicam parte de seu estudo de revisão ao mapeamento de ontologias, que diz respeito à possibilidade de reutilizar conhecimento por meio da interoperabilidade entre diferentes ontologias. Segundo os pesquisadores, essa reutilização pode se dar por meio de um processo denominado alinhamento, que ocorre

quando duas ontologias estão interligadas, ou por um processo chamado de combinação, que ocorre quando duas ontologias se misturam dando origem a uma única ontologia.

Retornando ao trabalho de Moreira (2003), cabe expor a resposta encontrada pela pesquisadora à pergunta que norteou sua investigação e que foi exposta no início deste tópico. A pergunta é: “seriam tesouros e ontologias termos que denotam o mesmo instrumento ou denotariam objetos diferentes e, portanto, seu uso indistinto seria inadequado?”. Moreira (2003) responde que, apesar de ambos os objetos (tesouros e ontologias) operarem no mesmo nível, nível epistemológico<sup>16</sup>, ambos apresentam objetivos distintos, pois, enquanto a ontologia está voltada para a inferência automatizada de conceitos de uma determinada área do saber, o tesouro se volta para a comunicação entre usuário e linguagem de especialidade.

Castoldi (2003), ao propor o uso de uma ontologia para a representação e distribuição de informação em bases de dados que operam como sistema de Governo Eletrônico, afirma que não existe uma única forma de se projetar ontologias, pois a compreensão que o projetista da ontologia tem com relação ao domínio do conhecimento que está trabalhando exerce forte influência nas decisões tomadas durante a elaboração do projeto. Uma possibilidade para se avaliar a qualidade de uma ontologia é por meio da verificação de seu uso e aplicação.

Sistema MASTER-Web é um sistema desenvolvido na Universidade Federal de Santa Catarina (UFSC) e tem como objetivo recuperar, classificar e extrair informações

---

<sup>16</sup> Embora o fato de as linguagens documentárias serem representações secundárias do conhecimento (ou seja, constituição de conceitos sobre o conhecimento que se tem dos ‘entes’), tanto o tesouro quanto a ontologia atuam como sistemas que controlam o uso da linguagem especializada que comunica o conhecimento por meio dos termos, pertencendo assim ao nível epistemológico.

da *Web*. Silva (2003) propõe, em sua pesquisa de Mestrado, um sistema de extração de informação baseado em ontologias para ser aplicado em páginas da *Web* previamente classificadas semanticamente pelo sistema MASTER-Web. O pesquisador justifica sua pesquisa afirmando que a precisão dos sistemas de recuperação na *Web* está seriamente comprometida devido à falta de semântica nas páginas da *Web* e à falta de critérios de busca nos sistemas de recuperação. Mais uma vez se evidencia a pertinência do uso de ontologias para os sistemas de informação disponibilizados no meio digital em rede.

Em se tratando de representação de conteúdo de informação no meio digital, as ontologias são os modelos de representação do conhecimento que mais vêm sendo utilizados por especialistas das mais variadas áreas, tendo sua arquitetura predominantemente desenvolvida por projetos da área da Ciência da Computação, sobretudo pela Inteligência Artificial, e sendo estudada por pesquisadores e pós-graduandos das áreas da Ciência da Computação, da Ciência da Informação, da Engenharia de Produção e da Linguística. Em consequência disso, são várias as situações em que estão sendo aplicadas ontologias para solucionar casos de representação e recuperação de informação. Além de ambientes como bibliotecas digitais, bancos e bases de dados, *sites* de busca na *Internet* entre outros, as ontologias estão sendo utilizadas dentro do ambiente de ensino a distância, como é apresentado por Prado (2004) em sua pesquisa de doutorado. A autora propõe uma arquitetura de ontologia que automaticamente forneça reforço de aprendizagem após a aplicação de provas objetivas. Outro ambiente em que o uso de ontologias se faz presente é na elaboração de catálogos. Brauner (2005) propõe uma arquitetura de catálogo de objetos baseada em ontologias com o intuito de oferecer recursos que viabilizam um vocabulário comum que facilita a interoperabilidade entre diferentes fontes de dados.

Em pesquisa sobre metodologias, ferramentas e linguagens para a construção de ontologias, Almeida (2003) elabora um roteiro para a construção de ontologias por meio de uma ferramenta automatizada, que utiliza uma linguagem baseada em lógica OIL (*Ontology Interchange Language*). Nesse trabalho, o autor afirma que embora as ontologias não possuam sempre a mesma estrutura, elas apresentam características e componentes em comum. Reportando-se a Gruber e Almeida (2003) reafirma que os componentes comuns nas ontologias são as classes (organizadas em uma taxionomia), as relações entre os conceitos, os axiomas (sentenças verdadeiras) e as instâncias (dados). Quanto à formalidade do vocabulário das ontologias, o autor, citando Uschold e Gruninger (1996), as classificam em: altamente informais (baseada em linguagem natural), semi-informais (baseada em linguagem natural de forma restrita), semi-formais (baseada em linguagem artificial formalmente definidas) e as rigorosamente formais (definidas com semântica formal, teoremas e provas) (USCHOLD; GRUNINGER, 1996 *apud* ALMEIDA, 2003). Ding (2001), pautado também em Uschold e Gruninger (1996), já havia apresentado ainda outros aspectos que diferenciam os diversos tipos de ontologias: a) quanto à proposta: comunicação, interoperabilidade, benefícios em engenharia de sistemas (reusabilidade, aquisição de conhecimento, especificação); e b) quanto ao assunto: assunto de matéria (tais como domínio), assunto de solução de problema (tarefas, métodos ou soluções ontológicas), e assunto de linguagem de representação do conhecimento (ontologia de representação ou meta-ontologia). Outro tipo de classificação possível é quanto ao detalhe de nível ontológico e a dependência de níveis, classificados em: ontologias de meta-nível, ontologias de compartilhamento, ontologias de referência, ontologias de domínio, ontologias de alto-nível, de tarefas e de aplicação. Guarino (1998) já havia concebido a seguinte distinção conceitual quanto ao

nível de especificidade das ontologias: a) ontologia de nível maior (*top-level*) – descreve os conceitos muito gerais como espaço, tempo, matéria, objeto, evento ação etc., que são independentes de um problema particular ou domínio (são indicadas para ontologias unificadas voltadas para grandes comunidades de usuários); b) ontologia de domínio e ontologia de tarefa – descrevem respectivamente o vocabulário relacionado a um domínio genérico (ex.: Medicina) ou uma tarefa ou uma atividade genérica (ex.: diagnosticar ou vender), especializando os termos introduzidos na ontologia de nível maior e; c) ontologia de aplicação – descreve conceitos dependendo de um domínio e de uma tarefa particular, que são freqüentemente especializações de ambas as ontologias mencionadas.

Mais recentemente, em artigo que visa esclarecer questões referentes à construção e ao compartilhamento de ontologias, Sowa (2006) afirma que

The subject of *ontology* is the study of the *categories* of things that exist or may exist in some domain. The product of such a study, called *an 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*. An uninterpreted logic is *ontologically neutral*: It imposes no constraints on the subject matter or the way the subject is characterized. By itself, logic says nothing about anything, but the combination of logic with an ontology provides a language that can express relationships about the entities in the domain of interest<sup>17</sup> (SOWA, 2006, p. 1-2).

---

<sup>17</sup> O objeto de estudo da ontologia é o estudo das categorias de coisas que existem ou podem existir em algum domínio. O produto deste estudo, conhecido como ontologia, é um catálogo dos tipos de coisas que são supostas existir em um domínio do interesse D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D. As ontologias 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. Uma lógica não interpretada é ontologicamente neutra: Não impõe nenhuma restrição no assunto ou na maneira que o assunto é caracterizado. Por si só, a lógica não diz nada sobre qualquer coisa, mas a combinação da lógica com uma ontologia fornece uma língua que pode expressar relacionamentos sobre as entidades no domínio de interesse. (Tradução nossa).

Nesse mesmo artigo, Sowa (2006) apresenta outra distinção entre tipos de ontologias, que diz respeito à ontologia terminológica e à ontologia formal. Segundo o autor, a primeira, também chamada de grande ontologia, consiste em uma hierarquia de formas e significados de palavras, sem muitos detalhes de relacionamento, pois as distinções feitas pelos axiomas e definições não são muito específicas. As ontologias formais apresentam maiores detalhes em seus axiomas e definições, sustentando maior complexidade em suas inferências. As ontologias formais são menos utilizadas que as terminológicas, principalmente ao que diz respeito ao processamento de linguagem natural.

Esbízaro (2006), em sua Dissertação de Mestrado, apresenta os seguintes propósitos relacionados aos objetivos das ontologias: permitir que múltiplos agentes compartilhem seu conhecimento; ajudar as pessoas a compreenderem melhor uma certa área do conhecimento; ajudar pessoas a atingir um consenso no seu entendimento sobre uma área de conhecimento; auxiliar o compartilhamento de um entendimento comum da estrutura da informação entre pessoas e agentes de software; possibilitar a reutilização do domínio de conhecimento; tornar explícita a compreensão do domínio; separar o domínio de conhecimento do domínio operacional; e analisar um domínio de conhecimento.

Com tantas potencialidades possíveis por meio das ontologias, a terceira geração da *World Wide Web*, conhecida como web semântica, que visa proporcionar uma web inteligente que opere com base também em significados de conteúdos, vê nas ontologias a possibilidade de tornar factível a inteligência de sistemas de computadores ligados em rede. Para Denny (2002), na web semântica ontologias podem ser usadas no sentido de

desambiguação de um diálogo entre aplicações remotas ou entre os agentes, combinando sempre de acordo com ontologias de nível e de domínio superiores. Deve-se considerar que cada ontologia é consensual e congruente com outras ontologias compartilhadas; o resultado é um discurso do domínio comum que possa ser interpretado mais precisamente com regras de inferência e de lógica de aplicação. É seguro afirmar que o avanço no desenvolvimento da web semântica depende impreterivelmente, dentre outras coisas, do avanço no desenvolvimento das ontologias, já que essas, usufruindo recursos da engenharia computacional, são responsáveis pela tarefa de viabilizar processos racionais realizados por máquinas, por meio de suas regras de inferência.

As regras de inferência das ontologias sevem para que máquinas realizem automaticamente deduções de acordo com o que for preestabelecido por especialistas humanos. A título de ilustração, pode-se citar o guia para construção de ontologias fomentado por Noy e McGuinness (2005), no qual as autoras apresentam uma ontologia (criada com o editor Protégé 2000)<sup>18</sup> para vinhos. Com as regras de inferência, a ontologia de vinhos apresentada pelas autoras consegue responder por exemplo qual o vinho mais adequado para acompanhar determinado prato, qual vinho é considerado branco, tinto ou rose, ou então, dos vinhos tintos entre a faixa de quinze e trinta reais, quais são recomendados para acompanhar uma lasanha, tudo isso automaticamente. Isso é possível porque além dos relacionamentos entre as classes (conceitos), e entre as propriedades e as instâncias, as ontologias operam com base no relacionamento de atributos que atuam como condições suficientes e necessárias. Por exemplo: se as condições suficientes e necessárias para que uma bebida seja considerada um vinho

---

<sup>18</sup> Ambiente de edição de ontologias desenvolvido pelo Departamento de Informática Médica da Universidade de Stanford (Stanford Medical Informatics).

branco for ‘ser um vinho’ e ter como atributo ‘a cor branca’, logo se infere que o *champagne* é um vinho branco, pois se trata de um tipo de vinho de cor branca. Outros exemplos de regras de inferência de ontologias são apresentados por Campos, Campos e Campos (2006)

Se uma laranjeira é uma árvore frutífera podemos inferir que laranjeira é também uma árvore, assumindo-se que a hierarquia de classes seja organizada como *árvore/ árvore frutífera/ laranjeira*. Em contrapartida, vamos supor que definimos que uma *árvore tropical* tem como condições necessárias e suficientes os atributos de ser árvore e ser originária dos trópicos e que a laranjeira possui esses atributos. Nesse caso, podemos então inferir que laranjeira é também uma árvore tropical, independentemente da hierarquia na qual esteja situada na ontologia (p. 62).

Se uma pizza doce é definida como sendo uma pizza que tem uma cobertura doce, e se uma pizza possui uma cobertura de chocolate, e ainda chocolate é definido como sendo doce, então a ontologia permite inferir que uma pizza de chocolate é uma pizza doce (p. 65).

Cabe afirmar ainda que toda a dinamicidade das ontologias só é possível devido aos recursos informáticos que são utilizados para seu desenvolvimento. Diferente dos tesouros, que podem ou não estar atrelados a aparatos informáticos, as ontologias indissociavelmente estão ligadas a recursos digitais complexos que viabilizam seu funcionamento. Tal complexidade é amenizada por ambientes ‘editores de ontologias’ que proporcionam, com base em recursos digitais da web (linguagens, protocolos e estruturas), a criação e gestão de ontologias. O Protégé, desenvolvido no Departamento de Informática Médica da Universidade de Stanford, é um exemplo de editor de ontologias. O Protégé ([protege.stanford.edu](http://protege.stanford.edu)) é um ambiente de edição de base de conhecimento compatível com o *Open Knowledge-Base Connectivity Protocol (OKBC)*, que é um protocolo de interoperabilidade que proporciona uma interface de consulta e construção comum de sistemas. Utiliza a linguagem Protégé Axiom Language (PAL), que especifica restrições por meio de lógica de primeira ordem, e, importa e exporta

representação em formato texto. Os objetivos do Protégé são: permitir a interoperabilidade com outros sistemas de representação do conhecimento; ser uma ferramenta de aquisição de conhecimento fácil de configurar e usar e; ser extensível. Outro exemplo de editor de ontologias é o Ontolingua ([www.ksl.stanford.edu/software/ontolingua/](http://www.ksl.stanford.edu/software/ontolingua/)). O Ontolingua fornece um ambiente colaborativo distribuído para navegação, criação, editoração, modificação e uso de ontologias, e, a exemplo do Protégé, foi desenvolvido pela Universidade de Stanford e é de acesso livre na web.

Com o exposto, é possível inferir que ontologia a) é uma definição formal das relações entre termos, b) é uma especificação de uma conceitualização, c) é uma visão abstrata simplificada do que se quer representar, d) exige formalismo informático, e) define conceitos e relações de alguma área do conhecimento de forma compartilhada e consensual, f) é composta de taxonomia e regras de inferência, g) tem como função principal a concepção de uma estrutura conceitual, e h) agrega maior profundidade semântica e tecnológica que as demais linguagens documentárias.

Apoiando-se na idéia de que ontologias são modelos de representação do conhecimento, operando no universo informático, marcadas pela formalidade, pelo caráter de explicitação, de conceitualização e de compartilhamento, não se perderá de vista sua função de facilitadora da comunicação técnica e científica.

### **3 MARCO TEÓRICO**

O marco teórico do presente trabalho está dividido em dois tópicos que apresentam respectivamente: a) um panorama das pesquisas relativas aos modelos de representação do conhecimento, com o intuito de identificar as principais teorias que vêm subsidiando tais estudos; e b) uma descrição da Teoria Comunicativa da Terminologia (TCT), que fundamenta esta pesquisa. Cabe ressaltar que esta pesquisa não faz uso da base metodológica definida pela TCT, que está voltada para o trabalho prático da terminologia (a terminografia), mas sim, lança mão da base teórica a qual descreve os princípios, os supostos e os fundamentos da TCT.

#### **3.1 Alguns Subsídios Teóricos que Fundamentam a Reflexão sobre as Linguagens Documentárias**

Para um estudo consistente de linguagens documentárias, deve-se lançar mão, além das disciplinas de Ciência da Informação, de duas práticas científicas, a Lingüística e a Terminologia. Enquanto a Lingüística, de uma maneira geral, centra seus esforços teóricos na ‘palavra’ como elemento de um processo de significação amplo (a palavra por si própria), a Terminologia visa utilizar a unidade lexical como ‘termo’, elemento de um processo de significação restrito a uma linguagem de especialidade. Estudando a palavra como elemento lingüístico, em sua origem, estrutura e significado<sup>19</sup>, a Lingüística se ocupa com a comunicação geral que faz uso de uma linguagem utilizada em âmbito geral. A Terminologia, que tem como objeto o termo, respeita as mesmas ‘regras’ das línguas gerais, porém, ocupa-se com a comunicação

---

<sup>19</sup> É importante frisar que esta definição das características da Lingüística se inscreve no estudo etimológico e principalmente na definição estruturalista de Ferdinand de Saussure, que serviu de suporte para Roland Barthes pensar os elementos da semiologia.

especializada. Na Lingüística, a palavra é estudada sob um olhar amplo e variado, perseguindo o entendimento de aspectos morfológicos, etimológicos, fonéticos, sintáticos, semânticos, pragmáticos, cognitivos, sociais etc. A ocupação dos estudos terminológicos estão direcionadas à articulação dos termos em uma estrutura conceitual, dedicando-se a olhar o termo como uma unidade transmissora de conhecimento em um âmbito específico. A palavra, como objeto da Lingüística, apresenta inúmeras possibilidades de significados, ao passo que o termo, objeto da Terminologia, visa a representação de um único conceito. Como exemplo, pode-se recorrer aos dicionários de língua geral (produto da prática lexicográfica), que para cada entrada (palavra) descreve uma porção de significados de acordo com os possíveis empregos da palavra. Já os dicionários especializados (produto da prática terminológica), para cada entrada (termo) descreve apenas um conceito.

Linguagens documentárias são constituídas de elementos delimitados por um determinado contexto ou por uma determinada área do conhecimento, melhor dizendo, são constituídas por termos. Com efeito, são subordinadas a uma terminologia, contextualizadas por um conteúdo documental ou por um domínio específico do conhecimento. Porém, a sintaxe e a semântica (objetos da Lingüística) são fatores relevantes na construção das linguagens documentárias. Outra noção lingüística que pode ser utilizada para melhor pensar a essência das linguagens documentárias (LD como sistema relacional de signos), é a noção semiológica de signos, significados e significantes, postulada por Ferdinand de Saussure e revisitada por Roland Barthes. Sem pormenorizar as idéias saussurianas é possível afirmar que sua iniciativa em propor uma ciência geral dos signos, precursora em pensar a língua como um sistema estruturado, em que um signo lingüístico é formado por significante e significado, pode constituir

em importante contribuição para o entendimento das unidades terminológicas, como será percebido no decorrer deste trabalho.

A Teoria do Conceito de Ingtraut Dahlberg, que analisa o conceito como a soma total dos enunciados verdadeiros de um objeto, pontuando informações a respeito de suas características, relações, categorias e definições, é a teoria que vem freqüentemente fornecendo subsídios teóricos aos estudos de representação do conhecimento na área da Ciência da Informação.

Não raro, encontram-se na literatura autores que, ao dissertarem sobre tesouros, ontologias, ou outros modelos de representação do conhecimento, se ancoram na Teoria do Conceito. O artigo de Alvarenga (2001) é um exemplo disso. No mencionado estudo, a autora afirma que a superação de limitações no que se refere à análise documentária, depende da criação de linguagens que possam ser entendidas por máquinas e homens, só assim a codificação de conceitos discursivos, gráficos e sonoros seria compreendida por ambos. Sem dúvida as ontologias são ferramentas que vêm colaborar para isso (pelo menos no que tangem os conceitos discursivos). Conseqüentemente, há cada vez mais a necessidade de investir esforços intelectuais que reflitam na construção e na aplicação desses instrumentos. Alvarenga (2001) afirma que antigas práticas e teorias, como a classificação facetada de Ranganathan, estavam a espera de recursos que possibilitassem sua frutificação.

Dentre as teorias da Terminologia que ancoram os estudos mais recentes de organização e representação do conhecimento, merecem destaque 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. A TGT, dotada de uma perspectiva prescritiva e normativa, é responsável pela concepção dos primeiros pilares do trabalho terminológico. A Socioterminologia dotada de uma perspectiva descritiva, dando ênfase ao uso social da língua, aceita a variação e a flexibilidade lexical e conceitual, o que proporciona uma maior aproximação entre mecanismos de informação e usuários. A TCT fundamenta-se em aspectos comunicativos das línguas naturais para melhor conduzir a comunicação entre especialistas. Esta teoria, que fundamenta o presente trabalho, é detalhadamente descrita na próxima seção deste capítulo. Paralelo a isso, a Teoria da Gramática Funcional de Simon Dik, também vêm sendo aplicada no campo da Ciência da Informação, postulando fundamentos preciosos aos estudos de linguagens documentárias.

Durante décadas, a TGT guiou os passos dos estudos terminológicos e, conseqüentemente, dos estudos de linguagens de controle terminológico. Ainda hoje, muitos dos trabalhos focados no respectivo tema se fundamentam, ou perpassam, pelos fundamentos da TGT de Wüster, que é precursora dos métodos científicos no campo terminológico. Porém, a partir da década de 1990 a TGT começa a ser questionada, e seus princípios passam a ser avaliados e criticados por não pautar suas aplicações em uma linguagem efetivamente utilizada no cotidiano técnico-científico. Embora os pilares levantados pela TGT inicialmente visassem dar conta da linguagem utilizada no campo da Engenharia, como de fato o fez durante muitos anos, seus postulados não são pautados em uma linguagem real, efetivamente utilizada no âmbito técnico-científico. Esse fato, atrelado ao caráter prescritivo e normativo, evidencia uma insuficiência teórica no que tange a comunicação de fato ocorrida no universo científico e profissional. Surgem então novas perspectivas voltadas ao caráter social e comunicativo

das linguagens de especialidade e, conseqüentemente, da terminologia. Essa transição da ótica prescritiva para a ótica descritiva e social será mais bem explicada na próxima seção (p. 76) deste trabalho.

No universo da representação do conhecimento torna-se perceptível, a partir das críticas dirigidas à TGT, uma predominância da visão funcionalista em contraposição à visão formalista. A perspectiva funcionalista, diferentemente da formalista, que metodologicamente se estrutura da função sintática para a semântica<sup>20</sup>, entende a linguagem como um instrumento de interação social utilizada para a comunicação de interlocutores reais. A perspectiva funcionalista busca entender a instrumentalidade da linguagem inserida nas situações sociais, como uma atividade cooperativa estruturada. Para tanto, a visão funcionalista parte do princípio que a função pragmática rege a semântica e a sintaxe nas expressões lingüísticas. Não se detendo às distintas correntes funcionalistas<sup>21</sup>, pode-se afirmar em concomitância com Castilho (1994), Pezatti (1994), Neves (1994 e 1997) que a gramática sob a égide da perspectiva funcional é uma atividade social.

Cabré (1993), com o pensamento direcionado à comunicação especializada, enfatiza que o que difere a Terminologia da Lexicologia não são as características de ordem gramatical, mas sim os aspectos pragmáticos característicos de sua vertente social, ou seja, os fatores pragmáticos que influenciam a relação comunicativa. Tais fatores são, segundo a autora, a função básica a que se propõe a comunicação, a

---

<sup>20</sup> O paradigma formalista, que considera a linguagem como um objeto abstrato baseado no conjunto de sentenças, acredita que só é possível entender o significado do discurso por meio da análise sintática, sendo esta a responsável em reger a semântica e o pragmatismo do texto.

<sup>21</sup> Neves (1997) discorre sobre o funcionalismo conservador (que apenas aponta os elementos inadequados do formalismo), o funcionalismo moderado (que além de ressaltar os pontos inadequados do formalismo, propõe uma análise funcionalista da estrutura) e, o funcionalismo extremado (que nega a estrutura e se baseia unicamente na função, desconsiderando a sintaxe).

temática tratada, os usuários, as situações comunicativas e os tipos de discurso. Considerando que a perspectiva funcionalista parte da função pragmática das expressões lingüísticas constantes nos discursos, e que além disso se ocupam dos fatores que influenciam o processo comunicativo, é seguro afirmar que a adoção de uma visão funcionalista é necessária para bem entender as implicações comunicativas de um discurso.

Faulstich (2003) faz uso da abordagem funcional para propor um modelo teórico que evidencia como os formativos lexicais se ordenam na cadeia dos termos complexos. Nesse estudo inédito, a autora demonstra que o conceito e a definição são elementos que se processam no momento em que o significado terminológico está sendo construído. Na pesquisa, Faulstich (2003) foca atenção nas regras de formação dos termos complexos, e enfatiza que “a extensão da forma e o conteúdo conceitual são entidades cujo funcionamento está de acordo com a gramática da(s) língua(s)” (FAULSTICH, 2003, p. 30/1).

Café (2003a), em trabalho que visa contribuir para a formação de base de conhecimento, muito pesquisada no tratamento automático da linguagem natural, propõe uma interpretação sob a luz da Teoria da Gramática Funcional (que abrange análise da sintaxe, da semântica e da pragmática dos termos) para elaborar uma metodologia de construção de *softwares* que se empenhem em reconhecer automaticamente segmentos lexicais em textos. Esse estudo, de visão funcionalista, é um exemplo categórico de uma análise qualitativa de termos-descritores. Nele, a autora, ancorada pela Teoria de Simon Dik, analisa descritores compostos (constituídos por mais de uma unidade lexical) de textos escritos em português do Brasil na área de

Biotecnologia de Cultura de Tecidos de Plantas. Esses descritores, denominados pela autora de *unidades terminológicas complexas* (UTCs)<sup>22</sup>, são submetidos a uma análise lingüística de natureza humana. A análise de UTCs apresentada por Café (2003b) é um esforço intelectual que converge com a atual tendência em recuperação automática da informação, que é a construção de sistemas de tratamento de texto integral.

Em uma aplicação de um modelo de predicação sintático-semântica para construção de linguagens documentárias facetadas, Café (2003b, 2006) novamente faz uso da Gramática Funcional para fundamentar sua análise de termos complexos. Nesse estudo, a autora realiza uma adaptação da Gramática Funcional para investigar as funções (sintáticas, semânticas e pragmáticas) que cada elemento lexical exerce numa expressão lingüística, visando contemplar os aspectos comunicativos que devem exercer as linguagens documentárias estruturadas em facetas. Café (2006) afirma que as pesquisas teóricas e aplicadas referentes à Terminologia, que sob uma ótica funcionalista, aceitam e trabalham as variações lingüísticas no âmbito da comunicação técnica e científica, são “de extrema importância para a análise temática, incluindo seus processos, produtos e instrumentos de organização do conhecimento aplicados à representação da informação” (CAFÉ, 2006, p.69).

Campos e Gomes (2003), estudando a organização e a representação dos domínios do conhecimento seguindo os princípios da teoria da classificação facetada de Ranganathan, afirmam que a importância fundamental da organização do conhecimento está ficando na superficialidade dos estudos voltados à representação de conteúdos de

---

<sup>22</sup> Para Faulstich (2003) e Café (2006), as UTCs são construídas por meio de um fenômeno de contínuo conceitual que se movem das unidades léxicas mais gerais para as mais específicas. Uma base predicada mais geral é redefinida pelos argumentos e satélites (unidades léxicas que compõem a predicação como um todo) que gradativamente vão a tornando mais específica.

informação no ambiente em rede, pois os princípios das classificações de assuntos são adotados nos dias atuais somente na organização da informação de suportes físicos tradicionais. As autoras ressaltam que

Os princípios apresentados por Ranganathan para a elaboração de uma base teórica sólida, visando a construção de classificação bibliográfica, como pode ser observado, é de fundamental importância para a organização de domínios de conhecimento. A representação de um domínio do saber, se configura como princípio norteador para a organização de documentos e informação. Dessa forma, consideramos que um repensar sobre estudos seminais, como o de Ranganathan, possam trazer novos rumos para conceitos atuais como o de organização de domínios do conhecimento. O desenvolvimento desses estudos, no âmbito de modelos teóricos de representação, permitirá ao profissional de informação a possibilidade de atuar cada dia mais num espaço interdisciplinar que englobe questões ligadas à epistemologia, à lógica, à teoria cognitiva, à computação e à terminologia (CAMPOS; GOMES, 2003, p. 162).

Com o objetivo de compreender melhor os conceitos terminológicos, buscando transcender o uso normalizado dos termos e dando ênfase à importância da Terminologia na construção de linguagens documentárias, Lara (2004b) afirma que a Ciência da Informação, ao assumir a responsabilidade de organizar e transferir informação, deve necessariamente inserir em suas atividades de representação do conhecimento, sobretudo na construção de linguagens documentárias, princípios descritos pela lingüística documentária e, principalmente, pelas teorias da terminologia.

Discutindo sobre estruturas classificatórias subjacentes às linguagens documentárias, Batista (2004), refletindo a respeito das redes de conceitos sob a égide da teoria dos jogos de linguagem de Wittgenstein, do labirinto tipo rede de Umberto Eco, e do modelo do rizoma de Deleuze e Guattari, afirma que, assim como o usuário da língua é convencionalmente treinado a utilizá-la, o usuário da informação pertencente à

comunidade científica também depende de linguagens pré-estabelecidas que explorem semelhanças entre conceitos de um domínio do conhecimento, determinando seu uso.

Com a proposta de elaborar uma metodologia para construção de linguagens documentárias, Cervantes (2004) busca na Terminologia, na Terminografia e na Teoria do Conceito elementos para fundamentar teoricamente sua pesquisa de Mestrado. A autora justifica sua fundamentação afirmando que a Terminologia, estudando signos de designação de linguagens de especialidade, é uma área cuja prática envolve “a criação de termos, sua análise, processamento, apresentação e uso na comunicação especializada” (CERVANTES, 2004, p. 57), sendo o recolhimento, a constituição, a organização e a difusão dos termos sob a forma de dicionário terminológico, uma incumbência da Terminografia. Cervantes (2004) lança mão do substantivo Terminótica, para designar o tratamento de termos com recursos automatizados, ou seja, a Terminótica é a Terminologia automatizada. Na Teoria do conceito, Cervantes (2004) busca bases mais consistentes para a compreensão do que é conceito no que tange representação e recuperação da informação.

Em estudo que determina uma metodologia mais adequada para a representação de domínios de conhecimento, levando em conta objetivos de aplicações ligados às áreas da Ciência da Informação, da Ciência da Computação e da Terminologia, Campos (2004) investiga e discute a problemática representacional nas referidas áreas, fato que permite a representação do conhecimento em diversos domínios, objetivando identificar um núcleo comum de conceitos fundamentais para modelar o conhecimento. Para tanto a autora resgata a teoria da classificação facetada de Ranganathan e a teoria do conceito de Dahlberg (na Ciência da Informação), o modelo orientado a objetos e a ontologia

formal (na Ciência da Computação) e a Teoria da Terminologia de Wüster. No trabalho de Campos (2004), a representação do conhecimento, em busca de uma modelização ideal de domínios, é abordada com base no método de raciocínio, nos objetos de representação e na relação entre os objetos de representação.

Conforme exposto neste breve panorama, limitado porém significativo, são diversas as teorias que estão freqüentemente fornecendo subsídios para a fundamentação teórica dos estudos de modelos de representação do conhecimento: a Teoria do Conceito, a Teoria da Classificação Facetada, a TGT, a Socioterminologia, a TCT e a Lingüística Funcional. Por serem os modelos de representação do conhecimento instrumentos de controle terminológico, não é descabido afirmar que sua fundamentação deve ser pautada em teorias cujo objeto seja a unidade terminológica, ou seja, o termo.

Na esfera da Terminologia, área em que esta pesquisa busca fundamentação teórica, evidencia-se duas correntes distintas de estudos voltados aos modelos de representação do conhecimento (linguagens documentárias). A primeira, liderada pela TGT, pauta-se na visão normativa e prescritiva de uma teoria que visa rotular e padronizar definitivamente o uso de termos e conceitos, objetivando a universalização dos termos específicos; e a segunda, liderada pela Socioterminologia e pela TCT, pauta-se em uma visão descritiva, com aceitação da flexibilidade lexical e conceitual, que se ancora em teorias voltadas não somente às estruturas e designações padronizadas, mas principalmente no uso social de seus objetos. A primeira corrente postula seus fundamentos em um campo dominado por uma linguagem “ideal”, sem considerar as inúmeras variações possíveis em um contexto comunicacional, ao passo que a segunda

corrente busca abranger aspectos e variáveis da linguagem natural existentes em situações de comunicação especializada, ou melhor, trabalham com uma linguagem “real”. A segunda tendência, voltada para o uso social da língua e para o aspecto comunicacional, propicia uma maior aproximação entre a informação e seu usuário.

O fato das linguagens documentárias serem empregadas em ambientes de comunicação especializada, seja de um determinado campo do saber seja de uma determinada área profissional, as tornam instrumentos requisitados em situações ‘reais’ de comunicação. Dessa forma, uma teoria de base lingüística, que leve em consideração as variáveis possíveis em um processo de comunicação, e que opere sob uma abordagem funcionalista, pautada no pragmatismo do discurso efetivamente proferido e consumido nos ambientes especializados, é seguramente adequada para fundamentar as reflexões a cerca desses instrumentos. Além de avançar os estudos terminológicos apontando as insuficiências da TGT, a TCT progride rumo a uma teoria que dialoga diretamente com a prática. Sendo assim, aponta-se a TCT como uma teoria capaz de fundamentar este estudo.

### **3.2 Teoria Comunicativa da Terminologia (TCT)**

Dentre as teorias existentes na Terminologia, a Teoria Comunicativa da Terminologia (TCT) fundamenta seus princípios no caráter comunicativo do discurso especializado, apresentando com isso uma coerente reflexão a respeito da linguagem efetivamente utilizada no âmbito especializado. Os fazeres profissionais e acadêmicos da Ciência da Informação parecem acompanhar um paradigma mais social do que cognitivo, mais voltado para uma coletividade (categoria ou comunidade) do que para

um indivíduo. Com base em orientações epistemológicas voltadas a dimensão comunicativa das línguas naturais, a TCT aponta para uma perspectiva funcionalista<sup>23</sup>, preocupada com o uso efetivo que os especialistas fazem da(s) terminologia(s) nos ambientes especializados. Isso faz com que a TCT se aproxime do paradigma social da Ciência da Informação. Assim como a Ciência da Informação direciona seus esforços para servir uma coletividade, a TCT direciona seus fundamentos para a comunicação entre membros de uma dada comunidade de especialistas.

Antes de discorrer sobre os fundamentos e princípios da TCT é necessário manifestar a importância da Teoria Geral da Terminologia (TGT), que foi a primeira a tratar sistematicamente a terminologia como uma disciplina, concebendo assim os primeiros alicerces da teoria moderna. Na década de 1930, o engenheiro austríaco Eugen Wüster (1898-1977), preocupado em criar um instrumento de trabalho que desfizesse a ambigüidade na comunicação técnico-científica, publica em 1938 o *The Machine Tool*, dicionário especializado na área de engenharia. Com uma preocupação puramente metodológica e normativa (ainda não teórica), Wüster traça as primeiras linhas de sua teoria terminológica (a TGT). Décadas depois, mais precisamente em 1968, Wüster deixa de lado a prática metodológica para estudar a fundo a natureza dos termos. Para Cabré (1993) a Terminologia é abordada por Wüster como uma matéria autônoma descrita pela TGT e definida como um campo próprio em relação a outras ciências e outras disciplinas. Melhor dizendo, a Terminologia é considerada pelo criador da TGT como uma disciplina de encruzilhada entre a Lingüística, a Ciência Cognitiva, a Ciência da Informação, a Comunicação e a Informática. Em síntese a TGT é uma teoria de caráter prescritivo, pautada em uma linguagem 'ideal' que desconsidera a

---

<sup>23</sup> A investigação terminológica proposta pela TCT não dá ênfase somente à estrutura dos termos, mas principalmente, ao funcionamento desses em uma situação comunicacional.

flexibilidade efetiva do processo comunicativo no ambiente especializado, e visa a normalização (padronização) internacional dos termos especializados.

Na década de 1990, Gaudin (1993), afirmando que a terminologia não pode ser estudada isoladamente, sem se considerar o contexto social, econômico e comercial, constrói os pilares da Socioterminologia. O autor lança mão da Sociolinguística, que considera o contexto social, cultural e político, para analisar o emprego da língua francesa, e afirma que a prática terminológica não pode de forma alguma ser dissociada do conhecimento do campo de atuação dessa prática, levando em consideração o texto produzido nesse campo, seu público, sua ação e sua utilidade. Segundo Gaudin (1993), as características da Socioterminologia possibilitam estudar o funcionamento dos léxicos especializados que efetivamente são utilizados na comunicação entre os pares de um domínio (científico, profissional etc.), contrapondo, com isso, a atribuição formal de rótulos postulados pela Teoria de Wüster. Outra característica marcante da Socioterminologia é se reportar à origem dos termos, verificando não somente sua recepção e aceitação na comunidade especializada, mas principalmente identificando as causas do fracasso ou as razões do sucesso de seu uso efetivo. Tal fato, faz com que a respectiva teoria transcenda a terminologia escrita e normativa. É evidente que a Socioterminologia atua nas práticas linguísticas e sociais concretas, e portanto, localiza-se nas leis que unem trabalho e linguagem.

Paralelo à abordagem social da teoria de Gaudin, surge uma perspectiva focada no caráter comunicativo dos produtos terminológicos, pautada em uma visão funcionalista (bem como a Socioterminologia). Na Espanha, esta corrente é liderada por

Maria Teresa Cabré da Universitat Pompeu Fabra (Barcelona), e no Brasil por Enilde Faulstich da Universidade de Brasília.

A criação da TCT ganha forças na segunda metade da década de 1990, momento em que a Teoria Geral da Terminologia recebe inúmeras críticas por seu caráter reducionista e idealista. Não deixando de considerar a importância da TGT para o estudo sistemático da terminologia, mas sim, evoluindo o pensamento terminológico de acordo com as necessidades reais de comunicação do universo técnico e científico, Cabré (1999) ressalta que a ótica prescritiva e normativa da TGT limita a comunicação profissional devido à inflexibilidade ao se tratar os termos, seus conceitos e características. Para a autora, esse reducionismo ocorre devido às crenças idealistas de que os conceitos preexistem às expressões; de que o conhecimento técnico-científico é uniforme e universal; de que a estruturação de um campo de conhecimento independe de seu contexto; de que os âmbitos especializados são neutros e consensuais; e devido à crença de que o termo normalizado apresenta as características mais significativas para todas as situações.

A TCT, diferentemente da TGT, que enaltece as diferenças entre Lingüística e Terminologia, visa considerar as concomitâncias existentes entre ambas disciplinas para traçar uma teoria de base lingüística que analise as unidades terminológicas tendo em conta a dimensão textual e discursiva dos termos. Em vez de restringir seus fundamentos, a TCT busca 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, pois

não há representação e comunicação eficientes do conhecimento sem uma terminologia).

Segundo Cabré (1999), 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.

Definindo a palavra terminologia, Cabré (1993) afirma haver pelo menos três conceitos possíveis: “a) o conjunto de princípios e de bases conceituais que regem o estudo dos termos; b) o conjunto de diretrizes que se utilizam no trabalho terminográfico<sup>24</sup>; e c) o conjunto de termos de uma determinada área de especialidade”<sup>25</sup> (CABRÉ, 1993, p. 82). Para a reflexão acerca das linguagens documentárias a primeira e a terceira definições são as mais indicadas, pois as LDs têm por objetivo geral controlar o uso do conjunto de termos de um determinado assunto, tendo em conta as bases conceituais.

A exemplo das demais teorias que fundamentam os estudos relacionados à representação do conhecimento e às linguagens documentárias, a TCT 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. O termo é visto na teoria de Cabré como uma unidade conceitual e denominativa, como uma unidade de conhecimento. Essa unidade de conhecimento, chamada de unidade terminológica ou termo, é composta por uma forma, ou denominação, e um conteúdo, sendo que a forma

---

<sup>24</sup> O trabalho terminográfico é a prática (aplicação) da terminologia.

<sup>25</sup> Tradução nossa.

é a unidade lexical que denomina o conceito (conteúdo). A título de ilustração é possível se reportar à concepção do signo lingüístico definida por Saussure. O lingüista suíço definiu o signo como sendo uma unidade constituída de duas faces, o significado e o significante, como uma moeda de dois lados, sendo a moeda o próprio signo lingüístico. Traçando uma analogia com a definição saussuriana, é possível dizer que o termo é a moeda, ou seja, o termo é o signo lingüístico, que, conforme o exemplo de Saussure, também é composto de dois elementos, o conceito, ou conteúdo (significado) e a unidade lexical, ou forma (significante). Essa ótica apresentada pela TCT foi decisiva na escolha teórica da fundamentação desta pesquisa, pois, concomitante a essa maneira de enxergar a unidade terminológica, pode-se avançar o pensamento de que o termo mostra (faz ver) um conceito, e não apenas o denomina. Essa noção de que o termo mostra um conceito será melhor explicada na seção destinada às interpretações e inferências.

Preocupada em conceber uma teoria que contemple o caráter flexível do processo comunicativo dentro do universo especializado, Cabré (1999) propõe uma visão alternativa de estudo das unidades terminológicas. Para tanto, apresenta três princípios e três condições que norteiam sua reflexão teórica:

1. *Princípio da poliedricidade do termo*: as unidades terminológicas são poliédricas, pois integram ao mesmo tempo aspectos lingüísticos, cognitivos e sociais.
2. *Princípio do caráter comunicativo da terminologia*: o termo persegue imediata ou remotamente a função comunicativa, seja comunicação direta (realizada entre

especialistas) seja comunicação indireta (realizada por meio de traduções ou interpretações) seja ainda por meio de linguagens documentárias.

3. *Princípio da variação*: no processo comunicativo existem variações do tipo sinonímicas, denominações distintas para um mesmo conceito, ou polissêmicas, conceitualizações distintas para uma mesma denominação. O que na Lexicografia é entendido por sinônimo (mais de uma palavra designando o mesmo significado), em Terminologia se assume a idéia de que diferentes termos estão em relação de sinonímia, assim como, o que em Lexicografia é entendido por polissemia (uma palavra possuir mais de um significado), em Terminologia se afirma que diferentes termos estão em relação de homonímia. Tais afirmações se devem ao fato de que para a Terminologia um termo é uma unidade composta por um único conceito<sup>26</sup> e uma única designação lexical, portanto, se uma mesma designação lexical possuir dois ou mais conceitos distintos, trata-se de dois ou mais termos em relação de homonímia. O mesmo se aplica para o caso de um único conceito que apresenta mais de uma designação lexical, trata-se de mais de um termo em relação de sinonímia. Tal princípio já havia sido esboçado quando Cabré (1993) afirmou que somente na esfera teórica é possível afirmar que termos são unidades unívocas (quando a relação entre forma em conceito é única) e monorreferenciais (quando um termo designa um único objeto). Na prática, afirma a autora, pode existir uma forma que tenha vários significados, e um significado pode apresentar várias formas.

---

<sup>26</sup> Neste trabalho, conceito é entendido parcialmente de acordo com Dahlberg (1978). Concorde-se com a afirmação de que conceito é a soma dos enunciados verdadeiros referentes a um objeto da realidade. Porém, discorda da afirmação de que o termo é um elemento do conceito que serve para sintetizar predicados de um referente. Prefere-se a idéia de que o conceito, juntamente com a unidade lexical que o nomeia, forma o termo. Emprega-se também a idéia de que o conceito é um construto mental, ou seja, é concebido, segundo Cabré (1999), por meio de um processo de percepção de um objeto real, passando posteriormente por um processo de abstração (conceitualização).

4. *Condição de linguagem natural*: a linguagem especializada é um subconjunto da linguagem natural, pois respeita o conjunto de regras (gramática) da linguagem natural. Embora a terminologia esteja sob a égide de uma linguagem de especialidade, e esta esteja restrita a um público reduzido, seu objeto é a unidade lexical, que originalmente não é nem palavra nem termo, pois é a situação comunicativa que irá decidir.
5. *Condição de comunicação especializada*: a terminologia é tematicamente marcada, produz-se em situações profissionais e não adquire seu significado diretamente do objeto da realidade, mas sim de estruturas consensuais e preestabelecidas. Portanto é formal e seletiva.
6. *Condição de especialização*: o grau de especialização de um texto é pautado no modo como este veicula sua temática dependendo de sua densidade terminológica e da variação expressiva dos conceitos referenciados.

Os princípios e condições expostos acima são os primeiros tijolos da construção de uma teoria de base lingüística que analisa a terminologia como um conjunto de unidades denominativo-conceituais extraídas da linguagem natural, representando e comunicando o conhecimento especializado dentro de uma situação profissional real.

Em síntese, os supostos da TCT, segundo Cabré (1999), são:

- Não conceber a terminologia como uma disciplina autônoma, mas sim, concebê-la como uma interdisciplina que integra aspectos de uma teoria da linguagem, uma teoria do conhecimento e uma teoria da comunicação, e que a referida teoria da linguagem aborda aspectos lingüísticos, cognitivos e sociais.

- Explicar as diferenças e concomitâncias entre conhecimento geral e especializado (considerando que ambos são competências do falante-especialista) mantendo as peculiaridades de cada um.
- Explicar a interdisciplinaridade dos termos, considerando a diversidade de perspectivas dos especialistas e das especialidades, pautada no caráter poliédrico e multidimensional do termo.
- Esclarecer como o conceito pode fazer parte de distintas áreas de conhecimento, mantendo, trocando ou acrescentando características que explicam que se trata ou não do mesmo conceito. Sem ignorar o fato de que um termo possa surgir genuinamente em um campo de especialidade, descarta-se a idéia de que um termo pertença genuinamente a um único âmbito de especialidade, pois o surgimento de termos nos mais variados âmbitos se dá significativamente por meio de transferências lexicais entre língua geral e linguagem de especialidade.
- Assumir o caráter polissêmico das unidades lexicais (fornecendo critérios que as identifiquem como tal) e a possível diversidade de especialidade da mesma unidade. Dar vazão para a banalização dos termos especializados e para a terminologização das palavras gerais.
- Admitir a existência real das sinonímias, fornecendo critérios que determinem seus valores como unidades terminológicas pertencentes a um contexto discursivo.
- A descrição dos termos deve se expandir às condições de combinação do discurso.
- Assumir a diversificação do discurso em função do tema, da perspectiva em que o tema está sendo tratado, dos interlocutores (emissor e destinatário), do nível de

especialização, do grau de formalidade, da situação, do propósito, e do tipo de discurso.

Os parâmetros (princípios, condições e supostos) da TCT são definidos com o intuito de cobrir quatro aspectos fundamentais, que segundo Cabré (1999) são os pilares para se construir uma teoria de base lingüística e de abordagem comunicativa: a poliedricidade (denominativa, cognitiva e funcional); a dupla função de representar e comunicar; a definição operacional (concepção da linguagem como real ou ideal, e da comunicação como uma atividade *in vivo* ou *in vitro*); e a diversidade aplicada, determinada pelas características pragmáticas da comunicação.

Sendo assim, Cabré (1999) apresenta os seguintes fundamentos da TCT:

- A terminologia é um campo interdisciplinar concebido com a contribuição de três teorias: uma teoria do conhecimento (que explique como se conceitualiza a realidade), uma teoria da comunicação (que descreva criteriosamente os tipos de situações que podem se dar o processo de comunicação, contemplando a amplitude e a diversidade desse processo), e uma teoria da linguagem (que considere as unidades terminológicas dentro da linguagem natural sem deixar de singularizar seu caráter terminológico).
- O objeto de estudo é o termo, unidade lexical que faz parte da língua geral e da gramática que descreve cada língua. Essa unidade pode exercer diferentes funções (como referencial, expressiva, conotativa) integradas em um discurso.
- Os termos são ativados singularmente por sua pragmática e adequação em um discurso. A forma é constante, mas seu conteúdo depende do âmbito, do tema,

da abordagem do tema, do tipo de texto, dos interlocutores e da situação discursiva.

- Os termos são unidades de forma e conteúdo, e o conteúdo sempre é simultâneo à forma. Se houver alguma relação de sinonímia ou de homonímia, trata-se de termos distintos relacionados. O conteúdo de um termo não é absoluto, mas sim relativo, dependendo do âmbito ou da situação de uso.
- Os conceitos de uma mesma especialidade mantêm relações de diferentes tipos. O conjunto dessas relações é denominado estrutura conceitual.
- O valor do termo é determinado pelo lugar que ele ocupa em uma estrutura conceitual de uma determinada matéria de acordo com um determinado trabalho.
- O objetivo da terminologia se divide em dois: teórico (descrever formal, semântica e funcionalmente as unidades que podem assumir papel de termos), e prático (recopilar os termos em um tema e em situações determinadas estabelecendo suas características de acordo com essa situação).
- As finalidades aplicadas das recopilações e análises dos termos são de representação do conhecimento aplicado e de sua transferência.

Embora esteja explícito que os fundamentos da TCT estão voltados para a aplicação da terminologia, ou seja, para o trabalho terminográfico, acredita-se que, além de adequado, é relevante a adoção desta teoria para analisar questões a respeito de instrumentos que, embora não sejam produtos da terminografia, são empregados com o objetivo de controlar uma dada terminologia.

Reafirmando que o propósito do presente trabalho é investigar semelhanças e diferenças teóricas entre dois modelos de representação do conhecimento, que por sua

vez estão a serviço de uma determinada terminologia, e, considerando a TCT como uma teoria coerentemente voltada para a comunicação efetiva entre especialistas, serão extraídos desta os fundamentos que nortearão a presente investigação. Como o foco do estudo está voltado para questões teóricas que envolvem os modelos de representação do conhecimento, serão utilizados como critérios de análise investigativa os seguintes fundamentos da TCT: a) o objeto (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 conteúdo, sem perder de vista o fato de que o valor do termo é determinado pelo lugar que ele ocupa em uma estrutura conceitual; 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, segundo as abordagens apresentadas na literatura. A idéia de utilizar os fundamentos da TCT como fonte inspiradora para a definição dos critérios de análise não se limita à extração de pontos fundamentais que servirão apenas de variáveis de investigação, mas também irá servir de referencial para acompanhar o raciocínio no processo de interpretações e inferências dos resultados obtidos.

No próximo capítulo são apresentados os procedimentos metodológicos que possibilitaram a investigação dos tesouros e ontologias com base nos fundamentos teóricos estabelecidos como relevantes.

#### **4 PROCEDIMENTOS METODOLÓGICOS: Método de Análise de Conteúdo**

Para caracterizar esta pesquisa é necessário apresentar os diversos pontos de vista que a envolvem e a tornam metodologicamente científica. Do ponto de vista da abordagem do problema, a pesquisa é qualitativa, pois está pautada em análises e interpretações de conteúdos para solucionar uma questão chave. Do ponto de vista de seus objetivos é uma pesquisa de caráter exploratório, e, na ótica dos procedimentos técnicos, trata-se de uma pesquisa documental, que emprega técnicas da Análise de Conteúdo para tratamento e análise dos dados. A escolha do método da Análise de Conteúdo, definido por Bardin (2003), se deve a dois motivos principais: a) seus procedimentos possibilitam uma análise com base em interpretações e inferências extraídas de conteúdos de documentos. A inferência, que é uma interpretação controlada por meio de variáveis ou indicadores, proporciona maior liberdade ao analista, sem perder o foco da investigação; e b) por se tratar de uma análise documental cujos documentos apresentam características homogêneas, o emprego de um método que se dirige fundamentalmente ao conteúdo de documentos parece o mais adequado. O método de Análise de Conteúdo é dividido em três fases, a saber:

1. Pré-análise;
2. Exploração do material;
3. Tratamento dos resultados, inferências e interpretações;

## 4.1 Pré-análise

Para Bardin (2003), a fase de pré-análise é a fase da organização, que visa operacionalizar e sistematizar as primeiras idéias. Esta fase auxilia na elaboração de um plano de análise que conduzirá o desenvolvimento das operações. A pré-análise possui três missões principais: *a escolha dos documentos* que serão submetidos à análise, caracterizada pela construção de um corpus de análise, *a formulação de hipóteses e objetivos*, e *a elaboração de indicadores* fundamentais à interpretação dos resultados.

Antecedendo as três missões mencionadas acima, Bardin (2003) aponta para a realização de uma leitura flutuante, que consiste em uma leitura primeira e descomprometida dos documentos a serem analisados para deles se extraírem impressões e orientações. É uma espécie de aproximação que visa à adaptabilidade do analista com o texto, a qual foi realizada neste estudo.

### 4.1.1 Construção do Corpus de Análise

A etapa da *escolha dos documentos* consiste na delimitação do universo (tipos de documentos a serem utilizados). O universo desta investigação é formado por artigos técnico-científicos e Teses. A área de abrangência é a Ciência da Informação e Ciência da Computação.

Seguindo a orientação de Bardin (2003), o próximo passo é a criação de um corpus de análise, que consiste no conjunto dos documentos que são submetidos à análise. Para a criação deste corpus, o método de Análise de Conteúdo sugere atenção

para quatro regras principais: a) *regra da exaustividade*: é preciso considerar todos os elementos do corpus de análise, não deixar nenhum elemento fora do processo de análise, para tanto, zela-se pela não-seletividade; b) *regra da representatividade*: a análise é realizada com base em uma amostra de significativa representação do universo inicial. Tal regra deve se ater para a questão da homogeneidade do universo, pois quanto menor a homogeneidade do universo maior deve ser a amostra; c) *regra de homogeneidade*: os documentos analisados devem ser homogêneos, assim como as técnicas de coleta de informações a respeito dos documentos. Demasiada particularidade em um determinado tipo de documento ou em determinado processo de coleta de informações pode deturpar (causar ruídos) os resultados; d) *regra de pertinência*: os documentos devem ser fontes de informações adequadas aos objetivos da análise.

No que diz respeito à construção do corpus de análise, cabe ressaltar que não foi possível zelar pela regra de *exaustividade*, que por sua vez zela pela não-seletividade. Tal afirmação se deve ao fato de que uma seleção dos documentos referentes à ontologia foi imprescindível para separar os textos que abordam tal tema como objeto da Filosofia daqueles que o aborda como objeto da Ciência da Informação ou da Computação. Outra seleção necessária diz respeito ao tipo de artigos levantados, pois, por se tratar de uma pesquisa teórica, a análise utilizou somente artigos de conteúdo teórico, ou seja, artigos que apresentem alguma reflexão teórica acerca do tema. Os artigos referentes às aplicações de tesauros ou ontologias em situações específicas, assim como propostas de avaliação e usabilidade, foram descartados.

Para a construção do corpus de análise foram necessários dois processos seletivos. O primeiro consistiu em selecionar todos os textos de reflexão teórica (referentes aos tesouros e às ontologias) recuperados nas buscas às bases de dados eleitas para a pesquisa (as bases de dados são descritas mais adiante). Essa primeira seleção foi realizada por meio das leituras dos resumos dos textos e acarretou na constituição de um primeiro corpus (Corpus 1). O segundo processo correspondeu em selecionar somente os textos que continham em seu conteúdo as variáveis (índices) da análise. Como se pode observar no item 4.1.1.2, os índices da análise são: *termos, conceitos e objetivos*. Melhor dizendo, foi realizada uma leitura integral dos textos formantes do Corpus 1 buscando identificar e selecionar somente aqueles que de alguma forma abordavam em seu conteúdo aspectos relativos aos termos e/ou aos conceitos presentes no modelo de representação do conhecimento (tesauro e/ou ontologia), bem como os textos que tratavam dos objetivos desses modelos. Tal processo acarretou na constituição de um segundo corpus (Corpus 2). O corpus de análise da pesquisa nada mais é que o Corpus 2 dividido tematicamente em três grupos: a) tesouros, b) ontologias e c) tesouros e ontologias. Este último grupo se deveu ao fato de que um dos textos abordava ambos os temas. Essa divisão por tema foi realizada apenas com o intuito de facilitar o tratamento e análise das informações contidas nos textos. O esquema abaixo ilustra a construção do corpus de análise:

CORPUS 1 → Textos teóricos (tesauro e ontologia) recuperados nas bases de dados

CORPUS 2 → Textos selecionados do Corpus 1 com base nos índices de análise

CORPUS DE ANÁLISE → Corpus 2 dividido por tema

Cabe ressaltar que um pré-teste para a construção do corpus foi realizado no início do mês de março de 2007 com o intuito de verificar a viabilidade do método aqui exposto. A estratégia de busca seguiu os passos descritos nos próximos parágrafos e obteve como resultado 81 (oitenta e um) títulos. Após uma depuração por meio da releitura dos resumos dos 81 textos, que aparentemente apresentavam reflexões teóricas, constatou-se um corpus inicial (Corpus 1) composto por 67 (sessenta e sete) títulos de documentos. Os números encontrados no pré-teste demonstraram a viabilidade de se utilizar tal estratégia.

A *representatividade* do corpus foi assegurada pela escolha das bases de dados utilizadas para o levantamento bibliográfico, pois as mesmas representam uma parcela significativa da literatura referente à temática. Como mencionado anteriormente, os documentos analisados foram artigos e relatórios de pesquisa (predominantemente *homogêneos* em estrutura e linguagem de especialidade), e as técnicas de análise também seguiram um padrão homogêneo. O respeito às regras descritas acima acarretou no zelo pela *pertinência* do material.

O levantamento bibliográfico foi realizado durante a última semana do mês de abril de 2007 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**<sup>27</sup> - BDTD. A LISA é um índice internacional de resumos de artigos projetado para profissionais bibliotecários e especialistas da informação. Cobre atualmente 440 periódicos de 68 países e em mais de 20 línguas diferentes. A Base Wilson consiste em uma coleção de periódicos, teses e

---

<sup>27</sup> Instituto Brasileiro de Informação em Ciência e Tecnologia

livros com textos integrais ou apenas resumos. Dentre as áreas do conhecimento cobertas pela Base está a Ciência da Informação. A BDTD integra os sistemas de informação de Teses e Dissertações existentes nas Instituições de Ensino Superior (IES) brasileiras. Os documentos que não estavam integralmente disponíveis *on-line* foram solicitados pela comutação bibliográfica por meio do sistema COMUT.

Como estratégia de busca foram utilizados os termos TESAURO, THESAURUS, THESAURI, ONTOLOGIA, ONTOLOGY e ONTOLOGIES nos dispositivos de consulta disponíveis nas respectivas bases de dados. Nas bases da Wilson e da BDTD a busca de cada termo foi realizada utilizando a opção de busca avançada por *Subject* (Assunto). Para a busca nas bases da LISA foi utilizada a busca avançada por *Keywords* (Palavras-Chave) para termos referentes ao tema tesouro, e, busca avançada por *Title* (Título) para os termos referentes ao tema ontologia, pois na realização do pré-teste foi constatado um número excessivo de ocorrências referente a esse tema na busca por *Keywords*. Outro dispositivo de busca que foi lançado mão para a consulta às bases da LISA foi o de *Peer-Reviewed Journals* (periódicos revisados pelos pares). O período de abrangência foi de dez anos (de 1998 a 2007) e os idiomas foram delimitados em português (para contemplar estudos realizados no Brasil), inglês (pelo fato de que a maior parte dos textos científicos da Ciência da Informação está escrita em inglês) e espanhol (pelo fato de que a teoria do presente estudo é espanhola). O idioma francês não foi incorporado para evitar um número excessivo de documentos, o que inviabilizaria a análise detalhada dos textos.

O resultado da busca foi muito próximo aos números obtidos na realização do pré-teste. Após a consulta às bases de dados descritas acima, foram identificados 78

(setenta e oito) documentos que, sob um primeiro olhar (leitura dos resumos), apresentavam abordagens teóricas a respeito dos tesouros e/ou ontologias. Para definir o Corpus 1 foi realizada uma leitura detalhada dos resumos, introduções e conclusões dos 78 documentos com o fim de selecionar os textos que efetivamente apresentavam uma abordagem teórica. É importante frisar que os textos que apresentavam estudos de caso de abordagem metodológica, mas, que de alguma maneira expunham reflexões de ordem teórica ou conceitual, podendo assim contribuir para o objetivo da pesquisa, foram também selecionados para compor o Corpus 1, constituído de 62 (sessenta e dois) documentos. A etapa seguinte consistiu na leitura integral dos textos do Corpus 1 visando selecionar somente aqueles que apresentavam em seu conteúdo alguma referência a qualquer dos índices da investigação (Termos, Conceitos e Objetivos). Como consequência foi construído um Corpus 2 composto por 34 (trinta e quatro) documentos, sendo trinta e três artigos técnico-científicos e uma Tese. Este resultado por si só já aponta um dado importante para a investigação. A incidência de textos de abordagem teórica referente aos modelos de representação do conhecimento foi relativamente pequena se comparada com o número de textos que apresentavam relatos de pesquisas aplicadas, haja vista que, na etapa de levantamento bibliográfico efetuada durante a construção do Corpus 1, centenas de documentos foram encontradas, tanto para tesouro quanto para ontologia. Mais uma vez se evidencia a relevância de um estudo de abordagem teórica para o tema proposto.

Dos trinta e quatro documentos selecionados para o Corpus 2, dezessete correspondem ao tema ‘tesouro’, sendo dezesseis artigos e uma tese, dezesseis artigos são relativos ao tema ‘ontologia’, e um artigo cobre ambos os temas. Com essa divisão foi definido o Corpus de Análise, que está arrolado no Apêndice A.

#### 4.1.2 Hipóteses, Índices e Indicadores

Na fase destinada à *formulação das hipóteses e objetivos*, Bardin (2003) define hipótese como sendo uma afirmação provisória passível de verificação e comprovação, sendo o objetivo a finalidade geral proposta pela análise. A autora ressalta que não há a necessidade de se criar um corpus de hipóteses previamente, pois a formulação dessas hipóteses muitas vezes consiste em explicitar direções de análise que funcionem durante o processo de análise (hipóteses implícitas).

O objetivo da análise, como já foi devidamente descrito, é apurar elementos teórico-conceituais que possibilitem a identificação de características que aproximam e que distanciam tesouros e ontologias. Optou-se em trabalhar com hipóteses implícitas que se manifestaram no decorrer da análise, principalmente na exploração do material.

A fase de *elaboração dos índices e indicadores* consiste na eleição de elementos que melhor explicitarão o conteúdo de acordo com os objetivos da análise. Considerando a afirmação de Bardin (2003) de que textos são índices informativos manifestados por meio de uma análise, tem-se que à referida fase compete a função de indicar quais são esses índices, e organizá-los sistematicamente com indicadores (frequência, em caso de análise quantitativa, e presença, em caso de análise qualitativa) relativos a eles.

Os índices da análise são as unidades lexicais consideradas primordiais para o alcance do objetivo deste estudo, que foram extraídas dos fundamentos da TCT,

conforme descrito na fundamentação do marco teórico desta dissertação. Os índices são: “termo”, “conceito” e “objetivo”. Por se tratar de uma análise qualitativa, os indicadores que regeram os trabalhos foram a ‘presença’ ou ‘ausência’ dos respectivos índices nos textos analisados. Ou seja, somente os textos que faziam alguma referência aos termos, conceitos e objetivos, foram incluídos na análise.

#### 4.1.3 Categorização

Como medida de classificação dos dados levantados, Bardin (2003) sugere o processo de categorização, que consiste no agrupamento em classes dos elementos convergentes em suas características. Segundo a autora, para uma eficiente categorização é necessário estar atento aos seguintes critérios: a) *exclusão mútua*: um elemento não pode existir em mais de uma classe; b) *homogeneidade*: uma única característica deve governar uma classe, ou seja, em uma categoria, somente uma característica pode servir como requisito para abrigar ou não determinado registro; c) *pertinência*: uma categoria para ser considerada pertinente necessita estar em sintonia com o material de análise e com a teoria que fundamenta a pesquisa; d) *objetividade e fidelidade*: devem-se aplicar os mesmos critérios de análise para cada parte de cada material; e e) *produtividade*: um processo de categorização é tido como produtivo quando fornece resultados férteis no que tange os índices de inferência.

Antecipando a escolha das categorias que forneceram subsídios para a análise dos conteúdos documentais, foram eleitas 04 (quatro) unidades de registro (UR) que serviram para descrever o documento como um todo, em sua forma e conteúdo,

identificando a produção bibliográfica analisada. As unidades de registro elencadas foram:

1. *Referencia bibliográfica do documento*: registro dos dados do autor, título, editor, data de publicação, volume, página, mês, ano, e demais dados complementares referentes à bibliografia.
2. *Indicação do tema do documento*: descreve se o documento aborda o tema tesouro, ontologia ou ambos.
3. *Resumo do conteúdo do documento*: registro dos elementos que possibilitam o entendimento geral do texto.
4. *Observações*: registro de informações relevantes, tais como aspectos estruturais, tipologias das LDs, teorias utilizadas e áreas de aplicação, que não foram contempladas nas demais categorias.

Respeitando os critérios descritos por Bardin (2003), este estudo apresenta 03 (três) categorias de análise que são baseadas nos índices da pesquisa, e que visam os seguintes registros:

1. *Termo (unidade terminológica)*: registra uma síntese contendo a perspectiva apresentada no documento com relação às unidades terminológicas 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.

Para o registro dos elementos correspondentes às categorias 1, 2 e 3, que contemplam os índices da análise, ou seja, são os elementos que possibilitam analisar as diferenças e semelhanças entre tesouros e ontologias, focaram-se os seguintes parâmetros baseados na ótica funcionalista da TCT: a) **Categoria 1 (Termo)** – as funções das unidades terminológicas, a relação forma-conteúdo do termo e, os níveis de relacionamento entre os termos; b) **Categoria 2 (Conceito)** – a relação com a designação do termo e a relação entre os próprios conceitos e; c) **Categoria 3 (Objetivo)** – finalidades das linguagens documentárias em questão.

É importante frisar que foi utilizada uma base de dados usando o aplicativo *Microsoft® Access 2003* para a operacionalização das unidades de registro e das categorias. A utilização desse aplicativo possibilitou a criação de uma base de dados que potencializou o manuseio das informações extraídas dos textos analisados, bem como a análise dessas informações. Ou seja, a base de dados foi utilizada no processo de coleta e registro das informações<sup>28</sup>, e também agilizou o processo de análise dos resultados. Esses processos se inscrevem efetivamente na etapa de Exploração do Material, descrita na seção seguinte.

---

<sup>28</sup> Opta-se pela expressão ‘coleta de informações’ em vez de ‘coleta de dados’ devido ao fato deste processo ter sido realizado por meio da técnica de fichamento de textos. Sendo assim, considera-se inadequada a idéia de que elementos extraídos de uma ação de fichamento sejam considerados apenas dados.

## 4.2 Exploração do Material

A exploração do material consiste na administração sistemática das decisões tomadas na fase da pré-análise e categorização. É o cumprimento das etapas preestabelecidas no projeto, ou seja, neste estudo constitui na operacionalização do processo de leitura dos textos do corpus de análise, visando à extração dos elementos necessários para analisar as variáveis (índices), e descrição desses elementos em suas respectivas unidades de registro e categorias.

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 minuciosa de cada documento foi devidamente fichada, com base nas ‘unidades de registro’ e ‘categorias’ descritas no item 4.1.2, e registrada na referida base de dados. Para cada texto analisado foi elaborada uma ficha (para fins de esclarecimento, cada página da base de dados corresponde a uma ficha) contendo registros relativos às unidades de registro (Referência, Tema, Resumo e Observações) e às categorias (Termo, Conceito e Objetivo). A Figura 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. 348-359, set./dez. 2006. Disponível em: < <a href="http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&amp;layout=abstract">http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&amp;layout=abstract</a> >. Acesso em: 29.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 conceituação 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. A 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.

**Figura 1 – Base de Dados Criada para o Fichamento**

Cabe afirmar que as informações registradas na base de dados seguiram o estilo e a linguagem de cada autor cujo texto foi analisado, com o intuito de respeitar a perspectiva apresentada em cada documento. Por se tratar de uma pesquisa de ordem qualitativa que se apropria de um método que dá vazão à interpretação e possibilita percorrer caminhos com base em hipóteses que surgem em meio à investigação, evitou-se a padronização das informações registradas. Como este estudo consiste em uma análise teórico-comparativa, o modo de cada autor expressar o conteúdo pode acarretar em significativa contribuição à análise realizada na etapa de tratamento e interpretação dos resultados.

O *Microsoft*® *Access* possibilita, dentre outros dispositivos, a geração automática de relatórios organizados conforme a necessidade requerida por quem o está utilizando. É possível gerar relatórios com quaisquer das categorias ou unidades que alimentaram sua base, além de definir flexivelmente a ordem de apresentação das informações. Lançando mão desse dispositivo, após o fichamento do material analisado, foram gerados dois relatórios para auxiliar o processo de análise dos resultados. O Relatório 1 (exposto no Apêndice A) descreve todas as informações registradas no processo de coleta de informações organizadas de acordo com a ordem de fichamento: Referência, Tema, Resumo, Termo, Conceito, Objetivo e Observações. O Relatório 2 (exposto no Apêndice B) descreve, em ordem temática, apenas as informações referentes às categorias de análise: Termo, Conceito e Objetivo. O Relatório 2 foi utilizado para analisar as variáveis que possibilitaram identificar as diferenças e semelhanças entre tesauros e ontologias, ao passo que o Relatório 1, além de apresentar as informações gerais de cada documento, auxiliou a redação deste estudo.

### **4.3 Tratamento dos Resultados Obtidos e Inferências**

Após elaboração dos Relatórios 1 e 2 iniciou-se a etapa de tratamento dos resultados obtidos com a exploração do material. Os resultados foram tratados com vistas a viabilizar a etapa final da análise do conteúdo, a inferência. Seguindo a orientação do método de Bardin (2003), que afirma que a Análise de Conteúdo possibilita uma leitura não-aderente (se é que isto seja possível) em que o leitor tem a oportunidade de se distanciar dos textos analisados e captar informações suplementares, foram eleitos o ‘pólo da análise’ e as ‘variáveis de inferência’. A autora, apoiada pelos elementos constitutivos da visão clássica da comunicação, afirma que os pólos de

observação para a interpretação da análise são: a mensagem (significação e código), o suporte (canal), e o interlocutor (emissor e receptor). A presente análise focou a ‘mensagem’ como pólo de observação.

Para a análise baseada na mensagem, Bardin (2003) afirma existir dois níveis possíveis: do código (significante) e da significação (significado). Apropriando-se da afirmação da autora de que o estudo formal do código nem sempre é necessário e que a análise pode ser realizada a partir das significações da mensagem, esta investigação foi baseada no pólo ‘mensagem’ no nível da ‘significação’.

As variáveis de inferência foram escolhidas de acordo com a percepção, ocorrida durante todos os processos anteriores, sobretudo durante a leitura e fichamento dos textos, de elementos recorrentes na grande maioria dos documentos analisados. Tais elementos foram selecionados e considerados como variáveis de inferência inseridas nas categorias da análise, a saber: a) Categoria Termo – variáveis: 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). Essas atribuições serão mais bem percebidas nos Quadros 1, 2, 3 e 4.

Com o intuito de facilitar a comparação dos aspectos que caracterizam os tesouros e as ontologias foram elaborados um Quadro 1, que descreve o conteúdo concernente aos tesouros, e um Quadro 2, que descreve o conteúdo referente às ontologias, ambos orientados pelas categorias de análise e variáveis de inferência. Os

conteúdos dos referidos quadros foram extraídos da análise realizada sobre o Relatório 2 e pautados na significação da mensagem, conforme segue.

**TESAURO: Análise da Mensagem (Nível: Significação)**

<b>CATEGORIA DE ANÁLISE</b>	<b>VARIÁVEL DE INFERÊNCIA</b>	<b>CONTEÚDO</b>
<b>Termo</b>	<b>Definição/função</b>	<ul style="list-style-type: none"> <li>• Representante lingüístico do conceito</li> <li>• Unidade lexical que se situa nos discursos especializados</li> <li>• Palavra ou frase de algum domínio de conhecimento. Preferencialmente substantivos e frases nominais</li> <li>• Signo verbal que designa um objeto/referente</li> <li>• Signo verbal que denota/denomina o conceito</li> <li>• Unidade lingüística e unidade de conhecimento especializado</li> <li>• Um elemento do conceito que o sintetiza como um todo</li> <li>• Definiendum. O definiendum (termo) + o definiens (atributos predicáveis do objeto) formam a definição do conceito</li> <li>• Designar e expressar um objeto do pensamento (conceitos e noções)</li> <li>• Evitar a superposição de conceitos e diminuir a flexibilidade da linguagem (TGT)</li> <li>• Minimizar ambigüidades</li> <li>• Descrever um conceito de maneira unívoca em um sistema de informação</li> <li>• Especificam forma e significado, expressam significados</li> <li>• Sintetizar um conceito e viabilizar sua comunicação</li> </ul>
	<b>Tipos de termos</b>	<ul style="list-style-type: none"> <li>• Termo equivalente</li> <li>• Termo preferido (UP), ou controlado, ou normalizado, (descriptor)</li> <li>• Termo não-preferido ou não-descriptor, ou ainda, termo proibido (USE)</li> <li>• Termo geral (genérico) TG</li> <li>• Termo específico TE</li> <li>• Termo relacionado TR</li> <li>• Termo simples e composto</li> <li>• Termos polissêmicos (homógrafos)</li> <li>• Identificadores (nomes próprios)</li> <li>• Termo qualificado ex.: Rádio (metal), Terra (planeta)</li> </ul>

	<b>Relação entre termos</b>	<ul style="list-style-type: none"> <li>• De equivalência (sinonímia), hierárquico (gênero/espécie, todo/parte) e associativo (não hierárquico).</li> <li>• Nota explicativa (NE)</li> <li>• Depende das características do conceito de cada termo</li> <li>• Dá-se por meio das características de divisão (ou qualidades essenciais, exclusivas e homogêneas) que os tornam singulares e pertencentes a uma única faceta</li> </ul>
	<b>Relação com o conceito</b>	<ul style="list-style-type: none"> <li>• Parte da literatura defende que o termo é uma unidade de conhecimento que designa um conceito</li> <li>• Parte da literatura defende que o termo é uma unidade lexical que remete a uma unidade de conhecimento</li> </ul>
<b>Conceito</b>	<b>Definição/função</b>	<ul style="list-style-type: none"> <li>• Representação abstrata percebida e interpretada de algo real</li> <li>• Representação mental e única</li> <li>• Unidade de conhecimento, incluindo a definição como elemento essencial para a fixação do conceito (Dahlberg)</li> <li>• Unidade de representação</li> <li>• Objeto do pensamento</li> <li>• Conjunto formado pelas características próprias de um objeto e por um termo que as sintetizam</li> </ul>
	<b>Organização dos conceitos</b>	<ul style="list-style-type: none"> <li>• São ordenados em classes cognitivamente pré-estabelecidas</li> <li>• Classes (agrupamento de elementos por semelhança)</li> <li>• Uso de categorias</li> </ul>
	<b>Relação entre conceitos</b>	<ul style="list-style-type: none"> <li>• Dá-se de acordo com as características (predicados do objeto/referente) do conceito</li> <li>• Relação entre conceitos de diversas línguas (voltado à comunicação entre tesouros multilingües): abstrato/genérico (é o tipo gênero/espécie); partitivo; descendência (um conceito B descende de um conceito A); instrumental (um conceito B é instrumento para alcançar um conceito A); causa/efeito (um conceito B é causa da existência de um conceito A); benefício (um conceito B é útil para um conceito A); prejuízo (causa de efeitos adversos); matéria (um conceito A consiste materialmente de um conceito B); aparência (enxerga o conceito como ele é reconhecível pela visão humana); processo (um conceito B indica um processo em um conceito A); estado (um conceito B é um estado de um conceito A)</li> <li>• Lógica: baseada no grau de semelhança entre os</li> </ul>

		<p>objetos e conceitos. Pode ocorrer por lógica de subordinação (do geral para o específico) e por superordenação (do específico para o geral). Pode ser gênero/espécie ou todo/parte (é possível encontrar na literatura a relação partitiva como pertencente ao relacionamento ontológico). É o resultado da própria compreensão dos conceitos</p> <ul style="list-style-type: none"> <li>• Ontológica: proximidade situacional dos elementos na realidade (contigüidade dos conceitos no espaço). Pode ocorrer por coordenação, por encadeamento, como causa e efeito ou seqüenciais. Relação entre conceito e realidade</li> <li>• Equivalência: quando um conceito é denominado por mais de uma denominação</li> <li>• O relacionamento é determinado pelo uso empregado pelo domínio e pelos fatores característicos do domínio</li> </ul>
	<b>Relação com o termo</b>	<ul style="list-style-type: none"> <li>• O conceito é representado lingüisticamente por um termo</li> <li>• O conceito é designado por um termo</li> <li>• Conceitualiza o termo, como o significado de uma palavra</li> <li>• Relação termos-conceitos de uma mesma língua (sinônimos e polissemias)</li> <li>• Univocidade, para cada conceito um termo</li> <li>• O conceito é formado por um termo (definiendum) mais os predicados do referente (definiens)</li> </ul>
<b>Objetivo</b>	<b>Teórico (relacionado com a terminologia em geral)</b>	<ul style="list-style-type: none"> <li>• Orientar qual o termo mais adequado para representar um conceito</li> <li>• Facilitar a inter-relação entre linguagem natural e linguagem artificial</li> <li>• Esclarecer barreiras lingüísticas</li> <li>• Controlar sinônimos e homógrafos</li> <li>• Sintetizar um conjunto (estrutura) conceitual</li> <li>• Propor um sistema de símbolos lingüísticos para agrupar informação singular relacionada ou direcionar para grupos mais gerais ou mais específicos de uma temática</li> <li>• Gerir e comunicar a linguagem específica usada nas organizações</li> <li>• Tesouros do tipo “taxonomia corporativa”: auxiliar o entendimento e uso de um assunto ou linguagem específica; facilitar o entendimento conceitual e comunicacional em um nível lingüístico que conceba um vocabulário corporativo</li> </ul>
	<b>Prático (relacionado com os sistemas de informação)</b>	<ul style="list-style-type: none"> <li>• Servir como ponte que liga necessidades de informação (busca) e sistemas de recuperação de informação</li> <li>• Organizar informação</li> <li>• Coordenar o vocabulário usado pelo usuário de um sistema de informação. Auxiliar a consulta do usuário expondo sistematicamente uma classificação que o direcione no momento da busca</li> </ul>

		<ul style="list-style-type: none"> <li>• Economizar o tempo do usuário</li> <li>• Potencializar a recuperação da informação atuando como interface entre informação e usuário</li> <li>• Auxiliar a indexação por assunto de documentos em um sistema de informação</li> <li>• Guiar a indexação baseada em linguagem natural</li> <li>• Auxiliar a representação do conteúdo documental, minimizando ruídos como ambigüidades</li> <li>• Maximizar a pertinência da recuperação da informação</li> <li>• Servir como vocabulário oficial que rege a indexação e a recuperação de documentos</li> <li>• Alcançar uma normalização da terminologia do sistema de informação a ser utilizado</li> <li>• Propor um conjunto estruturado de termos sob a base de um sistema de conceitos aptos a organizar os conteúdos de um sistema</li> </ul>
--	--	--

**Quadro 1 – Elementos que Caracterizam os Tesouros**

Como dito anteriormente, o conteúdo do Quadro 1, bem como do Quadro 2, foram extraídos das informações contidas no Relatório 2, portanto, a descrição dos referidos conteúdos é o resultado de uma síntese interpretativa do analista com base nas informações coletadas e registradas na base de dados (fichamento). Tal fato leva a constatação de que o processo de inferência já se faz presente na elaboração de ambos os quadros. Por se tratar de uma interpretação, não houve uma padronização rígida quanto ao julgamento daquilo que era ou não relevante para compor os conteúdos dos quadros, apenas se seguiu uma lógica interpretativa decorrente da leitura minuciosa do Relatório 2. Como esta análise está pautada na lógica de inferência, a composição dos conteúdos dos quadros foram constituídos pelas informações que, segundo o analista, sobressaíram quando da leitura do relatório.

**ONTOLOGIA: Análise da Mensagem (Nível: Significação)**

<b>CATEGORIA DE ANÁLISE</b>	<b>VARIÁVEL DE INFERÊNCIA</b>	<b>CONTEÚDO</b>
<b>Termo</b>	<b>Definição/função</b>	<ul style="list-style-type: none"> <li>• Etiqueta de conceitos</li> <li>• Viabilizar a formulação de consultas (usuário) fazendo uso de conceitos predeterminados por especialistas</li> <li>• São usados para definir formalmente `coisas` em domínio de interesse</li> <li>• Refere-se a um conceito</li> <li>• Representa um conceito ou classe</li> <li>• Expressa o conceito</li> </ul>
	<b>Tipos de termos</b>	<ul style="list-style-type: none"> <li>• Tipo (universal, geral)</li> <li>• Instância (particular, específico)</li> <li>• Relação (qualidade)</li> <li>• Entidade (substância)</li> </ul>
	<b>Relação entre termos</b>	<ul style="list-style-type: none"> <li>• De acordo com o significado/conceito de cada termo</li> <li>• Realizado pelos especialistas. Dá-se por meio do consenso de um determinado domínio</li> <li>• A relação semântica está diretamente ligada à apresentação sintática</li> </ul>
	<b>Relação com o conceito</b>	<ul style="list-style-type: none"> <li>• O termo contém o conceito</li> </ul>
<b>Conceito</b>	<b>Definição/função</b>	<ul style="list-style-type: none"> <li>• Entidades, atributos e processos</li> <li>• Visão abstrata e sintetizada de um ente</li> <li>• Conceitos são classes</li> <li>• Conceitos são representações mentais refletidas por meio de termos</li> </ul>
	<b>Organização dos conceitos</b>	<ul style="list-style-type: none"> <li>• Conceitos concretos principais (os quais apresentam propriedades do domínio, assim como seus relacionamentos) e conceitos abstratos (são as características)</li> <li>• Conceitos são organizados em taxonomias</li> </ul>
	<b>Relação entre conceitos</b>	<ul style="list-style-type: none"> <li>• Dá-se por meio dos significados dos conceitos</li> <li>• Dá-se por meio da intensão (conjunto de características) do conceito</li> </ul>

		<ul style="list-style-type: none"> <li>• As relações conceituais (intensionais) são definidas em um espaço do domínio, e podem ser representadas em grupos de mundos possíveis</li> <li>• Interligações entre conceitos mais refinados e conceitos mais periféricos formam as relações adicionais</li> <li>• Por meio de submissão (hierarquia)</li> <li>• Por meio de axiomas (expressam relacionamentos entre conceitos e definem a interpretação pretendida)</li> <li>• Relacionamento superordenado entre classes (conceitos)</li> <li>• Relacionamento sub-ordenado entre classes e propriedades, chamado também de relacionamento poli-hierárquico, pois uma classe pode ser uma sub-classe de uma ou mais classes superordenadas</li> <li>• A relação semântica entre conceitos se dá por meio das formas sintáticas de um discurso</li> </ul>
	<b>Relação com o termo</b>	<ul style="list-style-type: none"> <li>• O conceito é parte do termo</li> <li>• Um termo reflete um conceito</li> <li>• O conceito é expresso e personificado pelo termo</li> </ul>
<b>Objetivo</b>	<b>Teórico (relacionado com a terminologia em geral).</b>	<ul style="list-style-type: none"> <li>• Reduzir ou eliminar confusões terminológicas e conceituais</li> <li>• Conceber uma estrutura conceitual uniforme</li> <li>• Potencializar a comunicação e cooperação entre pessoas</li> <li>• Descrever conceitos necessários para falar sobre determinado assunto</li> <li>• Esclarecer o significado pretendido de um vocabulário por meio de axiomas</li> <li>• Fornecer um mapa semântico aos campos individuais e aos relacionamentos entre os campos, servindo como uma ferramenta de referência</li> <li>• Criar uma estrutura lógica, uma filosofia, uma classificação em uma disciplina (domínio)</li> </ul>
	<b>Prático (relacionado com os sistemas de informação).</b>	<ul style="list-style-type: none"> <li>• Identificar e definir um conjunto de conceitos relevantes que caracterize um dado domínio de aplicação</li> <li>• Melhorar a organização da informação nas empresas</li> <li>• Interoperabilizar sistemas</li> <li>• Proporcionar a interoperabilidade dos serviços ofertados na web, potencializando os níveis de serviços ofertados, ou seja, sustentar a web semântica</li> <li>• Propiciar benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação)</li> <li>• Possibilitar a consulta a um sistema automatizado de informação utilizando termos conceitualizados e especificados por especialistas</li> </ul>

		<ul style="list-style-type: none"> <li>• Ajudar a responder perguntas em um corpo de informação. Isto inclui as seguintes funções: relacionar conceitos aos termos e fornecer definições; esclarecer conceitos, contextualizando-os em uma classificação; relacionar conceitos e termos ou ícones por meio das disciplinas, das línguas e das culturas</li> <li>• Melhorar a comunicação e o aprendizado, assistindo escritores e leitores, fornecendo estruturas conceituais que sustentam o ensino</li> <li>• Fornecer base conceitual para projetos de pesquisa e execução</li> <li>• Sustentar a recuperação da informação em uma base de conhecimento</li> <li>• Servir como um dicionário para o uso humano e como um dicionário/base de conhecimento para o processamento de linguagem natural</li> <li>• Fornecer um entendimento comum e compartilhado de um determinado domínio</li> <li>• Possibilitar serviços baseados em operacionalizações semânticas</li> <li>• Ligar parte do conhecimento humano aos processamentos computacionais</li> <li>• Descrever a semântica de um domínio de modo que seja compreensível por homens e máquinas</li> <li>• Melhorar a consistência e o reuso da informação e o compartilhamento do conhecimento</li> <li>• Possibilitar, juntamente com aplicações lógicas, a construção de modelos computacionais para um determinado domínio de aplicação</li> <li>• Representar o mundo real (ou parte do mundo real) em um dado domínio</li> <li>• Recuperação da informação e raciocínio automático de um dado domínio</li> <li>• Capturar significados e relações entre os significados</li> <li>• Servir como um catálogo dos tipos de coisas que existem em um domínio D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D</li> <li>• 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</li> <li>• 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</li> <li>• Oferecer meios concisos e sistemáticos para definir a semântica dos recursos web</li> <li>• (Para a web semântica) proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação</li> </ul>
--	--	--

**Quadro 2 – Elementos que Caracterizam as Ontologias**

Com a caracterização dos elementos dos tesouros e das ontologias, devidamente registrados de acordo com as categorias de análise e variáveis de inferência, o passo seguinte foi o cruzamento dos conteúdos dos Quadros 1 e 2 com o fim de identificar as semelhanças e diferenças existentes ambos os modelos de representação do conhecimento.

#### *4.3.1 As Características Semelhantes*

Esta etapa consiste na comparação dos Quadros 1 e 2 com o propósito de identificar as características que convergem entre os tesouros e as ontologias. Para tanto, foram realizados três quadros (uma para cada categoria de análise) que apresentam os pontos de convergência<sup>29</sup> entre ambos modelos de representação do conhecimento: Quadro 3, relativo à Categoria Termo; Quadro 4, relativo à Categoria Conceito ; e Quadro 5, relativo à Categoria Objetivo. Cabe ressaltar que o conteúdo das informações relativas ao tesouro e à ontologia foi agrupado (colocado um ao lado do outro) de acordo com o grau de semelhança apresentado em cada característica descrita. Para maior clareza na descrição das semelhanças, seguem-se (páginas seguintes) os Quadros 3, 4 e 5 e seus respectivos comentários.

---

<sup>29</sup> Além das características que convergem diretamente, foram descritas também aquelas características que, embora não explicitem uma concordância direta, apresentem elementos que podem contribuir para a reflexão a cerca da aproximação de tesouros e ontologias.

### Semelhanças entre Tesouros e Ontologias no Tocante aos Termos

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	TESAURO	ONTOLOGIA
<b>Termo</b>	<b>Definição/ função</b>	<ul style="list-style-type: none"> <li>➤ Signo verbal que denota, denomina, expressa, designa um conceito</li> <li>➤ Representante lingüístico do conceito</li> <li>➤ Especifica forma e significado, expressa significado</li> <li>➤ Sintetiza um conceito e viabiliza sua comunicação</li> </ul>	<ul style="list-style-type: none"> <li>➤ Etiqueta que expressa o conceito</li> <li>➤ Representa um conceito ou classe</li> <li>➤ Refere-se a um conceito</li> </ul>
	<b>Tipos de termos</b>	<ul style="list-style-type: none"> <li>➤ Termo geral (genérico) TG</li> <li>➤ Termo específico TE</li> </ul>	<ul style="list-style-type: none"> <li>➤ Um Tipo (universal, geral)</li> <li>➤ Uma Instância (particular específico)</li> </ul>
	<b>Relação entre termos</b>	<ul style="list-style-type: none"> <li>➤ Dá-se por meio das características do conceito de cada termo</li> <li>➤ Hierárquico (gênero/espécie, todo/parte)</li> <li>➤ Dá-se por meio das características de divisão (ou qualidades essenciais, exclusivas e homogêneas) que os tornam singulares e pertencentes a uma única faceta</li> </ul>	<ul style="list-style-type: none"> <li>➤ Dá-se de acordo com o significado (conceito) de cada termo</li> <li>➤ Relação de um Tipo (geral) com uma Instância (particular)</li> <li>➤ Relação de qualidade (quando um termo representa uma qualidade de outro)</li> </ul>
	<b>Relação com o conceito</b>	Ver relação conceito-termo	Ver relação conceito-termo

**Quadro 3 – Inferências Relativas às Semelhanças da Categoria Termo**

*Comentário do Quadro 3*

- *Termo* é um signo verbal (lingüístico) que representa, denomina, designa, denota, sintetiza, expressa, especifica um *conceito*, ou seja, *o termo é um signo verbal que ‘mostra’ um conceito*.

Este estudo, apropriando-se da noção defendida pela TCT de que um termo é constituído por forma e conteúdo, opta em sintetizar os verbos utilizados na literatura por meio do verbo ‘mostrar’. Cada uma das ações manifestadas pelos verbos acima essencialmente exerce a mesma ação<sup>30</sup>, *mostrar um conceito*, no sentido de fazer ver um conceito. Quando um especialista faz uso de um tesauro ou de uma ontologia, ou mesmo quando consulta um texto, ele enxerga os conceitos especializados por meio dos termos especializados, fato este que torna segura a afirmação de que o conceito se manifesta por meio do termo, e que aquele está contido neste.

- Os termos genéricos e específicos dos tesouros podem ser considerados nas ontologias como os termos universais (Tipo) e particulares (Instâncias).
- *As características do conceito*, ou seja, os atributos predicáveis de cada objeto conceitualizado, *regem o relacionamento entre os termos*. Uma segunda semelhança referente ao relacionamento entre os termos está no fundamento da relação Gênero/Espécie (tesauro) e Tipo/Instância (ontologia). Outro ponto de encontro é a relação de qualidade, onde uma faceta ou categoria está atrelada a um valor de qualidade.

---

<sup>30</sup> O Dicionário Eletrônico Houaiss da Língua Portuguesa apresenta as seguintes definições para os verbos mencionados: A) Representar: [ser a imagem ou a reprodução de; trazer à memória; figurar como símbolo; aparecer numa outra forma; significar; tornar presente; substituir, estar no lugar de; fazer às vezes de]. B) Denominar: [nomear (algo ou alguém) de, designar]. C) Designar: [indicar (alguém ou algo) de maneira a distingui-lo de todos os demais; apontar, mostrar; ser marca, sinal, índice de; aplicar-se, representar, significar]. D) Denotar: [mostrar, indicar através de sinais ou indícios; representar, significar (um conceito, a idéia genérica de uma coisa concreta ou abstrata)]. E) Sintetizar: [tornar sintético, resumido; sumarizar; combinar (fatos, partes, elementos), de modo a compor um todo; congeminar, harmonizar]. F) Especificar: [indicar com precisão; precisar, apontar, discriminar].

### Semelhanças entre Tesouros e Ontologias quanto à Questão dos Conceitos

<b>CATEGORIA DE ANÁLISE</b>	<b>VARIÁVEL DE INFERÊNCIA</b>	<b>TESAURO</b>	<b>ONTOLOGIA</b>
<b>Conceito</b>	<b>Definição/ Função</b>	<ul style="list-style-type: none"> <li>➤ Objeto do pensamento</li> <li>➤ Representação abstrata percebida e interpretada de algo real</li> </ul>	<ul style="list-style-type: none"> <li>➤ Representação mental refletida por meio de um termo</li> <li>➤ Visão abstrata e sintetizada de um ente</li> </ul>
	<b>Organização dos conceitos</b>	<ul style="list-style-type: none"> <li>➤ Ordenados em classes pré-estabelecidas com base no agrupamento de elementos por semelhança</li> <li>➤ Organizados em categorias</li> </ul>	<ul style="list-style-type: none"> <li>➤ Organizados em taxonomias</li> </ul>
	<b>Relação entre conceitos</b>	<ul style="list-style-type: none"> <li>➤ Dá-se de acordo com os predicados do objeto/referente, ou seja, de acordo com o as características do conceito</li> <li>➤ Relacionamento hierárquico (gênero/espécie e todo/parte)</li> <li>➤ Lógica de superordenação (do específico para o geral)</li> <li>➤ Lógica de subordinação (do geral para o específico)</li> </ul>	<ul style="list-style-type: none"> <li>➤ Dá-se por meio da intensão (conjunto de características) do conceito</li> <li>➤ Relacionamento por meio de submissão, hierarquia</li> <li>➤ Relacionamento superordenado entre classes (subclasses)</li> <li>➤ Relacionamento subordinado entre classes e propriedades, chamado também de relacionamento poli-herárquico, pois uma classe pode ser uma subclasse de uma ou mais classes superordenadas</li> </ul>
	<b>Relação com o termo</b>	<ul style="list-style-type: none"> <li>➤ O conceito é representado, designado, sintetizado e comunicado lingüisticamente por um termo</li> <li>➤ Univocidade: para cada termo um conceito</li> </ul>	<ul style="list-style-type: none"> <li>➤ Um termo reflete, expressa e personifica um conceito</li> </ul>

**Quadro 4 – Inferências Relativas às Semelhanças da Categoria Conceito**

*Comentário do Quadro 4*

- *O conceito é uma representação mental (objeto do pensamento) de algo real (material ou imaterial) que é percebido, interpretado e mostrado pelo termo.*
- *Os conceitos são categorizados por semelhança.*
- *O relacionamento entre os conceitos, que é regido pelas características que possuem, é hierárquico, podendo se manifestar de forma superordenada (do conceito mais específico para o mais geral) e subordinada (do conceito geral para o específico).*
- *A relação do conceito com o termo é que este representa, designa, reflete, expressa, personifica, sintetiza e comunica aquele, ou seja, o conceito é visto por meio do termo.*

### Semelhanças entre Tesouros e Ontologias quanto aos Objetivos

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	TESAURO	ONTOLOGIA
<b>Objetivo</b>	<b>Teórico</b>	<ul style="list-style-type: none"> <li>➤ Esclarecer barreiras lingüísticas</li> <li>➤ Controlar sinônimos e homógrafos</li> <li>➤ Sintetizar uma estrutura conceitual</li>   <li>➤ Gerir e comunicar a linguagem específica usada nas organizações</li>   <li>➤ Auxiliar o entendimento e uso de um assunto ou linguagem específica</li> <li>➤ Facilitar o entendimento conceitual e comunicacional em um nível lingüístico que conceda um vocabulário corporativo</li> </ul>	<ul style="list-style-type: none"> <li>➤ Reduzir ou eliminar confusões terminológicas e conceituais</li> <li>➤ Conceber uma estrutura conceitual uniforme</li>   <li>➤ Potencializar a comunicação e a cooperação entre pessoas</li> <li>➤ Descrever conceitos necessários para falar sobre determinado assunto</li> </ul>
	<b>Prático</b>	<ul style="list-style-type: none"> <li>➤ Melhorar a organização da informação</li> <li>➤ Coordenar o vocabulário usado pelo usuário de um sistema de informação</li> <li>➤ Auxiliar a consulta do usuário expondo sistematicamente uma classificação que o direcione no momento da busca</li>   <li>➤ Potencializar a recuperação da informação atuando como interface entre informação e usuário</li> <li>➤ Maximizar a pertinência da recuperação da informação</li> </ul>	<ul style="list-style-type: none"> <li>➤ Melhorar a organização da informação nas empresas</li>   <li>➤ Possibilitar a consulta a um sistema automatizado de informação utilizando termos conceitualizados e especificados por especialistas</li> <li>➤ Sustentar a recuperação da informação em uma base de conhecimento</li> </ul>

**Quadro 5 – Inferências Relativas às Semelhanças da Categoria Objetivo**

### *Comentário do Quadro 5*

- *Tanto tesouros quanto ontologias têm como objetivos controlar terminologias especializadas, esclarecendo barreiras lingüísticas, concebendo uma estrutura conceitual, e, potencializar a comunicação especializada, gerindo a linguagem específica e concebendo um vocabulário compartilhado. É necessário frisar que a informação exposta no Quadro 3 (no item Objetivo Teórico), ao afirmar que dois dos objetivos do tesouro é “gerir e comunicar a linguagem específica usada nas organizações” e “facilitar o entendimento conceitual e comunicacional em um nível lingüístico que conceda um vocabulário compartilhado”, referem-se ao modelo conhecido como “tesouro corporativo”, muito semelhante às ontologias.*
- *Na prática, tanto tesouros quanto ontologias, têm como objetivos organizar informações especializadas, coordenar o vocabulário especializado, auxiliar a consulta do usuário de forma sistemática e, potencializar a recuperação da informação, atuando como interfaces entre informações e seus consumidores.*

#### *4.3.2 As Características Diferentes*

A exemplo dos passos seguidos na identificação dos elementos convergentes entre tesouros e ontologias, a visualização e o levantamento das características diferentes entre ambos os modelos se fez por meio da elaboração de três quadros (um para cada categoria de análise) que descrevem as características que distanciam tesouros de ontologias. Além dos elementos que caracterizam diferenças, descreve-se também, nos Quadros 6, 7 e 8, as características que estão presentes em um modelo de representação do conhecimento e que estão ausentes no outro. Os elementos de um modelo que não diferem diretamente dos elementos do outro modelo, mas que fazem

parte do grupo de características que o definem como tal, auxiliam significativamente a reflexão acerca das diferenças entre eles. Bem como os quadros das semelhanças, os quadros das diferenças também foram elaborados com base na análise dos Quadros 1 e 2. Porém, as características que distanciam os tesouros das ontologias são mais numerosas que as características que os aproximam. Por tal motivo, as inferências extraídas dos quadros das diferenças não são comentadas por meio de uma lista de características, como ocorrido com a descrição das semelhanças, mas sim por meio de pequenas redações explicativas que dão conta da complexidade de tal observação.

Seguem-se os Quadros 6, 7 e 8 e seus respectivos comentários.

### Diferenças entre Tesouros e Ontologias no Tocante aos Termos

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	TESAURO	ONTOLOGIA
<b>Termo</b>	<b>Definição/ função</b>	<ul style="list-style-type: none"> <li>➤ Unidade lexical que se situa nos discursos especializados</li> <li>➤ Palavra ou frase de algum domínio de conhecimento</li> <li>➤ Unidade lingüística e unidade de conhecimento especializado</li> <li>➤ Um elemento do conceito que o sintetiza como um todo</li> <li>➤ Definiendum de uma definição. O definiendum (termo) + o definiens (atributos predicáveis do objeto) formam a definição do conceito</li> <li>➤ Servem para evitar a superposição de conceitos e diminuir a flexibilidade e ambigüidade da linguagem</li> <li>➤ Serve para descrever um conceito de maneira unívoca em um sistema de informação</li> </ul>	<hr style="width: 10%; margin: 0 auto;"/> <ul style="list-style-type: none"> <li>➤ São usados para definir formalmente coisas em um domínio de interesse</li> <li>➤ Viabilizam a formulação de consultas (usuário) fazendo uso de conceitos pré-determinados por especialistas</li> </ul>

<b>Tipos de termos</b>	<ul style="list-style-type: none"> <li>➤ Termo relacionado TR</li> <li>➤ Termo equivalente</li> <li>➤ Termo simples (uma unidade lexical)</li> <li>➤ Termo composto (mais de uma unidade lexical)</li> <li>➤ Termo preferido (UP), ou controlado, ou normalizado (descriptor)</li> <li>➤ Termo não-preferido ou não-descriptor, ou ainda termo proibido (USE)</li> <li>➤ Termos polissêmicos (homógrafos)</li> <li>➤ Identificadores (nomes próprios)</li> <li>➤ Termo qualificado, ex.: Rádio (metal, Terra (planeta))</li> </ul>	<ul style="list-style-type: none"> <li>➤ Uma entidade (substância)</li> </ul>
<b>Relação entre termos</b>	<ul style="list-style-type: none"> <li>➤ De equivalência (sinonímia)</li> <li>➤ Associativo (não hierárquico)</li> <li>➤ Nota explicativa (para elucidar a respeito do uso e relacionamento do termo)</li> </ul>	<ul style="list-style-type: none"> <li>➤ A relação semântica está diretamente ligada à apresentação sintática de um discurso</li> <li>➤ O relacionamento é realizado pelos especialistas. Dá-se por meio do consenso de um determinado domínio</li> </ul>
<b>Relação com o conceito</b>	_____	_____

**Quadro 6 – Inferências Relativas às Diferenças da Categoria Termo**

*Comentário do Quadro 6*

*No que diz respeito ao termo, as diferenças encontradas nos textos residem predominantemente na função exercida por eles. Como pode ser observado, os textos relativos às ontologias focam mais nas funções desempenhadas pelos *termos*, sem se (pre)ocupar 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*. Observa-se que *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). Os demais tipos de termos, como visto anteriormente, apresentam alguma aproximação com aqueles constituintes dos tesouros (ver página 121).

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 de 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 *descriptor*, 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 descriptor correspondente); b) *Associativo* – apresenta relação semântica não hierárquica; c) *Nota*

*Explicativa* - orientação que elucida a respeito do emprego de determinado termo fornecendo informações como a definição do termo e sua relação com outros termos.

### Diferenças entre Tesouros e Ontologias quanto à Questão dos Conceitos

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	TESAURO	ONTOLOGIA
<b>Conceito</b>	<b>Definição/ Função</b>	<ul style="list-style-type: none"> <li>➤ Unidade de conhecimento, incluindo a definição como elemento essencial para a fixação do conceito</li> <li>➤ Unidade de representação</li> <li>➤ Conjunto formado pelas características próprias de um objeto e por um termo que as sintetizam</li> </ul>	<ul style="list-style-type: none"> <li>➤ Conceitos são classes</li> <li>➤ Entidades, atributos e processos</li> <li>➤ Unidades do vocabulário de uma ontologia</li> </ul>
	<b>Organização dos conceitos</b>	<p style="text-align: center;">_____</p>	<ul style="list-style-type: none"> <li>➤ Conceitos concretos principais (os quais apresentam propriedades do domínio, assim como seus relacionamentos) e conceitos abstratos (são as características)</li> </ul>
	<b>Relação entre conceitos</b>	<ul style="list-style-type: none"> <li>➤ Ontológico: proximidade situacional dos elementos na realidade (contigüidade dos conceitos no espaço), pode ocorrer por coordenação, encadeamento, seqüência. É a relação entre conceito e realidade</li> <li>➤ O relacionamento é determinado pelo uso empregado pelo domínio e pelos fatores característicos do domínio</li> <li>➤ Equivalência: quando o conceito é denominado por mais de uma de expressão (forma)</li> <li>➤ Descendência (um conceito B descende de um conceito A)</li> <li>➤ Instrumental (um conceito B é instrumento para alcançar um conceito A)</li> <li>➤ Causa/efeito (um conceito B é causa da existência de um</li> </ul>	<ul style="list-style-type: none"> <li>➤ As relações conceituais (intensionais) são definidas em um espaço do domínio, e podem ser representadas em grupos de mundos possíveis</li> <li>➤ As relações semânticas se dão por meio das formas sintáticas de um discurso</li> <li>➤ Relacionamento por meio de axiomas que expressam relacionamento entre</li> </ul>

	conceito A) ➤ Benefício (um conceito B é útil para um conceito A) ➤ Prejuízo (causa de efeitos adversos) ➤ Matéria (um conceito A consiste materialmente de um conceito B) ➤ Aparência (enxerga o conceito como ele é reconhecível pela visão humana) ➤ Processo (um conceito B indica um processo em um conceito A) ➤ Estado (um conceito B é um estado de um conceito A)	conceitos e definem a interpretação pretendida  ➤ Interligações entre conceitos mais refinados e conceitos mais periféricos formam as relações adicionais
<b>Relação com o termo</b>	_____	_____

**Quadro 7 – Inferências Relativas às Diferenças da Categoria Conceito**

*Comentário do Quadro 7*

*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 aplicação de seus produtos (nesse caso as ontologias). Isso explica a diferente maneira 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 que ambas as pesquisas, além de se complementarem, perpassam por outras áreas (como a Lingüística por exemplo) para fortalecer este ciclo.

*No tangente à organização dos conceitos*, os textos relativos ao tesouro não fazem menção alguma, ao passo que os textos relativos às ontologias 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 considerada 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.

### Diferenças entre Tesouros e Ontologias quanto aos Objetivos

CATEGORIA DE ANÁLISE	VARIÁVEL DE INFERÊNCIA	TESAURO	ONTOLOGIA
<b>Objetivo</b>	<b>Teórico</b>	<ul style="list-style-type: none"> <li>➤ Facilitar a inter-relação entre linguagem natural e linguagem artificial</li> <li>➤ Propor um sistema de símbolos lingüísticos para agrupar informação singular relacionada ou direcionar para grupos mais gerais ou mais específicos de uma temática</li> <li>➤ Orientar qual o termo mais adequado para representar um conceito</li> </ul>	<ul style="list-style-type: none"> <li>➤ Fornecer um mapa semântico aos campos individuais e aos relacionamentos entre os campos, servindo como uma ferramenta de referência</li> <li>➤ Criar uma estrutura lógica, uma filosofia, uma classificação em uma disciplina (domínio)</li> <li>➤ Esclarecer o significado pretendido de um vocabulário por meio de axiomas</li> </ul>
	<b>Prático</b>	<ul style="list-style-type: none"> <li>➤ Servir como ponte que liga necessidades de informação (busca) e sistemas de recuperação de informação</li> <li>➤ Economizar o tempo do usuário</li> <li>➤ Guiar a indexação baseada em linguagem natural</li> <li>➤ Servir como vocabulário oficial que rege a indexação e a recuperação de documentos</li> <li>➤ Alcançar uma normalização da terminologia do sistema de informação a ser utilizado</li> <li>➤ Propor um conjunto estruturado de termos sob a base de um sistema de conceitos aptos a organizar os conteúdos de um sistema</li> </ul>	<ul style="list-style-type: none"> <li>➤ Ajudar a responder perguntas em um corpo de informação. Isto inclui as seguintes funções: relacionar conceitos aos termos e fornecer definições; esclarecer conceitos, contextualizando-os em uma classificação; relacionar conceitos e termos ou ícones por meio das disciplinas, das línguas e das culturas</li> <li>➤ Servir como um dicionário para o uso humano e como um dicionário/base de conhecimento para o processamento de linguagem natural</li> <li>➤ Fornecer um entendimento comum e compartilhado de um determinado domínio</li> <li>➤ Identificar e definir um conjunto de conceitos relevantes que caracterize um dado domínio de aplicação</li> <li>➤ Possibilitar, juntamente</li> </ul>

	<ul style="list-style-type: none"> <li>➤ Auxiliar a representação do conteúdo documental, minimizando ruídos como ambigüidades</li> </ul>	<p>com aplicações lógicas, a construção de modelos computacionais para um determinado domínio de aplicação</p> <ul style="list-style-type: none"> <li>➤ Representar o mundo real (ou parte do mundo real) em um dado domínio</li> <li>➤ Interoperabilizar sistemas</li> <li>➤ Proporcionar a interoperabilidade dos serviços ofertados na web, potencializando os níveis de serviços ofertados, ou seja, sustentar a web semântica</li> <li>➤ Propiciar benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação)</li> <li>➤ Fornecer base conceitual para projetos de pesquisa e execução</li> <li>➤ Melhorar a comunicação e o aprendizado, assistindo escritores e leitores, fornecendo estruturas conceituais que sustentam o ensino</li> <li>➤ Possibilitar serviços baseados em operacionalizações semânticas</li> <li>➤ Ligar parte do conhecimento humano aos processamentos computacionais</li> <li>➤ Descrever a semântica de um domínio de modo que seja compreensível por homens e máquinas</li> <li>➤ Melhorar a consistência e o reuso da informação e o compartilhamento do conhecimento</li> <li>➤ Recuperação da informação e raciocínio automático de um dado domínio</li> <li>➤ Capturar significados e relações entre os significados</li> </ul>
--	---	---

		<ul style="list-style-type: none"> <li>➤ Servir como um catálogo dos tipos de coisas que existem em um domínio D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D</li> <li>➤ 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</li> <li>➤ 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</li> <li>➤ Oferecer meios concisos e sistemáticos para definir a semântica dos recursos web</li> <li>➤ (Para a web semântica) proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação</li> </ul>
--	--	---

**Quadro 8 – Inferências Relativas às Diferenças da Categoria Objetivo**

*Comentário do quadro 8*

*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 objetivo 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 uma disciplina (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 representação de assuntos especializados, os tesouros estão voltados ao elo que une a linguagem do usuário da informação (especialista ou não) com a linguagem utilizada pelas unidades e 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 seja, as ontologias de fato possuem como objetivo principal viabilizar uma base de conhecimento.

*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, baseado 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 a 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. Ao passo que 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, as ontologias visam a um entendimento comum e compartilhado de um determinado domínio, deixando clara sua função de responder perguntas em uma base de conhecimento. 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 tesauros. Dentre eles: a) representar o mundo real (ou parte do mundo real) em um dado domínio; b) interoperabilizar sistemas; c) proporcionar a interoperabilidade dos serviços ofertados na web, potencializando os níveis de serviços ofertados, ou seja, sustentar a web semântica; d) propiciar benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação); e) fornecer base conceitual para projetos de pesquisa e execução; f) melhorar a comunicação e o aprendizado, assistindo escritores e leitores, fornecendo estruturas conceituais que sustentam o ensino; g) possibilitar serviços baseados em operacionalizações semânticas; h) ligar parte do conhecimento humano aos processamentos computacionais; i) descrever a semântica de um domínio de modo que seja compreensível por homens e máquinas; j) melhorar a consistência e o reuso da informação e o compartilhamento do conhecimento; k) recuperação da informação e raciocínio automático de um dado domínio; l) capturar significados e relações entre os significados; m) servir como um catálogo dos tipos de coisas que existem em um domínio D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D; n) 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; o) 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; p) oferecer meios concisos e sistemáticos para definir a semântica dos recursos web e; q) proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação.

As características descritas acima, que diferenciam os modelos em questão, bem como aquelas que, embora não apresentem uma distinção direta demonstram ir mais além em relação as demais, tornam evidentes que o distanciamento entre tesouros e ontologias é melhor esclarecido na esfera das aplicações, haja visto que as diferenças de ordem conceitual, muitas vezes, correspondem a pontos de vistas. 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. É inevitável que o fato de as ontologias serem artefatos criados no e para o universo informático lhe concedem objetivos e façanhas mais audaciosos.

Fica evidente que as ontologias transcendem a questão da representação de conteúdos documentais, mais frequentemente atribuída aos tesouros, para assumir um papel de ferramenta elementar aos sistemas de informação automatizados, às bases de conhecimento e aos serviços ofertados pela web, sobretudo no tocante a web semântica. Não se quer com isso afirmar que as ontologias não cumprem o papel de representação do conhecimento, pelo contrário, assesta-se que as ontologias vão muito além, pois elas contemplam elementos das linguagens documentárias (sistemas de classificação bibliográfica, listas de cabeçalhos de assunto e tesouros) e resgatam as classificações filosóficas (agora sob um olhar coletivo), constituindo-se como genuíno – e evoluído – instrumento de organização e representação do conhecimento.

## 5 CONCLUSÕES

Como desfecho deste processo de investigação, cumpre-se por ora registrar algumas informações relevantes e conclusivas. De início foi possível perceber a evolução ocorrida com os modelos de representação do conhecimento no decorrer da história, que viu surgir desde sistemas de classificação de conhecimentos gerais e sistemas de classificação bibliográfica até a criação de artefatos computacionais dedicados a responder perguntas de um dado domínio de conhecimento operando como linguagens específicas em ambientes específicos. Como toda criação humana, o desenvolvimento das linguagens documentárias também acompanhou, e vem acompanhando, o avanço das tecnologias, neste caso, o avanço nas tecnologias de informação e comunicação. Este último, parece coibir, de maneira quase imposta e fora de controle, qualquer que seja o fato de não aumentar ainda mais a circulação da informação. Com isso todo esforço científico e tecnológico se volta ao melhoramento daquilo que já existe e que parece não parar de se multiplicar. A explosão informacional anunciada por Vannevar Bush na década de 1940 parece se atualizar e se potencializar a cada ano. Tal fato clama por uma constante e veloz adaptação dos produtos e serviços concebidos no universo da informação. Os modelos de representação do conhecimento, amplamente difundidos entre as comunidades da Ciência da Informação como linguagens documentárias, são instrumentos fundamentais nesse universo da informação, haja visto que são instrumentos que zelam pela comunicação da linguagem especializada, sem a qual não há desenvolvimento técnico-científico. Esta pesquisa surgiu da ansiedade de melhor definir e conhecer dois desses instrumentos, o tesouro e a ontologia. Porém, inviável seria mergulhar na literatura especializada sem antes estabelecer alicerces teóricos que proporcionassem critérios rigorosos para tal pesquisa.

Foi então que se julgou a TCT como uma teoria preciosa para o sucesso da investigação, pois se trata de uma teoria que vislumbra a ciência dos termos envolvida por uma linguagem em constante funcionamento e comunicação, nada mais apropriado para um contexto informacional que não cessa de se movimentar. Da TCT foram extraídos como elementos essenciais no universo da terminologia o termo, o conceito e o objetivo a ser alcançado pelo pensar terminológico. E foi com base nesses elementos, variáveis de análise desta pesquisa, que foram identificadas as semelhanças e diferenças entre os tesouros e as ontologias.

As características semelhantes e diferentes expostos mais acima, foram frutos 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, livre de qualquer neutralidade forjada, 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 a consideração de que as características que aproximam e 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 analisados nesta pesquisa são significativamente superiores aos pontos de convergência entre eles, algo que já poderia ter sido previsto pelo fato de ambos

pertencerem genuinamente a áreas de conhecimento distintas (tesauro – 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.

É notório o fato de que as diferenças entre tesauros e ontologias são mais numerosas que as semelhanças entre eles. 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 tesauros 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 de controle terminológico que auxiliam o processo de classificação, indexação e recuperação de documentos por assunto, uma definição muito aquém das possibilidades de execução de uma ontologia.

Para deixar mais claro o que foi discorrido no parágrafo anterior, ressalta-se que as ontologias não são “somente” linguagens documentárias, pois elas vão mais além daquilo que é comumente afirmado nas definições literárias referentes às linguagens documentárias, a saber: instrumentos que controlam o processo de classificação, indexação e recuperação de documentos por assunto. Ora, é evidente que as ontologias

cumprem esses papéis desempenhados pelas linguagens documentárias, mas fica claro também que elas transcendem essas funções, e não somente na dimensão tecnológica, mas no resgate conceitual das classificações filosóficas, que antes eram perspectivas individuais e agora, com base em ontologias, estão sob um olhar coletivo de uma dada comunidade. Com isso, é possível assertar que a ontologia se constitui como um genuíno e evoluído instrumento de organização e representação do conhecimento.

O que é possível dizer, diante das semelhanças por ora descritas, é que ambos os modelos trabalham para esclarecer implicações relativas aos termos e aos conceitos de um dado domínio, e que por vezes compartilham de objetivos iguais. Embora claramente distintos, ambos operam em ambientes de linguagens específicas com o intuito de facilitar a comunicação especializada. Com isso, não é descabido afirmar que tesouros e ontologias são objetos distintos operando em ambientes semelhantes, com funções e capacidades que ora se aproximam ora se distanciam.

As características de 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 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 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.

Com a hipótese de que as ontologias potencialmente legitimam o que foi definido pela TCT surge uma semente que poderá germinar um pensamento direcionado à idéia de que a teoria que melhor vislumbra o funcionamento da linguagem especializada pode ser formalizada por meio de uma ferramenta computacional. Tem-se com isso uma teoria das Ciências Humanas sendo formalmente aplicada por um artefato proveniente das Ciências Tecnológicas contribuindo para o avanço da comunicação técnica e científica, conseqüentemente, contribuindo para o avanço da representação do conhecimento nos mais variados ambientes estudados no universo da Ciência da Informação.

A despeito da hipótese levantada, que obviamente requer estudos futuros para deixar de ser uma “hipótese” e se tornar uma “tese”, a contribuição principal da presente investigação está na identificação, devidamente comentada, das diferenças e semelhanças 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 e suas finalidades, auxiliando assim sua adequada aplicação. Por fim, vê-se este trabalho como uma fonte para o entendimento conceitual e prático dos tesouros e das ontologias.

## REFERÊNCIAS

AITCHISON, Jean; CLARKE, Stella Dextre. The thesaurus: a historical viewpoint, with a look to the future. **Cataloging & Classification Quarterly**, v. 37, n. 3/4, p. 5-21, 2004.

ALMEIDA, Maurício B. Roteiro para construção de uma ontologia bibliográfica através de ferramenta automatizada. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, p. 164-179, jul./dez. 2003. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=141&layout=abstract>>. Acessado em: 22 maio 2006.

ALMEIDA, Maurício B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20 set./dez. 2003. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=36&layout=abstract>>. Acessado em: 26 mar.2006.

ALVARENGA, Lídia. A Teoria do Conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramaZero: Revista de Ciência da Informação**. v. 2, n. 6, dez/2001. Disponível em <[http://www.dgz.org.br/dez01/Art\\_05.htm](http://www.dgz.org.br/dez01/Art_05.htm)>. Acessado em: 17 maio 2005.

\_\_\_\_\_. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 15, 1º sem. 2003. Disponível em <<http://www.encontros-bibli.ufsc.br/regular.html>> Acessado em: 25 ago.2006.

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

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

BARQUÍN, Beatriz A. R.; GONZÁLEZ, José A. M.; PINTO, Adilson L. Construção de uma ontologia para sistemas de informação empresarial para a área de telecomunicações **DataGramaZero: Revista de Ciência da Informação** - v.7 , n.2, abr/2006. Disponível em <[http://www.dgz.org.br/abr06/F\\_I\\_art.htm](http://www.dgz.org.br/abr06/F_I_art.htm)> Acessado em: 02 abr.2006.

BARRETO, Aldo de Albuquerque. Uma história da ciência da informação. In: TOUTAIN, Lídia Maria Batista Brandão (Org.). **Para entender a ciência da informação**. Salvador: EDUFBA, 2007. p. 13-34. (Saladeaula, 5).

BARTHES, Roland. **Elementos de semiologia**. Tradução de Izidoro Blinkstein. 3. ed. São Paulo: Cultrix, 1974, 116 p.

BATISTA, Gilda Helena Rocha. Redes de conceitos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 1, p. 6-17, jan./jun. 2004. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=55>>. Acessado em: 09 mar.2006.

BATRES, Eduardo Jaime Queirós *et alli*. Uso de ontologias para a extração de informações em atos jurídicos em uma instituição pública. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n.19, p.73-88, 1º sem. 2005. Disponível em <<http://www.encontros-bibli.ufsc.br/sumario.htm>>. Acessado em: 10 mar.2006.

BERNARD, Michel. **De quoi parle ce livre?** Elaboration d'un thésaurus pour l'indexation thématique d'oeuvres littéraires. Paris: Honoré Champion, 1994, 365 p.

BRÄSCHER, Marisa. A ambigüidade na recuperação da informação. **DataGramZero: Revista de Ciência da Informação**, v. 3, n. 1, fev., 2002. Disponível em <[http://www.dgz.org.br/fev02/F\\_I\\_art.htm](http://www.dgz.org.br/fev02/F_I_art.htm)>. Acessado em: 17 jul.2006.

BRAUNER, Daniela Francisco. **Uma arquitetura para catálogos de objetos baseados em ontologias**. Rio de Janeiro, 2005. 141f. Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica, 2005. Disponível em <[http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG\\_0599.EXE/7071\\_1.PDF?NrOcoSis=20064&CdLinPrg=en](http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/PRG_0599.EXE/7071_1.PDF?NrOcoSis=20064&CdLinPrg=en)>. Acessado em: 12 mar.2006.

BUFREM, Leilah Santiago. A relação inescusável entre lingüística e documentação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n.19, 1º sem. 2005. Disponível em <<http://www.encontros-bibli.ufsc.br/sumario.htm>>. Acessado em: 04 maio 2006.

CABRÉ, M. Teresa. **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.

CAFÉ, Lígia. Aplicação do modelo de predicação sintático-semântico na construção de linguagens documentárias facetadas. In: CUNHA, Miriam Veira da; SOUZA, Francisco das Chagas de. (Org.). **Comunicação, gestão e profissão: abordagens para o estudo da Ciência da Informação**. Belo Horizonte: Autêntica, 2006. 176 p.

\_\_\_\_\_. Contribuições da Gramática Funcional na delimitação de segmentos descritores de informação. In: RODRIGUES, Georgete Medleg; LOPES, Ilza Leite (Org.) **Organização e representação do conhecimento na perspectiva da ciência da informação**. Brasília: Thesaurus, 2003a, p. 118-140. (Estudos Avançados em Ciência da Informação, v.2)

\_\_\_\_\_. Terminologia: aplicação do (re)modelo de Simon Dik. In: FAULSTICH, Enilde; ABREU, Sabrina P. (Org.). **Lingüística aplicada à terminologia e à lexicologia: cooperação internacional: Brasil e Canadá**. Porto Alegre: UFRGS, Instituto de Letras, NEC, 2003b. p. 59-82.

CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, jan./abril 2004. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=77&layout=html>>. Acessado em: 02 maio 2006.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Organização de domínios de conhecimento e os princípios ranganathianos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, p. 150-163, jul./dez. 2003. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=140&layout=abstract>>. Acessado em: 19 maio 2006.

CAMPOS, Maria Luiza Machado; CAMPOS, Maria Luiza de Almeida; CAMPOS, Linair Maria. Web semântica e gestão de conteúdos digitais. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lídia Brandão; SAYÃO, Luís (Org). **Bibliotecas Digitais: saberes e práticas**. 2ed. Salvador: EDUFBA; Brasília: IBICT, 2006. p. 55 - 74.

CASTILHO, Ataliba Teixeira de. Um ponto de vista funcional sobre a predicação. **ALFA: Revista de Lingüística**. São Paulo, UNESP, v. 38, p. 75-95, 1994.

CASTOLDI, André Vinícius. **Uma ontologia para enlaces de unidades de informação em plataformas de governo eletrônico**. Florianópolis, 2003. 91 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, 2003. Disponível em <<http://teses.eps.ufsc.br/Resumo.asp?4726>>. Acessado em: 15 abr.2006.

CAVALCANTI, Cordelia R. **Indexação & tesouro: metodologia e técnicas**. Brasília: Associação de Bibliotecários do Distrito Federal, 1978. 89 p. (Edição Preliminar).

CERVANTES, Brígida Maria Nogueira. **Contribuição para a terminologia do processo de inteligência competitiva: estudo teórico e metodológico**. Marília, 2004. 183f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade do Estado de São Paulo, Campus de Marília, 2004.

CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Cabeçalho de assunto como linguagem de indexação. **Revista da Escola de Biblioteconomia**, UFMG, Belo Horizonte, ano 7, n. 2, p. 268-88, set. 1978.

CHAUMIER, Jacques. **Analyse et langages documentaires: le traitement linguistique de l'information documentaire**. Paris : Entreprise Moderne d'Édition, 1982.

CINTRA, Ana Maria et alii. **Para entender as linguagens documentárias**. 2 ed. São Paulo: Polis, 2002. 92 p. (Coleção Palavra-Chave, 4).

DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**. Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.

DENNY, Michael. **Ontology building**: a survey of editing tools. XML.com., p. 1-7, 2002. Disponível em <<http://www.xml.com/lpt/a/2002/11/06/ontologies.html>>. Acessado em: 01 jun.2007.

DING, Ying. A review of ontologies with the semantic web in view. **Journal of Information Science**, v. 27, n.6, p. 377-84, 2001. Disponível em <<http://jis.sagepub.com/cgi/reprint/27/6/377>>. Acessado em: 28 maio 2007.

DING, Ying; FOO, Schubert. 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](http://homepage.uibk.ac.at/~c703205/download/01jis01_final_revision.pdf)>. Acessado em: 13 mar.2006.

\_\_\_\_\_. Ontology research and development, Part 2 : a review of ontology mapping and evolving. **Journal of Information Science**, v. 28, n.5, p. 375-388, 2002. Disponível em <<http://wotan.liu.edu/doi/data/Articles/julbaefggy:2002:v:28:i:5:p:375-388.html>>. Acessado em: 28 maio 2007.

DODEBEI, Vera Lucia Doyle. **Tesouro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119 p.

ESBÍZARO, André Luiz Dias. **Recuperação de informações sobre log de eventos apoiada em ontologia**. 2006. 186 f. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília, 2006. cap.2.5, p. 75-134.

FAULSTICH, Enilde. Formação de termos: do constructo e das regras às evidências empíricas. In: FAULSTICH, Enilde; ABREU, Sabrina P. (Org.). **Linguística aplicada à terminologia e à lexicologia**: cooperação internacional: Brasil e Canadá. Porto Alegre: UFRGS, Instituto de Letras, NEC, 2003. 222 p.

FAYEN, Emily. A new standard for controlled vocabularies. **The Indexer**, v. 24, n. 2, p. 62-65, october/2004. Disponível em <<http://wilsonxt.hwwilson.com/pdf/full/02825/at8k6/dsv.pdf>>. Acessado em 30 maio.2007.

FEITOSA, Ailton. **Organização da informação na web**: das tags à web semântica. Brasília: Thesaurus, 2006. 132 p.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em <<http://server01.bc.unicamp.br/seer/ojs/viewarticle.php?id=9&layout=abstract>>. Acessado em: 08 jan.2008.

GARCÍA GUTIÉRREZ, A. **Estructura lingüística de la documentación**: teoria y método. Murcia: Ed. Un. Murcia, 1990.

GARDIN, Jean-Claude. Document analysis and linguistic theory. **Journal of Documentation**, v. 29, n. 2, p. 137-168, jun. 1973.

GAUDIN, François. **Socioterminologie**: des problèmes sémantiques aux pratiques institutionnelles. Rouen: Publications de L'Université de Rouen, 1993.

GILCHRIST, Alan. Thesauri, taxonomies, ontologies: an etymological note. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003.

GOMES, Hagar Espanha (Coord.). **Manual de elaboração de tesouros monolíngües**. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990. 78 p.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida. Tesouro e normalização terminológica: o termo como base para intercâmbio de informações. **DataGramZero: Revista de Ciência da Informação**. Rio de Janeiro, v. 5, n. 6, Dez./2004. Disponível em <[http://www.dgz.org.br/dez04/F\\_I\\_art.htm](http://www.dgz.org.br/dez04/F_I_art.htm)>. Acessado em: 02 maio 2006.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Appeared in Knowledge Acquisition**, v. 5, n. 2, p.1-26, 1993a. Disponível em <<http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>>. Acessado em: 07 fev.2007.

\_\_\_\_\_. **Toward principles for the design of ontologies used for knowledge sharing**. p. 1-22, 1993b. Disponível em <<http://tomgruber.org/writing/onto-design.pdf>>. Acessado em: 13 jan.2007.

GUARINO, Nicola. Formal ontology and information systems. In: FOIS'98, 1998, Trento, Italy. **Proceedings...** Trento, Italy. p. 3-15, 1998. Disponível em <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Acessado em: 05 fev.2007.

GUARINO, Nicola; GIARETTA, Pierdaniele. **Ontologies and knowledge bases: towards a terminological clarification**. P. 1-7. 1995. Disponível em <<http://www.loa-cnr.it/Papers/KBKS95.pdf>>. Acessado em: 12 jan.2007.

INTERNATIONAL STANDARD ORGANIZATION ISO, Genova. ISO 704:2000; **Terminology work: principles and methods**. 2. ed. Genova, 2000.

\_\_\_\_\_, Genova. ISO 1087-1:2000; **Terminology work: vocabulary**. Part 1: theory and application. Genova, 2000.

JACOB, Elin K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**. v. 29, n. 4, p. 19-22, abr./maio 2003.

KOBASHI, Nair Yomiko. **A elaboração de informações documentárias: em busca de uma metodologia**. São Paulo, 1994. 195 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, 1994.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à terminologia: teoria e prática**. São Paulo: Contexto, 2004. 223p.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e recuperação de informação textual: os sintagmas nominais. **Ciência da Informação**, Brasília, v. 25, n.

2, 1996. Disponível em <<http://www.ibict.br/cionline/250296/25029605.pdf>>. Acessado em: 09 mar.2006.

LAAN, Regina H. van der. **Terminologia**: uma inter-relação lógica. Porto Alegre, 2002. 185 f. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras. Porto Alegre, 2002.

LA BANQUE DE DONNÉES D'HISTOIRE LITTÉRAIRE. Université de Paris III. France. Disponível em <<http://www.cavi.univ-paris3.fr/phalese/bdhl/bdhl.htm>>. Acessado em: 19 mar.2006.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Tradução de Antônio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 2004. 452 p.

LARA, Marilda Lopez Ginez de. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 91-96, maio/ago. 2004a. Disponível em <<http://www.ibict.br/cienciadainformacao/viewaarticle.php?id=304&layout=html>> Acessado em: 05 abr.2006.

\_\_\_\_\_. Linguagem documentária e terminologia. **Transinformação**. Campinas. v. 16, n. 3, p. 231-240. Setembro/Dezembro, 2004b. Disponível em <<http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=8>> Acessado em: 20 abr.2006.

LEVACOV, Marília et. alii. O tesouro eletrônico do mundo do trabalho: produto de um esforço interdisciplinar. **DataGramZero**: Revista de Ciência da Informação. v. 3, n. 4, ago./2002. Disponível em <[http://www.dgz.org.br/ago02/F\\_I\\_art.htm](http://www.dgz.org.br/ago02/F_I_art.htm)>. Acessado em: 18 jan.2007.

LIMA, Vânia Mara Alves. **Da classificação do conhecimento científico aos sistemas de recuperação de informação**: enunciação de codificação e enunciação de decodificação da informação documentária. São Paulo, 2004. 148 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, 2004.

LISA: Library and Information Science Abstracts. Disponível em <<http://www.csa.com/factsheets/lisa-set-c.php>>. Acessado em: 02 mar.2007.

MENDONÇA, Ercilia Severina. A lingüística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**, Brasília, v. 29, n. 3, p. 50-70, set./dez. 2000. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=259&layout=abstract>> Acessado em: 02 abr.2006.

MISSIKOFF, Michele; VELARDI, Paola; FABRIANI, Paolo. Text mining techniques to automatically enrich a domain ontology. **Applied Intelligence**, v. 18, n. 3, p. 323-40, 2003. Disponível em <<http://portal.acm.org/citation.cfm?id=762622.762657>>. Acessado em: 29 maio 2007.

MOREIRA, Alexandra; ALVARENGA, Lídia; OLIVEIRA, Alcione de Paiva. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias.

**DataGramZero** – Revista de Ciência da Informação. Rio de Janeiro, v.5, n.6, dez. 2004. Disponível em <[http://www.dgz.org.br/dez04/F\\_I\\_art.htm](http://www.dgz.org.br/dez04/F_I_art.htm)> Acessado em: 20 abr.2006.

MOREIRA, Alexandra. **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)>. Acessado em: 16 abr.2006.

MOREIRO, José; MARZAL, Miguel Angel; BELTRÁN, Pilar. Desarrollo de un Método para la creación de mapas conceptuales. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6, 2003, Belo Horizonte. **Anais...** Belo Horizonte. CD-ROM.

NEVES, Maria Helena de Moura. **A Gramática funcional**. São Paulo: Martins Fontes, 1997. 160 p. (Texto e Linguagem).

NIELSEN, Marianne L. Thasaurus construction: key issues and selected readings. **Cataloging & Classification Quarterly**, v. 37, n. 3/4, p. 57-74, 2004.

NOY, Natalya F.; MCGUINNESS, Deborah L. **Desarrollo de ontologías – 101**: guia para crear tu primera ontología. Tradução de Erick Antezana, 29 p., 2005. Disponível em <[http://protege.stanford.edu/publications/ontology\\_development/ontology101-es.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf)>. Acessado em: 06 nov.2007.

ONTOLINGUA. Disponível em <<http://www.ksl.stanford.edu/software/ontolingua/>>. Acessado em: 30 nov.2007.

PEREIRA, Edmeire Cristina; BUFREM, Leilah Santiago. Princípios de organização e representação de conceitos em linguagens documentárias. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, n. 20, p. 21-37, 2º semestre de 2005. Disponível em <<http://www.encontros-bibli.ufsc.br/sumario.htm>>. Acessado em: 04 maio 2006.

PEZATTI, Erotilde Goreti. Uma abordagem funcionalista da ordem de palavras no português falado. **ALFA**: Revista de Lingüística. São Paulo, UNESP, v. 38, p. 37-56, 1994.

PIEDADE, Maria Antonieta Requião. **Introdução à teoria da classificação**. Rio de Janeiro: Interciência, 1977. 190 p.

\_\_\_\_\_. **Introdução à teoria da classificação**. 2.ed.rev.aum. Rio de Janeiro: Interciência, 1983.

PRADO, Simone das Graças Domingues. **Um experimento no uso de ontologias para reforço da aprendizagem em educação a distância**. São Paulo, 2004. 190f. Tese (Doutorado em Engenharia) – Escola Politécnica, Universidade de São Paulo, 2004. Disponível em <<http://www.teses.usp.br/teses/disponiveis/3/3141/tde-08062005-160610/>>. Acessado em: 15 mar.2006.

PROTÉGÉ. Disponível em <<http://protege.stanford.edu/>>. Acessado em: 30 nov.2007.

ROSA, Antonio de la. Tesauros, tesauros automaticos, tesauros automaticos online. El Profesional de la Información: **Revista Internacional Científica y Profesional**, Barcelona, v. 8, n. 7-8, jul-ago/1999. Disponível em <[http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros\\_tesauros\\_automáticos\\_tesauros\\_automáticos\\_online.html](http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros_tesauros_automáticos_tesauros_automáticos_online.html)>. Acessado em: 01 jun.2007.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun.1996.

SAUSSURE, Ferdinand de. **Curso de lingüística geral**. Tradução de José Victor Adragão. 6. ed. Lisboa: Dom Quixote, 1992, 392 p.

SCHMITZ-ESSER, Winfried. Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval. **Knowledge Organization**, v. 26, n. 1, p. 10-22, 1999.

SCHNEIDER, Jesper W.; BORLUND, Pia. Introduction to bibliometrics for construction and maintenance of thesauri: methodical considerations. **Journal of Documentation**, v. 60, n. 5, 2004. Disponível em <[http://www.db.dk/binaries/schneider\\_borlund%20\(2004\).pdf](http://www.db.dk/binaries/schneider_borlund%20(2004).pdf)>. Acessado em: 01 jul.2007.

SHIRI, Ali Asghar; REVIE, Crawford. Thesauri on the WEB: current developments and trends. **Online Information Review**, Tucson, United States, v. 24, n. 4, p. 273-279, 2000. Disponível em <<http://dlist.sir.arizona.edu/163/01/thesauri.pdf>>. Acessado em: 01 jun.2007.

SLYPE, Georges Van. **Los lenguajes de indización**: concepción, construcción y utilización en los sistemas documentales. Tradução de Pedro Hípola e Félix de Moya. Madri: Fundación Germán Sánchez Ruipérez; Pirámide, 1991. 200 p. (Biblioteca del Libro).

SOUZA, Renato Rocha. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6, 2005, Florianópolis. **Anais...** Florianópolis. CD-ROM.

SILVA, Tércio de Moraes Sampaio. **Extração de informação para busca semântica na Web baseada em ontologias**. Florianópolis, 2003. 79 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Santa Catarina, 2003.

SMIT, Johanna W. (Coord.). **Análise documentária: a análise da síntese**. 2. ed. Brasília: IBICT, 1989. 135 p.

SOUZA, Renato Rocha. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6, 2005, Florianópolis. **Anais...** Florianópolis. CD-ROM.

SOWA, John F. **Building, sharing, and merging ontologies**. 2006. Disponível em <<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>>. Acessado em: 16 jan.2007.

STRAIOTO, Ana Claudia. **A análise em facetas como dimensão teórica e prática na organização do conhecimento**. 2001. 16f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2001.

THOMAS, Alan R. Teach yourself thesaurus: exercises, readings, resources. **Cataloging & Classification Quarterly**, v. 37, n. 3-4, p. 23-34, 2004.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p.161/171, maio/ago. 2004. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=120&layout=abstract>> Acessado em: 26 abr.2006.

VIZCAYA ALONSO, Dolores. **Lenguajes documentarios**. Mendoza: Nuevo Paradigma Ediciones, 1997. 160 p.

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>>. Acessado em: 07 mar.2007.

### Referência do Corpus de Análise

AITCHISON, Jean; CLARKE, Stella Dextre. The thesaurus: a historical viewpoint, with a look to the future. **Cataloging & Classification Quarterly**, v. 37, n. 3/4, 2004, p. 5-21.

ALMEIDA, Maurício B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, set./dez. 2003, p. 7-20. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=36&layout=abstract>>. Acessado em: 01 jun.2007.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**,

Belo Horizonte, v.11, n.3, p. 348-359, set./dez. 2006. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&layout=abstract>>. Acessado em: 29 maio 2007.

DENNY, Michael. **Ontology building: a survey of editing tools**. XML.com. 2002, p. 1-7. Disponível em <<http://www.xml.com/lpt/a/2002/11/06/ontologies.html>>. Acessado em: 01 jun. 2007.

\_\_\_\_\_. **Ontology tools survey, revisited**. XML.com. 2004, p. 1-8. Disponível em <<http://www.xml.com/lpt/a/2004/07/14/onto.html>>. Acessado em: 01 jun. 2007.

DING, Ying. A review of ontologies with the semantic web in view. **Journal of Information Science**, v. 27, n.6, 2001, p. 377-84. Disponível em <<http://jis.sagepub.com/cgi/reprint/27/6/377>>. Acessado em: 28 maio 2007.

DING, Ying; FOO, Schubert. **Ontology research and development part 1: a review of ontology generation**. Object-oriented Systems Modeling Laboratory - OSM, 200?. Disponível em <<http://osm.cs.byu.edu/CS652s04/DF02OntologyGeneration.pdf>>. Acessado em: 26 maio 2007.

\_\_\_\_\_. Ontology research and development, Part 2 : a review of ontology mapping and evolving. **Journal of Information Science**, v. 28, n.5, 2002, p. 375-388. Disponível em <<http://wotan.liu.edu/does/data/Articles/julbaefggy:2002:v:28:i:5:p:375-388.html>>. Acessado em: 28 maio 2007.

FAYEN, Emily. A new standard for controlled vocabularies. **The Indexer**, v. 24, n. 2, october/2004, p. 62-65. Disponível em <<http://wilsonxt.hwwilson.com/pdf/full/02825/at8k6/dsv.pdf>>. Acessado em 30 maio 2007.

GILCHRIST, Alan. Thesauri, taxonomies, ontologies: an etymological note. **Journal of Documentation**, v. 59, n. 1, 2003. p. 7-18.

GUARINO, Nicola. Formal ontology and information systems. In: FOIS'98, 1998, Trento, Italy. **Proceedings...** Trento, Italy. Disponível em <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Acessado em: 01 jun.2007.

JACOB, Elin K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**. v. 29, n. 4, abr./mai. 2003. p. 19-22.

KENT, Robert E. The IFF foundation for ontological knowledge organization. Knowledge Organization and Classification in International Information Retrieval. The Haworth Press, Inc., 2003, p. 187-203.

KNAPP, Sara D.; COHEN, Laura B.; JUEDES, D. R. A natural language thesaurus for the humanities: the need for a database search aid. **Library Quarterly**, v. 68, n. 4, 1998. p. 406-430.

LAAN, Regina H. van der. **Terminologia**: uma inter-relação lógica. Porto Alegre, 2002. 185 f. Tese (Doutorado). Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras. Porto Alegre, 2002.

LAME, Guirau de. Using NLP techniques to identify legal ontology components: concepts and relations. **Artificial Intelligence and Law**, n. 12, 2004. p. 379-396.

MARTÍNEZ, Ana María. Tesauros y listas de epígrafes: hacia una integración? **Información, Cultura y Sociedad**, n. 9, 2003. p.9-27.

McILWAINE, I. C. Trends in knowledge organization research. **Knowledge**, v. 30, n. 2, 2003, p. 75-86.

MISSIKOFF, Michele; VELARDI, Paola; FABRIANI, Paolo. Text mining techniques to automatically enrich a domain ontology. **Applied Intelligence**, v. 18, n. 3, 2003, p. 323-40. Disponível em <<http://portal.acm.org/citation.cfm?id=762622.762657>>. Acessado em: 29 maio 2007.

NAUMIS-PEÑA, Catalina. El tesoro en ambiente digital. **Investigación Bibliotecológica**, Ciudad de México, DF, v. 15, n. 31, jul./dez. 2001, p. 5-30. Disponível em <<http://www.ejournal.unam.mx/iibiblio/vol15-31/IBI03102.pdf>>. Acessado em: 01 jun.2007.

\_\_\_\_\_. Análisis de la confluencia entre término y descriptor em la elaboración de tesauros. **Investigación Bibliotecológica**, Ciudad de México, DF, v. 14, n. 29, jun./dez.2000, p. 95-113. Disponível em <<http://www.ejournal.unam.mx/iibiblio/vol14-29/IBI02906.pdf>>. Acessado em: 01 jun.2007.

NIELSEN, Marianne L. Thasaurus construction: key issues and selected readings. **Cataloging & Classification Quarterly**, v. 37, n. 3/4, 2004. p. 57-74.

PRIETO-DÍAZ, Rubén. **A faceted approach to building ontologies**. Universiteit Utrecht. Department of Information and Computing Sciences, 2002. Disponível em <<http://www.cs.uu.nl/docs/vakken/ks/BulidOntologiesRPD-ER2002.pdf>>. Acessado em: 01 jul.2007.

PULIDO, J.R.G. et al. Ontology languages for the semantic web: a never completely update review. **Knowlegde-Based Systems**, n. 19, 2006. p. 489-497.

ROSA, Antonio de la. Tesauros, tesauros automaticos, tesauros automaticos online. El Profesional de la Información: **Revista Internacional Científica y Profesional**, Barcelona, v. 8, n. 7-8, jul-ago/1999. Disponível em <[http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros\\_tesauros\\_automaticos\\_tesauros\\_automaticos\\_online.html](http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros_tesauros_automaticos_tesauros_automaticos_online.html)>. Acessado em: 01 jun.2007.

SCHMITZ-ESSER, Winfried. Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval. **Knowledge Organization**, v. 26, n. 1, 1999, p. 10-22.

SCHNEIDER, Jesper W.; BORLUND, Pia. Introduction to bibliometrics for construction and maintenance of thesauri: methodical considerations. **Journal of Documentation**, v. 60, n. 5, 2004. Disponível em <[http://www.db.dk/binaries/schneider\\_borlund%20\(2004\).pdf](http://www.db.dk/binaries/schneider_borlund%20(2004).pdf)>. Acessado em: 01 jul.2007.

SHIRI, Ali Asghar; REVIE, Crawford. Thesauri on the WEB: current developments and trends. **Online Information Review**, Tucson, v. 24, n. 4, 2000, p. 273-279. Disponível em <<http://dlist.sir.arizona.edu/163/01/thesauri.pdf>>. Acessado em: 01 jun.2007.

SOERGEL, Dagobert. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**. v. 50, n. 12, Oct./1999, p. 1119-20. Disponível em <<http://www.dsoergel.com/cv/B70.pdf>>. Acessado em: 30 Maio 2007.

SPITERI, Louise F. The essential elements of faceted thesauri. **Cataloging & Classification Quarterly**, v. 28, n. 4, 1999. p. 31-52.

SOWA, John F. **Building, sharing, and merging ontologies**. 2006. Disponível em <<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>>. Acessado em: 29 maio 2007.

THOMAS, Alan R. Teach yourself thesaurus: exercises, readings, resources. **Cataloging & Classification Quarterly**, v. 37, n. 3-4, 2004. p. 23-34.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004. Disponível em <<http://www.ibict.br/cienciainformacao/viewarticle.php?id=120&layout=abstract>>. Acessado em: 29.mai.2007.

WEBER, Ron. Conceptual modelling and ontology: possibilities and pitfalls. **Journal of Database Management**, v. 14, n. 3, jul-set/ 2003. p. 1-20.

## APÊNDICES

## **Apêndice A**

# RELATÓRIO 1

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>ROSA, Antonio de la. Tesauros, tesauros automaticos, tesauros automaticos online. El Profesional de la Información: Revista Internacional Científica y Profesional, Barcelona, v. 8, n. 7-8, jul-ago/1999. Disponível em &lt;<a href="http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros_tesauros_automaticos_tesauros_automaticos_online.html">http://www.elprofesionaldelainformacion.com/contenidos/1999/julio/tesauros_tesauros_automaticos_tesauros_automaticos_online.html</a>&gt;. Acessado em: 01/jun./2007.</p>	<p>Apresenta uma análise de tesouro como um instrumento de classificação levando em conta sua estrutura e gestão. Chama a atenção para as transformações ocorridas no conceito de tesouro com as mudanças de meio de atuação e implementação. Apresenta também as transformações ocorridas em seus objetivos e o processo de adaptação deste instrumento na WEB. Propõe diretrizes para a construção de softwares de elaboração de tesauros automáticos na WEB. Não enfoca especificidades sobre termos e conceitos, o foco está voltado para a nova</p>	<p>Relação entre os próprios termos: sinonímia, hierárquica e por associação. Porém, reclama pelo fato de que as relações deveriam ser estabelecidas pelos usuários no momento da consulta, estabelecendo a longitude e os números de campos dos termos. Não fica claro a definição de termo e nem sua relação com o conceito, embora fique implícita a idéia de que o termo é o representante lingüístico do conceito.</p>	<p>Conceitos são representações abstratas percebidas e interpretadas de algo real, e ordenadas em classes já existentes no âmbito cognitivo de quem os apreendem. Todo conhecimento apreendido é classificado de acordo com classes (agrupamento de elementos por semelhança) cognitivamente pré-estabelecidas. E é a ordenação dos conceitos que rege a gestão da informação que, por sua vez, rege a gestão do conhecimento. Não fica clara a relação existente entre conceito e termo, embora fique implícita a idéia de que o termo é o representante lingüístico do conceito.</p>	<p>Orientar qual o termo mais adequado para representar um conceito. Servir de ponte entre necessidades informacionais e sistemas de recuperação de informação. Deve ser um instrumento que, embora controlado, seja um recurso coordenável (no que tange coordenação de vocabulário, relação entre termos e apresentação) pelo usuário. Fato esse que facilitaria a inter-relação linguagem natural e linguagem artificial.</p>	<p>Enfatiza que a construção de tesauros automáticos devem melhor explorar as possibilidades da WEB, visando a elaboração de tesauros compartilhados integrados a sistemas de informação. Devem ser mais flexíveis, possibilitando que os usuários estabeleçam vocabulários (com base em estatísticas de uso e frequência), definam relações entre os termos e escolham a apresentação do tesouro, ou seja, o usuário desenharia sua estratégia de busca de informação manipulando o próprio procedimento do tesouro, inserindo e apagando termos, como em um sistema Wiki. O tesouro seria periodicamente atualizado de acordo com termos e relações estabelecidas por usuários no momento da busca, e teria sua consistência garantida automaticamente pelo software. Para o vocabulário do tesouro, seria possível agregar termos extraídos de documentos da WEB. [Isto é uma proposta para a construção de softwares</p>

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		SHIRI, Ali Asghar; REVIE, Crawford. Thesauri on the WEB: current developments and trends. Online Information Review, Tucson, United States, v. 24, n. 4, 2000, p. 273-279. Disponível em < <a href="http://dlist.sir.arizona.edu/163/01/thesauri.pdf">http://dlist.sir.arizona.edu/163/01/thesauri.pdf</a> >. Acessado em: 01/jun./2007.	O artigo apresenta uma vista geral de desenvolvimentos recentes relacionados à aplicação de tesouros na organização e na recuperação da informação na Web. Descreve alguns projetos recentes de tesouro empreendidos para facilitar o acesso aos recursos da informação na Internet. Tipos de tesouros disponíveis na Web, Tesouros integrados nas bases de dados e nos sistemas de recuperação da informação, e os sistemas de tesouros para busca em bases de dados cruzados também são assuntos discutidos neste texto.	Não trata.	Não trata.	Tesouro WEB tem por objetivo potencializar a recuperação da informação na Internet atuando como uma interface entre informação e usuário.	Tipos de tesouros WEB quanto a estrutura e formato: tesouro em formato de texto simples e estático (ASFA); tesouros em formato HTML (ainda estático, sem uso efetivo de hiperlinks) (Inforterm); tesouros em formato dinâmico HTML (com uso de hiperlinks) (MeSH); tesouros com interfaces visuais e gráficas (Plumb Design Visual Thesaurus); e tesouros em formato XML (virtual HyperGlossary). Tipos de tesouros WEB quanto à funcionalidade e usabilidade: tesouro autônomo (não é parte de um sistema de informática); e tesouros totalmente integrados a banco de dados ou sistemas de recuperação da informação.
X		SCHNEIDER, Jesper W.; BORLUND, Pia. Introduction to bibliometrics for construction and maintenance of thesauri: methodical considerations. Journal of Documentation, v. 60, n. 5, 2004. Disponível em < <a href="http://www.db.dk/binaries/schneider_borlund%20(20">http://www.db.dk/binaries/schneider_borlund%20(20</a>	O artigo introduz métodos bibliométricos na área da pesquisa de organização do conhecimento, especialmente no que diz respeito à construção e à manutenção dos tesouros. Revê os trabalhos que inspiraram a abordagem semi-automática, baseado na bibliometria, para a construção e manutenção de tesouros. Discute também as considerações metódicas referentes à citada abordagem. Eventualmente, a abordagem semi-automática é usada para verificar a	Termo é palavra ou frase de algum domínio de conhecimento. Preferencialmente substantivos e frases de substantivos. Relacionamento entre termos: equivalência, hierárquico e associativo.	Não trata	Não trata	Tesouros são fundamentalmente lingüísticos e conceituais. Para a seleção de termos candidatos é necessário uma abordagem semi-automática, que faça uso de métodos intelectual manual e

aplicabilidade de métodos bibliométricos como um suplemento à construção e a manutenção dos tesouros. No contexto da organização do conhecimento, o trabalho esboça duas abordagens fundamentais à organização do conhecimento, a abordagem intelectual manual e a abordagem algorítmica automática. Métodos bibliométricos pertencem à segunda abordagem, embora apresentem características especiais que são substancialmente diferentes de outros métodos dentro

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. Perspectivas em Ciência da Informação, Belo Horizonte, v.11, n.3, p. 348-359, set./dez. 2006. Disponível em< <a href="http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&amp;layout=abstract">http://www.eci.ufmg.br/pcionline/viewarticle.php?id=491&amp;layout=abstract</a> >. Acessado em: 29.mai.2007.	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 lingüística 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	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. A 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.	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.	Serve para auxiliar a indexação e a recuperação em um sistema de recuperação de informação.	Teorias utilizadas: Teoria do Conceito, Teoria da Classificação Facetada. Defende que primeiro se determina o conceito e depois o termo.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		MARTÍNEZ, Ana María. Tesoros y listas de epígrafes: hacia una integración? <i>Información, Cultura y Sociedad</i> , n. 9,	Analisa diferenças e semelhanças existentes entre os tesouros e as listas de cabeçalhos de assunto, considerando o plano de revisão de 2003 da norma estadunidense Z39.19. Apresenta pontos relativos à importância da revisão de normas, sobretudo as de alcance internacional, a necessidade de rigor em se tratando de controle de terminologia, a importância de uma definição correta das relações hierárquicas e associativas no que concerne a gestão pós-coordenada.	<p>Relação com o próprio</p> <p>termos de indexação dos tesouros: a) descritores ou termos preferidos; b) identificadores (nomes próprios), c) não descritores ou termos não preferidos (que são os termos não autorizados para indexação).</p> <p>Existem a) termos simples, ex.: biblioteca; b) termos compostos, ex.: (frase adjetivada) Organizações Culturais Internacionais, (frase preposicional) Dinâmica em grupo; c) termo qualificado, ex.: Rádio (metal), Terra (planeta). Relações: Equivalência (USE, UP), Hierárquica (gênero/espécie, todo/parte, enumerativo, polihierárquico), Associativa (termo</p>	<p>Relação com o próprio conceito: se dá de acordo com as características do conceito (os predicados de um objeto/referente).</p> <p>Não trata.</p>	Não Trata.	<p>Pauta-se na ANSI/NISO Z39.19 e na ISO 2788. Os tesouros podem utilizar notas de alcance para esclarecer questões referentes aos significados dos termos. Tipos de nota de alcance: a) definição do termo, b) limitação do alcance do termo [que pode ser positiva (o termo inclui...) ou negativa (o termo exclui...)], c) instrução para os indexadores, d) nota histórica, e) nota de datas (data em que um termo foi modificado ou substituído).</p>
X		McILWAINE, I. C. Trends in knowledge organization research. <i>Knowledge</i> , v. 30, n. 2, 2003, p. 75-86.	Apresenta as tendências correntes na pesquisa em organização do conhecimento, focando os sistemas universais, mapeando vocabulários e interoperabilidades concernentes, problemas de padrões de qualidade, a	Não trata.	Não trata.	Auxiliar o processo de recuperação da informação, devido a sua uniformidade terminológica que possibilita esclarecer bareias lingüísticas. Auxiliar a consulta do usuário, expondo sistematicamente uma	

Internet e motores de busca, recursos de busca, tesouros e apresentação visual. Discute alguns problemas voltados aos pesquisadores atuais.

classificação que o direcione no momento da busca.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		FAYEN, Emily. A new standard for controlled vocabularies. The Indexer, v. 24, n. 2, october/2004, p. 62-65. Disponível em < <a href="http://wilsontxt.hwwilson.com/pdf/02825/at8k6/ds.v.pdf">http://wilsontxt.hwwilson.com/pdf/02825/at8k6/ds.v.pdf</a> >. Acessado em 30/mai./2007.	O texto apresenta uma revisão do padrão ANSI/NISO Z39.19. Afirmando que a referida norma é ainda relevante em muitas etapas da comunidade da informação, ressalta que o padrão foi revisado duas vezes desde sua incipiência (elaborado em 1974, foi revisado em 1980 e 1993). Porém, vem enfatizar as limitações existentes do padrão e o escopo das revisões que estão em andamento. Tais revisões são imprescindíveis devido ao fato do exponencial desenvolvimento tecnológico dos mecanismos de informação, sobretudo da Internet. A atualização de 1993 da Z39.19 foi fortemente pautada na Norma Internacional ISO 2788 e na Norma Britânica BS 8723. O Padrão revisado basicamente amplia a cobertura (de documentos para conteúdos de objetos digitais) e amplia os tipos de vocabulários controlados, além de ampliar os formatos de exposição e a interoperabilidade, possível no meio digital.	Relacionamento entre termos: Equivalência (uso/usado para) - indica o termo preferido. Hierarquia - indica um termo mais geral e mais específico. Associação - indica outros tipos de relacionamentos entre termos.	Não trata.	Não trata.	Propriedades dos tesouros (segundo as revisões do Padrão): Uma lista consiste somente em termos preferidos. É um grupo simples de termos. Um termo preferido é selecionado para a inclusão em um vocabulário controlado entre sinônimos ou quase-sinônimos. Um grupo de sinônimos consiste somente em termos que têm um relacionamento de equivalência. Uma taxonomia consiste somente em termos preferidos que têm um relacionamento hierárquico. É um grupo de termos preferidos, conectado por uma hierarquia ou por uma poli-hierarquia. Um tesouro possui todas as propriedades descritas acima (lista de termos, conjunto de sinônimos e taxonomia) e mais alguns tipos de relacionamento (conforme descrito no campo Termo). O tesouro tem como vantagem sobre os demais vocabulários controlados a

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		SCHMITZ-ESSER, Winfried. Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval. Knowledge Organization, v. 26, n. 1, 1999, p. 10-22.	Descreve uma proposta para uma nova abordagem no planejamento e construção de tesouro com vistas na troca de informações entre tesouros multilíngues. A ênfase é nos diversos tipos de relacionamentos encontrados entre conceitos, noções e linguagens universais. A abordagem dada aos relacionamentos é analisada com base em tesouros existentes. A proposta está situada na esfera das discussões críticas (científica) e prática, e tem por objetivo central fornecer bases para a melhoria no planejamento e integração de tesouros multilíngues. No nível científico o autor reflete duas linhas de pensamento distintas: abordagem computacional para questões de linguagem, e, abordagem fenomenológica.	Designa um objeto do pensamento (conceitos e noções). Expressa claramente um objeto de pensamento de uma determinada língua. Esses termos únicos em um sistema também são chamados de Descritores. Os descritores são termos preferidos. Após a escolha de descritores deve-se ater aos termos equivalentes ou não-equivalentes. Na proposta do autor, os termos, ou seqüência de termos, considerados de igual significados e não preferidos, devem ser Expressões de Acesso Adicional	Objeto do pensamento. É designado por um termo e este termo deve ser único no sistema complexo de um tesouro. Relação Classe I: relação termos-conceitos de uma mesma língua (sinônimos e polissemias). Relação Classe II: relação conceitos-conceitos independente da língua utilizada, chamado pelo autor de Apropriadas Relações de Conceitos (Concept Relations Proper). Na classe II há a necessidade de uma linguagem intermediária para dar conta dos relacionamentos entre os conceitos de diversas línguas. Uma cadeia de descritores equivalentes (ECD), constituída de termos de diferentes línguas, tem seus elementos identificados por meio de um número identificador de meta-linguagem (MLIN). Nesta classe II é possível treze tipos de relacionamentos: a) abstrato/genérico, b) partitivo, c) todo/parte jurídico, d) partitivo geográfico, e) descendência, f) instrumental, g) causa/efeito, h) benefício, i) prejuízo, j) matéria, l) aparência, m) processo e, n)	Sua proposta tem por objetivo: alta força de articulação no nível descritivo quanto à manutenção do controle metodológico; síntese do conjunto conceitual e consistência na análise das partes constituintes e; definição de uma plataforma de referência comum para permitir construção e manutenção distribuída de tesouros lingüísticos, bem como o intercâmbio dos dados do tesouro.	

estado.

a) É do tipo gênero e espécie. Tanto o genérico (conceito A) quanto o específico (conceito B) são concebidos

como classes características.

b) Relação entre partes de um todo material ou imaterial. c)

Relação entre partes de uma todo (o todo sendo uma Jurisprudência ou uma Lei, por exemplo. d) Relação

entre partes de um todo geográfico. e) Relação em que o conceito B descende do conceito A. f) O conceito

B é um instrumental para alcançar um resultado,

conceito A. g) O conceito B é a causa da ocorrência, ou existência, do conceito A. h)

O conceito B é útil para o conceito A. i)

Relacionamento com base em causas e efeitos adversos

(por extensão pode-se dizer que se trata de uma relação de adversidade). j) O

conceito A consiste do conceito B quando visto sob aspecto de matéria ou uso

material. l) Enxerga o conceito A como ele é reconhecível pela visão

humana: forma, configuração ou cor. m)

Conceito B indica um processo complexo

(processado) no conceito de A. n) O conceito A pode ser

visto sob o aspecto de seu estado.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>NAUMIS-PEÑA, Catalina. Análisis de la confluencia entre término y descriptor en la elaboración de tesauros. Investigación Bibliotecológica, Ciudad de México, DF, v. 14, n. 29, jun./dez.2000, p. 95-113. Disponível em &lt;<a href="http://www.ejournal.unam.mx/iibiblio/vol14-29/IBI02906.pdf">http://www.ejournal.unam.mx/iibiblio/vol14-29/IBI02906.pdf</a>&gt;. Acessado em: 01/jun./2007.</p>	<p>Traça uma relação entre termo e descriptor entendendo que o primeiro exerce sua função no âmbito da comunicação científica e, o segundo exerce sua função no âmbito da indexação e recuperação da informação. O termo está localizado no marco da terminologia e o descriptor está localizado no marco da organização do conhecimento. Defende a idéia de que o termo serve para outorgar clareza e precisão ao descriptor na comunicação especializada.</p> <p>Apresenta traços que distingue termos de descritores.</p>	<p>Pautando-se na TGT afirma que uma das finalidades dos termos é evitar a superposição de conceitos e buscar a diminuição da flexibilidade da linguagem. A terminologia não estuda o termo levando em consideração aspectos psicológicos ou filosóficos, mas sim sua expressão por meio da linguagem. O ideal é que haja um termo para cada conceito. O termo deve se situar na estrutura de um dado conhecimento. Os termos integram uma estrutura agrupada por princípios cognitivos. Termos são unidades lingüísticas e unidades de conhecimento especializado. Termo é uma unidade lingüística cuja função principal é descrever um conceito de maneira unívoca em um sistema de informação. Deve-se levar em conta o 'contexto' que envolve o termo. O contexto é tido como toda realidade (verbal, física, histórica, cultural e ocasional) que rodeia o signo. É o contexto (o campo de conhecimento, neste caso) que irá outorgar o conceito de determinado termo. O descriptor, que é um termo controlado, representa um conceito. O termo atua na comunicação especializada, o descriptor</p>	<p>Pauta-se nas afirmações de que o conceito é construído por meio de um processo de agrupação e ordenação dentro de categorias abstratas dos entes percebíveis, e se relaciona com outros conceitos por meio das proximidades de suas características. Relação com o termo: o termo representa um conceito. Relação com o próprio conceito: proximidade das características.</p>	<p>Possibilitar uma recuperação pertinente da informação, de modo a economizar o tempo do usuário, minimizando ruídos, como ambigüidades, no momento da representação do conteúdo documental.</p>	

atua na indexação e recuperação  
 O termo normalizado é uma fonte para validar o descritor.  
 Relação com o conceito: designação.  
 Relação com o próprio termo: os termos se ordenam de acordo com seu nível de especificidade.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>NAUMIS-PEÑA, Catalina. El tesouro en ambiente digital. Investigación Bibliotecológica, Ciudad de México, DF, v. 15, n. 31, jul./dez. 2001, p. 5-30. Disponível em &lt;<a href="http://www.ejournal.unam.mx/iibiblio/vol15-31/IBI03102.pdf">http://www.ejournal.unam.mx/iibiblio/vol15-31/IBI03102.pdf</a>&gt;. Acessado em: 01/jun./2007.</p>	<p>Apresenta questões relacionadas à estrutura dos tesouros em um ciclo documental inserido no ambiente digital. Pautando-se no fato de que o meio informático proporciona significativas vantagens na representação e recuperação da informação, defende a idéia de que os descritores de um tesouro substituí o uso de palavras da linguagem natural.</p>	<p>O termo serve para expressar um conceito. Em um tesouro, a clareza do conceito está implícita no termo. A combinação de termos simples (termo compostos) aumenta a precisão do conceito expresso. A autora admite que existem termos polissêmicos, homógrafos. [Tal fato leva a crer que sua posição é de que termo e conceito estão separados (no sentido de serem elementos que não pertencem um ao outro), porém relacionados com base em seu contexto e uso]. Os termos são formados basicamente por substantivos e, quando necessário, por adjetivos qualificadores. São empregados também, porém com menos frequência, adjetivos substantivos (participio), advérbios, preposições e artigos para a composição de descritores compostos. Considera que termos formam descritores. [Vários termos podem formar um único</p>	<p>A rede de conceitos rege a estrutura de um tesouro. Ora trata como significado de palavra ora como conceito de termo. Não apresenta uma definição para conceito, porém deixa implícito que se trata de uma abstração de um 'ente' real.</p>	<p>Serve para dar maior consistência à indexação e potencializar o acesso do usuário (recuperação da informação). A autora (NAUMIS-PEÑA, 2001, p. 8) enumera os seguintes objetivos:          1) Servir de vocabulário oficial para coordenar dois processos: indexação e recuperação de documentos digitais de um sistema;          2) Propor um sistema de símbolos lingüísticos para agrupar informação similar relacionada ou direcionar para grupos mais gerais ou mais específicos de uma temática;          3) Alcançar uma normalização da terminologia do sistema de informação a ser utilizado;          4) Propôr um conjunto estruturado de termos sob a base de um sistema de conceitos aptos para organizar os conteúdos</p>	

descritor].  
 Relação com o conceito:  
 designação  
 Relação com o próprio  
 termo: equivalência  
 (sinônimo), hierárquica  
 (gênero/espécie), associativa  
 (termos relacionados  
 horizontalmente e, notas de  
 esclarecimento).

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
<b>X</b>		LAAN, Regina H. van der. Terminologia: uma inter-relação lógica. Porto Alegre, 2002. 185f. Tese (Doutorado). Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras. Porto Alegre, 2002.	Estudo referente à inter-relação entre tesouro e Terminologia sob a ótica da Teoria Comunicativa da Terminologia. Essa inter-relação se dá por meio da aproximação dos descritores com os termos. Estudando os tesouros sob a perspectiva de organização estrutural, enfatiza os descritores como elementos de representação e recuperação de informação em uma determinada especialidade do conhecimento. Parte do pressuposto que a relação entre sistema de recuperação e usuário é uma relação de comunicação especializada.	Sugere uma aproximação entre descritores (unidade de representação e recuperação de informação) e termos, ou seja, enxergar o descritor como uma unidade terminológica. O termo é a denominação de um conceito, serve para representar o conceito. "Termo é a unidade lexical básica da Terminologia que expressa um conceito, um objeto ou um processo. É uma unidade de designação de elementos do universo percebido ou concebido [...] termo é a designação de um conceito por uma expressão lingüística em uma linguagem de especialidade"(LAAN, 2002, p.63). A unidade lexical assume papel de termo dentro de um contexto, dentro de uma situação comunicacional real. Os termos são unidades lexicais que se situam nos discursos especializados, por isso apresentam sinonímia e variação.	Conceitos são representações mentais, são únicos, não podem ser compostos, o que ocorre são diferentes traços semânticos que integram um determinado conceito. Relação com o próprio conceito: lógica e ontológica. Relação lógica: baseada no grau de semelhança entre os objetos e conceitos. Este tipo de relação pode ocorrer por meio de uma lógica de subordinação (é partir do conceito mais geral para o mais específico, ou seja, do conceito de maior extensão e menor intensão para o conceito de menor extensão e maior extensão) ou por uma lógica de superordenação (o oposto da subordinação, do conceito mais específico para o mais geral). Esta relação pode ser de gênero/espécie ou todo/parte. Relação ontológica: não são baseadas na semelhança entre os conceitos, mas sim na proximidade situacional	Auxiliar o processo de representação e recuperação da informação em um SRI. Controlar as ambigüidades próprias do léxico comum. Facilitar a comunicação entre SRI e usuário.	Teoria utilizada: TGT e TCT

Função e relação com o conceito: a definição é um tanto quanto obscura, pois ora menciona termo como um signo lingüístico [possuidor de significante (denominação) e significação (conceito)] ora menciona o termo como uma unidade lexical que designa um conceito, e ora como designação de conceito, objeto ou processo. O que fica evidente é a definição de que o termo "é uma unidade lexical terminológica que se encontra em funcionamento nos discursos especializados" (LAAN, 2002, p.48).  
 Relação com o próprio termo: Conforme a relação entre os descritas na próxima categoria.

dos elementos em uma realidade. Este tipo de relacionamento, que baseado na contigüidade dos conceitos no espaço, pode ser de coordenação, ou encadeamento, como causa e efeito ou sequenciais.  
 Relação de equivalência: ocorre quando um conceito é representado por mais de uma denominação.  
 Relação com o termo: o termo serve como designação do conceito. É único. Para cada termo um conceito.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		THOMAS, Alan R. Teach yourself thesaurus: exercises, readings, & resources. Cataloging & Classification Quarterly, v. 37, n. 3-4, 2004. p. 23-34.	Apresenta uma base lógica para auto-instruções de elaboração de tesouro. Além de apresentar definições de tesauros, fornece recursos apropriados para uma auto-instrução adequada. Descreve pontos relativos à classificação, à análise facetada, e à catalogação por assunto. Propõe e explica uma abordagem para desconstruir e reconstruir seções de sistemas de classificação e tesauros. Detalha procedimentos para	Remete a tutoriais de elaboração de tesauros que contemplam os seguintes tipos de relacionamentos: termos gerais (TG), específicos (TE), relacionados (TR), termos preferidos (UP) e não-preferidos (USE). Além de notas explicativas (NE).	Não trata.	Não trata.	Para explicar o sistema de auto-instrução para construção de tesauros o autor faz menção às classificações sistemáticas (CDD, CDU e LC) e facetadas (BLISS Bibliographic Classification).

empregar exercícios na construção de tesouros. São sugeridos meios de avaliação e utilização de softwares livres. Foca na possibilidade de criar uma auto-aprendizagem de elaboração de tesouros.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		SPITERI, Louise F. The essential elements of faceted thesauri. <i>Cataloging &amp; Classification Quarterly</i> , v. 28, n. 4, 1999. p. 31-52.	<p>O objetivo do artigo é avaliar, comparar e contrastar como análise facetada é empregada para a exposição (apresentação) sistemática ou facetada de tesouros de recuperação da informação. Visa examinar quais princípios da análise facetada são empregadas nos tesouros, e quais os tesouros que aplicam esses princípios. Com um instrumento de medida foram avaliados a estrutura de quatorze tesouros facetados. O estudo revela que os tesouros não compartilham uma definição comum do que constitui uma faceta. Há tesouros que aplicam simultaneamente o arranjo enumerativo e o arranjo por termos indexados. Os tesouros por vezes adotam facetas que não são homogêneas e nem excludentes. O princípio de síntese é empregado em somente 50% dos casos e não há uma ordem consistente de citação.</p>	<p>Relação entre termos se dá por meio das características de divisão (ou qualidades essenciais, exclusivas e homogêneas) que os tornam singulares e pertencentes a uma única faceta.</p>	<p>Não trata.</p>	<p>Não trata.</p>	<p>Análise facetada é definida como classe de termos de um determinado campo de conhecimento. As facetas são homogêneas, mutuamente exclusivas, derivadas de um universo comum por características singulares de divisão. Uma faceta consiste em um conjunto de termos que representa somente uma característica de divisão de um determinado campo de assunto.</p> <p>Componentes típicos de um tesouro facetado: uma introdução, uma apresentação sistemática ou facetada, e uma apresentação alfabética. Lança mão dos princípios teóricos de Ranganathan e do CRG para examinar o tratamento que os tesouros de recuperação de informação estão dando à questão da análise facetada (e sistema analítico-sintético).</p>

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		KNAPP, Sara D.; COHEN, Laura B.; JUEDES, D. R. A natural language thesaurus for the humanities: the need for a database search aid. Library Quartely, v. 68, n. 4, 1998. p. 406-430.	Pretende determinar o valor potencial de um tesauro que integra termos livres das Ciências Humanas e Sociais para a área das humanidades. Realiza uma comparação entre um cabeçalho de assunto da área de humanidades (Humanities Index) com um tesauro das Ciências Sociais (The Contemporary Thesaurus of Social Science Terms and Synonyms: A Guide for Natural Language Computer Searching) para verificar o ponto de coincidência entre ciências humanas e sociais. O resultado aponta que é grande e significativo o ponto de contato entre as duas áreas, e um tesauro que integre as duas terminologias é relevante para ambas as disciplinas. Os resultados demonstram também o baixo índice de utilização de quase-sinônimos no cabeçalho de assuntos. Foi elaborada também uma busca realizada por estudantes da área de humanidades utilizando tanto de vocabulários controlados quanto termos de texto-livre, para verificar a relevância dos resultados encontrados. O resultado indica que a combinação das duas abordagens é mais pertinente, e, que o tesauro de termos de texto-livre proposto na pesquisa	Se refere a termos de texto-livre (se comparado aos termos (descritores) dos vocabulários controlados). Segundo as autoras: a) termos especificam forma e significado. b) termos expressam significados. c) termos descrevem conceitos. Podem ser empregados para minimizar ambiguidades, já que um termo de texto-livre define um conjunto relevante de termos registrados em uma base de dados. Os aspectos relativos ao 'termo', que aqui são apresentados, se refere aos termos de texto-livre atuando em ambientes de recuperação de documentos em bases de dados automáticas. Os termos de texto-livre são aqueles colocados nos campos de assunto no momento da busca.	Não trata.	Não trata.	

complementa o vocabulário controlado.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
<b>X</b>		NIELSEN, Marianne L. Thasaurus construction: key issues and selected readings. <i>Cataloging &amp; Classification Quarterly</i> , v. 37, n. 3/4, 2004. p. 57-74.	Apresenta uma bibliografia selecionada com o intuito de introduzir questões e problemas relacionados à construção de tesouros e, apresenta também um conjunto de leituras que podem ser utilizadas na prática de elaboração de tesouros. Discute o conceito de tesouros e como este conceito tem se envolvido nas tecnologias de recuperação de informação. Apresenta ainda diferentes abordagens de construção de tesouros e detalha uma série de leituras que auxiliam no desenvolvimento dessas linguagens documentárias que são pautadas em termos e conceitos. Reflete tanto as técnicas manuais quanto as automáticas para a elaboração de tesouros.	O relacionamento associativo dos termos depende de como o ambiente informacional, em que o tesouro está operando, lida com a utilização dos conceitos. O modo como o domínio relata e relaciona os termos é mais relevante que as próprias diretrizes normatizadoras. A forma do termo é controlada (em um tesouro) por uma gramática formal que determina a ortografia, a forma singular ou plural, abreviações e formas compostas. Acordos e restrições são utilizadas pela adição de notas de escopo e, no caso de homógrafos, são empregadas definições e frases qualificativas. Tipos de relacionamentos: equivalente, hierárquico e associativo. Percebe-se uma indefinição do texto, pois ora parece evidenciar que o termo é constituído por forma e conceito ora afirma que o termo apenas expressa um conceito.	O significado do conceito é controlado (em um tesouro) e sinônimos são selecionados para expressar conceitos. O uso dado aos conceitos em um determinado domínio, e os fatores característicos do domínio, determinam mais eficazmente o relacionamento conceitual de um tesouro do que os padrões internacionais de elaboração.	Guiar o processo de indexação e recuperação controlada da informação, bem como, o processo de indexação baseado em linguagem natural. Além de seu uso tradicional voltado à indexação e à recuperação, os tesouros estão sendo empregados como ferramentas para gerir e comunicar a linguagem específica usada nas organizações. Tesouros do tipo 'taxonomia corporativa' tem como objetivos: a) assegurar e sustentar a recuperação da informação; b) auxiliar no entendimento e uso de um assunto ou linguagem específica e; c) facilitar o entendimento conceitual e comunicacional em um nível lingüístico que conceba um vocabulário corporativo.	Tesouros de busca são diferentes dos tesouros tradicionais que visam a padronização terminológica. Os tesouros de busca operam como um vocabulário voltado à busca da informação. Também conhecido como tesouro de usuário-final, esse tipo de tesouro oferece uma grande lista de termos sinônimos e equivalentes (respeitando os relacionamentos semânticos) para orientar o usuário no momento da busca, ou seja, funciona como uma ferramenta que permitirá o usuário fazer uso dos mesmos termos que foram pensados pelos indexadores do sistema. Um tipo mais recente de tesouro é a taxonomia corporativa, que incorpora elementos dos sistemas de classificação e dos tesouros. Esse tipo de tesouro ordena e etiqueta metadados, fornecendo acesso a dados e informações primárias em uma estrutura hierárquica, permitindo o gerenciamento e manipulação sistemático dos dados. É utilizado para a comunicação sistemas de empresa.
<b>X</b>		TRISTÃO, A. M. D.; FACHIN, G. R. B.;	Tem o objetivo de aprofundar questões teóricas	O termo (terminum) é um elemento que sintetiza o	Representação mental das relações que ocorrem entre	Organizar e recuperar informação.	Teorias utilizadas: Teoria do Conceito, Sistema de

<p>ALARCON, O. E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. Ciência da Informação, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004. Disponível em &lt;<a href="http://www.ibict.br/cienciainformacao/viewarticle.php?id=120&amp;layout=abstract">http://www.ibict.br/cienciainformacao/viewarticle.php?id=120&amp;layout=abstract</a>&gt;. Acessado em:</p>	<p>e conceituais que fundamentam a concepção de sistema de classificação facetada e tesouro. Apresenta aspectos teóricos referentes aos meios utilizados para organização e recuperação da informação. Descreve princípios, natureza e tipos de sistema de classificação, enfatizando o sistema de classificação facetada e o tesouro.</p>	<p>conceito como um todo, possibilitando sua comunicação verbal. O termo denota um referente. O tratamento terminológico de um termo está no fato de vê-lo como um representante de um referente. O termo é o Definiendum (as características do conceito ou os atributos predicáveis do objeto/referente são o Definiens). O definiendum (termo) + o definiens (predicados do objeto) formam a definição do conceito dentro de um sistema, e essa definição serve para fixar um conceito em um sistema de conceitos.</p>	<p>objetos na realidade empírica. A definição do conceito é o conjunto formado pelas características próprias de um objeto e por um termo que as sintetizam. No nível conceitual as relações entre conceitos se dão de maneira lógica (resultado da própria compreensão dos conceitos) e ontológica (relação entre conceito e realidade). Relação com o termo: um termo + os predicados do referente formam a definição do conceito. Relação com o próprio conceito: Tipos de relacionamento: a) lógico - quando existem características comuns na definição dos conceitos, podendo ser gênero/espécie, analítico e de oposição. b) ontológico - são relações indiretas, pois resultam das propriedades do objeto empírico, podendo ser partitivo, de sucessão e material/produto. c) de efeito - associação indicando o efeito, podendo ser de causalidade, instrumental, descendência, genealógica, estágios de substâncias, antogênicas.</p>	<p>Classificação Facetada.</p>
--	--	---	---	--------------------------------

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>AITCHISON, Jean; CLARKE, Stella Dextre. The thesaurus: a historical viewpoint, with a look to the future. Cataloging &amp;</p>	<p>O artigo apresenta uma visão histórica dos tesouros, percorrendo a respeito dos padrões voltados à construção de tesouros,</p>	<p>Tipos de relacionamento: hierárquico (gênero-espécie e todo-parte), associativo (ou não hierárquico), equivalência (considera os</p>	<p>Não trata.</p>	<p>Auxiliar a recuperação da informação nos processos de indexação (indexadores) e busca (usuários). Em um primeiro momento,</p>	<p>Tipos de estrutura de tesouros: abordagem alfabética e abordagem sistemática (esquema de classificação, arranjos por notações). Há</p>

Classification Quarterly, v. 37, n. 3/4, 2004, p. 5-21.

assim como a respeito de seus diversos formatos. Descreve as abordagens que predominaram em determinados períodos, no que tange a construção de tesouros, sobretudo as abordagens alfabética e sistematicamente classificada. Aborda a construção de tesouros na era computacional, focando a questão do acesso à informação e a interoperabilidade.

não descritores, embora determine o termo preferido, 'descriptor'. Uso de tags USE, UF, BT, NT e RT para remissivas).

auxiliar a recuperação da informação controlando sinônimos e homógrafos e apresentando relações hierárquicas e associativas entre termos. Em um segundo momento, acrescenta-se a relação de equivalência, incluindo termos sinônimos e quase sinônimos, termos preferidos ou descritores (considerando a existência dos não descritores).

também a abordagem ancorada pelo sistema de classificação facetada. Nesse estilo de tesouro as principais subdivisões são os campos de assunto que, inseridos em uma análise facetada, determinam relações de termos em cada campo de assunto ou entre outros campos. O tesouro baseado em análise facetada pode ainda não ser subdividido por campos de assuntos, mas sim por categorias fundamentais (entidades, ações, espaços, tempo etc.)

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>MISSIKOFF, Michele; VELARDI, Paola; FABRIANI, Paolo. Text mining techniques to automatically enrich a domain ontology. Applied Intelligence, v. 18, n. 3, 2003, p. 323-40. Disponível em &lt;<a href="http://portal.acm.org/citati on.cfm?id=762622.762657">http://portal.acm.org/citati on.cfm?id=762622.762657</a>&gt; . Acessado em:</p>	<p>No que se refere à construção de ontologias, um assunto complicado é a identificação, a definição e a entrada de conceitos definidos. Em domínios de aplicações complexas, essas tarefas demandam muito tempo, custos e controvérsias conceituais. A redução de tempo, custos, e até mesmo de ásperas discussões é imprescindível para a construção e atualização de ontologias. Ferramentas de mineração de textos podem auxiliar esta difícil tarefa. Este artigo apresenta as diretrizes da SymOntos (sistema de gerenciamento de ontologias), e a abordagem de mineração de textos (em inglês "text mining") para sustentar a construção de ontologias. Esses</p>	<p>Os significados dos termos são as definições. Não faz distinção entre termo e conceito, ora fala em conjunto de termos ora em conjunto de conceitos relacionados de acordo com seus significados. Afirma que o que existe é um vocabulário formado por termos (conceitos) de um determinado domínio. Termos são etiquetas de conceitos.</p>	<p>Conceitos são: entidades, atributos e processos. Conceitualização é o inter-relacionamento entre os significados dos conceitos.</p> <p>Afirma que ontologias descrevem conceitos e tesouros descrevem termos.</p>	<p>Reduzir ou eliminar confusões terminológicas e conceituais. Identificar e definir corretamente um conjunto de conceitos relevantes que caracterize um dado domínio de aplicação. Visa uma estrutura conceitual uniforme. Comunicação e cooperação entre pessoas. Melhor organização das empresas. Interoperabilidade entre sistemas. Benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação). Uma ontologia pode ser vista como um estágio preliminar de uma base de conhecimento, cujo objetivo é a descrição de conceitos</p>	<p>Adota a definição de que ontologia é 'especificação explícita e formal de uma conceitualização compartilhada' de Gruber (1993). Necessariamente uma ontologia é formada por um vocabulário de termos com seus significados (definições) e seus relacionamentos. Uma ontologia pode ser vista como um tesouro enriquecido que permite ir além do relacionamento entre termos. Por meio de um relacionamento semântico mais rico, a ontologia é capaz de representar um conhecimento conceitual. A abordagem de "text mining" visa a extração de conceitos relevantes de um dado domínio relatados na literatura e a descoberta de relações</p>

operadores de mineração são importantes para a questão da extração de conceitos e do relacionamento semântico entre conceitos de um dado domínio.

necessários para falar sobre determinado assunto. A base de conhecimento por sua vez consiste no conhecimento necessário para idealizar um problema, derivado de um novo conhecimento, com o fim de responder questões.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	ALMEIDA, Maurício B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. <i>Ciência da Informação</i> , Brasília, v. 32, n. 3, set./dez. 2003, p. 7-20. Disponível em < <a href="http://www.ibict.br/cienciadainformacao/viewarticle.php?id=36&amp;layout=abstract">http://www.ibict.br/cienciadainformacao/viewarticle.php?id=36&amp;layout=abstract</a> >. Acessado em: 01.jun.2007.	Focando a questão da recuperação da informação e o uso de ontologias para tal questão, o artigo apresenta um panorama geral sobre o estado-da-arte no estudo de ontologias. Apresenta definições e discussões acerca dos termos e significados atribuídos às ontologias, além de tipos de ontologias para deferentes domínios do conhecimento. Expõe uma série de metodologias, linguagens e ferramentas utilizadas para a construção de ontologias encontradas na literatura.	O artigo não foca aspectos relativos ao termo. Afirma que a relação entre os termos é feita pelos especialistas, o que possibilita que o usuário formule consultas usando conceitos especificados. Para os autores a ontologia define regras para a combinação e relação entre os termos, define o conjunto de termos (linguagem) que o usuário utilizará para formular a busca. É possível dizer com isso que, sob a ótica do artigo, o conceito está contido no termo especificado pelos especialistas. Função: viabilizar a formulação de consultas (por parte do usuário) fazendo uso de conceitos predeterminados pelos especialistas. Relação com o conceito: o termo contém conceito. Relação com o próprio termo: consenso de um determinado domínio.	O artigo não é enfático quanto à definição do conceito. Simpatizando com uma definição de conceitualização sob a ótica intensional, em que afirma ser a conceitualização uma lista de características do conceito, e uma visão abstrata e sintetizada de um ente real, fica implícito uma proximidade com a Teoria do Conceito no que se refere a afirmação de que 'conceito é a soma dos enunciados verdadeiros referentes a um objeto real'. Relação com o termo: o conceito é parte do termo. Relação com o próprio conceito: a intensão do conceito, ou seja, o conjunto de características do conceito, é responsável pelo relacionamento entre eles.	Possibilitar a consulta a um sistema automatizado de informação (recuperação da informação) utilizando termos conceitualizados e especificados por especialistas	Estruturas que se organizam a partir da utilização de termos são os arquivos de autoridade, glossários e dicionários. Estruturas que se organizam com a classificação e a criação de categorias são os cabeçalhos de assunto e os esquemas de classificação (ou taxonomias). As estruturas que se organizam a partir de conceitos e de seus relacionamentos são as ontologias, os tesouros e as redes semânticas. (ALMEIDA; BAX, 2003, p.1). Os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados) (GRUBER,1996; NOY & GUINNESS, 2001 apud ALMEIDA; BAX, 2003, p.3).

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
X		<p>DENNY, Michael. Ontology building: a survey of editing tools. XML.com. 2002, p. 1-7. Disponível em &lt;<a href="http://www.xml.com/lpt/a/2002/11/06/ontologies.html">http://www.xml.com/lpt/a/2002/11/06/ontologies.html</a>&gt; Acessado em: 01/07/08.</p>	<p>Diferentemente dos formatos superficiais e dos recursos dos bancos de dados relacionais e do XML, a ontologia, pode frequentemente fornecer uma especificação objetiva da informação do domínio, representando um consenso nos conceitos e nas relações que caracterizam o conhecimento da maneira que o domínio está expressado. Esta especificação pode ser a primeira etapa nos sistemas de informação (orientados semanticamente) para empresas, governos, e atividades pessoais. Apresenta etapas da construção de uma ontologia. Detalha ferramentas utilizadas para a construção de ontologias.</p>	<p>As ontologias possuem uma peça chamada componente terminológico, que define os termos e a estrutura da ontologia. O componente assertivo apresenta exemplos e indivíduos que manifestam a definição terminológica. Os termos são usados para definir formalmente 'coisas' em um domínio de interesse.</p>	<p>Há uma decisão ontológica específica que determina a divisória entre o tratamento de uma coisa como um conceito e o tratamento de uma coisa como indivíduo. Para projetar a estrutura conceitual é necessário identificar os conceitos concretos principais e suas propriedades do domínio, assim como seus relacionamentos, e criar conceitos abstratos como características.</p>	<p>Pode atuar nas seguintes situações: na WEB-semântica; na medicina para controlar a saúde do paciente por meio da criação de guias medicinais, e, traçando os genomas das plantas e dos animais; procurando por recursos públicos específicos da informação; projeto colaborativo da engenharia; análise de segurança in-depth; e a troca automatizada da informação eletrônica entre sócios comerciais. Ajudam a responder perguntas em um corpo de informação.</p>	<p>Na web-semântica, ontologias podem ser usadas no sentido de desambiguação de um diálogo entre aplicações remotas ou entre os agentes, combinando sempre de acordo com ontologias superiores do nível e do domínio. Deve-se considerar que cada ontologia é consensual e congruente com outras ontologias compartilhadas (por exemplo, as ontologias incluem rotineiramente uma outra); o resultado é um discurso do domínio comum que possa ser interpretado mais precisamente com regras de inferência e de lógica da aplicação. Construir uma ontologia significa atrelar coisas diferentes à práticas diferentes. Atua na descrição de algo, por isso é um avanço com relação aos vocabulários controlados, tesouros organizados em categorias, taxonomias em que os termos são organizados hierarquicamente, pois as ontologias podem atribuir propriedades que definem conceitos novos e novos relacionamentos entre os conceitos, como por exemplo: "mudanças, efeitos" ou "compras de ". Para o desenvolvimento de ontologias são utilizadas ferramentas (softwares) específicos que executam todas as especificações pré-estabelecidas. Essas ferramentas possibilitam a</p>

ligação de ontologias externas e internas ao processo, comparando-as e reconciliando-as em formulários específicos. Quanto a interoperabilidade, deve-se ater à compatibilidade formal de nível sintático e semântico.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	DENNY, Michael. Ontology tools survey, revisited. XML.com. 2004, p. 1-8. Disponível em < <a href="http://www.xml.com/lpt/a/2004/07/14/onto.html">http://www.xml.com/lpt/a/2004/07/14/onto.html</a> >. Acessado em: 01/jun./2007.	Ontologia é uma maneira de especificar a estrutura do domínio de conhecimento em uma lógica formal projetada para processar no meio informático. Seu principal papel tecnológico é gerenciar o índice dos dados das operações processuais dos algoritmos e das regras de representação dos próprios dados. O artigo consiste em uma síntese de uma pesquisa relativa aos editores de ontologias, que são os ambientes que possibilitam o desenvolvimento desses modelos de representação do conhecimento.	Não trata.	Na construção de ontologias as estruturas de esqueleto de conceitos do núcleo são estendidas com conceitos mais refinados e mais periféricos, e estes são interligados elaborando relações adicionais.	As ontologias estão sendo usadas para guiar a extração do índice semântico das coleções dos originais (textos) que descrevem pesquisas médicas, produtos de consumidores, tópicos de negócios, e ambientes governamentais. Buscadores Web como Google e Yahoo também fazem uso de abordagens baseadas em ontologias para organizar conteúdos na Web.	OWL (baseada em RDF) é uma linguagem computacional utilizada (recomendada pela W3C) na construção de ontologias, assim como o predecessor DAML+OIL. O World Wide Web Consortium (W3C) é um consórcio de empresas de tecnologia que desenvolve protocolos abertos que garantem a interoperabilidade na Web. As linguagens de ontologias usam estruturas de especificação semântica (que permitem a interpretação semânticas dos dados) , e as linguagens baseadas em objetos (object-oriented) são utilizadas como mecanismos que permitem a reutilização de códigos.
TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	GUARINO, Nicola. Formal ontology and information systems. In: FOIS'98, 1998, Trento, Italy. Proceedings... Trento, Italy. Disponível em < <a href="http://www.loa-">http://www.loa-</a>	A pesquisa sobre o ontologia vem sendo desenvolvida em diversas áreas do conhecimento, em especial na Ciência da Computação. As áreas de	Não trata.	Para a IA uma ontologia refere-se a um artefato da engenharia computacional, constituído por um vocabulário específico usado para descrever alguma	O papel de uma ontologia pode ser entendido considerando ontologia como um conjunto de axiomas lógicos projetados para esclarecer o significado	As ontologias estão sendo amplamente estudadas pela inteligência artificial, lingüística computacional, e teoria da base de dados. E também em campos de pesquisa

cnr.it/Papers/FOIS98.pdf>. Acessado em: 01.jun.2007.

aplicação que mais vêm focando estudos sobre ontologias são engenharia do conhecimento, projeto de base de dados e integração, recuperação e extração de informação. O artigo discute a base metodológica e as peculiaridades da arquitetura das ontologias. No que tange a metodologia, destaca a possibilidade de adotar uma aproximação altamente interdisciplinar. No que tange a arquitetura, destaca a centralidade que um sistema de informação gerenciado por uma ontologia pode alcançar. Focaliza a aplicação de ontologias em sistemas de informação.

realidade, mais um conjunto de suposições explícitas a respeito do significado pretendido das palavras do vocabulário. Este conjunto de suposições tem geralmente a forma de uma teoria lógica de primeira ordem, onde as palavras do vocabulário aparecem como nomes unários ou binários do predicado, respectivamente chamados de conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia dos conceitos relacionados por submissão; em casos mais sofisticados, os axiomas apropriados são adicionados a fim de expressar outros relacionamentos entre conceitos e confinar sua interpretação pretendida. Conceitualização: relação matemática ordinária (extensional), relações conceituais (intensional). Uma maneira padrão de representar relações conceituais é vê-las como funções em grupos de mundos possíveis. Quando as relações ordinárias forem definidas em um determinado domínio, as relações conceituais serão definidas em um espaço do domínio. Define-se um espaço do domínio como uma estrutura  $\langle D, W \rangle$ , onde  $D$  é um domínio e  $W$  é um conjunto de estados de relações de tal domínio (chamado também

pretendido de um vocabulário. Não é fácil (talvez nem possível) encontrar o conjunto de axiomas perfeitos. Conseqüentemente, uma ontologia especifica uma conceitualização somente de maneira muito indireta, porque 1) ela pode somente aproximar um conjunto de modelos pretendidos; 2) tal conjunto de modelos pretendidos é somente uma caracterização fraca de uma conceitualização. Afirma-se que uma ontologia  $O$  para uma linguagem  $L$  se aproxima de uma conceitualização  $C$  se existir um compromisso ontológico  $K$  (interpretação intensional incluído nos modelos de  $O$ .

como: engenharia de conhecimento, representação do conhecimento, modelo qualitativo, engenharia da língua, projeto de base de dados, informação modelar, integração da informação, análise object-oriented, recuperação e extração de informação, gerência e organização do conhecimento. As áreas de aplicações atuais são: integração de empresa, tradução de linguagem natural, medicina, mecânica projetada, padronização do conhecimento e de produto, comércio eletrônico, sistemas de informação geográficas, sistema de informação legal e sistema de informação biológica. Tipos de ontologias com base no nível de especificidade: ontologias de nível maior (Top-level) - descrevem os conceitos muito gerais como espaço, tempo, matéria, objeto, evento, ação, etc., que são independentes de um problema particular ou domínio (indicadas para ontologias unificadas voltadas para grandes comunidades de usuários); ontologias de domínio e ontologias de tarefa - descrevem, respectivamente, o vocabulário relacionado a um domínio genérico (ex.: Medicina) ou uma tarefa ou uma atividade genérica (ex.: diagnosticar ou vender), especializando os termos introduzidos na ontologia de nível maior; ontologias de aplicação - descrevem

mundos possíveis). Por exemplo, D pode ser um conjunto de blocos em uma tabela e W pode ser o conjunto de todos os arranjos espaciais possíveis destes blocos. Em um espaço do domínio  $\langle D, W \rangle$ , podemos dizer que uma conceitualização é um conjunto das relações conceituais definidas em um espaço do domínio.

conceitos dependendo de um domínio e de uma tarefa particular, que são frequentemente especializações de ambas as ontologias relacionadas.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	SOERGEL, Dagobert. The rise of ontologies or the reinvention of classification. <i>Journal of the American Society for Information Science</i> . v. 50, n. 12, Oct./1999, p. 1119-20. Disponível em <a href="http://www.dsoergel.com/cv/B70.pdf">http://www.dsoergel.com/cv/B70.pdf</a> . Acessado em: 30/Mai./2007.	Afirmando que sistemas de classificação, tesouros e ontologias são tipos de bases de conhecimento desenvolvidas para distintas finalidades, o artigo apresenta a idéia de que ontologia é apenas uma nomenclatura (nova) para um sistema de classificação, e é adotada pela IA, pelas tecnologias de programação e engenharia computacional. O texto foca as funções desempenhadas (ou que pelo menos devem desempenhar) esses tipos de ferramentas.	Não trata	Não trata	As estruturas lexicais e ontológicas são o sustentamento do trabalho científico e acadêmico, da aprendizagem, e da inteligência de máquinas. Servem para muitas funções críticas no pensar e no comunicar, no organizar, e na recuperação da informação por pessoas e máquinas. As funções das ferramentas que fornecem tais estruturas incluem o seguinte: a) Fornecer um mapa semântico de caminhos aos campos individuais e aos relacionamentos entre campos, assim fornecendo a orientação e servindo como uma ferramenta de referência. Isto inclui as seguintes funções específicas: relacionar conceitos aos termos e fornecer definições; esclarecer conceitos, contextualizando-os em uma classificação/ontologia;	

relacionar conceitos e termos ou ícones através das disciplinas, das línguas, e das culturas. b) Melhorar a comunicação e o aprendizado: assistindo escritores e leitores; sustentar o ensino fornecendo estruturas conceituais, com estudantes se esforçando para produzir tais estruturas; sustentação da aprendizagem da linguagem; e suportar o desenvolvimento de materiais instrutivos. c) Fornecer a base conceitual para o projeto da pesquisa e da execução ótimas: ajudar investigadores na exploração do contexto conceitual de um projeto de pesquisa, de uma política, de uma planta, ou de um projeto da execução e estruturar o problema; suportar a definição consistente das variáveis/medidas para uns resultados mais comparáveis e mais cumulativos da pesquisa e da avaliação. d) Fornecer a classificação para a ação: uma classificação das doenças para o diagnóstico, de habilidades da equipe de funcionários para as atribuições da tarefa, dos productos para o uso. e) Sustentar a recuperação da informação em uma base de conhecimento; suportar a busca hierárquica expandida; suportar exposições bem-estruturadas

de resultados da busca; fornecer uma ferramenta para a indexação (problem-oriented) e controlar o vocabulário do indexador. f) Fornecer a base conceitual para sistemas de bases de conhecimento. g) Fornecer a base conceitual para a definição do elemento de dados e objetar hierarquias em sistemas de software. h) Efetuar todas as funções descritas acima por meio das disciplinas, das linguagens, e das culturas. i) Servir como um dicionário para o uso humano e como o dicionário/base de conhecimento para processar linguagem natural - tradução de máquina e compreensão de linguagem

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	DING, Ying; FOO, Schubert. <i>Ontology research and development part 1: a review of ontology generation. Object-oriented Systems Modeling Laboratory - OSM, 200?</i> . Disponível em < <a href="http://osm.cs.byu.edu/CS652s04/DF02OntologyGeneration.pdf">http://osm.cs.byu.edu/CS652s04/DF02OntologyGeneration.pdf</a> >. Acessado em: 26/mai./2007.	Ontologia é uma disciplina que tem alcançado grandes avanços no que tange a organização, o entendimento e a gestão da informação. Ela tem um papel crucial no acesso baseado em conteúdo, interoperabilidade, comunicação, e oferece maior qualidade nos níveis de serviço ofertados pela WEB. A presente pesquisa é apresentada em duas partes. A primeira apresenta uma revisão do estado-da-arte de trabalhos e pesquisas referentes a geração automática e semi-automática de ontologias. A segunda é	O sentido principal de um termo pode ser descoberto por meio da associação entre termos diferentes.	Conceitos são classes. Conceitos são representações mentais refletidas por meio de termos.	Ontologia é um modelo de referência definido explicitamente de aplicação de domínios com a finalidade de melhorar a consistência e o reuso da informação, a interoperabilidade dos sistemas e o compartilhamento do conhecimento. Descreve a semântica de um domínio em uma maneira humano-compreensível e computador-processável.	Adota a definição de 'especificação explícita e formal de uma conceitualização compartilhada de Gruber (1993). Ontologia é concebida como um conjunto de classes (conceitos), relações, funções, axiomas e instancias (GRUBER, 1993). O texto não aborda aspectos referentes aos termos e conceitos, apresenta uma revisão do estado-da-arte de trabalhos e pesquisas referentes a geração automática e semi-automática de ontologias.

dedicada ao mapeamento e manutenção de ontologias. Identificou-se que a extração de informação e o processamento de linguagem natural são desdobrados para a extração de conceitos ou classes de textos livres ou dados semi-estruturados. Tal fato faz com que a extração de relações se torne assunto muito complexo. Essa pesquisa pretende encontrar meios apropriados para detectar ou identificar relações de modo automático ou semi-automático. Este artigo corresponde somente a primeira parte da mencionada pesquisa.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	<b>X</b>	DING, Ying. A review of ontologies with the semantic web in view. <i>Journal of Information Science</i> , v. 27, n.6, 2001, p. 377-84. Disponível em < <a href="http://jis.sagepub.com/cgi/reprint/27/6/377">http://jis.sagepub.com/cgi/reprint/27/6/377</a> >. Acessado em: 28/mai./2007.	A World Wide Web está transpondo-se da primeira geração web para a segunda geração web, a web semântica. As ontologias são a espinha dorsal dessa nova geração. Este artigo fornece uma visão geral a respeito das ontologias. Discute as definições de ontologia, tipos de ontologias, ferramentas e linguagens de ontologias, além de apresentar importantes projetos encontrados na literatura corrente.	Não trata.	Não trata.	Fornecer um entendimento comum e compartilhado de um determinado domínio. Possibilitar serviços baseados em operacionalizações semânticas. Auxiliar pessoas e computadores no que diz respeito ao acesso à necessidade informacional.  Proporcionar a interoperabilidade dos serviços ofertados na WEB, potencializando os níveis de serviços ofertados, ou seja,	Adota a definição de que ontologia é a 'especificação explícita e formal de uma conceitualização compartilhada', de Gruber (1993). Uma ontologia típica apresenta uma taxonomia que define classes e suas relações, e, um conjunto de regras de inferência operadoras de funções de raciocínio. Destaca como áreas de desenvolvimento de ontologias a inteligência artificial, engenharia do conhecimento e representação

sustentar a web-semântica. Ligar parte significativa do conhecimento humano aos processamentos computacionais, potencializando assim o processo de comunicação entre especialistas de um determinado domínio.

do conhecimento. Embora este texto não aborde aspectos relativos aos termos e conceitos, ele surge como preciosa, e muito citada, fonte para definir tipologia, ferramentas e linguagens de ontologias, além de descrever os principais projetos desenvolvidos.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	<b>X</b>	PRIETO-DÍAZ, Rubén. A faceted approach to building ontologies. Universiteit Utrecht. Department of Information and Computing Sciences, 2002. Disponível em < <a href="http://www.cs.uu.nl/docs/vakken/ks/BulidOntologiesRPD-ER2002.pdf">http://www.cs.uu.nl/docs/vakken/ks/BulidOntologiesRPD-ER2002.pdf</a> >. Acessado em: 01/jul./2007.	Adotando a definição de que ontologia é uma conceitualização explícita de um domínio de discurso, proporcionando um conhecimento comum e compartilhado do domínio, o artigo apresenta a idéia de que métodos de análise de domínio podem ser usados para a construção de ontologias. A análise do domínio visa os modelos genéricos que representam grupos de sistemas similares dentro de um domínio da aplicação. Segundo os métodos de análise de domínio, a categorização de objetos e operações comuns proporciona maior clareza quanto as definições das ambigüidades e as definições de relações.	Termo faz referência a um conceito. Representa um conceito ou classe.	Não trata.	Criar uma estrutura lógica, uma filosofia, uma classificação, desenvolver um conhecimento (entendimento) comum em uma disciplina. Determinar a extensão e a complexidade dos processos. Fornecer um entendimento básico de um domínio, o que requer menos esforço do que sustentar argumentos lógicos e provas de um domínio.	Afirma que uma terminologia é primeiro desenvolvida fornecendo um vocabulário controlado para o assunto da área ou domínio de interesse, então ela é organizada em uma taxonomia onde conceitos-chave são identificados, definidos e relacionados para criar uma ontologia.
	<b>X</b>	SOWA, John F. Building, sharing, and merging ontologies. 2006. Disponível em	Este artigo consiste em um extrato de artigos publicados anteriormente por John F. Sowa. Está	Não trata.	Não trata.	Possibilitar, juntamente com aplicações lógicas, a construção de modelos computacionais para um	Faz distinção entre ontologia terminológica e ontologia formal. A primeira, também chamada de grande ontologia,

<<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>  
>. Acessado em: 29/mai./2007

atualizado com novos materiais referentes a projetos atuais sobre ontologias e suas implicações para bases de dados, bases de conhecimento e web semântica

determinado domínio de aplicação. Tais modelos computacionais visam ao cumprimento de tarefas de representação do conhecimento de um domínio.  
Servir como um catálogo dos tipos de coisas que existem em um domínio do interesse D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D.  
Representar os predicados, os sentidos das palavras, ou os tipos de conceito e relação da língua L quando usados para discutir tópicos no domínio D.

consiste em uma hierarquia de formas e significados de palavras, sem muitos detalhes de relacionamento, pois as distinções feitas pelos axiomas e definições não são muito específicos. As ontologias formais apresentam maior detalhes em seus axiomas e definições, sustentando maior complexidade em suas inferências. As ontologias formais são menos utilizadas que as ontologia terminológicas, principalmente no que diz respeito ao processamento de linguagem natural.  
O texto se detém mais aos aspectos de arquitetura e combinação entre ontologias. A combinação e compartilhamento de ontologias é ponto forte nestes artefatos computacionais que operam como modelos de representação do conhecimento.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	JACOB, Elin K. Ontologies and the semantic web. Bulletin of the American Society for Information Science and Technology. v. 29, n. 4, abr./mai. 2003. p. 19-22.	Adota a idéia de que uma ontologia é um entendimento semântico (a estrutura conceitual do conhecimento) compartilhado entre indivíduos de um mesmo domínio. Entende que uma ontologia pode ser uma estrutura conceitual informal com tipos de conceitos e seus relacionamentos nomeados e definidos em linguagem natural, ou, pode ser uma estrutura semântica formal de um dado domínio com tipos de conceitos e seus relacionamentos sistematicamente definidos em uma linguagem lógica e organizado hierarquicamente. Foca atenção nas ontologias inseridas no ambiente web, onde segundo a autora, "uma ontologia não é simplesmente uma estrutura conceitual, mas sim uma estrutura sintática que modela a semântica de um domínio - a estrutura conceitual - em linguagem inteligível por máquinas". Detalha ainda as funções dos recursos RDF e RDFS e aponta as características de ontologias que utilizam de recursos RDFS.	Não trata.	Todo elemento de uma ontologia é ou um tipo de classe ou um tipo de propriedade. Relacionamento super-ordenado entre classes Relacionamento sub-ordenado entre classes e propriedades. O relacionamento entre classes ou propriedades são poli-hierárquico, pois uma classe pode ser uma sub-classe de uma, duas, três ou mais classes super ordenadas.	No ambiente da web semântica uma ontologia serve para fornecer uma conceitualização parcial de um determinado domínio, de forma compartilhada entre usuários comuns e que é formalmente definida por uma linguagem processada por máquinas. Tem como objetivo também o compartilhamento de informações semânticas por meio de sistemas automáticos. E, oferecer meios concisos e sistemáticos para definir a semântica dos recursos web.	<p>Lança mão da definição de Sowa, que diz que ontologia é a investigação das coisas que existem ou podem existir em um dado domínio, e produz um catálogo detalhado dos tipos de coisas e relacionamentos entre elas, que são relevantes para o domínio. Faz uma breve crítica ao uso não contextualizado da definição de Gruber, o que a torna incompleta. A ontologia especifica conceitos relevantes de um domínio, bem como as propriedades desses conceitos e possíveis relacionamentos entre conceitos e propriedades. Isso faz com que as ontologias concebam suportes para processar recursos baseados em interpretação de significados de grandes conteúdos e características sintáticas como ordem sequencial por exemplo.</p> <p>Afirma que qualquer esquema de metadados é uma ontologia especificando um conjunto de características físicas ou conceituais de recursos considerados importantes para um domínio. Por exemplo o Dublin Core [define um conjunto de atributos (tipos de elementos fundamentais) e sub-tipos de elementos qualificados]. Uma ontologia é um sistema representacional único que integra dentro de uma única estrutura características de</p>

abordagens tradicionais como de classificações hierárquicas,

tesauros facetados e vocabulários controlados. Uma ontologia concebe bases semânticas para esquemas de metadados, facilita a comunicação entre sistemas e agentes e, determina uma padronização conceitual para comunidades de usuários.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	<b>X</b>	KENT, Robert E. The IFF foundation for ontological knowledge organization. Knowledge Organization and Classification in International Information Retrieval. The Haworth Press, Inc., 2003, p. 187-203	Discute uma abordagem axiomática para a integração semântica de ontologias. Essa abordagem é representada na Information Flow Framework (IFF), uma infraestrutura para organização de informação de bibliotecas digitais que operam com bases de dados e ontologias. Argumenta que a integração semântica de ontologias ocorre por meio de dois tipos de processo: alinhamento e unificação. O primeiro consiste no compartilhamento terminológico e semântico comum por meio de mediação ontológica, e o segundo está focado em uma ontologia virtual de conexões comuns. Esta é a fusão do processo de alinhamento de ontologias de comunidades participantes. Adota as linguagens ontológicas segundo o conceito de um	Com base nas distinções de Aristóteles que o autor usa analogicamente para tratar dos elementos da IFF, um termo pode ser: um tipo (universal, geral) uma instância (particular, específico) uma relação (qualidade) uma entidade (substância).	Não trata.	Não trata.	As expressões em uma ontologia usam uma linguagem contida de predicacões relevantes (entidades e tipos de relações). Faz menção à classificação facetada de Ranganathan. Uma ontologia é como um espectro formal. Explica dois tipos de ontologias: a) terminológica (taxonomia) - situada no extremo informal do espectro: axiomatiza a hierarquia de categorias utilizando as meta-relações de subtipos e disjunções, geralmente usado para processamento de linguagem natural. b) axiomatizada - situada no extremo formal do espectro: usada para modelos de conhecimento de senso-comum e raciocínio. Lança mão das categorias de Aristóteles, e principalmente das teorias de predicacão (essencial: qualquer coisa que é dito de um assunto; e acidental: qualquer coisa que

tipo de linguagem IFF

existe/é em um assunto) para explicar a arquitetura semântica IFF. O modelo IFF representa "semântica interpretativa". As distinções feita por Aristóteles entre 'universal' versus 'particular' e 'qualidade' versus 'substancia' são usadas para a IFF ao distinguir 'tipo versus instância' e 'relação versus entidade'.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	PULIDO, J.R.G. et al. Ontology languages for the semantic web: a never completely update review. Knowledge-Based Systems, n. 19, 2006. p. 489-497.	Apresenta uma revisão das abordagens utilizadas pelas comunidades de pesquisas em representação do conhecimento. Destacando a importância de uma abordagem dividida em camadas (estratos, níveis) e do uso de padrões, começa mostrando os esforços realizados inicialmente pelos pesquisadores da inteligência artificial. Descreve abordagens recentes direcionadas à web semântica. Apresenta exemplos de abordagens encontrados na literatura. Expõe uma previsão para o processo de criação de ontologias na web semântica.	Não trata.	Não trata.	A construção de linguagens de marcação, imprescindíveis para a realização da web semântica, estão sendo desenvolvidas com base em uma abordagem arranjada em camadas (níveis), sobretudo no nível do vocabulário ontológico. A ontologia, ou, o vocabulário resultante da ontologia, é o nível responsável pelas bases de raciocínio e inferência da web semântica. A finalidade das ontologias na esfera da web semântica é proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação.	Descreve linguagens de representação do conhecimento na web: abordagens anteriores: Knowledge Interchange Format (KIF); F-Logic; Dublin Core; CYC. Abordagens recentes: XML; RDF; Knowledge Annotation (KA); Simple HTML Ontology Extensions; OIL; DAML; OWL.
	X	WEBER, Ron. Conceptual modelling and ontology: possibilities and pitfalls.	Teorias ontológicas estão sendo desenvolvidas para a representação de estruturas e	Não trata.	Não trata.	Teorias ontológicas pretendem conceber uma representação útil da	Para a prática do modelamento conceitual é de suma importância a distinção entre

Journal of Database Management, v. 14, n. 3, jul-set/ 2003. p. 1-20.

comportamento do mundo real. O autor discute como essas teorias podem ser usadas para subsidiar a pesquisa prática e pedagógica de modelamento conceitual. Modelamento conceitual é uma atividade realizada durante o desenvolvimento de um sistema de informação para construir uma representação semântica de um domínio real do mundo. Apresenta um exemplo de como uma teoria ontológica específica pode melhorar o entendimento seguro da prática e da gramática do modelamento conceitual. Discute também algumas possibilidades e dificuldades para se associar esse uso de teorias ontológicas para o modelamento conceitual.

estrutura geral e dinâmica do mundo real como ele existe/é. Proporciona uma lente que nos permite enxergar o mundo real e impor a ordem dos fenômenos, concebendo um consenso desses fenômenos. Por tal motivo as teorias ontológicas são úteis para os sistemas de informação (agindo na atividade de modelamento conceitual), já que esses têm como objetivo representar um sistema do mundo real. Ontologias concebem para os sistemas de informação uma representação do mundo real em dado domínio.

coisa e propriedade de coisa, para evitar interferências na comunicação de um sistema de informação, como receber e enviar mensagens sem distinções claras do que é coisa ou propriedade de coisa, ex.: casamento é uma coisa ou uma propriedade de coisa (propriedade de pessoa)? Coisas (no sentido de entes) têm propriedades. Uma propriedade sempre deve estar ligada a uma coisa, e jamais a uma outra propriedade. Propriedades adicionais (atributos e relacionamentos/associações) não devem ser usados para modelamento conceitual, apenas se recomenda o uso de propriedades obrigatórias. Pauta-se na teoria ontológica de Mario Bunge, a qual está estruturada em fundamentos de coisas, propriedades de coisas, leis, estados, eventos, combinações, sistemas e nível de estrutura, para sustentar a reflexão e a avaliação de gramáticas e práticas de modelamento conceitual.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	LAME, Guiraude. Using NLP techniques to identify legal ontology components: concepts and relations. Artificial Intelligence and Law, n. 12, 2004. p. 379-396.	Apresenta um método para identificar componentes de ontologias. Tal método faz uso de técnicas de processamento de linguagem natural para extrair conceitos e relações entre os conceitos. Foca na construção de ontologias para recuperação da	O termo expressa o conceito. O termo é a etiqueta do conceito. Termos legais (jurídicos) são etiquetas de conceitos legais específicos. Ex.: contrato, responsabilidade. Ou também etiquetas de objetos do mundo apreendidos por Lei, ex.:	Conceitos são expressados (personificados) por meio de termos que ligados semanticamente entre conceitos são personificados em relações sintáticas entre estes termos. Os conceitos são definidos um a um. As definições dos conceitos podem ser mais ou menos	Recuperação da informação e raciocínio automático de um dado domínio.	Ontologias são úteis tanto na recuperação da informação quanto em sistemas de consulta interativa. O autor busca na análise sintática, por meio de métodos estatísticos, as relações semânticas em textos jurídicos.

informação no campo jurídico, sugerindo que tal método se estende à aplicação em qualquer área

passageiro, armas e drogas. Observando a relação sintática entre termos é possível identificar a relação semântica entre eles. Afirma que, estatisticamente, termos relacionados semanticamente ocorrem com muita frequência com a mesma apresentação sintática.

explícitas. Ex.: a definição do registro de nascimento das pessoas nascidas no exterior é explícito, pois está sub-entendido que condições um registro de nascimento é estabelecido e que elementos os faz ser de pessoas estrangeiras. A definição do conceito de divórcio é menos explícita, pois se refere à ruptura do casamento, que pode ser dada pelo falecimento de um dos envolvidos, ou, pelo pedido legal de separação. Mas, uma ontologia pode definir o divórcio apenas como a ruptura do casamento. A relação semântica entre conceitos se dá por meio das formas sintáticas em um discurso. Em uma ontologia conceitos são definidos pela semântica das relações estabelecidas entre cada conceito e entre outros conceitos.

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	GILCHRIST, Alan. Thesauri, taxonomies, ontologies: an etymological note. Journal of Documentation, v. 59, n. 1, 2003. p. 7-18.	Apresenta como os termos tesouro, taxonomia e ontologia vêm sendo utilizados nas áreas de ciência da informação, prática de inteligência artificial e trabalhos que focam a web semântica. Não se trata de uma revisão de técnicas, mas sim uma contribuição que pretende minimizar a confusão existente no emprego dos	Tesouro: termo é visto como descritor, palavra-chave. Tipos de relacionamentos: hierárquico, associativo e equivalência. Ontologia: não trata.	Tesouro: não trata. Ontologia: conceitos organizados em taxonomias.	Tesouro: auxiliar a indexação e a recuperação da informação. Ontologias: descrever um entendimento comum e compartilhado de um determinado domínio.	Há também os chamados 'tesouros de busca', que são compostos de combinações de listas de tesouro e focam a atenção na recuperação em meio informático. No âmbito da inteligência artificial pesquisadores desenvolveram a chamada 'estrutura semântica', que é uma derivação do tesouro. Porém, as estruturas semânticas apresentam os termos conforme o contexto e

respectivos termos.

seu relacionamento semântico, e, usufruem de aparatos gráficos e espaços multidimensionais, ou seja, sua navegação é mais elaborada que a dos tesouros. As estruturas semânticas são muito utilizadas em portais de informação de empresas.

Ontologia: Adota a definição de que ontologia é a 'especificação explícita e formal de uma conceitualização compartilhada', de Gruber (1993). É adepto à idéia de que uma ontologia descreve assunto de um determinado domínio usando noções de conceitos instâncias, relações, funções e axiomas.

São utilizadas em tradução automática, modelamento de sítios de empresas, reuso de conhecimento e recuperação da informação. São mais ambiciosamente utilizadas na gestão do conhecimento e na web semântica.

Pauta-se também nas definições de Berners-Lee que afirmam ser ontologias " um documento que formalmente define as relações entre termos. O tipo mais comum de ontologia para a web é composto por uma taxonomia e por um conjunto de regras de inferência".

TES.	ONT.	REFERÊNCIA	RESUMO	TERMO	CONCEITO	OBJETIVOS	OBSERVAÇÕES
	X	DING, Y.; FOO, S. Ontology research and development, Part 2 : a review of ontology mapping and evolving. Journal of	Este artigo é a segunda parte da revisão de pesquisas e desenvolvimentos de ontologias. Esta parte está dedicada ao mapeamento e	Não trata.	Não trata.	Fornecer um entendimento comum e compartilhado de um domínio que pode ser comunicado entre pessoas e sistemas de aplicação.	Adota a definição de que ontologia é 'especificação explícita e formal de uma conceitualização compartilhada' de Gruber

Information Science, v. 28, n.5, 2002, p. 375-388. Disponível em <<http://wotan.liu.edu/doi/data/Articles/julbaefggy:2002:v:28:i:5:p:375-388.html>>. Acessado em:

manutenção de ontologias. Pauta-se na definição de que ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. Ela fornece um entendimento comum e compartilhado de um domínio que pode ser comunicado entre pessoas e sistemas de aplicação. Ontologias têm grande potencial no que tange captura e trocas de significados e relações, fato que permite dizer que ontologias não são modelos estáticos. Estudar o mapeamento e a manutenção de ontologias é uma tarefa essencial na compreensão de

Capturar significados e relações entre os

(1993). Ontologias são descrições de conceitos estruturados semanticamente, representados como regras de interpretação. O texto não aborda implicações referentes aos termos e aos conceitos, se detém a explicar aspectos sobre o mapeamento e a manutenção de ontologias por meio da apresentação de projetos deste assunto. Termo em inglês para manutenção de ontologias: "ontology evolving". A parte da manutenção de ontologias, a exemplo da parte do mapeamento, é explicada por meio da apresentação de projetos desenvolvidos e por meio da explicação estrutural de ontologias.

## **Apêndice B**

# RELATÓRIO 2

## TESAURO ONTOLOGIA TERMO

X

Relação entre os próprios termos: sinonímia, hierárquica e por associação. Porém, reclama pelo fato de que as relações deveriam ser estabelecidas pelos usuários no momento da consulta, estabelecendo a longitude e os números de campos dos termos. Não fica claro a definição de termo e nem sua relação com o conceito, embora fique implícita a idéia de que o termo é o representante lingüístico do conceito.

X

Não trata.

X

Termo é palavra ou frase de algum domínio de conhecimento. Preferencialmente substantivos e frases de substantivos.  
Relacionamento entre termos: equivalência, hierárquico e associativo.

X

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. A 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

Conceitos são representações abstratas percebidas e interpretadas de algo real, e ordenadas em classes já existentes no âmbito cognitivo de quem os apreendem. Todo conhecimento apreendido é classificado de acordo com classes (agrupamento de elementos por semelhança) cognitivamente pré-estabelecidas. E é a ordenação dos conceitos que rege a gestão da informação que, por sua vez, rege a gestão do conhecimento.  
Não fica clara a relação existente entre conceito e termo, embora fique implícita a idéia de que o termo é o representante lingüístico do conceito.

Não trata.

Não trata

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

## OBJETIVOS

Orientar qual o termo mais adequado para representar um conceito.  
Servir de ponte entre necessidades informacionais e sistemas de recuperação de informação.  
Deve ser um instrumento que, embora controlado, seja um recurso coordenável (no que tange coordenação de vocabulário, relação entre termos e apresentação) pelo usuário. Fato esse que facilitaria a inter-relação linguagem natural e linguagem artificial.

Tesouro WEB tem por objetivo potencializar a recuperação da informação na Internet atuando como uma interface entre informação e usuário.

Não trata

Serve para auxiliar a indexação e a recuperação em um sistema de recuperação de informação.

TESAURO ONTOLOGIA	TERMO	CONCEITO	OBJETIVOS
X	<p>Termos de indexação dos tesouros: a) descritores ou termos preferidos; b) identificadores (nomes próprios), c) não descritores ou termos não preferidos (que são os termos não autorizados para indexação). Existem a) termos simples, ex.: biblioteca; b) termos compostos, ex.: (frase adjetivada) Organizações Culturais Internacionais, (frase preposicional) Dinâmica em grupo; c) termo qualificado, ex.: Rádio (metal), Terra (planeta). Relações: Equivalência (USE, UP), Hierárquica (gênero/espécie, todo/parte, enumerativo, poliherárquico), Associativa (termo relacionado, TR)</p>	Não trata.	Não Trata.
X	Não trata.	Não trata.	<p>Auxiliar o processo de recuperação da informação, devido a sua uniformidade terminológica que possibilita esclarecer barreiras lingüísticas. Auxiliar a consulta do usuário, expondo sistematicamente uma classificação que o direcione no momento da busca.</p>
X	<p>Relacionamento entre termos: Equivalência (uso/usado para) - indica o termo preferido. Hierarquia - indica um termo mais geral e mais específico. Associação - indica outros tipos de relacionamentos entre termos.</p>	Não trata.	Não trata.
X	<p>Designa um objeto do pensamento (conceitos e noções). Expressa claramente um objeto de pensamento de uma determinada língua. Esses termos únicos em um sistema também são chamados de Descritores. Os descritores são termos preferidos. Após a escolha de descritores deve-se ater aos termos equivalentes ou não-equivalentes. Na proposta do autor, os termos, ou sequência de termos, considerados de igual significados e não preferidos, devem ser Expressões de Acesso Adicional</p>	<p>Objeto do pensamento. É designado por um termo e este termo deve ser único no sistema complexo de um tesouro. Relação Classe I: relação termos-conceitos de uma mesma língua (sinônimos e polissemias). Relação Classe II: relação conceitos-conceitos independente da língua utilizada, chamado pelo autor de Apropriadas Relações de Conceitos (Concept Relations Proper). Na classe II há a necessidade de uma linguagem intermediária para dar conta dos relacionamentos entre os conceitos de diversas</p>	<p>Sua proposta tem por objetivo: alta força de articulação no nível descritivo quanto à manutenção do controle metodológico; síntese do conjunto conceitual e consistência na análise das partes constituintes e; definição de uma plataforma de referência comum para permitir construção e manutenção distribuída de tesouros lingüísticos, bem como o intercâmbio dos dados do tesouro.</p>

## TESAURO ONTOLOGIA TERMO

## CONCEITO

## OBJETIVOS

línguas. Uma cadeia de descritores equivalentes (ECD), constituída de termos de diferentes línguas, tem seus elementos identificados por meio de um número identificador de meta-linguagem (MLIN). Nesta classe II é possível treze tipos de relacionamentos: a) abstrato/genérico, b) partitivo, c) todo/parte jurídico, d) partitivo geográfico, e) descendência, f) instrumental, g) causa/efeito, h) benefício, i) prejuízo, j) matéria, l) aparência, m) processo e, n) estado.

a) É do tipo gênero e espécie. Tanto o genérico (conceito A) quanto o específico (conceito B) são concebidos como classes características. b) Relação entre partes de um todo material ou imaterial. c) Relação entre partes de um todo (o todo sendo uma Jurisprudência ou uma Lei, por exemplo. d) Relação entre partes de um todo geográfico. e) Relação em que o conceito B descende do conceito A. f) O conceito B é um instrumental para alcançar um resultado, conceito A. g) O conceito B é a causa da ocorrência, ou existência, do conceito A. h) O conceito B é útil para o conceito A. i) Relacionamento com base em causas e efeitos adversos (por extensão pode-se dizer que se trata de uma relação de adversidade). j) O conceito A consiste do conceito B quando visto sob aspecto de matéria ou uso material. l) Enxerga o conceito A como ele é reconhecível pela visão humana: forma, configuração ou cor. m) Conceito B indica um processo complexo (processado) no conceito de A. n) O conceito A pode ser visto sob o aspecto de seu estado (B).

X

Pautando-se na TGT afirma que uma das finalidades dos termos é evitar a superposição de conceitos e buscar a diminuição da flexibilidade da linguagem. A terminologia não estuda o termo levando em consideração aspectos psicológicos ou filosóficos, mas sim sua expressão por meio da linguagem. O ideal é que haja um termo para cada conceito. O termo deve se situar na estrutura de um dado conhecimento. Os termos integram uma estrutura agrupada por princípios cognitivos. Termos são unidades

Pauta-se nas afirmações de que o conceito é construído por meio de um processo de agrupação e ordenação dentro de categorias abstradas dos entes percebíveis, e se relaciona com outros conceitos por meio das proximidades de suas características. Relação com o termo: o termo representa um conceito. Relação com o próprio conceito: proximidade das características.

Possibilitar uma recuperação pertinente da informação, de modo a economizar o tempo do usuário, minimizando ruídos, como ambigüidades, no momento da representação do conteúdo documental.

**TESAURO ONTOLOGIA TERMO**

lingüísticas e unidades de conhecimento especializado. Termo é uma unidade lingüística cuja função principal é descrever um conceito de maneira unívoca em um sistema de informação. Deve-se levar em conta o 'contexto' que envolve o termo. O contexto é tido como toda realidade (verbal, física, histórica, cultural e ocasional) que rodeia o signo. É o contexto (o campo de conhecimento, neste caso) que irá outorgar o conceito de determinado termo. O descritor, que é um termo controlado, representa um conceito. O termo atua na comunicação especializada, o descritor atua na indexação e recuperação da informação. O termo normalizado é uma fonte para validar o descritor. Relação com o conceito: designação. Relação com o próprio termo: os termos se ordenam de acordo com seu nível de especificidade.

**CONCEITO**

O termo serve para expressar um conceito. Em um tesauro, a clareza do conceito está implícita no termo. A combinação de termos simples (termo compostos) aumenta a precisão do conceito expresso. A autora admite que existem termos polissêmicos, homógrafos. [Tal fato leva a crer que sua posição é de que termo e conceito estão separados (no sentido de serem elementos que não pertencem um ao outro), porém relacionados com base em seu contexto e uso]. Os termos são formados basicamente por substantivos e, quando necessário, por adjetivos qualificadores. São empregados também, porém com menos frequência, adjetivos substantivos (particípio), advérbios, preposições e artigos para a composição de descritores compostos. Considera que termos formam descritores. [Vários termos podem formar um único descritor]. Relação com o conceito: designação Relação com o próprio termo: equivalência (sinonímia), hierárquica (gênero/espécie), associativa (termos relacionados horizontalmente, notas explicativas).

**OBJETIVOS**

A rede de conceitos rege a estrutura de um tesauro. Ora trata como significado de palavra ora como conceito de termo. Não apresenta uma definição para conceito, porém deixa implícito que se trata de uma abstração de um 'ente' real.

Serve para dar maior consistência à indexação e potencializar o acesso do usuário (recuperação da informação). A autora (NAUMIS-PEÑA, 2001, p. 8) enumera os seguintes objetivos:

- 1) Servir de vocabulário oficial para coordenar dois processos: indexação e recuperação de documentos digitais de um sistema;
- 2) Propôr um sistema de símbolos lingüísticos para agrupar informação similar relacionada ou direcionar para grupos mais gerais ou mais específicos de um temática;
- 3) Alcançar uma normalização da terminologia do sistema de informação a ser utilizado;
- 4) Propôr um conjunto estruturado de termos sob a base de um sistema de conceitos aptos para organizar os conteúdos desse sistema.

**X**

**TESAURO ONTOLOGIA TERMO****X**

Sugere uma aproximação entre descritores (unidade de representação e recuperação de informação) e termos, ou seja, enxergar o descritor como uma unidade terminológica. O termo é a denominação de um conceito, serve para representar o conceito. "Termo é a unidade lexical básica da Terminologia que expressa um conceito, um objeto ou um processo. É uma unidade de designação de elementos do universo percebido ou concebido [...] termo é a designação de um conceito por uma expressão lingüística em uma linguagem de especialidade"(LAAN, 2002, p.63).

A unidade lexical assume papel de termo dentro de um contexto, dentro de uma situação comunicacional real. Os termos são unidades lexicais que se situam nos discursos especializados, por isso apresentam sinonímia e variação.

Função e relação com o conceito: a definição é um tanto quanto obscura, pois ora menciona termo como um signo lingüístico [possuidor de significante (denominação) e significação (conceito)] ora menciona o termo como uma unidade lexical que designa um conceito, e ora como designação de conceito, objeto ou processo. O que fica evidente é a definição de que o termo "é uma unidade lexical terminológica que se encontra em funcionamento nos discursos especializados" (LAAN, 2002, p.48).

Relação com o próprio termo: conforme a relação entre os conceitos descritas a seguir.

**X**

Remete a tutoriais de elaboração de tesauros que contemplam os seguintes tipos de relacionamentos: termos gerais (TG), específicos (TE), relacionados (TR), termos preferidos (UP) e não-preferidos (USE). Além de notas explicativas (NE).

**X**

Relação entre termos se dá por meio das características de divisão (ou qualidades essenciais, exclusivas e homogêneas) que os tornam singulares e pertencentes a uma única faceta.

**CONCEITO**

Conceitos são representações mentais, são únicos, não podem ser compostos, o que ocorre são diferentes traços semânticos que integram um determinado conceito.

Relação com o próprio conceito: lógica e ontológica.

Relação lógica: baseada no grau de semelhança entre os objetos e conceitos. Este tipo de relação pode ocorrer por meio de uma lógica de subordinação (é partir do conceito mais geral para o mais específico, ou seja, do conceito de maior extensão e menor intensão para o conceito de menor extensão e maior extensão) ou por uma lógica de superordenação (o oposto da subordinação, do conceito mais específico para o mais geral). Esta relação pode ser de gênero/espécie ou todo/parte.

Relação ontológica: não são baseadas na semelhança entre os conceitos, mas sim na proximidade situacional dos elementos em uma realidade. Este tipo de relacionamento, que baseado na contigüidade dos conceitos no espaço, pode ser de coordenação, ou encadeamento, como causa e efeito ou seqüenciais.

Relação de equivalência: ocorre quando um conceito é representado por mais de uma denominação.

Relação com o termo: o termo é a designação de um conceito. É único. Para cada termo um conceito.

**OBJETIVOS**

Auxiliar o processo de representação e recuperação da informação em um SRI.

Controlar as ambigüidades próprias do léxico comum. Facilitar a comunicação entre SRI e usuário.

Não trata.

Não trata.

**TESAURO ONTOLOGIA TERMO****X**

Se refere a termos de texto-livre (se comparado aos termos (descritores) dos vocabulários controlados). Segundo as autoras:

- a) termos especificam forma e significado.
- b) termos expressam significados.
- c) termos descrevem conceitos.

Podem ser empregados para minimizar ambigüidades, já que um termo de texto-livre define um conjunto relevante de termos registrados em uma base de dados. Os aspectos relativos ao 'termo', que aqui são apresentados, se refere aos termos de texto-livre atuando em ambientes de recuperação de documentos em bases de dados automáticas. Os termos de texto-livre são aqueles colocados nos campos de assunto no momento da busca.

**X**

O relacionamento associativo dos termos depende de como o ambiente informacional, em que o tesouro está operando, lida com a utilização dos conceitos. O modo como o domínio relata e relaciona os termos é mais relevante que as próprias diretrizes normatizadoras.

A forma do termo é controlada (em um tesouro) por uma gramática formal que determina a ortografia, a forma singular ou plural, abreviações e formas compostas. Acordos e restrições são utilizadas pela adição de notas de escopo e, no caso de homógrafos, são empregadas definições e frases qualificativas.

Tipos de relacionamentos: equivalente, hierárquico e associativo.

Percebe-se uma indefinição do texto, pois ora parece evidenciar que o termo é constituído por forma e conceito ora afirma que o termo apenas expressa um conceito.

**CONCEITO**

Não trata.

O significado do conceito é controlado (em um tesouro) e sinônimos são selecionados para expressar conceitos.

O uso dado aos conceitos em um determinado domínio, e os fatores característicos do domínio, determinam mais eficazmente o relacionamento conceitual de um tesouro do que os padrões internacionais de elaboração.

**OBJETIVOS**

Não trata.

Guiar o processo de indexação e recuperação controlada da informação, bem como, o processo de indexação baseado em linguagem natural.

Além de seu uso tradicional voltado à indexação e à recuperação, os tesouros estão sendo empregados como ferramentas para gerir e comunicar a linguagem específica usada nas organizações.

Tesouros do tipo 'taxonomia corporativa' tem como objetivos: a) assegurar e sustentar a recuperação da informação; b) auxiliar no entendimento e uso de um assunto ou linguagem específica e; c) facilitar o entendimento conceitual e comunicacional em um nível lingüístico que conceba um vocabulário corporativo.

**TESAURO ONTOLOGIA TERMO****X**

O termo (terminum) é um elemento que sintetiza o conceito como um todo, possibilitando sua comunicação verbal. O termo denota um referente. O tratamento terminológico de um termo está no fato de vê-lo como um representante de um referente. O termo é o Definiendum (as características do conceito ou os atributos predicáveis do objeto/referente são o Definiens). O definiendum (termo) + o definiens (predicados do objeto) formam a definição do conceito dentro de um sistema, e essa definição serve para fixar um conceito em um sistema de conceitos. Função: sintetizar o conceito e viabilizar sua comunicação.  
 Relação com o conceito: elemento que designa um conceito  
 Relação com o próprio termo: depende das características do conceito (predicados do referente).

**X**

Tipos de relacionamento: hierárquico (gênero-espécie e todo-parte), associativo (ou não hierárquico), equivalência (considera os não descritores, embora determine o termo preferido, 'descriptor'. Uso de tags USE, UF, BT, NT e RT para remissivas).

**CONCEITO**

Representação mental das relações que ocorrem entre objetos na realidade empírica. A definição do conceito é o conjunto formado pelas características próprias de um objeto e por um termo que as sintetizam.

No nível conceitual as relações entre conceitos se dão de maneira lógica (resultado da própria compreensão dos conceitos) e ontológica (relação entre conceito e realidade).

Relação com o termo: um termo + os predicados do referente formam a definição do conceito.

Relação com o próprio conceito: Tipos de relacionamento: a) lógico - quando existem características comuns na definição dos conceitos, podendo ser gênero/espécie, analítico e de oposição. b) ontológico - são relações indiretas, pois resultam das propriedades do objeto empírico, podendo ser partitivo, de sucessão e material/produto. c) de efeito - associação indicando o efeito, podendo ser de causalidade, instrumental, descendência, genealógica, estágios de substâncias, antogênicas.

Não trata.

**OBJETIVOS**

Organizar e recuperar informação.

Auxiliar a recuperação da informação nos processos de indexação (indexadores) e busca (usuários). Em um primeiro momento, auxiliar a recuperação da informação controlando sinônimos e homógrafos e apresentando relações hierárquicas e associativas entre termos. Em um segundo momento, acrescenta-se a relação de equivalência, incluindo termos sinônimos e quase sinônimos, termos preferidos ou descritores (considerando a existência dos não descritores).

**TESAURO ONTOLOGIA TERMO****X**

Os significados dos termos são as definições. Não faz distinção entre termo e conceito, ora fala em conjunto de termos ora em conjunto de conceitos relacionados de acordo com seus significados. Afirma que o que existe é um vocabulário formado por termos (conceitos) de um determinado domínio. Termos são etiquetas de conceitos.

**X**

O artigo não foca aspectos relativos ao termo. Afirma que a relação entre os termos é feita pelos especialistas, o que possibilita que o usuário formule consultas usando conceitos especificados. Para os autores a ontologia define regras para a combinação e relação entre os termos, define o conjunto de termos (linguagem) que o usuário utilizará para formular a busca. É possível dizer com isso que, sob a ótica do artigo, o conceito está contido no termo especificado pelos especialistas. Função: viabilizar a formulação de consultas (por parte do usuário) fazendo uso de conceitos predeterminados pelos especialistas. Relação com o conceito: o termo contém conceito. Relação com o próprio termo: consenso de um determinado domínio.

**CONCEITO**

Conceitos são: entidades, atributos e processos. Conceitualização é o inter-relacionamento entre os significados dos conceitos. Afirma que ontologias descrevem conceitos e tesouros descrevem termos.

O artigo não é enfático quanto à definição do conceito. Simpatizando com uma definição de conceitualização sob a ótica intensional, em que afirma ser a conceitualização uma lista de características do conceito, e uma visão abstrata e sintetizada de um ente real, fica implícito uma proximidade com a Teoria do Conceito no que se refere a afirmação de que 'conceito é a soma dos enunciados verdadeiros referentes a um objeto real'. Relação com o termo: o conceito é parte do termo. Relação com o próprio conceito: a intensão do conceito, ou seja, o conjunto de características do conceito, é responsável pelo relacionamento entre eles.

**OBJETIVOS**

Reduzir ou eliminar confusões terminológicas e conceituais. Identificar e definir corretamente um conjunto de conceitos relevantes que caracterize um dado domínio de aplicação. Visa uma estrutura conceitual uniforme. Comunicação e cooperação entre pessoas. Melhor organização das empresas. Interoperabilidade entre sistemas. Benefícios na engenharia de sistemas (reusabilidade, confiabilidade e especificação). Uma ontologia pode ser vista como um estágio preliminar de uma base de conhecimento, cujo objetivo é a descrição de conceitos necessários para falar sobre determinado assunto. A base de conhecimento por sua vez consiste no conhecimento necessário para idealizar um problema, derivado de um novo conhecimento, com o fim de responder questões de um determinado domínio.

Possibilitar a consulta a um sistema automatizado de informação (recuperação da informação) utilizando termos conceitualizados e especificados por especialistas

**TESAURO ONTOLOGIA TERMO**

**X** As ontologias possuem uma peça chamada componente terminológico, que define os termos e a estrutura da ontologia. O componente assertivo apresenta exemplos e indivíduos que manifestam a definição terminológica. Os termos são usados para definir formalmente 'coisas' em um domínio de interesse.

**X** Não trata.

**CONCEITO**

Há uma decisão ontológica específica que determina a divisória entre o tratamento de uma coisa como um conceito e o tratamento de uma coisa como indivíduo. Para projetar a estrutura conceitual é necessário identificar os conceitos concretos principais e suas propriedades do domínio, assim como seus relacionamentos, e criar conceitos abstratos como características.

Na construção de ontologias as estruturas de esqueleto de conceitos do núcleo são estendidas com conceitos mais refinados e mais periféricos, e estes são interligados elaborando relações adicionais.

**OBJETIVOS**

Pode atuar nas seguintes situações: na WEB-semântica; na medicina para controlar a saúde do paciente por meio da criação de guias medicinais, e, traçando os genomas das plantas e dos animais; procurando por recursos públicos específicos da informação; projeto colaborativo da engenharia; análise de segurança in-depth; e a troca automatizada da informação eletrônica entre sócios comerciais. Ajudam a responder perguntas em um corpo de informação.

As ontologias estão sendo usadas para guiar a extração do índice semântico das coleções dos originais (textos) que descrevem pesquisas médicas, produtos de consumidores, tópicos de negócio, e ambientes governamentais. Buscadores Web como Google e Yahoo também fazem uso de abordagens baseadas em ontologias para organizar conteúdos na Web.

## TESAURO ONTOLOGIA TERMO

**X**

Não trata.

## CONCEITO

Para a IA uma ontologia refere-se a um artefato da engenharia computacional, constituído por um vocabulário específico usado para descrever alguma realidade, mais um conjunto de suposições explícitas a respeito do significado pretendido das palavras do vocabulário. Este conjunto de suposições tem geralmente a forma de uma teoria lógica de primeira ordem, onde as palavras do vocabulário aparecem como nomes unários ou binários do predicado, respectivamente chamados de conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia dos conceitos relacionados por submissão; em casos mais sofisticados, os axiomas apropriados são adicionados a fim de expressar outros relacionamentos entre conceitos e confinar sua interpretação pretendida.

Conceitualização: relação matemática ordinária (extensional), relações conceituais (intensional). Uma maneira padrão de representar relações conceptuais é vê-las como funções em grupos de mundos possíveis. Quando as relações ordinárias forem definidas em um determinado domínio, as relações conceituais serão definidas em um espaço do domínio. Define-se um espaço do domínio como uma estrutura  $\langle D, W \rangle$ , onde  $D$  é um domínio e  $W$  é um conjunto de estados de relações de tal domínio (chamado também mundos possíveis). Por exemplo,  $D$  pode ser um conjunto de blocos em uma tabela e  $W$  pode ser o conjunto de todos os arranjos espaciais possíveis destes blocos. Em um espaço do domínio  $\langle D, W \rangle$ , podemos dizer que uma conceitualização é um conjunto das relações conceituais definidas em um espaço de domínio.

## OBJETIVOS

O papel de uma ontologia pode ser entendido considerando ontologia como um conjunto de axiomas lógicos projetados para esclarecer o significado pretendido de um vocabulário. Não é fácil (talvez nem possível) encontrar o conjunto de axiomas perfeitos. Conseqüentemente, uma ontologia específica uma conceitualização somente de maneira muito indireta, porque 1) ela pode somente aproximar um conjunto de modelos pretendidos; 2) tal conjunto de modelos pretendidos é somente uma caracterização fraca de uma conceitualização. Afirma-se que uma ontologia  $O$  para uma linguagem  $L$  se aproxima de uma conceitualização  $C$  se existir um compromisso ontológico  $K$  (interpretação intensional) incluído nos modelos do  $O$ .

## TESAURO ONTOLOGIA TERMO

**X**

Não trata

## CONCEITO

Não trata

## OBJETIVOS

As estruturas lexicais e ontológicas são o sustentamento do trabalho científico e acadêmico, da aprendizagem, e da inteligência de máquinas. Servem para muitas funções críticas no pensar e no comunicar, no organizar, e na recuperação da informação por pessoas e máquinas. As funções das ferramentas que fornecem tais estruturas incluem o seguinte: a) Fornecer um mapa semântico de caminhos aos campos individuais e aos relacionamentos entre campos, assim fornecendo a orientação e servindo como uma ferramenta de referência. Isto inclui as seguintes funções específicas: relacionar conceitos aos termos e fornecer definições; esclarecer conceitos, contextualizando-os em uma classificação/ontologia; relacionar conceitos e termos ou ícones através das disciplinas, das línguas, e das culturas. b) Melhorar a comunicação e o aprendizado: assistindo escritores e leitores; sustentar o ensino fornecendo estruturas conceituais, com estudantes se esforçando para produzir tais estruturas; sustentação da aprendizagem da linguagem; e suportar o desenvolvimento de materiais instrutivos. c) Fornecer a base conceitual para o projeto da pesquisa e da execução ótimas: ajudar investigadores na exploração do contexto conceitual de um projeto de pesquisa, de uma política, de uma planta, ou de um projeto de execução e estruturar o problema; suportar a definição consistente das variáveis/medidas para uns resultados mais comparáveis e mais cumulativos da pesquisa e da avaliação. d) Fornecer a classificação para a ação: uma classificação das doenças para o diagnóstico, de habilidades da equipe de funcionários para as atribuições da tarefa, dos productos para o uso. e) Sustentar a recuperação da informação em uma base de conhecimento; suportar a busca hierárquica expandida; suportar exposições bem-estruturadas de resultados da busca; fornecer uma ferramenta para a indexação (problem-oriented) e controlar o vocabulário do indexador. f) Fornecer a base conceitual para sistemas de bases de conhecimento. g) Fornecer a base conceitual para a definição do elemento de dados e objetar hierarquias em sistemas de software. h) Efetuar todas as funções

TESAURO ONTOLOGIA TERMO	CONCEITO	OBJETIVOS
<p><b>X</b> O sentido principal de um termo pode ser descoberto por meio da associação entre termos diferentes.</p>	<p>Conceitos são classes. Conceitos são representações mentais refletidas por meio de termos.</p>	<p>descritas acima por meio das disciplinas, das linguagens, e das culturas. i) Servir como um dicionário para o uso humano e como o dicionário/base de conhecimento para processar linguagem natural - tradução de máquina e compreensão de linguagem natural para a extração e indexação automáticas de dados.</p> <p>Ontologia é um modelo de referência definido explicitamente de aplicação de domínios com a finalidade de melhorar a consistência e o reuso da informação, a interoperabilidade dos sistemas e o compartilhamento do conhecimento. Descreve a semântica de um domínio em uma maneira humano-compreensível e computador-processável.</p>
<p><b>X</b> Não trata.</p>	<p>Não trata.</p>	<p>Fornecer um entendimento comum e compartilhado de um determinado domínio. Possibilitar serviços baseados em operacionalizações semânticas. Auxiliar pessoas e computadores no que diz respeito ao acesso à necessidade informacional. Proporcionar a interoperabilidade dos serviços ofertados na WEB, potencializando os níveis de serviços ofertados, ou seja, sustentar a web-semântica. Ligar parte significativa do conhecimento humano aos processamentos computacionais, potencializando assim o processo de comunicação entre especialistas de um dado domínio.</p>
<p><b>X</b> Termo faz referência a um conceito. Representa um conceito ou classe.</p>	<p>Não trata.</p>	<p>Criar uma estrutura lógica, uma filosofia, uma classificação, desenvolver um conhecimento (entendimento) comum em uma disciplina. Determinar a extensão e a complexidade dos processos. Fornecer um entendimento básico de um domínio, o que requer menos esforço do que sustentar argumentos lógicos e provas de um domínio.</p>

## TESAURO ONTOLOGIA TERMO

**X**

Não trata.

## CONCEITO

Não trata.

## OBJETIVOS

Possibilitar, juntamente com aplicações lógicas, a construção de modelos computacionais para um determinado domínio de aplicação. Tais modelos computacionais visam ao cumprimento de tarefas de representação do conhecimento de um domínio. Servir como um catálogo dos tipos de coisas que existem em um domínio do interesse D da perspectiva de uma pessoa que use uma língua L com a finalidade de falar sobre D. Representar os predicados, os sentidos das palavras, ou os tipos de conceito e relação da língua L quando usados para discutir tópicos no domínio D.

**X**

Não trata.

Todo elemento de uma ontologia é ou um tipo de classe ou um tipo de propriedade.  
Relacionamento super-ordenado entre classes  
Relacionamento sub-ordenado entre classes e propriedades.  
O relacionamento entre classes ou propriedades são poli-hierárquico, pois uma classe pode ser uma sub-classe de uma, duas, três ou mais classes super-ordenadas.

No ambiente da web semântica uma ontologia serve para fornecer uma conceitualização parcial de um determinado domínio, de forma compartilhada entre usuários comuns e que é formalmente definida por uma linguagem processada por máquinas. Tem como objetivo também o compartilhamento de informações semânticas por meio de sistemas automáticos. E, oferecer meios concisos e sistemáticos para definir a semântica dos recursos web.

**X**

Com base nas distinções de Aristóteles que o autor usa analogicamente para tratar dos elementos da IFF, um termo pode ser:  
um tipo (universal, geral)  
uma instância (particular, específico)  
uma relação (qualidade)  
uma entidade (substância).

Não trata.

Não trata.

**X**

Não trata.

Não trata.

A construção de linguagens de marcação, imprescindíveis para a realização da web semântica, estão sendo desenvolvidas com base em uma abordagem arranjada em camadas (níveis), sobretudo no nível do vocabulário ontológico. A ontologia, ou, o vocabulário resultante da ontologia, é o nível responsável pelas bases de raciocínio e inferência da web semântica. A finalidade das ontologias na esfera da web semântica é proporcionar regras de inferência e deduções racionais para que sistemas sejam

**TESAURO ONTOLOGIA TERMO**

**X** Não trata.

**X** O termo expressa o conceito. O termo é a etiqueta do conceito.  
Termos legais (jurídicos) são etiquetas de conceitos legais específicos. Ex.: contrato, responsabilidade. Ou também etiquetas de objetos do mundo apreendidos por Lei, ex.: passageiro, armas e drogas. Observando a relação sintática entre termos é possível identificar a relação semântica entre eles. Afirma que, estatisticamente, termos relacionados semanticamente ocorrem com muita frequência com a mesma apresentação sintática.

**X** Tesouro: termo é visto como descritor, palavra-chave. Tipos de relacionamentos: hierárquico, associativo e equivalência.

**CONCEITO**

Não trata.

Conceitos são expressados (personificados) por meio de termos que ligados semanticamente entre conceitos são personificados em relações sintáticas entre estes termos. Os conceitos são definidos um a um. As definições dos conceitos podem ser mais ou menos explícitas. Ex.: a definição do registro de nascimento das pessoas nascidas no exterior é explícito, pois está subentendido que condições um registro de nascimento é estabelecido e que elementos os faz ser de pessoas estrangeiras. A definição do conceito de divórcio é menos explícita, pois se refere à ruptura do casamento, que pode ser dada pelo falecimento de um dos envolvidos, ou, pelo pedido legal de separação. Mas, uma ontologia pode definir o divórcio apenas como a ruptura do casamento. A relação semântica entre conceitos se dá por meio das formas sintáticas em um discurso. Em uma ontologia conceitos são definidos pela semântica das relações estabelecidas entre cada conceito e entre outros conceitos.

Tesouro: não trata.  
Ontologia: conceitos organizados em taxonomias.

inteligíveis e processados automaticamente por meio das várias linguagens de marcação.

**OBJETIVOS**

Teorias ontológicas pretendem conceber uma representação útil da estrutura geral e dinâmica do mundo real como ele existe/é. Proporciona uma lente que nos permite enxergar o mundo real e impor a ordem dos fenômenos, concebendo um consenso desses fenômenos. Por tal motivo as teorias ontológicas são úteis para os sistemas de informação (agindo na atividade de modelamento conceitual), já que esses têm como objetivo representar um sistema do mundo real.

Ontologias concebem para os sistemas de informação uma representação do mundo real em um dado domínio.

Recuperação da informação e raciocínio automático de um dado domínio.

Tesouro: auxiliar a indexação e a recuperação da informação.  
Ontologias: descrever um entendimento comum e

Ontologia: não trata.

compartilhado de um determinado domínio.

**X**

Não trata.

Não trata.

Fornecer um entendimento comum e compartilhado de um domínio que pode ser comunicado entre pessoas e sistemas de aplicação.  
Capturar significados e relações entre os significados.