

UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE CIÊNCIA DA INFORMAÇÃO

ELIANA ANTONIA DEMARQUES

**Estudo exploratório sobre compatibilização semântica de vocabulário utilizando ontologia: fundamentação teórica-metodológica visando ao aperfeiçoamento do MHTX.**

Belo Horizonte

2011

**ELIANA ANTONIA DEMARQUES**

**Estudo exploratório sobre compatibilização semântica de vocabulário utilizando ontologia: fundamentação teórica-metodológica visando ao aperfeiçoamento do MHTX.**

Dissertação apresentada ao PPGCI - Programa de Pós-Graduação em Ciência da Informação da ECI - Escola de Ciência da Informação da UFMG - Universidade Federal de Minas Gerais para obtenção do grau de título de Mestre em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação.

Orientadora: Dra. Gercina Ângela Borém de Oliveira Lima.

Belo Horizonte

2011

Demarques, Eliana Antonia.

D372e      Estudo exploratório sobre compatibilização semântica de vocabulário utilizando ontologia: fundamentação teórica-metodológica visando ao aperfeiçoamento do MHTX/ Eliana Antonia Demarques. – 2011. 172 f.; enc.

Orientadora: Gercina Angela Borém de Oliveira Lima.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.  
Referências: f. 158-168.  
Anexos: f. 169-172.

1. Ciência da Infomação – Teses. 2. Recuperação da infomação – Teses. 3. Ontologias (Recuperação da informação) – Teses. 4. Web semântica – Teses. 5. Bibliotecas digitais. – Teses. I. Título. II. Lima, Gercina Ângela Borém de Oliveira. III. Universidade Federal de Minas Gerais. Escola de Ciência da Infomação.

CDU: 025.4.03



UFMG

Universidade Federal de Minas Gerais  
Escola de Ciência da Informação  
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO

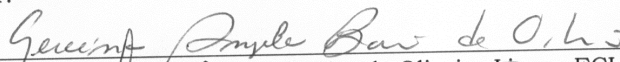
"ESTUDO EXPLORATÓRIO SOBRE COMPATIBILIZAÇÃO SEMÂNTICA DE VOCABULÁRIO UTILIZANDO ONTOLOGIA: FUNDAMENTAÇÃO TEÓRICA-METODOLÓGICA VISANDO AO APERFEIÇOAMENTO DO MHTX

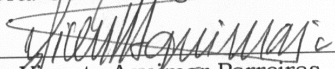
Eliana Antônia Demarques

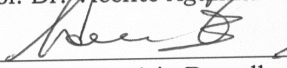
Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de **"Mestre em Ciência da Informação"**, Linha de Pesquisa: **"Organização e Uso da Informação - OUI"**.

Dissertação aprovada em: 26 de abril de 2011.

Por:

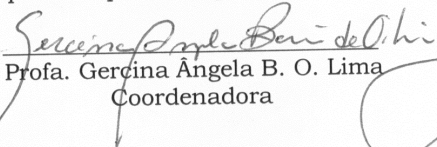
  
\_\_\_\_\_  
Prof. Dra. Gercina Ângela Borém de Oliveira Lima - ECI/UFMG (Orientadora)

  
\_\_\_\_\_  
Prof. Dr. Vicente Aguiar Parreiras - CEPET/MG

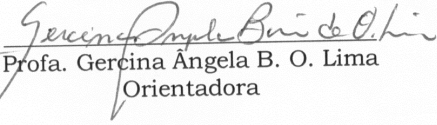
  
\_\_\_\_\_  
Prof. Dr. Maurício Barcellos Almeida - ECI/UFMG

  
\_\_\_\_\_  
Prof. Dra. Lídia Alvarenga - ECI/UFMG

Aprovada pelo Colegiado do PPGCI

  
\_\_\_\_\_  
Prof. Gercina Ângela B. O. Lima  
Coordenadora

Versão final Aprovada por

  
\_\_\_\_\_  
Prof. Gercina Ângela B. O. Lima  
Orientadora



UFMG

Universidade Federal de Minas Gerais  
Escola de Ciência da Informação  
Programa de Pós-Graduação em Ciência da Informação

ATA DA DEFESA DE DISSERTAÇÃO DE **ELIANA ANTÔNIA DEMARQUES**,  
matrícula: 2007669751

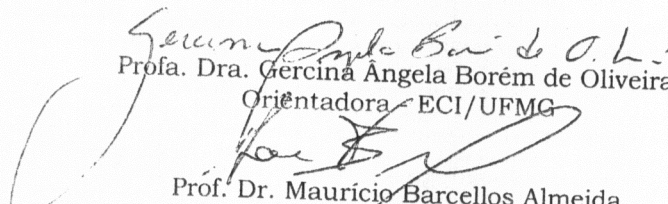
Às 14:30 horas do dia 26 de abril de 2011, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada *ad referendum* pela Coordenadora do Programa de Pós-Graduação em Ciência da Informação em 08/04/2011, para julgar, em exame final, o trabalho intitulado **Estudo exploratório sobre compatibilização semântica de vocabulário utilizando ontologia: fundamentação teórica-metodológica visando ao aperfeiçoamento do MHTX**, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, Área de Concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação - OUI. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Gercina Ângela Borém de Oliveira Lima, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a argüição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

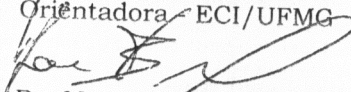
Profa. Dra. Gercina Ângela Borém de Oliveira Lima - Orientadora	APROVADA
Prof. Dr. Vicente Aguiar Parreiras	APROVADA
Prof. Dr. Maurício Barcellos Almeida	APROVADA
Profa. Dra. Lídia Alvarenga	APROVADA

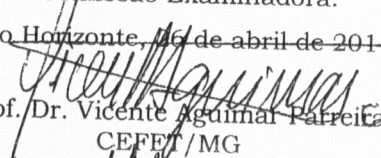
Pelas indicações, a candidata foi considerada APROVADA.

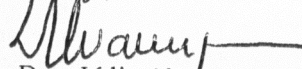
O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 26 de abril de 2011.

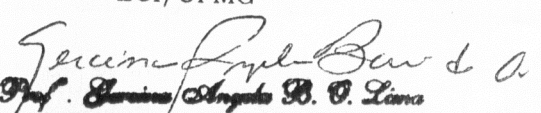
  
Profa. Dra. Gercina Ângela Borém de Oliveira Lima  
Orientadora ECI/UFMG

  
Prof. Dr. Maurício Barcellos Almeida  
ECI/UFMG

  
Prof. Dr. Vicente Aguiar Parreiras  
CEFET/MG

  
Profa. Dra. Lídia Alvarenga  
ECI/UFMG

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.

  
Prof. Gercina Ângela B. O. Lima

## AGRADECIMENTO

Aproveito esta oportunidade para agradecer às pessoas mostrando-lhes que o conhecimento não se constrói sozinho. Por isso, agradeço primeiramente a Deus por ter me orientado nos árduos caminhos da vida, fortalecendo meus passos para que minhas ideias, aparentemente impossíveis, pudessem ser aplicadas na construção de um modelo totalmente inovador, superando dificuldades com sabedoria. Inovar, para mim, é poder ver as facetas de um mesmo objeto com diferentes olhares multifuncionais...

Agradeço a meus pais, Moisés Demarques (*in memoriam*) e Antonieta Baldon, pelo amor incondicional, pelos exemplos, ensinamentos e perseverança frente aos problemas em busca de soluções inovadoras. Com quem compartilhei tantos olhares diferenciados de momentos inesquecíveis e de quem herdei a sensibilidade e o gosto pela área educacional. Agradeço ao meu esposo e fiel companheiro, Djalma Santos Ribeiro, pelo amor, apoio e paciência. Isso também se estende aos nossos filhos, Fernanda Izabele e Igor Henrique pela compreensão dos isolados momentos que se fizeram necessários na escrita e re-escrita deste trabalho para formar todo o conteúdo deste documento.

Aos profissionais da ECI-UFMG, agradeço pelo apoio aos difíceis problemas de saúde que enfrentei durante este período implicando no adiamento da defesa de minha dissertação. Em especial, à Prof<sup>a</sup>. Gercina que me mostrou princípios, métodos e teorias da Ciência da Informação, destacando a importância das pesquisas sociais aplicadas. À CAPES, pelo incentivo financeiro concedido durante um ano de bolsa de Mestrado.

Agradeço a meus amigos, companheiros de trabalho, profissionais da Educação e Computação. Em especial, àqueles que direta ou indiretamente contribuíram na organização deste trabalho, incentivando e ajudando a superar dificuldades e aos colegas do grupo MHTX que possibilitaram a troca de experiências de valiosos conhecimentos. Sou eternamente grata por todos vocês fazerem parte da história desta “obra aberta” à crítica e sugestões de melhorias.

Sem mais palavras para agradecer...

*“[...] quando a palavra emudece e se torna escrita, é aí que fala mais alto e fica ainda mais bonita”.*

Eloí Callage

*"Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim."*

Chico Xavier

## RESUMO

Estudo exploratório sobre compatibilização semântica de vocabulário, utilizando uma ontologia de domínio no MHTX<sup>1</sup>, procurando refletir sobre as práticas de leitura e escrita de documentos publicados em bibliotecas digitais. Nesses ambientes, a ambiguidade do vocabulário é apontada como um dos aspectos cruciais para a compatibilização semântica de bases heterogêneas. A integração de modelos conceituais é uma das técnicas mais usadas para compartilhar informações com entendimento comum entre significados conceituais terminológicos. O MHTX favorece a criação de um ambiente interativo. A ideia é analisar a representação de diferentes publicações em um único mapa como um sistema de conceitos integrados, tendo como base a fundamentação teórica-metodológica sobre compatibilização de vocabulário, visando o aperfeiçoamento do MHTX. Propõe-se uma nova técnica de modelagem informacional de objetos digitais, considerando os mesmos princípios de classificação facetada. Esta técnica permitiu usar as funcionalidades do conceito para explicitar relações de equivalência em um único mapa, indo ao encontro dos pressupostos deste estudo que considera o 'conceito' um metadado descritivo, administrativo e estrutural. Para isso, tanto a indexação quanto a recuperação deve ser baseada em uma terminologia consensual para o estabelecimento de categorias formais diretamente ligadas ao uso da linguagem. A combinação dessas categorias representadas em facetadas de uma ontologia de domínio mostrou conexões semânticas em que as características expressas nos objetos hipertextuais estão presentes na definição e nas relações entre conceitos. Os resultados confirmam a viabilidade de inserir uma ontologia de domínio no MHTX, utilizando a modelagem proposta com mecanismos automatizados de indexação para conduzir compatibilização semântica do vocabulário.

**Palavras-chave:** Ontologia; Modelagem Informacional; Compatibilização Semântica de Vocabulário; Biblioteca Digital; Modelo de Interatividade em Recuperação de Informação.

---

<sup>1</sup> O protótipo do Modelo Hipertextual para Organização de Documentos (MHTX) pode ser visitado no endereço <http://gercinalima.com/mhtx>



## ABSTRACT

This is an exploratory study regarding semantic vocabulary compatibilization utilizing the MHTX<sup>[1]</sup> ontology, contemplating reading and writing practices of published documents in digital libraries. Within these environments, the vocabulary ambiguity is stressed as one of the crucial aspects for the semantic compatibilizing of heterogeneous bases. The integration of conceptual models is one of the most used techniques to share information with a common understanding among terminological conceptual meanings. The MHTX supports the creation of an interactive environment. The idea is to analyze the representation of different publications in a unique map with an integrated conceptual system, using as an underlying concept the methodological-theory about vocabulary compatibilization, aiming at the improvement of the MHTX. A new technique of digital objects' representation is presented considering the very same principles of faceted classification. This technique allowed using the concept's functionalities to explain the equivalence relationships in a unique map, meeting the assumed ideas of this study that stresses a concept of descriptive, administrative and structural metadata. Therefore both indexation and retrieval shall be based in a terminological agreement for the establishment of formal categories directly connected to the language's use. The combination of these categories represented in domain ontology's facets showed semantic connections, in which the expressed characteristics in hyper-textual objects are present in the definition and relationships among these hyperlinked objects. The results confirm the viability of inserting MHTX domain ontology, using the proposed modeling with automated mechanisms of indexation to conduct the compatibilization semantic vocabulary.

**Key-words:** Ontology; Informational Modeling; Vocabulary Semantic Compatibilization; Digital Library; Interactivity in the retrieval of information's Model.

## LISTA DE FIGURAS

Figura 1	Mapa da organização da dissertação.....	22
Figura 2	Elementos constitutivos de um documento.....	55
Figura 3	Interação entre tecnologia e conhecimento registrado.....	58
Figura 4	Evolução tecnológica da biblioteca.....	60
Figura 5	Sistema de Armazenamento e Recuperação de Informação.....	63
Figura 6	Categorias de Aristóteles.....	67
Figura 7	Espiral do método científico de Ranganathan.....	69
Figura 8	Compartimento do cérebro: formação da ideias.....	72
Figura 9	Triângulo conceitual e modelo para a construção dos conceitos.....	79
Figura 10	Triângulo de significado estendido.....	82
Figura 11	Adaptação do Triângulo de Significado ao Triângulo Conceitual.....	83
Figura 12	Objeto, nome, símbolo e codificação.....	85
Figura 13	Modelo conceitual de componentes de uma ontologia.....	89
Figura 14	Arquitetura genérica de um sistema de PLN.....	91
Figura 15	Raciocínio Baseado em Casos.....	92
Figura 16	Indexação automática e índice invertido. ....	94
Figura 17	Modelo de uma rede semântica.....	95
Figura 18	Apresentação em HTML e XML e exemplo de anotação semântica.....	98
Figura 19	Arquiteturas da Web Semântic.....	100
Figura 20	Evolução da Web. Fonte: Santos e Alves, 2009.....	103
Figura 21	Resumo das etapas de desenvolvimento de um software.....	107
Figura 22	Etapas processo de construção de ontologias agrupadas em um ciclo .....	108
Figura 23	Recuperação de informação dirigida por semântica.....	118
Figura 24	Primeira tela do mapa conceitual hiperbólico do MHTX.....	122
Figura 25	Modelo estratificado de interação em RI e comportamento do usuário .....	130
Figura 26	Gestão inteligente e as necessidades informacionais.....	140
Figura 27	Posicionamento da inteligência na busca por um conceito relevante .....	141
Figura 28	ONTOCI e hiperdocumento aberto com a ferramenta SMORE.....	149
Figura 29	Exemplo do funcionamento da modelagem informacional proposta. .....	153

## LISTA DE QUADROS E TABELAS

QUADRO 1 – CONJUNTO DE ELEMENTOS DUBLIN CORE.....	62
QUADRO 2 – PLANOS E PRINCÍPIOS DO MODELO SIMPLIFICADO DE SPITERI .....	72
QUADRO 3 – PRINCÍPIOS DO PMEST UTILIZADOS POR DAHLBERG. .	76
QUADRO 4 – CARACTERÍSTICAS E RELACIONAMENTOS ENTRE CONCEITOS .....	81
QUADRO 5 – MAPEAMENTO ENTRE AS CATEGORIAS DE ARISTÓTELES E RANGANATHAN.....	89
QUADRO 6 – SÍNTESE DE ALGUMAS FERRAMENTAS PARA O USO DE ANOTAÇÃO SEMÂNTICA. ....	101
QUADRO 7 – NÍVEL DE GERENCIAMENTO DO CONHECIMENTO.....	118
QUADRO 8 – EXEMPLO DE SEÇÃO DA MATRIZ DE COMPATIBILIZAÇÃO CONCEITUAL DE VOCABULÁRIOS ...	127
QUADRO 9 – ESQUEMA RESULTANTE DA UNIFICAÇÃO MANUAL DAS ESTRUTURAS FACETADAS DAS 2 TESES.....	149
QUADRO 10 – EXEMPLO DE IMPLEMENTAÇÃO EM RDF/XML COM DUBLIN CORE.....	155
QUADRO 11 – ESTRUTURA HIERÁRQUICA OBTIDA A PARTIR DA NAVEGAÇÃO NO MAPA DE TÓPICOS RESULTANTE DA FUSÃO AUTOMATIZADA DE DUAS TESES NO MHTX .....	173

## LISTA DE ABREVIATURAS

BDICI - Biblioteca Digital Interativa da Ciência da Informação  
CC - Ciência da Computação  
CI - Ciência da Informação  
CRG - Classification Research Group  
DC - Dublin Core Metadata Element  
DTD - Document Type Definition  
HTML - Hypertext Markup Language  
IA - Inteligência Artificial  
MARC - Machine-Readable Cataloging  
MHTX - Modelo Hipertextual para Organização de Documentos.  
NDLTD ETD MS - Networked Digital Library of Theses and Dissertation e An Interoperability  
Metadata Standard for Theses and Dissertations  
NISO - National Information Standards Organization  
OAI-PMH - The Open Archives Initiative Protocol for Metadata Harvesting  
OAIS - Open Archival Information System  
ONTOCI – Ontologia da Ciência da Informação  
OUI - Organização e Uso da Informação  
OWL -Web Ontology Language  
PPGCI - Programa de Pós-Graduação em Ciência da Informação  
RDF - Resource Description Framework  
RI - Recuperação de Informações  
SGML - Standard Generalized Markup Language  
SOIF -Summary Object Interchange Format  
SRI - Sistema de Recuperação de Informações  
TIC - Tecnologias de Informação e Comunicação  
UFMG - Universidade Federal de Minas Gerais  
URI - Uniform Resource Identifier  
URI - Uniform Resource Indicator  
URL - Uniform Resource Locator  
URN - Uniform Resource Name  
W3C - World Wide Web Consortium  
XML -Extensible Markup Language

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>15</b>
1.1 OBJETIVOS GERAIS E ESPECÍFICOS .....	21
1.2 MÉTODOS, TÉCNICAS E MATERIAIS DA PESQUISA .....	22
1.3 ORGANIZAÇÃO E MAPA DA DISSERTAÇÃO .....	23
<b>2 FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS.....</b>	<b>24</b>
2.1 DOCUMENTO TEXTUAL: LINGUAGEM, INFORMAÇÃO E CONHECIMENTO.....	24
2.1.1 A estruturação de um texto científico.....	26
2.1.2 Campo conceitual para estruturação de vocabulários.....	28
2.1.3 A gramática generativa transformacional e a formação de conceitos espontâneos e científicos em um campo conceitual.....	31
2.1.4 Aspectos multifuncionais da linguagem na escrita textual .....	34
2.1.5 A ambiguidade do vocabulário humano e os desafios do leitor- interpretador de conceitos ou termos desconhecidos .....	39
2.1.6 A importância do contexto para desambiguação do vocabulário .....	43
2.1.7 Cognição, linguagem e princípios computacionais: modelos conceituais de modelos mentais .....	45
2.2 CIÊNCIA DA INFORMAÇÃO E OS DESAFIOS DA RECUPERAÇÃO NOS HIPERDOCUMENTOS CIENTÍFICOS NA CONTEMPORANEIDADE .....	51
2.3 BIBLIOTECAS E OS DESAFIOS DA ORGANIZAÇÃO DA INFORMAÇÃO.....	60
2.4 ONTOLOGIAS E SEUS BENEFÍCIOS NO PLANEAMENTO DE SISTEMAS .....	68
2.4.1 Teoria da Análise facetada e suas relações com os princípios filosóficos de categorias formais nos sistemas de classificação ...	69
2.4.2 Teoria do Conceito e os princípios de classificação facetada .....	77
2.4.3 Palavra, termo, conceito, definição e suas relações com os princípios filosóficos de uma ontologia .....	83
2.4.4 Uso de ontologias no processamento da linguagem natural.....	92
2.4.5 Padrões de metadados digitais propostos pelo W3C com o uso de anotações semânticas e ontologias.....	98
2.4.6 Ferramentas, linguagens e modelagem de ontologias .....	106

<u>2.5 BIBLIOTECAS DIGITAIS ESPECIALIZADAS E A NECESSIDADE DE PADRÕES DE METADADOS W3C PARA CONDUZIR INTEROPERABILIDADE SEMÂNTICA.....</u>	<u>113</u>
<u>2.6 BIBLIOTECAS DIGITAIS INTERATIVAS E AS PROPOSTAS PARA O MHTX.....</u>	<u>121</u>
<u>2.6.1 Linguagens documentárias e compatibilização do vocabulário.....</u>	<u>126</u>
<u>2.6.2 Sistemas hipertextuais organizados para interatividade em RI .....</u>	<u>131</u>
<u>2.6.3 Funcionalidades da linguagem e potencialidades do “conceito” .....</u>	<u>138</u>
<b><u>3 ESTUDO EXPLORATÓRIO NO MHTX PARA COMPATIBILIZAÇÃO SEMÂNTICA DO VOCABULÁRIO UTILIZANDO ONTOLOGIA</u></b>	<b><u>146</u></b>
<u>3.1 DETALHAMENTO DAS IMPLEMENTAÇÕES CONCEITUAIS E TECNOLÓGICAS.....</u>	<u>148</u>
<u>3.2 ANÁLISE DOS RESULTADOS.....</u>	<u>154</u>
<b><u>4 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....</u></b>	<b><u>157</u></b>
<b><u>REFERÊNCIAS.....</u></b>	<b><u>161</u></b>
<b><u>ANEXO I.....</u></b>	<b><u>173</u></b>

# 1 INTRODUÇÃO

Este capítulo apresenta a justificativa, o problema, os objetivos, os pressupostos, os métodos e materiais de pesquisa que têm como meta central um estudo exploratório sobre compatibilização semântica de vocabulário, utilizando uma ontologia de domínio para representar um sistema de conceitos integrados em uma base de dados<sup>2</sup>. Esta meta envolve temas diversificados, todos eles de extensa abrangência, e por isto, não se delimitam fronteiras, assumindo uma ampla visão de pesquisas interdisciplinares, relacionando a Ciência da Informação (CI) com outras ciências tais como a Ciência da Computação (CC), Biblioteconomia, Filosofia, Ciências Cognitivas, Linguística, entre outras.

A CI sempre trabalhou numa perspectiva interdisciplinar, ajudando a organizar e sistematizar o conhecimento de diversas áreas. Os objetivos de criar um sistema de conceitos são: a) organizar efetivamente o conhecimento de uma área específica; b) representar claramente as relações entre conceitos; e c) revelar conceitos ainda inexistentes ou redundantes, possibilitando uma normalização terminológica (CAMPOS, 2001).

Muitas pesquisas buscam similaridades no processo de construção de ontologias na Inteligência Artificial (IA), subárea da CC com métodos e técnicas utilizados na CI. Uma ontologia na IA pode ser definida como uma “especificação explícita e formal de uma conceitualização compartilhada”, em que ‘especificação explícita e formal de uma conceitualização’ permite explicitar relações entre conceitos compreensíveis por humanos e máquinas; e ‘compartilhada’ por se tratar de um vocabulário comum definidos por uma comunidade científica (BORST, 1997).

Silva, Souza e Almeida (2008) constatam que os princípios advindos da Teoria da Análise Facetada proposta por Ranganathan (1967) e da Teoria do Conceito proposta por Dahlberg (1978) são pertinentes nos processos de construção de ontologias na IA. Esses princípios serão investigados nesta pesquisa, visando o aperfeiçoamento do MHTX - Modelo Hipertextual para Organização de Documentos que foi proposto por Lima (2004), utilizando quatro princípios básicos: a Teoria da Análise Facetada, a Teoria dos Mapas Conceituais, a estrutura semântica de links hipertextuais e as normas gramaticais da Associação Brasileira de Normas Técnicas (ABNT).

O intuito principal de Lima (2004) com o MHTX foi minimizar o problema da sobrecarga cognitiva na leitura de sistemas hipertextuais. Lima alerta que muitos desses sistemas vêm sendo criados a partir de textos convencionais, sem qualquer preocupação com a organização semântica da informação. Essa organização e estruturação de

<sup>2</sup> Uma base de dados é uma coleção coerente de dados relacionados com algum significado próprio, mantendo modelos de algum universo de discurso (ELMASRI e NAVATHE, 2002).

hiperlinks conceituais encontram-se distanciadas do autor do hiperdocumento, implicando na existência de diversos problemas na área de Recuperação da Informação (RI). A área de RI trata dos aspectos intelectuais da descrição da informação e sua especificação para busca em qualquer sistema, além de verificar técnicas ou máquinas mais adequadas que são empregadas para realizar esta operação (MOOERS, 1951).

Essa preocupação com a preservação de conteúdos documentais, de modo a permitir posteriormente sua recuperação, vem desde tempos antigos, em bibliotecas tradicionais, relacionando índices como apontadores às informações. Assim, por séculos, índices foram selecionados, categorizados, classificados e catalogados manualmente (indexação manual). Mas, com o advento dos computadores, outras técnicas e soluções foram propostas, buscando criar novos mecanismos de indexação automática.

No âmbito da CI, antes mesmo do surgimento dos sistemas de hipertextos, a organização e o tratamento da informação já haviam sido afetados pelo distanciamento entre o autor do documento textual e os sistemas de RI (LIMA, 2004). Buckland (1999) alerta para a existência de múltiplos vocabulários representados nesses sistemas: o vocabulário do produtor do documento; o vocabulário do profissional da informação e o vocabulário do usuário.

Essa multiplicidade de vocabulários pode provocar ineficácia e inconsistências na recuperação, considerando que o universo da palavra é dinâmico, e seu significado pode mudar, dependendo do contexto. A ambiguidade de um vocabulário é um dos problemas mais difíceis para que um sistema encontre compatibilidade semântica entre as múltiplas fontes de informação para posterior recuperação. O uso de uma terminologia padronizada é essencial para uma comunicação sem ambiguidades (ALMEIDA, 2006).

As ontologias desenvolvidas na IA, têm sido utilizadas como uma das possibilidades mais promissoras para garantir a compatibilização e interoperabilidade semântica de aplicações na Web. Um sistema possui compatibilidade e interoperabilidade semântica quando é capaz de trabalhar em conjunto com outros sistemas, compartilhando informações e com entendimento de um vocabulário comum entre os significados dos seus termos (BURANARACH, 2001). Entretanto, a compatibilização semântica é um aspecto crucial em qualquer sistema, natural ou artificial, devido à complexidade do vocabulário humano usado na escrita de qualquer documento textual.

Tudo isso compromete a eficácia de um modelo de sistema que possui em seu acervo importantes descobertas descritas em teses e dissertações que precisam ser compartilhadas por uma comunidade científica. Muitas vezes, a informação pode estar armazenada, mas não há mecanismos adequados que ajudem na RI (FURGERI, 2006). Devido à complexidade de funcionalidades e interações necessárias, os atuais modelos de



sistemas padecem de graves problemas de usabilidade<sup>3</sup>: seu aprendizado é difícil, sua eficácia é ruim, causam muitos erros, a memorização de seu modo de usar é árdua e não são, de modo algum, agradáveis de usar (CRUZ, 2008).

Considerando que o planejamento de um modelo computacional é a atividade que define o modo como o sistema deverá ser suportado pelas novas Tecnologias de Informação e Comunicação (TIC), planejar um novo sistema para o MHTX é delinear um futuro desejado, conhecendo processos eficazes que possam ajudar na concretização desse plano. Segundo Kent (1972), planejar um sistema de RI seria um problema trivial se: a) todo assunto de um documento provocasse o mesmo fluxo de observações; b) se todo usuário-sistema utilizasse palavras idênticas para descrever a sua necessidade de informação sobre este assunto; e c) se cada usuário que estivesse interessado em aprender algo sobre esse assunto fizesse perguntas usando a mesma terminologia.

Portanto, não basta apenas ter tecnologia, é preciso estabelecer critérios de organização para o tratamento documentário<sup>4</sup> como requisito essencial para que, em uma etapa posterior, usuários ou ferramentas de busca alcancem maior eficácia na RI. Esta organização deve preceder a fase de construção do sistema, iniciando, preferencialmente, na fase de modelagem de dados para que a navegação em contexto semântico proposta pelo MHTX possa ser acessada pelos nós do mapa conceitual (LIMA, 2004).

A princípio, o MHTX abrigou em sua base de dados a tese de Lima (2004) em processos manuais, demandando muito tempo para agregar pontos de acesso, utilizando um mapa conceitual de interface hiperbólica<sup>5</sup>. Esta pesquisa busca novos recursos que possam contribuir para o aperfeiçoamento do MHTX como um modelo de biblioteca digital para interatividade em RI<sup>6</sup>: o MHTX pode ser um sistema gerenciador de conteúdo semântico de teses e dissertações para compor o acervo do Programa de Pós-Graduação em Ciência da Informação (PPGCI) e da Escola de Ciência da Informação (ECI) através de interfaces que sustentam suas funções em uma rede semântica (LIMA, 2007, p. 6).

Rezende (2003) define uma rede semântica como um grafo<sup>7</sup> rotulado, formado por objetos que podem representar conceitos de um domínio, denominados nós e um

<sup>3</sup> Usabilidade é a medida na qual um produto pode ser usado para alcançar objetivos com eficácia, eficiência e satisfação em contextos específicos de uso (ABNT, 2002, p. 3).

<sup>4</sup> Neste trabalho, o termo tratamento documentário é utilizado como sinônimo de tratamento da informação. Assim como, o termo representação documentária é utilizado como sinônimo de representação da informação.

<sup>5</sup> A interface hiperbólica permite visualização detalhada de conceitos hiperlinkados, sem perder a referência da sua função em relação à estrutura global. Outras partes do hiperdocumento (tese em formato digital) podem ser expandidas e lidas arrastando links, ou quando se apresentam completamente ocultas, por meio de pesquisa direta por conceitos (LIMA, 2004, p. 28).

<sup>6</sup> De acordo com Saracevic (1997), interatividade em RI é o diálogo entre usuário e computador por meio de uma interface que deve ter o objetivo principal de influenciar o estado cognitivo do usuário no uso efetivo da informação em conexão com uma aplicação.

esquema de arcos, chamados *links*. As bibliotecas digitais têm como base de dados conteúdos documentais em textos completos armazenados em formatos eletrônicos e disponíveis para acesso, seguindo processos padronizados e distribuídos via rede de computadores (MARCONDES, KURAMOTO, TOUTAIN e SAYÃO, 2005).

Uma biblioteca digital interativa visa à produção cultural, na qual as pessoas possam ser membros criadores e, não apenas, consumidores de cultura, possuindo uma concepção dinâmica de espaço dentro da própria unidade acadêmica para permitir interação entre os indivíduos que a acessam (PERROTTI, 2001). A ideia é que o mapa conceitual do MHTX utilize a ontologia da CI para explicitar “conceitos” integrados em um único mapa, visando à compatibilização e interoperabilidade semântica do vocabulário.

A organização e estruturação de determinada área de conhecimento, para fins de criação de um sistema de conceitos integrados em uma rede semântica pode ser favorecida por um mapa ou grafo conceitual de uma ontologia que compõe o mapeamento de um domínio, explicitando relações de equivalência. Esta percepção foi realizada nos estudos de Silva (2008), usando mapas de tópicos<sup>8</sup> na tentativa de compatibilização de vocabulário controlado no MHTX. Silva propõe o inter-relacionamento semântico entre os mapas conceituais de duas teses que já estavam com o conteúdo devidamente estruturado em forma de facetas<sup>9</sup> para possibilitar o processo de fusão. Entretanto, esse processo apresentou grande complexidade, exigindo um alto nível de interferência manual.

Os resultados dos estudos de Silva (2008) são de grande contribuição para o desenvolvimento desta pesquisa. A ocorrência de poli-hierarquias no mapa de tópicos resultante da fusão dessas duas teses mostrou-se bastante eficiente no sentido de mostrar diferentes definições conceituais para um mesmo termo que possa fazer parte da estrutura taxonômica facetada de uma ontologia de domínio. Por isso, essa fusão (Anexo I) serve de amostra como estudo exploratório desta pesquisa, considerando que esses resultados foram analisados à luz das teorias de compatibilização de linguagens desenvolvidas no âmbito da CI e da Biblioteconomia, a partir da década de 60.

Para transformar o MHTX em um sistema de biblioteca digital interativa, outros documentos científicos precisam ser inseridos neste repositório, necessitando de

---

<sup>7</sup> Um grafo conceitual é uma linguagem formal de representação do conhecimento, baseado nos trabalhos de Charles Sanders Peirce sobre grafos existenciais e nas redes semânticas da IA. Segundo Sowa, os grafos conceituais são capazes de expressar de forma logicamente precisa, o conhecimento de maneira a serem interpretados tanto por homens como por máquinas. Fonte: <http://www.jfsowa.com/cg>

<sup>8</sup> Mapas de Tópicos são coleções de tópicos que representam conceitos relacionados uns aos outros por associações (OLIVEIRA; MARCOS; VAASAN, 2000, p.6).

<sup>9</sup> Uma faceta é a coleção de termos que apresentam um relacionamento hierárquico. Dentro de cada faceta, os termos que a constituem são suscetíveis a novos agrupamentos, pela aplicação de outras características divisionais (LIMA, 2002, p.190).

compatibilização e interoperabilidade semântica entre as fontes de informação para automatizar processos com padrões de metadados digitais que condicionem resultados de estratégia de busca. No entanto, o MHTX não aborda a estrutura de padrões de metadados de preservação digital quando novos documentos forem inseridos neste repositório.

Esses padrões implicam em um compromisso entre usuários e provedores de definições de uma terminologia consensual (FREITAS, 2006). Os padrões de metadados digitais sugeridos pelo W3C (World Wide Web Consortium) exercem as principais funções para solucionar problemas de compatibilização e interoperabilidade semântica na Web, utilizando ontologias e anotações semânticas. Kiryakov (2003) define essas anotações como um esquema específico para geração e uso de metadados. O esquema refere-se a entidades nomeadas que podem ser descritas e referenciadas pelo nome em documentos de texto. Por isso, é comum encontrar outros termos relacionados a anotações semânticas em documentos como metadados semânticos, rótulos, *tags*, marcas, marcação de procedimento, etiquetas, etiquetagem, rotulagem, etc.

Uma estrutura de padrões de metadados de preservação digital deve ser genérica, abrangente e aplicável a uma vasta gama de recursos, unindo diversos componentes constituintes de um objeto complexo em uma rede semântica (FREITAS, 2006). A importância da compatibilização e interoperabilidade semântica entre as comunidades produtoras e/ou usuárias para catalogação de objetos em repositório digital é evidente por meio de padrões de metadados. Ao se compartilhar dados, minimiza-se o tempo no desenvolvimento deste tipo de tecnologia, reduzindo também o tempo de tratamento das informações geradas (DZIEKANIAK, 2006).

Este trabalho propõe um novo tipo de modelagem informacional para que o conteúdo de um hiperdocumento científico possa ser inter-relacionado pelo próprio autor no momento da publicação de uma tese ou dissertação com conceitos representados em um único mapa de uma ontologia de domínio. Assim, o recurso hipertextual, legível por pessoas, pode ser anotado e associado às suas descrições semânticas na ontologia da CI para a compatibilização do vocabulário. Busca-se que o próprio hiperdocumento científico possa ser anotado para indexação direta na fonte, propondo um tratamento documentário mais adequado para a estrutura de padrões de metadados de preservação digital.

Acredita-se que um 'conceito' representado na ontologia de domínio possa ser anotado e relacionado ao contexto semântico da base de dados do MHTX, fazendo com que esse 'conceito' funcione como um metadado descritivo, administrativo e estrutural de objetos informacionais digitais. Para Freitas (2006), há três tipos de metadados nos ambientes digitais: a) os descritivos que utilizam elementos como título, resumo, autor e palavras-chave com a finalidade de busca e identificação dos dados; b) os administrativos que contêm

informações de gerenciamento e outros dados para arquivologia; e c) os estruturais que indicam a ordenação das páginas e capítulos.

Segundo Alvarenga (2001), o que na realidade se classifica em uma biblioteca tradicional ou digital não são os documentos, mas os conceitos contidos nesses documentos. Os conceitos se constituem em elementos dos quais os modelos de representação do conhecimento são construídos e, por este motivo, eles são conhecidos como modelos conceituais. Neste trabalho, acredita-se que o “conceito” seja o principal metadado digital deste conjunto de dados atômicos para estender a toda uma classe de objetos informacionais digitais para o estabelecimento de relações de equivalência.

Gomes (1990) aponta uma direção para o estabelecimento de relações entre conceitos, que são as características usadas na comparação de termos, devendo buscar a definição ou a classificação mais adequada para atender os propósitos da área de RI. Para isso, tanto a indexação quanto a recuperação deve ser baseada em um sistema de ‘conceitos’ terminológicos consensuais e não em qualquer palavra, ressaltando inúmeras vantagens em se trabalhar dessa forma, em especial, quando se estabelecem categorias, classes e diversas relações lógicas e ontológicas diretamente ligadas ao uso da linguagem.

Essa direção apontada por Gomes vai ao encontro dos pressupostos deste trabalho que considera o ‘conceito’ um metadado descritivo, administrativo e estrutural para a organização semântica da informação no MHTX. A utilização de uma terminologia consistente é um importante passo para a consolidação e desenvolvimento de um campo científico (MARCONDES, 2009). Uma ontologia da CI pode oferecer uma terminologia comum da representação do conhecimento registrado em teses e dissertações, evitando interpretações imprecisas desse vocabulário em um modelo de biblioteca digital.

Tudo isto necessita de muita motivação para encontrar alternativas de fundamentação teóricas e metodológicas decorrentes das contradições existentes entre as práticas tradicionais e os novos mecanismos exigidos nos ambientes digitais. Os atuais sistemas de bibliotecas digitais precisam ser (re)pensados sob diferentes perspectivas, considerando a organização semântica da informação fundamental para garantir interatividade em RI. Modelos mais dinâmicos e interativos apresentam-se como desafio à necessidade de singularização contextual na reconstrução do conhecimento, a partir da determinação de requisitos de qualidade e relevância das informações (RAMALHO, VIDOTTI e FUJITA, 2007).

Motivações não faltam para o desenvolvimento deste trabalho. Além dos diversos desafios a enfrentar para quem se propõe a escrever sobre um tema que envolve questões interdisciplinares, existem poucos estudos relacionados ao uso de ontologias em bibliotecas digitais, o que motiva mais ainda o desenvolvimento deste trabalho. A principal

motivação é que uma ontologia da CI possa ser inserida na arquitetura do MHTX, contribuindo em melhorias do atual sistema, vislumbrando-o como um modelo de biblioteca digital interativa para que possa ser utilizado por alunos e professores do PPGCI. Conseqüentemente, a recuperação pode ser mais eficaz, facilitando todo o processo do tratamento documentário que, ainda hoje, é adotado em modelos de bibliotecas digitais.

Apesar de construir uma ontologia na IA ser um processo muito complexo, um campo em evolução como é a CI pode se beneficiar com essa proposta. Ontologias são normalmente construídas por um grupo de pessoas e se constituem como ferramentas intelectuais muito utilizadas desde os primórdios da ciência por objetivarem capturar o conhecimento de um domínio, podendo ser reusáveis em diversas aplicações computacionais. Segundo Campos (2007), ontologias precisam ser construídas em língua portuguesa para serem utilizadas em bibliotecas digitais, viabilizando processos de integração de informações entre pesquisadores como auxílio na busca de um modo geral, e, principalmente, na melhoria dos processos do tratamento da informação para a área de RI.

Em síntese, são muitos problemas e desafios introduzidos até aqui para justificar o desenvolvimento dessa pesquisa. Questiona-se, neste trabalho, como compatibilizar a multiplicidade de vocabulários existentes em hiperdocumentos científicos que podem compor o acervo do MHTX. Outras questões subjacentes podem ser: Quais critérios de padronização de metadados digitais estão sendo aplicados a ontologias, visando compatibilização e interoperabilidade semântica? Quais mecanismos podem aprimorar o MHTX, usando uma única estrutura que ajude a otimizar o tratamento documentário?

## **1.1 Objetivos gerais e específicos**

Os objetivos gerais visam o aperfeiçoamento do MHTX, tendo como base a fundamentação teórica-metodológica sobre compatibilização semântica de vocabulário, utilizando uma ontologia de domínio. Os objetivos específicos são:

- Compreender as características e componentes básicos comuns presentes em uma ontologia e um sistema de conceitos através da bibliografia da área;
- Estudar mecanismos e ferramentas propostos pelo W3C para compatibilização e interoperabilidade semântica de metadados digitais, agregando aos subsídios teóricos e metodológicos da CI com processos de construção de ontologias, através de um estudo exploratório; e
- Contribuir nas pesquisas da área de RI, bibliotecas digitais e ontologias.

## 1.2 Métodos, Técnicas e Materiais da Pesquisa

A presente pesquisa apresenta um estudo exploratório, utilizando-se de observação sistemática de fatos e acontecimentos apreendidos em um contexto que permite inferências e previsões acerca de outros acontecimentos em mesmas condições. O domínio empírico é o da CI, por possuir uma longa tradição teórica e metodológica, constituída por tradições de pesquisa sistemática que podem facilitar a busca por um modelo que represente o conhecimento contido em documentos científicos (MARCONDES, 2008).

Quanto à sua natureza, ela é classificada como pesquisa aplicada, tendo como base o método científico com abordagens qualitativas para a interpretação, organização, compreensão e descrição dos resultados. A fundamentação teórica possui referências interdisciplinares, convergindo-se em mais de uma área do conhecimento e seus resultados podem contribuir para o avanço das ciências envolvidas através da transferência de métodos de uma para a outra que podem gerar novos conhecimentos (CAPES, 2003, p. 3).

O universo pesquisado é a base de dados do protótipo MHTX e as estruturas facetadas de duas de suas teses que servirão de amostras para a aplicação do experimento. Essas amostras representam um cenário ideal para o desenvolvimento deste trabalho, pois elas já foram analisadas por Silva (2008) pelo processo de fusão dos mapas conceituais de seus tópicos estruturados em forma de facetas. Na primeira amostra foi utilizada a Teoria da Análise Facetada proposta por Ranganathan (1967) e na segunda amostra, as categorias e subcategorias formais da Teoria do Conceito proposta por Dahlberg (1978).

Este trabalho dá continuidade aos estudos de Silva, procurando relacionar a Teoria do Conceito a outras teorias, principalmente, a Teoria da Linguagem para verificar a compatibilização semântica do vocabulário utilizando uma ontologia de domínio na IA. Os procedimentos da presente pesquisa ultrapassam a sua natureza aplicada. A ampliação de sua abrangência e validação por meio de testes experimentais adicionais para implementação no protótipo MHTX são ações que foram desenvolvidas para viabilizar o estudo exploratório e que podem contribuir em etapas posteriores. A abrangência está relacionada a outras áreas do conhecimento para modelagem de dados e catalogação de recursos de objetos digitais para indexação na fonte e extração de conceitos válidos.

Esta implementação foi subdividida em duas partes: uma conceitual e outra tecnológica. Na conceitual, foram utilizados referenciais teóricos e metodológicos sugeridos por Lima (2004), Silva (2008) e Campos (2001 e 2004) para a construção de um sistema de conceitos e modelagem terminológica, considerando o hipertexto como um sistema interativo composto de objetos informacionais de metadados digitais. Na tecnológica, esses

recursos digitais foram relacionados a métodos e teorias da CC para construção do modelo implementado para o mapa conceitual representado por uma ontologia de domínio construída na IA. As ferramentas usadas na prototipação do modelo de sistema proposto estão de acordo com os padrões de metadados W3C (Protégé<sup>10</sup> e Smore<sup>11</sup>).

### 1.3 Organização e mapa da dissertação

Esta pesquisa está organizada em quatro capítulos principais, incluindo a introdução. O segundo capítulo apresenta os fundamentos históricos, teóricos e metodológicos, contendo conceitos gerais interdisciplinares sobre o tema. O terceiro capítulo apresenta os procedimentos metodológicos para o estudo exploratório que analisa a viabilidade da ontologia como um único mapa com conceitos integrados aos documentos do MHTX. O quarto capítulo apresenta as reflexões decorrentes dos resultados do estudo proposto e sugestões de trabalhos futuros. Por fim, são apresentadas as referências e anexos que contribuíram com valiosas informações. A Figura 1 apresenta um mapeamento dessa organização com objetivo de ajudar a leitura interdisciplinar abordada na dissertação.

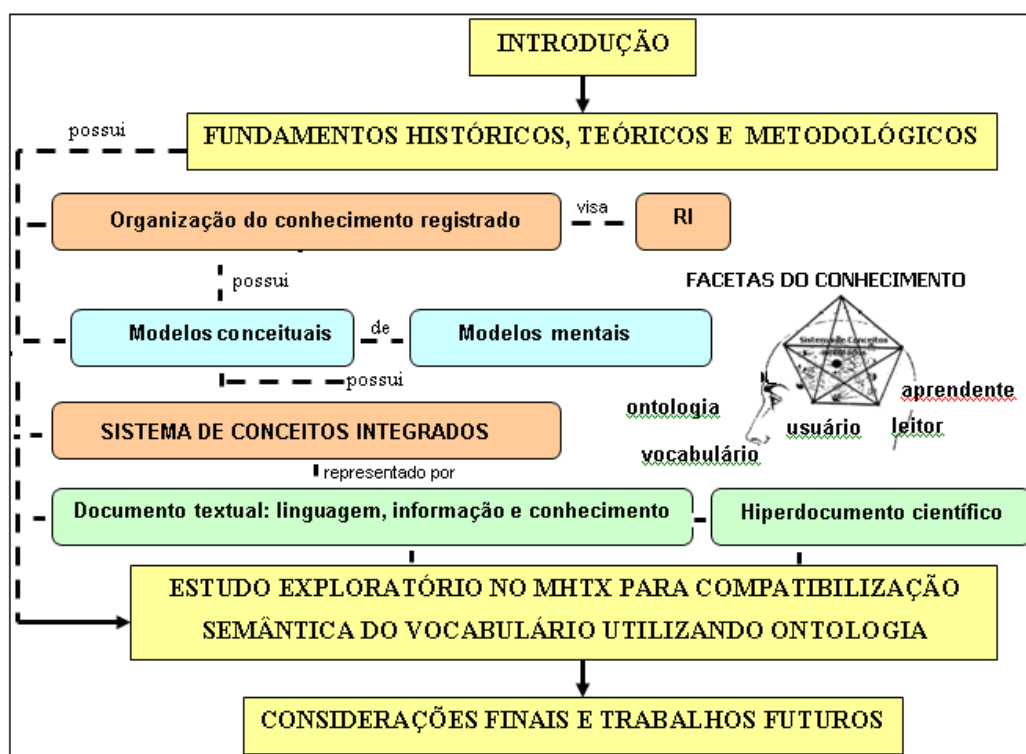


Figura 1 – Mapa da organização da dissertação.

<sup>10</sup> O Protégé e o plug-in OWL (Web Ontology Language) estão disponíveis em <http://protege.stanford.edu>.

<sup>11</sup> A ferramenta Smore está disponível em <http://www.mindswap.org/2005/SMORE/smores50.zip>.

## 2 FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS

A meta central desta pesquisa é um estudo exploratório sobre compatibilização semântica de vocabulário, utilizando uma ontologia de domínio no MHTX. Para a revisão de literatura, faz-se necessário pensar na diversidade de produtos, tais como teses e dissertações que compõem o repositório do MHTX, impondo a necessidade de categorização para depois iniciar a divisão de etapas do trabalho. Esses produtos podem estar vinculados ao PPGCI da ECI-UFMG como uma organização de um campo temático<sup>12</sup> da CI. Com a inserção de uma ontologia da CI na arquitetura do MHTX, busca-se uma visão mais ampla desses produtos como uma biblioteca digital especializada que possa atender a uma vasta rede de publicação desta comunidade científica.

Na tentativa de categorizar esses produtos como “documentos científicos”, esbarra-se em incoerências conceituais sobre o termo documento. Um documento é encontrado ora como equivalente a suporte de dados, ora como equivalente ao meio de comunicação e ora associado ao termo mensagem (GALVÃO, 1998). Então, inicia-se este capítulo procurando esclarecer o complexo universo do vocabulário humano existente na escrita de qualquer texto de um documento, seja ele manuscrito, impresso ou digital.

### 2.1 Documento textual: linguagem, informação e conhecimento

Um documento textual no sentido de informação registrada possui características físicas e intelectuais de objetos informacionais registrados. As características físicas podem ser identificadas pelo tamanho, peso, forma de produção e as características intelectuais incluem tipo, tema, assunto, autoria, difusão, fonte, dentre outras (DIAS, 2001). Como a informação pode estar registrada em diferentes suportes, Robredo (2005) utiliza o termo ‘objetos informacionais’ que podem fazer com que a informação tenha diferentes significações, ao referenciar os respectivos conteúdos de um documento textual.

A informação sempre esteve presente através da técnica e da linguagem, ou seja, da maneira de como escrever sobre determinados objetos informacionais e da forma de transmitir o conhecimento sobre esse fazer (FREIRE, 2006). O vocabulário humano é constituído pelo conjunto de palavras que uma pessoa conhece e se refere ao uso de uma língua em discurso na comunicação. A representação material de externalizar uma palavra recebe o nome de vocábulo e possui um núcleo semântico estrutural que é sua base mórfica ou raiz além de outros elementos periféricos. Por exemplo, uma palavra pode ser composta

---

<sup>12</sup> Um campo temático constitui-se de conjuntos de termos funcionalmente possíveis no interior de um determinado tema ou assunto de um documento cuja organização interna depende de parâmetros emprestados à atividade psicossocial.



apenas pela raiz, como é o caso de “feliz”, ou possuir morfema periférico como “infeliz”. Segundo Azeredo (2008), há quatro tipos de morfemas gramaticais: bases, afixos, vogais temáticas e desinências ou morfemas flexionais.

A ideia fundamental da Teoria da Linguagem é um sistema fechado de signos linguísticos formados pelo conceito e pelo som que estão intrinsecamente relacionados. Assim, a linguagem recorta simultaneamente um significante na massa informe dos sons (palavras pronunciadas) e um conceito (uma ideia) na massa dos conceitos. O significado designa o conceito de um objeto e o significante designa a imagem e a ‘acústica’ da palavra. A relação entre significante e significado é arbitrária (TOUTAIN, 2007).

A língua é um conjunto de convenções necessárias para permitir a prática da linguagem. De acordo com Freire (2006), tanto a técnica, como a linguagem, são desenvolvidas em uma determinada cultura, de acordo com necessidades econômicas e sociais, e irão evoluir de acordo com o contexto em que se encontram. O estudo do significado de cada vocábulo recebe o nome de semântica e possui um tempo determinado pelas circunstâncias históricas de seu uso (VIEIRA e LIMA, 2001).

A semântica estuda o significado das palavras e a pragmática estuda suas relações, relativo a quem usa e ao contexto de uso. Assim, quando nos referimos à palavra enquanto um conjunto de sons, ela é parte da linguagem falada. Quando nos referimos à palavra enquanto um conjunto de letras, ela é parte da linguagem escrita ou gráfica. Quando nos referimos à palavra enquanto índice de uma ideia ou representação imaterial, este aspecto interno e funcional de uma língua recebe o nome de termo.

O léxico pode ser considerado o patrimônio vocabular de uma comunidade linguística através de sua história como um acervo que é transmitido de uma geração para a geração seguinte, traduzindo a experiência cultural acumulada por uma sociedade através do tempo. Por exemplo, o léxico da língua portuguesa é o conjunto de todas as palavras compreensíveis nesta língua e que uma pessoa tem à sua disposição para expressar-se, oralmente ou por escrito. Este conjunto desenvolveu-se a partir da língua tardo-latina com interferências do vernáculo. Foi assim que ela se tornou uma língua dotada de uma tradição literária riquíssima e de variadas características linguísticas em novos territórios descobertos e colonizados pelos portugueses como é o caso do Brasil.

A Linguística deve se ocupar dos vários aspectos da linguagem, entre eles o léxico-gramatical que prevê uma exigência de formalização. Para verificar a validade de sequências de palavras numa certa língua, o sistema natural precisa que a língua seja especificada por um léxico-gramatical. Os resultados da descrição das regras de derivação do léxico devem ser formais o suficiente para que possam ser verificados por conferição

com a realidade do uso e postos em prática no processamento automático através da realização de analisadores sintáticos (DIK, 1997).

Segundo Dik (1997), se não houver qualquer regra produtiva que explique a derivação de um item com forma e significado particulares, sua formação será considerada não produtiva e, portanto, tais itens pertencerão ao léxico de uma língua. Se os itens complexos puderem ser considerados predicados derivados da aplicação de regras de formação e/ou expressão produtivas na língua, então tais itens já não pertencem ao domínio do léxico, mas ao das regras de derivação como um dos componentes da gramática.

Geraldi (2006) destaca que a língua é um processo inacabado. Enquanto produto e condição de produção da história presente, ela vem marcada pelos espaços sociais de seus usos das construções do passado e passou a ser incorporada às análises textuais. Então, com base na compreensão da linguagem como prática social, a linguagem não pode corresponder a um sistema fechado e pronto. Ao lado da fonologia, da morfologia e da sintaxe, a Linguística precisa ficar atenta às mudanças do vocabulário falado e documentado por uma comunidade.

Marcuschi (1998) destaca que, no final da década de 1970, o enfoque deixa de ser somente da competência dos falantes de uma língua natural local. Os estudos linguísticos passaram a incluir a noção de textualidade que estabelecia múltiplas conexões ativadas na comunicação. Assim, pela primeira vez, o texto foi concebido como objeto central de pesquisa, procurando estabelecer um sistema de regras estruturais recorrente dos aspectos transformacionais de uso de uma língua natural local numa visão sistêmica e multifuncional para compreender a linguagem como um “sistema” interativo.

### **2.1.1 A estruturação de um texto científico**

Existe uma interação entre língua e cultura que se entrelaçam, sendo uma dependente da outra para formar o texto de um documento científico como uma dissertação de mestrado ou tese de doutorado. O texto deve convergir em três ações: linguísticas, cognitivas e sociais (MARCUSCHI, 2003). Nessas ações de interatividade e textualidade, tanto a linguagem verbal e não verbal devem funcionar juntas para a função de toda estruturação, ou seja, o conteúdo do documento na sua visão mais lógica de texto completo.

Todo texto se constrói segundo alguns princípios tipológicos. A tipologia mais comumente utilizada é narração, descrição e informação. Dentre os textos informativos, pode-se reconhecer o texto científico e, nesse tipo de texto, o conteúdo é quase inteiramente determinado pelo autor por ser construído numa estrutura convencional que reflete o arranjo de composicionalidade textual. Tais arranjos apresentam-se como

esquemas formais de organização de ideias, a partir dos quais se podem identificar traços básicos que os caracterizam (NAVES, 2000).

A estruturação de um texto científico exige domínio de uma língua culta com técnicas de redação, normalização e gramaticalização. Na redação de um texto científico, a escrita de forma bem estruturada, com suas citações e referências seguindo regras estabelecidas por uma norma, confere a ele um caráter de cientificidade e confiabilidade, fundamental para a seriedade do trabalho proposto (MARCONDES, 2009). Para isso, a linguagem verbal utilizada deve ser formal para descrever objetos informacionais, narrar fatos e ocorrências, dissertar sobre as facetas de um tema, explicar descobertas, argumentar soluções de problemas ou refletir sobre teorias, hipóteses e pressupostos.

Para Curty e Boccato (2005), a normalização não tem o propósito de limitar a criatividade e a liberdade dos autores de um documento, mas, sim, o de facilitar aos diferentes leitores das diversas culturas o acesso às suas ideias e concepções científicas. Dessa forma, a linguagem utilizada pelo autor deve ser precisa, pautada de clareza e coerência pelo uso da norma culta escrita, permitindo a quem o lê reproduzir o trabalho realizado, tal qual ele foi realizado pelo autor.

Para compor o texto completo de uma tese ou dissertação, o autor precisa ler e (re)ler, escrever e (re)escrever, estruturar e (re)estruturar o vocabulário diversas vezes até atrelar uma informação à outra para obter um todo organizado e facetado. O texto científico também pode ser enriquecido com elementos não verbais como o uso de imagens, por exemplo. Uma imagem possui uma linguagem própria, ou seja, um conjunto de convenções quanto à forma, arranjo e associação, que permite combinar traços isolados em uma unidade semântica (SIMÕES, 2008).

Documentos como teses e dissertações possuem relatos de experiências em busca de inovações e soluções científicas. Nesse sentido, o texto científico não pode ser limitado a um texto qualquer, pois seu alcance pode ser ampliado para transmitir muito mais que informação, ou seja, possibilita transmitir conhecimento ao leitor. A estrutura desse tipo de documento pode ser considerada como um conjunto de elementos multifuncionais que formam um todo ordenado e com princípios lógicos de unidade, coerência e coesão.

Martins (2004) destaca que é consenso que a coesão e a coerência estão intimamente relacionadas aos processos de produção, compreensão e interpretação. Contudo, a coesão não garante a coerência, pois é preciso contar com conhecimentos exteriores ao texto. A coerência depende tanto do conhecimento do leitor quanto do autor além da situação em que o texto é produzido. Por isso, a coerência deixou de ser vista como mera propriedade ou qualidade do texto, e passou a ser vista ao modo como o leitor-interpretador interage com o texto e o reconstrói como uma configuração veiculadora de

sentidos. Essa interação na interlocução que promove a intertextualidade caracteriza o diálogo texto-sujeitos, estabelecendo entre si múltiplas relações.

Para Naves (2000) é mediante a interação de diversos níveis de conhecimento que o leitor consegue construir o sentido do texto. Durante o processo de leitura, a compreensão se caracteriza pela utilização do conhecimento prévio que se refere ao estoque de conhecimento armazenado na memória do leitor, que se configura como uma espécie de “quadro de referência” formado por uma rede multidimensional de unidades conceituais. Nesse estoque de conhecimento, incluem-se referências a entidades linguísticas (estoque de palavras – vocabulário – frases, forma de organização textual, tipologias textuais) e conceituais (CINTRA, 1983).

Nesse complexo universo de uso do vocabulário humano para formar um texto completo, formalizam-se transformações constituídas por grandezas-signos que constituem processos de produção de significação e informação, por meio dos quais preponderantemente diversos elementos podem ser combinados em campos para formar um conjunto organizado de objetos-em-ação. Esses objetos podem ser estruturados por uma palavra-chave com um denominador semântico comum que se delimitam reciprocamente, e são delimitados pelos elementos periféricos de outros campos do conhecimento que podem ser agrupados em torno de uma noção-chave.

### **2.1.2 Campo conceitual para estruturação de vocabulários**

O vocabulário de um texto de um documento pode combinar uma série de elementos representados por uma palavra-chave de um campo conceitual para formar um “sistema de conceitos”. Esse sistema deve ser representado por um conjunto de termos funcionalmente possíveis no interior de uma determinada situação temática em discurso e cuja organização interna depende de parâmetros emprestados à atividade psicossocial para que sejam estruturados em um sistema de campos conceituais de um domínio específico.

Um campo conceitual refere-se à organização estrutural de uma área conceitual por um sistema linguístico particular, partindo do pressuposto de que se encontra uma substância de significado não estruturada subjacente ao vocabulário de todas as línguas. Este campo distingue-se de campo lexical, que designa especificamente o conjunto de lexemas de que uma língua se serve para abranger e estruturar uma área conceitual, em função das relações de sentido existentes entre os mesmos. Assim, é possível falar no campo conceitual das cores, que linguisticamente se realiza num campo lexical que inclui lexemas como "preto", "branco", "azul" (LYONS, 1977).

A noção de campo foi introduzida na chamada Teoria de Campo e no Dicionário de Termos Linguísticos<sup>13</sup> referindo-se ao vocabulário enquanto sistema de cadeias parciais articuladas entre si. Isso reforça o fato de existir uma estreita relação entre os diversos elementos que podem ser agrupados em torno de uma palavra-chave de um campo conceitual na escrita de um texto de um documento, tais como:

- **campo lexical:** refere-se ao conjunto de lexemas que, organizados em função das relações de sentido existentes entre si, abrangem uma determinada área de significação, estruturada num campo conceitual;
- **campo semântico:** refere-se a um conjunto de lexemas ou outras unidades linguísticas que se encontram ligadas semanticamente;
- **campo de aplicação:** refere-se ao domínio de atividade social no qual um item lexical é utilizado numa acepção particular;
- **campo associativo:** designa um conjunto de unidades lexicais que apresenta uma semelhança particular de formas e sentidos entre os seus componentes. Este campo está relacionado a fatores afetivos, intelectuais, culturais, e ao domínio da experiência de cada indivíduo;
- **campo morfológico:** baseia-se nas analogias no plano do significante (semelhanças formais). Assim, as palavras que têm o mesmo prefixo ou sufixo ou radical pertencem a um mesmo campo morfológico;
- **campo morfossemântico:** combina relações de forma (significante) e de sentido (significado), numa dupla perspectiva sincrônica e diacrônica agrupando em torno de uma determinada palavra a totalidade dos seus derivados e compostos, tendo em conta expressão e conteúdo.

Naturalmente, todos esses campos estão inter-relacionados e um pode ser importante para a compreensão do outro. A Teoria dos Campos Conceituais de Gérard Vergnaud considera útil falar em distintos campos conceituais se eles puderem ser consistentemente descritos. Nesse sentido, para formar um texto completo de um documento científico, diversos conceitos podem ser estruturados como se o conteúdo fosse entrelaçado por ideias conectadas à análise conceitual de um domínio e transformadas em grandezas-signos. Para Vergnaud um campo conceitual é definido, em primeiro lugar, como um conjunto de situações cujo domínio requer, por sua vez, o domínio de vários conceitos, procedimentos e representações de naturezas distintas (MOREIRA, 1999).

Entre os fatores determinantes deste conjunto de situações, originou-se a função da própria Terminologia que veio da necessidade manifestada pelos especialistas de

<sup>13</sup> [http://www.ait.pt/recursos/dic\\_term\\_ling/dtl\\_pdf/C.pdf](http://www.ait.pt/recursos/dic_term_ling/dtl_pdf/C.pdf)

organizar as denominações de seus “sistemas de conceitos” e se estabelecer uma comunicação unívoca dentro de um domínio do conhecimento (CABRÉ, 1999). Marcuschi (2003) lembra que o ser humano é um ser que fala e não um ser que lê e escreve naturalmente. Apesar de a leitura e escrita serem um bem social indispensável, para que os indivíduos possam interagir, tanto o autor, quanto o leitor enfrentam dificuldades nos processos de formação de conceitos para organizar um texto completo.

O objetivo da Terminologia é identificar, em um domínio, primeiramente os conceitos para depois lhes atribuir um termo, controlando o vocabulário e, ao mesmo tempo, dando a cada conceito uma definição rigorosa e normalizada para ser representada em um sistema. A unidade de estudo do modelo ideológico textual é o evento de letramento, isto é, situações que são praticadas na escrita de qualquer texto. Essas práticas exigem diversas situações e relações conceituais na escolha do vocabulário mais adequado, constituindo parte essencial de uso de uma língua em relação à interação entre os participantes e as estratégias interpretativas para dar sentido à situação e ao texto completo.

A relação entre oralidade e letramento era considerada dicotômica, atribuindo-se à escrita valores cognitivos intrínsecos no uso da língua, não se vendo nelas duas práticas sociais. Entretanto, hoje, predomina a posição de que se pode conceber oralidade e letramento como atividades interativas e complementares no contexto das práticas sociais e culturais (MARCUSCHI, 2003). Vergnaud considera essas atividades como esquema-situação para formar conceitos-em-ação e teoremas-em-ação em torno de um campo.

Esquemas têm como ingredientes essenciais conceitos-em-ação e teoremas-em-ação. Teorema-em-ação é uma proposição considerada como verdadeira sobre o real; conceito-em-ação é uma categoria de pensamento tida como pertinente. Segundo Vergnaud um conceito torna-se significativo através de uma variedade de situações, mas o sentido não está nas situações em si mesmas, assim como não está nas palavras nem nos símbolos. O sentido é uma relação do sujeito com situações e significantes. Mais precisamente, são os esquemas, as ações e sua organização, evocados no sujeito por uma situação ou por um significante que constituem o sentido dessa situação ou desse significante para esse indivíduo (MOREIRA, 1999).

Ao transformar informação potencial contínua da experiência em significação, as grandezas-signos tornam-se elas mesmas depositárias de potencial de informação, concernente aos valores autorais ou comunidades do grupo que a criou. Essas práticas sociais surgem da necessidade de analisar, classificar e organizar o potencial de informações dos dados da experiência de indivíduos para elaborar e sustentar seus

sistemas de valores, sua visão de mundo, linguística e semiótica<sup>14</sup> construída (BARBOSA, 1998). Para isso, a Terminologia deve refletir o conceito aplicado a um termo puro, traduzível e perfeitamente coerente para ser representado em um sistema construído e estruturado de maneira unívoca e explícita como teoremas-em-ação (CINTRA, 1983).

O autor de um texto científico tem dificuldades em explicar conceitos-em-ação, mas isso não significa que tal conhecimento não possa ser explicitado. Através desses processos de explicitação é que teoremas-em-ação e conceitos-em-ação se formam verdadeiros teoremas e “conceitos científicos”. Essas dificuldades de formação de conceitos relacionada aos processos de ler e escrever para formar um texto lembram as investigações introduzidas por Noam Chomsky com o tema “gramática generativa transformacional”, a partir de 1965, e que fomentou um vasto conjunto de pesquisas científicas de visão multidisciplinar das Ciências Cognitivas e Linguísticas; bem como as diferenças entre conceitos espontâneos e científicos apresentadas em diversas obras de Vygotsky.

### **2.1.3 A gramática generativa transformacional e a formação de conceitos espontâneos e científicos em um campo conceitual**

A visão de gramática generativa transformacional proposta por Chomsky se caracteriza pela tentativa de estabelecer relações conceituais entre o uso da linguagem e o raciocínio lógico do pensamento humano. Chomsky julgava que os dados nunca falariam por si só. Por isso, deve-se tentar entender que tipo de “sistema” é a linguagem e expor as conclusões em termos de um sistema formal. Tal análise deve levar à postulação de regras que possam explicar a produção de qualquer sentença gramatical concebível e, ao mesmo tempo, essas regras não devem “gerar” nenhuma sentença incorreta ou agramatical. Uma vez estabelecido este “sistema”, deve-se examinar expressões particulares para determinar se podem ser adequadamente geradas através da adesão às regras sistêmicas por vários processos integrativos estruturais da cognição humana (GORSKI, 2007).

Chomsky acreditava que a complexidade da linguagem era própria do cérebro humano e seguia regras e princípios que transcendiam todos os povos. Como se as pessoas fossem dotadas de uma “faculdade de linguagem” altamente específica que Steven Pinker, neurolinguista, prefere chamar de “instituto”, procurando refletir sobre o modo de como a “mente” cria a linguagem. Para Pinker este “instituto” é composto de muitas partes e alerta que a limitação da base de conhecimento humano está relacionada à falta de um estudo de lógica matemática e dos níveis de linguagem propriamente dita, como sintaxe,

---

<sup>14</sup> A Semiótica estuda os signos e a forma como eles funcionam na produção de significados relacionando qualquer marca, símbolo ou sinal à construção de um signo usado para indicar e veicular pensamentos e informação.

semântica e pragmática. A sintaxe funciona como seu sistema combinatório discreto que constrói as estruturas sintagmáticas. A morfologia funciona como um segundo sistema combinatório que constrói palavras em um espaço léxico, dando um tratamento vocal renovado com regras e estruturas fonológicas, percepção da fala, algoritmos de análise e algoritmos de aprendizagem. (GORSKI, 2007).

Assim, a gramática impõe injunções de ordem prática dentro da sociedade para uso sistematizado de controle e estruturação do vocabulário em campos conceituais para organizar o contexto de um documento científico. Essas concepções estão presentes nas expressões simbólicas do sujeito, registradas em um conjunto de informações que são praticadas numa dada situação-problema. Para Vergnaud, o conhecimento está organizado em “campos conceituais”, que se desenvolvem no sujeito, num certo período de tempo, por meio do pensamento, experiência, maturação e aprendizagem (MOREIRA, 1999).

Essas competências e habilidades começam a se desenvolver a partir das primeiras ações do sujeito, espontaneamente, quando está inserido numa dada situação-problema como, por exemplo, na leitura e interpretação de “conceitos científicos”. Para Vygotsky (1998), conceitos científicos fazem parte de um sistema lógico e organizado para um indivíduo letrado que está pronto para ler, interpretar e observar fenômenos naturais, passando a um ser crítico de suas práticas sociais e questionador de teorias, hipóteses e pressupostos. A análise, a interpretação e a descrição dos mecanismos de formação de conceitos, segundo uma abordagem sócio-cultural-cognitiva das obras de Vygotsky coloca em evidência as relações existentes entre conceitos espontâneos e científicos.

Nas obras de Vygotsky, a história da sociedade e o desenvolvimento do ser humano estão totalmente relacionados com a linguagem e suas complexas inter-relações com o pensamento. Para Vygotsky, as crianças, desde que nascem, têm constante interação com os adultos e, através deste contato, os processos psicológicos mais complexos vão se ligando socialmente com sua cultura. No início, esses processos são intersíquicos, passando posteriormente para intrapsíquicos. A fala adquire multifunções e conceitos espontâneos que são gerados nas situações concretas de relacionamentos, desenvolvidos no cotidiano com as práticas e interações sociais (OLIVEIRA, 1997).

Este sistema lógico-sócio-cultural-cognitivo tem uma peculiaridade que precisa ser observada, que é a posição da palavra e do conceito que rege a organização das relações léxico-gramaticais. Vygotsky deixa claro em todas as suas obras que os conceitos científicos e espontâneos se desenvolvem dinamicamente em direções contrárias e, gradativamente, os conceitos espontâneos vão evoluindo até que eles possam se encontrar quando o indivíduo possui certa autonomia para direcioná-los. Esses relacionamentos, naturalmente hierárquicos são muito semelhantes aos arranjos de um sistema de conceitos.



Essa visão reforça que os sujeitos, agentes de sua própria inteligência, podem possuir dois sistemas de formação conceitual, um baseado em categorias difusas e o outro em conceitos clássicos, ou logicamente definidos e classificados em taxonomias<sup>15</sup> (POZO, 2005).

Os processos de categorização e classificação são muito semelhantes. A diferença básica entre os dois processos está na ideia de que na classificação, o objeto é alocado em uma única classe, enquanto que na categorização o objeto pode estar associado a várias categorias (LOH, 2001). Esses princípios são os mais comuns de uma taxonomia. Por isso, existem diferentes técnicas de estruturação e organização do conhecimento para representar objetos e dispor os termos em qualquer tipo de lei ou princípio de um sistema classificatório.

Esses processos intelectuais por meio dos quais objetos passam a ser analisados em um domínio, podem ser considerados como uma abstração e generalização perceptiva espacial, estrutural, temporal, entre outras (BASTOS, 2005). Uma abstração e generalização é uma operação mental que consiste em estender a toda uma classe de objetos ou fenômenos os elementos essenciais, gerais, universais, constatados num certo número de entidades ou fenômenos da mesma classe com características comuns (OLIVEIRA, 1997). Além disso, quanto mais abstrato e genérico for um conceito, ele poderá ser representado em um termo puro (DAHLBERG, 1978).

Para Vergnaud, há uma relação dialética entre conceitos-em-ação e teoremas-em-ação, uma vez que conceitos são ingredientes de teoremas e teoremas são propriedades que dão aos conceitos seus conteúdos. Moreira (1999) alerta que seria um erro confundir conceitos-em-ação e teoremas-em-ação porque conceitos não são teoremas. Conceitos em ação são ingredientes necessários das proposições. Proposições podem ser verdadeiras ou falsas. Mesmo assim, não existem proposições sem conceitos assim como também não existem conceitos sem proposições.

Conceitos podem ser apenas relevantes ou irrelevantes. Um teorema-em-ação é uma proposição tida como verdadeira sobre o real e o conceito-em-ação é um objeto, um predicado, ou uma categoria de pensamento tida como relevante. Uma proposição explícita pode ser debatida, uma proposição tida como verdadeira de maneira totalmente implícita, não. Assim, o caráter do conhecimento muda se for comunicável, debatido e compartilhado por uma comunidade científica. Cada comunidade científica possui seu próprio vocabulário e, antes de se lançar a publicar um documento, o autor precisa apropriar-se do vocabulário

---

<sup>15</sup> Embora as taxonomias tenham suas raízes nos trabalhos de Aristóteles, Linnaeus e Darwin, o significado deste termo tem sido amplamente usado para outras propostas na CI e CC, visando a organização, armazenagem e recuperação de informação no que se refere aos sistemas de classificação e representações do conhecimento como instrumentos que organizam logicamente os conteúdos informacionais (CARLAN, 2010).

de sua área de interesse. Isto significa ter acesso à interpretação e compreensão compartilhada em um domínio com uma terminologia aceita na comunidade científica, implicando em tarefas complexas mesmo para um indivíduo letrado.

Esses procedimentos de naturezas distintas conceituais são práticas sociais que exigem diversos aspectos multifuncionais linguísticos para formar um texto completo que ajudam a contornar dificuldades conceituais de interpretação e compreensão de forma a adequar, organizar e estruturar o vocabulário dentro de uma determinada situação de uso e contextualização. De fato, os processos ler, (re)ler, escrever e (re)escrever sem dúvida são algumas das melhores maneiras de aumentar o vocabulário. Por meio deles diversos elementos podem ser combinados em campos para formalização e transformações de grandezas-signos. Esses processos mecânicos e intelectuais são necessários para dar sentido ao texto completo e formar um todo organizado de objetos-em-ação, ou melhor, objetos-informacionais-em-ação.

#### **2.1.4 Aspectos multifuncionais da linguagem na escrita textual**

Os aspectos multifuncionais da linguagem natural humana podem ser entendidos como um elemento integrante de um “sistema” como uma enorme rede de opções inter-relacionadas e de significado paradigmático potencial e não linear (SIMÕES, 2006). Esses aspectos têm sido estudados por diversos pesquisadores, procurando, explicar processos e meios pelos quais as pessoas se comunicam entre si, uniu pesquisadores para tentar explicar as diversas potencialidades da linguagem, incluindo visões léxico-gramaticais do Português culto falado no Brasil. Essas pesquisas procuram atribuir os aspectos sistêmico-funcionais num sentido suficientemente amplo para abarcar vários tipos de interatividade na escrita textual (CASTILHO, 2007).

Esses pesquisadores deixaram de aderir à aplicação de uma teoria única da linguagem e passaram a operar com princípios de variada ordem como nas ideias da Linguística Textual de Halliday com a gramática sistêmico-funcional e contextualizada em práticas sociais, situacionais e culturais. Halliday (1994) enfoca a natureza sistêmico-funcional e interativa da linguagem, considerando o texto como um complexo de significados ideacionais, interpessoais e textuais. A gramática é o mecanismo linguístico que liga essas metafunções a determinadas situações de percepção informacional da sentença em uma situação de relações significativas numa forma estrutural unificada para realizar a troca de ação e informação (CASTILHO, 2007).

A metafunção ideacional encontra-se voltada ao campo do discurso, abrangendo o conhecimento e experiências de mundo dos falantes/escritores. A metafunção interpessoal enfatiza as relações entre os participantes e seus

posicionamentos. A metafunção textual refere-se ao modo e tessitura da mensagem, ou seja, é por meio da textual, que as outras duas metafunções, ideacional e interpessoal, se presentificam para trazer coerência a um texto oral ou escrito, possibilitando a construção de mensagens em um sistema temático. As variáveis de registros (campo-relações-modo) dentro de cada metafunção (ideacionais-interpessoais-textuais) vão expressar as escolhas semânticas de um sistema temático para formar o conteúdo de um texto completo.

De acordo com Halliday (1994), este sistema é que dá ao texto completo o caráter de mensagem informacional em seus contextos e usos. Essa estrutura pode ser dividida em duas partes: o tema e o rema. O tema é o ponto de partida da mensagem, enquanto que o rema é o elemento em que o tema é desenvolvido. Dessa forma, na metafunção ideacional, um campo se associa aos significados por meio dos padrões de transitividade da gramática. As relações exploram o significado da oração como troca para interação entre os participantes do discurso. O modo de organizar e estruturar o texto é que dá a coerência ao sistema temático.

Os padrões de transitividade envolvem processos material, comportamental, mental, verbal, relacional e existencial. Para Vygotsky, a linguagem simplifica e generaliza a experiência humana, ordenando instâncias do mundo real em categorias conceituais cujo significado é compartilhado pelos usuários dessa linguagem desde a fase de uma criança até a fase de um adulto. Os relacionamentos hierárquicos ou esquemas de complexidade conceitual (generalização/especialização) ocorrem quando, a partir da observação de duas classes, se abstrai delas uma classe ou uma categoria mais genérica.

Para formar o conceito de algo, faz-se uso de quatro processos mentais inseparáveis: análise, síntese, abstração e generalização. A análise é a operação mental que consiste em separar em partes, decompor, fragmentar um todo (objeto ou fenômeno) em seus elementos constituintes. A síntese é a operação mental que, ao contrário da análise, consiste em recompor um todo a partir de seus elementos constituintes, a fim de compreendê-lo em sua totalidade (OLIVEIRA, 1997).

Analisar o gênero sob os aspectos multifuncionais da linguagem é entender o texto completo de um documento como prática social, vendo o discurso em diferentes contextos situacionais e culturais. A linguagem funciona como um “sistema” mediador que se organiza na forma de rede de escolhas léxico-gramaticais (HALLIDAY, 1994). Para que essa análise seja bem sucedida é fundamental o conhecimento de gramática normativa e descritiva. A gramática prescritiva ou normativa trata de uso de padrões de uma língua e a gramática descritiva se propõe a descrever essas regras de como uma língua é realmente falada, a despeito da prescrição “correta” do vocabulário (PERINI, 2006). Por isso, a gramática normativa não pode se anular diante da gramática descritiva.

Luft (2003) destaca que fazer uso de uma língua, ou de uma modalidade, ou de um nível de língua exige o conhecimento essencial de sua respectiva gramática. A gramática usa um sistema de regras, unidades e estruturas que o autor/escritor/falante de uma língua tem programado em sua memória e que sustenta as práticas documentárias de leitura e interpretação de línguas com suas diversas possibilidades transformacionais. Os pontos básicos dessa visão refletem o texto como a unidade maior de funcionamento da língua, em cujo interior se dá a criação dos significados com os seus componentes multifuncionais, atuando em níveis diferenciados (CASTILHO, 2007).

As ligações léxico-gramaticais prevêm uma exigência de formalização que pode expressar significados distintos para compor um texto completo de um documento dotado de unidades linguísticas e seus aspectos multifuncionais. Castilho (2007) destaca que, a partir dos anos 80, diversos linguistas retomaram os estudos sobre a gramaticalização do português falado no Brasil, incluindo aspectos cognitivistas e pragmáticos. Esses estudos entendem por gramaticalização a alteração das propriedades de um item lexical ser modificado em suas propriedades sintáticas, morfológicas, fonológicas e semânticas.

Os princípios de composicionalidade de conteúdo de um texto científico completo (dado, informação, conhecimento) revelam a importância das relações entre os elementos léxico-gramaticais de objetos informacionais registrados, tanto no tempo real quanto no tempo aparente, ao combinar elementos na escrita de um texto. Os registros descrevem o contexto situacional no qual o texto é produzido combinando elementos de campo, relação e modo. Esses elementos podem se agrupar em distintos campos conceituais como uma unidade 'neutra', apta, portanto, a se adequar a diferentes padrões culturais em que a rede lexical forma o seu universo semântico estrutural de relações intertextuais (SIMÕES, 2006).

Quando essas combinações ligam elementos textuais, tem sido usual denominá-las "relações semânticas", mas quando não é necessária a distinção, usa-se o termo "relações semântico-lexicais", podendo ser paradigmáticas ou sintagmáticas. As relações paradigmáticas associam palavras através do significado, como "nadar" e "água". As sintagmáticas conectam palavras que são frequentemente encontradas no mesmo discurso, como "água" e "poça" (LYONS, 1977). Alguns exemplos podem ser:

- **sinonímia e antonímia:** A relação de sinonímia entre duas palavras é dada entre seus sinônimos, ou seja, palavras que possuem o mesmo sentido e podem ser substituídas em uma expressão sem alterar o significado da mesma. A relação de antonímia é a relação de contraste entre duas palavras que possuem sentidos opostos, ou seja, as palavras exprimem ideias

contrárias. Exemplos: as palavras “assunto” e “tema” possuem relação sinonímica e a palavra “fácil” possui uma relação antonímica com “difícil”;

- **hiperonímia e hiponímia:** A relação de hiperonímia é a relação entre dois lexemas em que um denota uma generalização do outro. A hiponímia se dá entre dois lexemas, onde o significado de um está contido no significado do outro, ou seja, quando um termo denota ser subclasse do outro. Exemplos: o termo animal apresenta relação hiperonímica para generalizar todas as espécies de animais, sejam eles racionais ou irracionais. O termo pessoa possui uma relação hiponímica com o termo animal racional;
- **meronímia e holonímia:** A meronímia é a relação semântica entre dois termos, quando um é identificado como parte do outro. Esta relação pode ser entendida como uma relação “é parte de”. A holonímia é a relação que ocorre entre dois lexemas quando um contém o outro entre as partes que o constituem. Pode-se dizer que a holonímia é a relação inversa à meronímia, por exemplo, o pé possui uma relação meronímica com a perna. O pé e a perna possuem uma relação de holonímia com o corpo, pois o corpo é formado por pé, perna, braço e outros membros.
- **homonímia e polissemia:** A homonímia é a relação entre duas ou mais palavras que possuem significados distintos, mas pertencem à mesma estrutura fonológica. Por exemplo, conserto, concerto. Por polissemia entende-se a qualidade que faz com que uma palavra possa ter diversos sentidos no contexto de utilização. Por exemplo, a palavra ‘braço’ possui sentido diferente nas expressões: “O braço do rio” e “O corpo é formado por braço, perna e outros membros”. No caso, da palavra “braço”, o que era inicialmente uma metáfora, na realidade, passa a ser mais um caso de polissemia, principalmente se a palavra ficar de uso corrente na língua. Por exemplo, “o braço do rio” no uso corrente na língua portuguesa adquire o sentido de “afluente”.
- **metonímia e metáfora:** A metonímia tem uma função referencial ao permitir usar uma entidade para representar outra. Exemplo: “O Planalto Federal não se pronunciou.” A metáfora tem muito mais que a função de compreensão, ao conceber uma coisa em termos de outra por isso ela foi citada no final da lista de exemplos de relações semânticas lexicais. Como existe uma vasta literatura especializada sobre este assunto, apresento nas próximas linhas uma breve reflexão de acordo os propósitos deste trabalho.

Nos últimos anos, a metáfora vem sendo estudada como um poderoso instrumento do “pensamento complexo” como um mecanismo conceitual de interpretação, comparação, translação e inter-relações. Lakoff e Johnson (1980) mostram que a metáfora não consiste apenas num recurso literário, assumindo antes uma função metafórica e analógica, fundamental nos aspectos cognitivos por operar nos conceitos sistematicamente organizados e refletidos no uso da língua. Nessa concepção, a linguagem sofre uma interação mútua com as restantes capacidades cognitivas como a percepção, categorização, memória para estabelecer relações de correspondência entre elementos distintos, ou seja, ao mesmo tempo em que a linguagem serve para integrar conhecimento, ela também é responsável pela estruturação do pensamento.

Castilho (2007) nos lembra que a referenciação e a predicação constituem duas atividades fundamentais na organização das sentenças quando denominamos um participante via referenciação e lhe atribuímos uma propriedade via predicação. Por exemplo, podemos com o signo de herói denominar uma criança ou selecionar no panteão dos heróis da pátria alguém que mereça essa denominação ou ainda, um grande patife, quando queremos ironizar. As investigações têm mostrado entre outras coisas que, em nosso recorte do mundo físico, podemos destacar as propriedades íntimas de referência de um objeto (intensionalidade), sua capacidade de remeter a conjuntos e indivíduos desses conjuntos (extensionalidade), também nossa habilidade de apresentar “coisas” especificamente definidas (especificidade) ou genericamente indefinidas (genericidade).

Essas diferentes formas de representações têm sido um grande problema para a Semântica. As metáforas presentes na língua são manifestações de conceitualização, ou seja, a forma como entendemos e representamos determinados conceitos. Trata-se de uma operação cognitiva que emprega dois domínios conceituais diferentes: um domínio experiencial mais concreto, estreitamente ligado à experiência e vivência para compreender um domínio abstrato, cuja natureza não permite uma representação direta e estática e sim, dinâmica com diferentes dimensões, que, sendo recorrentes e co-ocorrentes, geram as metáforas. Nesta operação, ligam-se elementos imagéticos ou mapeamentos conceituais que funcionam como correspondências ontológicas entre domínios (LIMA, 2001).

Lakoff (1987) usa a palavra “viagem” como exemplo prototípico de esquema imagético metafórico, consistindo de elementos de integração origem-percurso-destino. Esses elementos se agrupam em distintos campos conceituais para se adequar a diferentes padrões culturais de relações intertextuais. Ao contrário da metáfora conceitual, em que os elementos fazem relações de similaridade, a metonímia estabelece uma relação de co-ocorrência. Os exemplos de relações semântico-lexicais mostram a ambiguidade do vocabulário humano que pode expressar significados distintos de acordo com os princípios

de composicionalidade de um texto representados em esquemas, em sua maioria de natureza espacial, funcionando como “modelos cognitivos idealizados<sup>16</sup>” (CASTILHO, 2007).

Segundo Miranda e Simeão (2002), os percursos gerativos de codificação e decodificação dos elementos textuais de um documento permitem transmissão de conhecimento ao leitor quando este consegue adentrar nas estruturas composicionais do texto, conhecendo seus meandros, descobrindo as ideologias que o permeiam e a visão de mundo do seu autor. Isso exige diversas estratégias de leitura e interpretação para compreender os elementos léxico-gramaticais em uso de uma língua local e, ao combinar esses elementos, o autor pode, naturalmente, utilizar um vocabulário ambíguo, transferindo o problema de compreensão, tradução, interpretação para o leitor.

### **2.1.5 A ambiguidade do vocabulário humano e os desafios do leitor-interpretador de conceitos ou termos desconhecidos**

O texto de um documento não se caracteriza como um aglomerado de palavras ou expressões que se distribuem aleatoriamente. Para formar o sentido de um texto completo, uma informação precisa se atrelar à outra em várias partes menores a fim de se obter um todo inter-relacionado. Isso exige estratégias de leitura que são postas em ação em situações concretas de interação social. Segundo Koch (2004), essa “inter-textualidade” é um recurso argumentativo do autor que pode estar explícito ou implícito, mas que requer estratégias para ativação do texto-fonte na mente do leitor no intuito de orientar a compreensão daquilo que se lê.

A leitura de um documento é como se o leitor estivesse conversando consigo mesmo como na tela do computador ou na leitura de um livro impresso em que o autor não está presente para esclarecer palavras desconhecidas ou termos ambíguos. Bräscher (2002) cita diferentes tipos de ambiguidade que podem ocorrer durante a produção de um documento e, muitas vezes, o autor nem percebe a existência de termos sinônimos que possam confundir o leitor na interpretação de conceitos ao escrever sobre determinado conteúdo experiencial ou expressar suas ideias. Alguns tipos de ambiguidade podem ser:

- **ambiguidade lexical:** ocorre quando há mais de uma interpretação possível do significado de uma unidade lexical. Este tipo de ambiguidade é provocado por homografia e polissemia;
- **ambiguidade sintática:** ocorre na estruturação da expressão em constituintes hierarquizados, quando se definem as ligações que se estabelecem entre os sintagmas. As expressões preposicionais são uma das

---

<sup>16</sup> Segundo Castilho (2007), os modelos cognitivos idealizados são construções conceituais destinadas a enquadrar situações, processo mediante o qual formulamos nossa compreensão do mundo, consolidando as categorias que o descrevem e fixando o semantismo das expressões da língua.

fontes mais frequentes de ambiguidade sintática como, por exemplo: “a professora de dança espanhola” pode ser interpretado como a professora é espanhola ou a dança é espanhola;

- **ambiguidade predicativa:** ocorre na interpretação das relações temáticas que articulam predicado, argumentos e participantes. Exemplo: “a crítica deste autor” o autor pode ser o objeto da crítica ou o agente da crítica;
- **ambiguidade morfológica:** ocorre quando não é possível classificar determinada forma quanto à categoria gramatical. Este tipo de ambiguidade é ocasionado pela policategorização de palavras pertencentes a mais de uma categoria gramatical como substantivo, adjetivo ou verbo;
- **ambiguidade semântica:** ocorre quando há mais de uma interpretação possível para o relacionamento dos termos, como, por exemplo, no cálculo dos operadores de negação e de quantificação: “ela não chora mais porque ele partiu” pode ser interpretada como ela chorava porque ele havia partido ou ela parou de chorar uma vez que ele já havia partido;
- **ambiguidade pragmática:** relaciona-se ao cálculo dos valores enunciativos, à reconstrução destes valores, que estão ligados à situação do falante no momento da enunciação, por exemplo: “animais de dois pés” pode ser uma referência geral ou específica.

Tradicionalmente, a ambiguidade deve ser evitada durante a escrita de um documento científico procurando clareza, objetividade e concisão na exposição das ideias. Entretanto, a necessidade de considerar o conhecimento intuitivo de uma pessoa na construção do sentido global do enunciado e no estabelecimento das relações semânticas lexicais entre frases, e o fato de vínculos coesivos não assegurarem unidade ao texto conduziram novas pesquisas na Linguística Textual.

Leffa (1996) alerta que o autor tem bem claro em sua mente o que deseja expressar e muitas vezes não percebe a existência de ambiguidade passando o problema para o leitor-interpretador, para quem teoricamente três soluções são possíveis: a) a ambiguidade é resolvida pelo leitor por algum mecanismo desambiguador, usando dados que estão além da frase; b) a ambiguidade não é resolvida pelo leitor, com prejuízo da compreensão do texto; e c) a ambiguidade é resolvida erroneamente pelo leitor, com a adoção do significado não pretendido pelo autor.

Uma das propostas deste trabalho é que o novo sistema do MHTX ajude o leitor-interpretador a solucionar problemas de ambiguidade do vocabulário, funcionando como um mecanismo artificial desambiguador de termos sinônimos evitando as duas últimas



alternativas alertadas por Leffa no parágrafo anterior. Por exemplo, na CI usa-se o conceito “análise de assunto” representado por outros termos tais como “análise documentária”, “análise temática”, entre outros. Todos esses termos são sinônimos, e representam o mesmo conceito. Assim, o mapeamento proposto poderá indicar todos os termos sinônimos compatíveis do domínio da CI. Esse “mapa” precisa ter o léxico organizado em um sistema de conceitos integrados ao contexto semântico e ontologicamente estruturados.

A leitura é a única maneira de se ter acesso ao conteúdo do documento, sendo um processo simultâneo e não simplesmente simétrico à escrita. KATO (1985) aponta três fatores básicos para a legibilidade de um texto: a qualidade do texto, o conhecimento prévio do leitor e o tipo de estratégias que o texto exige para ser produzido. Por isso, o contexto é fundamental para a legibilidade, ou seja, é preciso saber para quem o autor escreve, objetivo e situação que se assume no momento da produção.

O conteúdo de um documento científico usa um conjunto de condições internas e externas à língua e necessários para a produção, interpretação e interação que devem ser compartilhados entre autor e leitor, procurando identificar informações e significados conceituais. Essa relação pode ser dada por princípios gerais como concepção e raciocínio. Tais relações se servem tanto dos aspectos internos quanto dos aspectos externos para estabelecer conexões entre a expressão linguística e a maneira pela qual seus falantes organizam o sistema linguístico circundante de seu universo. As relações semântico-lexicais só podem ser identificadas como tais a partir deste universo contextual.

As operações de construção de um documento completo para formar o sentido de um texto resultam de várias ações praticadas que envolvem sempre mais de um indivíduo, pois são ações conjuntas e coordenadas. O próprio escritor, no momento de autoria, tem consciência de que dirige o texto a algum tipo de leitor e, para isso, deve selecionar bem o vocabulário mais adequado. Essas operações não são simples e exigem diferentes escolhas e estratégias para formar um documento completo no sentido de informação registrada. É importante assinalar que uma das melhores oportunidades de enriquecimento e ampliação de vocabulários é o processo de ler e escrever, interpretar termos desconhecidos e escolher termos mais adequados para que uma informação possa se atrelar à outra e formar um todo bem organizado e estruturado.

O entendimento de um texto é um processo dinâmico e a interpretação de termos desconhecidos ou significados conceituais de cada vocábulo de uma língua está sempre sendo criado e recriado durante a leitura do conteúdo de um documento. Geralmente, a interpretação se inicia de uma forma ingênua em que o leitor identifica significados do texto em seu conjunto, partindo para um entendimento mais profundo por meio do reconhecimento do relacionamento das partes com o todo. Na verdade, a

interpretação é uma forma explícita de entendimento e, a tarefa do leitor-interpretador é adaptar o texto à situação concreta à qual o texto realmente se expressa (BEPPLER, 2008).

O leitor entende o todo a partir de concepções sobre o significado das partes e suas inter-relações. As concepções de cada pessoa influenciam o entendimento do significado de palavras ou conceitos desconhecidos como na ideia do vai e vem da metáfora do círculo hermenêutico. Segundo Gadamer (1989), a Hermenêutica propõe uma conexão entre a teoria e a prática na qual a interpretação, entendimento e aplicação estão ligados como uma unidade. Por isso, Gadamer adicionou aplicação e interatividade na metáfora do funcionamento do círculo hermenêutico. O entendimento acontece em um contexto no qual os conceitos envolvidos entram em uma espécie de prática que requer um julgamento sobre as implicações por possíveis alternativas em situações locais e a interatividade refere-se ao vai e vem natural do círculo hermenêutico (BEPPLER, 2008).

Os “interesses hermenêuticos” estão em explicar processos articulados à historicidade presente da ação comunicativa e se revelam na capacidade do sujeito de valer-se de necessidades sociais, problemas circunstanciais e limites individuais de leitura. O entendimento envolve um diálogo contínuo que visa compreender o todo e as dimensões de suas partes. Segundo Oliveira (2001), o círculo hermenêutico-dialético é uma técnica que pode ser utilizada como um procedimento metodológico bastante dinâmico e interativo no qual se opera a expressão do conhecimento mais fundamental e intuitiva de análise-síntese.

O círculo hermenêutico-dialético ajuda a interpretar a fala e depoimentos dos atores sociais e analisar conceitos em direção a uma visão sistêmica como um conjunto de componentes em interação. No momento da interpretação dos dados, o sentido do vai e vem se refere ao nível das determinações pelo qual se busca descobrir as conexões que a experiência empírica mantém com o plano das relações essenciais. Para isso, é necessário utilizar três princípios organizadores prévios de forma integrada como dimensões da mente sistêmica: complexidade, instabilidade e intersubjetividade (OLIVEIRA, 2001).

Assim, o leitor-interpretador para relacionar “pensamento”, “ideias” e “palavras” exige movimentação como no vai e vem do círculo hermenêutico-dialético. Capurro (2003) cita o círculo hermenêutico como uns dos princípios filosóficos que melhor estabelece a relação entre conhecimento e a ação na busca pela informação e aponta que a organização dos sistemas de informação precisa considerar três dimensões: a) a hermenêutica dos usuários; b) a hermenêutica da coleção de documentos; e c) a hermenêutica do sistema computacional sob o paradigma físico da representação. Assim, a organização de um sistema hipertextual deve ser vista em diferentes aspectos: a) informação-como-fenômeno; b) informação-como-processo; e c) informação-como-coisa (BUCKLAND, 1991).

Morin (2001) também reflete a dificuldade de se observar o conhecimento, considerando somente uma dimensão. O “pensamento complexo” está sempre repleto de relações e interpretações. Por isso, ao considerar a complexidade dos processos de organização e representação da informação na tela do computador, é importante considerar alguns desses princípios, tais como o dialógico, que possibilita manter a dualidade na unidade conceitual e associar termos ao contexto. O princípio hologramático, que vê o todo a partir das partes, remetendo ao princípio de Pascal, de que não se pode conceber o todo sem conceber as partes e nem conceber as partes sem conhecer o todo, ou seja, o contexto.

### **2.1.6 A importância do contexto para desambiguação do vocabulário**

A ambiguidade do vocabulário expressa significados conceituais distintos que podem ser compreendidos de diferentes maneiras por um receptor envolvendo diversos fatores como, por exemplo, casos de polissemia, homonímia, metáfora e metonímia (CINTRA, 2002). A própria palavra “navegação” é uma metáfora utilizada para descrever como os usuários se movimentam, desenhando um percurso em uma rede semântica de escolhas de símbolos de objetos informacionais digitais.

As relações semântico-lexicais de uso de uma língua natural mostram a ambiguidade do vocabulário humano que pode expressar significados distintos na escrita de um texto, possibilitando a comunicação linguística prestar-se a mais de uma interpretação. Essas relações são características de um processo ativo e contínuo de sentidos. O processo pelo qual a ambiguidade de um vocabulário é solucionada recebe o nome de desambiguação, exigindo diferentes níveis de conhecimentos linguísticos e extralinguísticos. Bräscher (2002) cita alguns exemplos como a ambiguidade morfológica que pode ser solucionada pela análise do co-texto imediato que circunda a palavra policategorial, recorrendo-se apenas a conhecimento morfossintático como categoria gramatical, concordância e combinações sintáticas entre constituintes da frase, entre outros.

A noção de co-texto foi proposta pela Linguística Textual para dar conta da intervenção das unidades que fixam a significação de diferentes formas linguísticas. De acordo com Halliday (1994), isso significa dizer que o sistema linguístico se organiza de tal forma que o contexto social serve para antecipar o texto e exemplifica com a chegada de uma pessoa a uma reunião, podendo situar-se rapidamente porque houve uma captação de um campo conceitual. Essas relações são estabelecidas em determinados modo da situação, de tal maneira, podendo até ter uma ideia do que está sendo dito. Isso significa compreender a linguagem em seus diversos aspectos multifuncionais.

Com base em tais aspectos, Halliday questiona: “de que outro modo pode-se considerar a linguagem senão em seu contexto social?” e acrescenta que são os usos do dia a dia da linguagem que servem para transmitir ao indivíduo as qualidades essenciais da sociedade e a natureza do ser. Nesses aspectos sistêmico-funcionais de uso, a linguagem deve ser fundamentada para entender o ambiente (função ideacional ou reflexiva); influir sobre os outros (função interpessoal). Assim, cada elemento da língua em uso é explicado por referência à sua função sistêmica, por isso, a gramática funcional constrói e descreve todas as unidades da língua-objeto como configuração orgânica de funções (SIMÕES, 2006).

Halliday entende os “usos da linguagem” como uma “seleção de sentidos” dentro do sistema linguístico real. A relação entre texto e contexto está fundamentada na interação do uso da linguagem que se organiza na forma de redes de escolhas léxico-gramaticais dentre os quais significados podem ser construídos e reconstruídos. Para entender essas relações, Martins (2004) alerta sobre a necessidade de confrontá-las com as demais partes que constituem o todo, ou seja, deve-se sempre levar em consideração o contexto em que está inserida para que não ocorra uma leitura equivocada.

Para compreender a linguagem como um sistema interativo e a importância do contexto no processo de desambiguação de vocabulário, Janczura (2005) estudou associações produzidas em dois contextos semânticos diferentes e comparou com associações produzidas sem contexto. Os resultados apontaram uma redução em média de 50% no campo semântico devido ao número de erros cometidos na ausência de contexto ter sido maior do que em sua presença. Estes estudos estão no bojo de interesse desta pesquisa no que tange às estratégias de associações entre significados conceituais e suas relações ao contexto de uso do vocabulário. O contexto provoca a redução do campo semântico revelado pelo número de palavras associadas, reduz o número de erros associativos e gera um primeiro membro fortemente associado à categoria possibilitando estabelecer critérios de aceitabilidade classificatórios (JANCZURA, 2005).

Segundo Bastos (2005), a semântica estrutural trabalha com os significados que constituem estruturas dentro do uso de uma língua natural e demonstra por meio dos estudos realizados na lexicologia que cada unidade léxica incorporada no interior de um campo semântico não provoca alteração da estrutura semântica global. Nos estudos de Bastos, a teoria dos campos semânticos e campos conceituais em conjunto com a Terminologia e a análise de domínio na CC contribuíram para a formação de grupo de subáreas. A teoria dos campos conceituais de estruturação de vocabulário pode fornecer subsídios metodológicos para trabalhar a delimitação de uma área do conhecimento pela

identificação dos traços semânticos mínimos e distintivos das palavras, envolvendo a transformação das palavras da linguagem natural para a construção de categorias.

Como são as situações-problemas que dão sentido aos conceitos, é natural definir campo conceitual como sendo, sobretudo, um conjunto de situações conectadas para estruturar um sistema dinâmico e funcional de significado paradigmático. Organizados em campos conceituais, os conceitos de um domínio podem ser definidos por três conjuntos de situações: o primeiro é um conjunto de situações que constituem o referente do conceito, o segundo é um conjunto de invariantes operatórios (teoremas e conceitos-em-ação) que dão o significado do conceito, e o terceiro é um conjunto de representações simbólicas que compõem seu significante (MOREIRA, 1999).

A psicologia cognitiva vê o armazenamento de informação na memória humana altamente organizada como uma espécie de hierarquia conceitual. O contexto que permite ao leitor-interpretador fazer conexões significativas entre conceitos semanticamente organizados e estruturados. Para entender esses processos, modelos conceituais que se aproximem de modelos mentais podem ajudar a simular a organização do comportamento imposto por práticas sócio-culturais para que esses modelos sejam sistêmico-funcionais.

### **2.1.7 Cognição, linguagem e princípios computacionais: modelos conceituais de modelos mentais**

Um modelo conceitual é um esquema criado para permitir a atividade de abstração e explicitação do conhecimento (MOREIRA, 1997). Cognição significa a aquisição de um conhecimento através de processos mentais do complexo pensamento humano. Simplificando, pode-se dizer que o sistema cognitivo é a forma como o cérebro humano aprende, memoriza e processa uma informação captada de sentido e percepção. A visão de mente-computador foi introduzida pelo psicólogo infantil Jean Piaget como uma função simbólica da linguagem, centrada na gênese do conhecimento em construção de campos conceituais, estudando aspectos iniciais do raciocínio de estruturas lógicas.

Essa visão de mente-computador ajuda a entender o processamento da linguagem natural e os mecanismos neurais de como o cérebro humano recebe códigos, estrutura e processa dados para apresentar resultados. Para Severino (1980), o conhecimento humano inicia-se com a formação de conceitos, considerando conceito como uma imagem mental por meio da qual se representa um objeto. O conceito é o sinal imediato do objeto representado que garante uma referência direta ao objeto informacional registrado em um documento. Essa referência é dita intencional no sentido em que o conceito adquirido por processos especiais de apreensão das coisas pelo intelecto se refere a objetos, a seres, a ideias, de uma maneira representativa e substitutiva, simbolizado por um termo ou palavra.

Na realidade, o cérebro humano processa dados com modelos mentais, mas modelos conceituais podem ajudar na construção de modelos mentais ao explicar consistentemente o conhecimento aceito em um domínio. Estes conceitos foram introduzidos por Johnson-Laird para designar esquemas utilizados por pessoas para entender o mundo real, físico e natural dos sistemas conceituais, distinguindo-se quatro tipos principais de modelos conceituais domínio (MOREIRA, 1997).

- **modelo monádico:** representa afirmações do raciocínio silogístico sobre individualidades, propriedades e identidades entre elas;
- **modelo relacional:** é aquele que agrega um número finito de relações entre as entidades representadas em um modelo monádico;
- **modelo metalinguístico:** é o que contém elementos (“tokens<sup>17</sup>”) correspondentes a certas frases linguísticas e certas relações abstratas entre elas e elementos de um modelo de qualquer tipo;
- **modelo conjunto teórico:** é aquele que contém um número finito de elementos (“tokens”) que representam diretamente conjuntos, incluindo identidade e não-identidade entre os elementos que representam conjuntos.

Wilson (1965), estudioso dos mecanismos do cérebro humano, afirma que tanto o cérebro como o computador são capazes de processar informações tais quais as recebem e separar das demais para uma determinada situação específica. A concepção de raciocínio inteligente computacional visa simular como o cérebro humano processa informações, lembrando que o computador é dependente de uma inteligência artificialmente programada. Então, para que a máquina processe informações, é necessária toda uma configuração e mecanismos de codificação/decodificação de metalinguagens compiladas, ou seja, linguagens que são interpretadas por um sistema tradutor.

Um dos primeiros estudos filosóficos sobre o raciocínio humano foi sistematizado pelo filósofo Aristóteles que influenciou a sintaxe da linguagem lógica formal dos primeiros computadores. Aristóteles, ao conceituar uma “coisa”, se preocupou em separar os atributos específicos para encontrar suas características comuns e essenciais. Pela abstração das coisas do mundo, Aristóteles estabeleceu categorias gerais na ação de classificar e buscar um silogismo com a linguagem. O silogismo forma o conhecimento por meio de uma regra lógica, por exemplo, considerando duas proposições como verdadeiras:

---

<sup>17</sup> O termo tokens pode ser substituído pelo termo instâncias se interpretado como tudo aquilo que tem existência num espaço e tempo determinado. Na CC, um token é um padrão de caracteres com um significado específico em um código fonte de uma base de dados.

“Aristóteles é um homem” e “Todo homem é mortal”; pode-se formar uma terceira proposição verdadeira “Aristóteles é mortal” (CHAUÍ, 2000).

A linguagem na visão de Aristóteles é criação do “Ser” que tem caráter simbólico de seus usos sociais, podendo ser construída para representar um tipo de discurso racional que precisa ser verificável como lógico, ou seja, verdadeiro ou falso como nos modelos conceituais por prototipação, representando modelos mentais. Aristóteles cercou essa afirmação dos devidos cuidados, quando mostra que os nomes são símbolos das coisas, mas a relação entre o conceito (nóema) e o sinal (semêion) ou entre a coisa (prâgma) e o nome (ónoma) não é sempre de congruência (NEVES, 2007).

Segundo Castilho (2007), a Teoria dos Protótipos sublinha a dinamicidade das categorias gramaticais. Essa teoria tem fundamento na psicologia cognitiva de atributos graduáveis. Os graus de relevância traçados dentro das categorias correspondem ao fato de que estas se organizam em torno de um centro cognitivo exemplar com entidades ordenadas, produzindo o que se conhece como efeitos de protótipo. Neste sentido, categorizar significa eleger protótipos como um modelo conceitual que permite explicar mais facilmente métodos cognitivos de classificação em categorias.

As características estruturais da categorização linguística, tais como prototipicidade, polissemia, modelos cognitivos, metáforas e imagens mentais são os princípios funcionais da organização linguística (iconicidade e naturalidade). A interface conceitual entre sintaxe e semântica, a pragmática ligada à experiência da linguagem-no-uso e a relação entre linguagem e pensamento são fenômenos verdadeiramente conceituais e constituem importantes modelos cognitivos (SILVA, 1997). Esses fenômenos em transformação e constituintes desses modelos nos quais a linguagem interage com outros domínios mostram que o conhecimento possui múltiplas facetas de representação simbólica, podendo ativar diferentes categorias formais e referências a objetos informacionais.

Essas múltiplas potências de expressar o raciocínio humano possuem conexões léxico-gramaticais de diferentes modos de categorizar, classificar e comparar situações lógicas em facetas. A linguagem além de ter o poder de ser instrumento do conhecimento, também é substrato em que se instaura o ser pensante como o elo de comunicação do solo constituidor da cultura que se estabelece na memória. A capacidade de compor o raciocínio em forma estruturada é uma dessas potencialidades da linguagem (FONSECA, 2009).

O que está no som é símbolo do pensamento, mas não necessariamente o conceito que está no som. O significado é congruente com o conceito, embora só sob as formas de linguagem possam ser apreendidos os conteúdos de modelos mentais não recobrando inteiramente conceito e palavra. Nessa concepção, se a língua oral é a

representação primária dos "estados de alma", a escrita é sua representação secundária, ou seja, símbolo das palavras (re)apresentadas (NEVES, 2007).

Segundo Montague (1974), a linguagem natural humana não se distingue essencialmente da artificial, assumindo a hipótese de descrever estruturas mediante formalismos precisos que permitam explicitá-la sintaticamente e investigá-la como na linguagem lógica matemática. Tal base está fundamentada na forma representacional considerando o conhecimento objetivo. Isso não seria assumido como propriedade interna de representações, mas como relação de um sistema simbólico, externo às condições puramente cognitivas, cuja organização deste conhecimento é uma entidade formal pressupostamente construída em si mesma.

Todas essas abordagens contribuíram muito para avanços de modelos conceituais de sistemas computacionais, principalmente no formalismo de representação de uma informação a ser processada pelo computador, utilizando códigos da linguagem natural representados em um conjunto do sistema binário que utiliza apenas dois símbolos da matemática o zero e o um. Para chegar a uma conclusão verdadeira, o silogismo lógico computacional deve obedecer a um conjunto complexo de regras escrito em uma sintaxe formal de uma linguagem de programação de baixo nível<sup>18</sup> para ser gravado em uma memória artificial. Este conjunto de regras utiliza a linguagem lógica matemática e booleana, originalmente desenvolvida por George Boole, que permite mapear objetos em bits e bytes.

Um bit ou um dígito binário é o nome dado aos impulsos elétricos representados pelos símbolos 1 ou 0. O byte é o nome dado ao conjunto de oito bits que serve para representar no computador um único caracter alfanumérico (uma letra, um número ou um símbolo). Uma aplicação interessante da linguagem lógica matemática e booleana é o processo de realimentação da informação e significa que o valor de saída de uma porta retorna à sua entrada, ou seja, um caracter ao entrar por um circuito eletrônico pode ser codificado, agrupado e desagrupado, controlado, guardado e processado. Isso possibilita criar uma variedade de circuitos úteis para executar a lógica de programação computacional.

Nessa dinâmica de representação simbólica, concebendo a linguagem como estruturas de categorias instáveis, o matemático, Alan Turing, propôs, em 1936, um protótipo conhecido como "Máquina de Turing" que serviu como modelo do processamento linguístico de criação dos primeiros computadores. A base lógica do raciocínio inteligente computacional dos primeiros equipamentos combinava estruturas de algoritmos numa

---

<sup>18</sup> Linguagens de baixo nível são mais próximas da linguagem de máquina (dígitos binários).



sintaxe formal, demonstrando condições lógicas como ordem de parada para interromper o processamento mecânico do programa computacional.

A ordem de parada dos primeiros algoritmos foi baseada nos "Teoremas da Incompletude" escritos por Kurt Friedrich Gödel. Esses teoremas mostram que existem limites passíveis de serem processados na matemática, ou seja, se o último número pudesse ser produzido, haveria um sempre maior, como não pode haver um conjunto de todos os números possíveis, nunca há um último número na série. Esses algoritmos surgiram como uma equação matemática na resolução de cálculos traduzindo caracteres alfanuméricos utilizando linguagem de máquina decodificada em dígitos binários como um jogo dinâmico de alocação de memória computacional condicionada por apontadores.

Esses princípios ainda estão presentes no tratamento da informação dos computadores modernos em que o diálogo humano-computador, denominado de interface, passou a ser descrito de forma serial representados por sequências específicas de códigos binários, podendo manipular base de dados em forma de arquivos eletrônicos. Como o computador é limitado a regras lógicas que processa caracteres codificados/decodificados, houve a necessidade de criar novos mecanismos, metalinguagens de alto nível<sup>19</sup> para descrever o conteúdo de um documento legível para os humanos.

A expansão da massa documental impressa produzida pela ciência evidencia a necessidade do uso do computador para ajudar nos processos do tratamento documentário. A representação de um texto no computador passou a ser organizada em forma de arquivos. O documento legível para os humanos foi traduzido automaticamente para o computador em forma de códigos binários. Lyytinen (1987) destaca que no início da aplicação dos computadores, houve um profundo impacto na maneira de tratar e organizar a informação em forma de arquivos para ser representada na forma lógica.

Com o surgimento da Engenharia de Software, muitas técnicas foram sistematizadas iniciando-se pela modelagem de dados, surgindo métodos de automatização com índices físicos<sup>20</sup> e *frameworks*<sup>21</sup>. Logo depois, apareceram os Sistemas de Gerenciamento de Banco de Dados (SGBD<sup>22</sup>) com recursos de representação da informação e na organização de bases de dados no computador. Este paradigma de (re)

---

<sup>19</sup> Linguagens de alto nível são metalinguagens que tentam aproximar mais as linguagens de máquina com a linguagem natural humana. Por exemplo: FORTRAN, C, COBOL, JAVA, etc.

<sup>20</sup> Esses índices são estruturas implementadas em meios computacionais para acelerar o acesso a unidades de textos (por palavras-chave) como forma de armazenar a posição do elemento no documento e acessar rapidamente termos e conceitos em base de dados (LOH, 2001).

<sup>21</sup> Em desenvolvimento de sistemas computacionais, um framework é uma abstração que une códigos comuns entre vários projetos de software provendo uma funcionalidade genérica podendo atingir uma funcionalidade específica durante a programação de uma aplicação.

<sup>22</sup> Um SGBD é uma coleção de programas que permite ao usuário definir e manipular bases de dados para as mais diversas finalidades.

apresentação da informação em forma de palavras-chave sobre o conteúdo de um documento na sua visão mais lógica (texto completo) está relacionado a limitações da capacidade inicial dos primeiros computadores em armazenar e processar informações textuais. Este paradigma ficou conhecido como barreira de palavras-chave que ordena e aponta para termos-chave, possibilitando consultas em bases de dados e sistemas de RI.

Para a concepção de um modelo de sistema de RI é necessário determinar quais são as operações que podem ser executadas durante um processo de busca e definir a fonte de informação, ou seja, a relação de documentos a serem recuperados. Em seguida, deve ser construído um índice para armazenar os termos contidos nos documentos podendo, assim, iniciar um processo de recuperação. Esse processo necessita que as informações estejam registradas em uma base de dados. Essas bases de dados são utilizadas para indicar a coleção de dados, servindo de suporte ao sistema, sendo que as principais bases incluem textos completos de artigos de periódicos, jornais ou outras modalidades de documentos textuais e referências bibliográficas (VAN DER LAAN e ALVORCEM, 2007).

A forma lógica de representação no computador exige formalismos precisos que permitam explicitar o vocabulário sintaticamente. É esse formalismo que codifica os possíveis sentidos de palavras-chave, independente do contexto. A gramática formal impõe regras para que o vocabulário possa ser descrito em uma sintaxe lógica de programação estruturada. Essa representação pode ser de primeira ordem, proposicional ou descritiva. A lógica de primeira ordem e proposicional se baseia apenas no uso de constantes simbólicas e conectivos lógicos. A lógica descritiva procura definir primeiro os termos-chave e depois utiliza esses termos como conceitos de um domínio, especificando propriedades de objetos.

Muitos modelos conceituais por prototipação foram desenvolvidos para auxiliar a área de RI. Modelos booleanos, vetoriais, probabilísticos passaram a auxiliar processos já consolidados de análise bibliométricas para mensurar a ciência com indicadores da produção científica em determinada área do conhecimento. Muitos desses processos foram auxiliados pelos profissionais da CI que vieram de diferentes áreas do conhecimento, com uma visão multidisciplinar (inter e transdisciplinar) de organização e uso da informação com o foco na recuperação (TOUTAIN, 2007). A CI mantém uma estreita relação com a CC, a Biblioteconomia, a Ciência Cognitiva, Comunicação, entre outras (SARACEVIC, 1996).

Não se pode negar o fato de que a tela do computador tornou-se a principal base de leitura de documentos científicos na contemporaneidade. A Web é um grande repositório de informação que possibilita parte integrante do processo de pesquisa com recursos midiáticos para armazenamento de todo tipo de objetos informacionais digitais em qualquer língua, gráfica ou sonora. Esses objetos podem ser: a) não estruturados por não

possuir uma estrutura de armazenamento definida; b) semi-estruturados por não ter uma estrutura de armazenamento muito rígida; e c) estruturados que possuem uma estrutura rígida de armazenamento. São essas grandezas-signos que desempenham o papel de privilegiado instrumento gerador dessa 'realidade' (BARBOSA, 1998). Portanto, esses recursos devem ser usados como ferramentas intelectuais e possibilitar o desenvolvimento de redes de inteligência coletiva numa visão de trabalho colaborativo (LÉVY, 1993).

## **2.2 Ciência da Informação e os desafios da recuperação nos hiperdocumentos científicos na contemporaneidade**

A CI surge no horizonte de transformações das sociedades contemporâneas que passaram a considerar o conhecimento, a comunicação, os sistemas de informação e significados conceituais como objetos de pesquisa científica e domínios de intervenção tecnológica (GONZÁLEZ de GÓMEZ, 2000). De acordo com Wersig (1993), a origem da CI veio da necessidade de se documentar a informação disponível a partir de uma rede conceitual interdisciplinar com a preocupação dos efeitos da tecnologia sobre o uso do conhecimento e que pode contribuir para o desenvolvimento da "Filosofia do Conhecimento".

Uma ciência é formada por sistemas e subsistemas. A informação é indissociável de qualquer "sistema". Robredo (2005) define um sistema de informação como uma entidade complexa e organizada que armazena, processa, fornece, usa e distribui a informação. Para recuperar uma informação em qualquer "sistema", primeiramente se faz necessário organizá-la porque a informação dispersa não constitui inteligência e somente a partir da sua organização e estruturação é que a inteligência passa a existir (TARAPANOFF, JÚNIOR e CORMIER, 2000).

González de Gómez (2000) destaca que a recuperação, as novas TIC e a ênfase na informação científica seriam os objetos, participantes deste "sistema", a incorporarem ideias paradigmáticas que funcionam como um sistema formal intermediário de armazenamento e recuperação. Dentro desse sistema haveria um fluxo de informação entre os dois pólos (geradores e usuários), buscando otimizar a coleta, armazenagem, organização, representação, recuperação e disseminação. Um dos paradigmas da CI seria estudar o processo de comunicação presente neste fluxo sob o ponto de vista não só dos fatores internos organizacionais necessários para a recuperação, como também dos fatores externos de condições iniciais de uso desde a criação da informação até a assimilação pelo receptor (BARRETO, 2002).

Para Saracevic (1992), a CI é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação e de seus registros.

No tratamento desses problemas, deve-se tirar maior vantagem possível das modernas TIC. O problema da expansão da massa documental, por exemplo, que é impulsionado pelo poder proliferativo do uso do vocabulário humano nas práticas documentárias, com o uso das novas TIC, tende a aumentar cada vez mais. O Dr. Vannevar Bush foi um dos primeiros pesquisadores a alertar sobre este problema. Bush (1945) destaca a existência de uma montanha crescente de pesquisa com conclusões importantes que ele não consegue encontrar tempo para compreender, muito menos memorizar devido à velocidade como as pesquisas científicas surgem e soluções são tratadas.

Segundo Bush, os mecanismos de revisar resultados dessas pesquisas estavam inadequados, chamando a atenção para a forma de extrair dados em um sistema de RI quando eles são colocados em ordem alfabética ou numérica e a informação é encontrada descendo-se de uma subdivisão para outra em um único lugar. Para isso, o pesquisador precisa conhecer as regras que conduzirão pelo caminho certo, além disso, essas regras são muito complicadas e dificultam a recuperação relevante da informação.

Para Bush, a seleção é uma questão muito maior do que apenas a extração de dados para fins de pesquisa científica. A seleção ou extração de um dado envolve todo o processo pelo qual um usuário-leitor-sistema pode tirar proveito do conhecimento registrado com o uso das novas TIC. Capurro (2003) alerta que é preciso compreender a distinção entre uma oferta de sentido e um processo de seleção. O destaque de selecionar uma informação relevante em um sistema é ação primordial de seu uso. A informação não se fecha como uma “cápsula cognitiva” transportada de um lugar para outro.

Lima (2004) destaca que o hipertexto foi inspirado nas ideias de Bush com o Memex que tentava imitar a estrutura da memória humana de organizar documentos em trechos e combiná-los conforme necessidades de compreensão de leitura em um contexto semântico. Na visão de Bush, a mente humana, ao encontrar um conceito relevante, agarra-se imediatamente a outro que deve ser sugerido por meio de um mecanismo artificial que possibilita associar pontos de acesso para recuperação mais eficaz.

É importante destacar que os pressupostos deste trabalho estão de acordo com a visão de Bush para que a organização semântica de uma base de dados possa ajudar o leitor-interpretador a identificar significados conceituais ou termos sinônimos desconhecidos. Um dos grandes problemas na área de RI é a questão de identificar a palavra-chave que o conceito válido foi indexado pelo sistema computacional para que o usuário obtenha sucesso na busca. Em uma base de dados, o documento relevante que o usuário procura pode não possuir a palavra-chave digitada e sim outros termos sinônimos.

Pode-se dizer que Bush foi o idealizador das primeiras interfaces gráficas presentes até hoje no computador, forçando a representação da informação por meio de artefatos de engenharia. Para encontrar na interface alguma referência, o usuário de qualquer sistema computacional deve primeiramente lidar com a seleção de artefatos (OLIVEIRA e BARANAUSKAS, 1998). Esses artefatos devem possibilitar a interatividade em RI, como se a interface dialogasse com o leitor-interpretador na busca por uma informação relevante, em uma base de dados, ao encontrar termos sinônimos explicitados por um mapa.

Um problema comum, nas interfaces de sistemas computacionais, é o direcionamento ineficaz do usuário-leitor-sistema na busca pela informação relevante. Na prática de leituras hipertextuais, frequentemente, os usuários se esquecem de como chegaram à determinada parte do texto. Essa situação é comumente causada pela falta de critérios claros e consistentes na modelagem dos sistemas de hipertextos, especialmente, nas fases de escolha e estruturação das informações relevantes e na criação de mapeamento de *links* na interface. De acordo com Lima (2004), navegar significa direcionar uma mudança de foco ou movimentação em relação ao conhecimento disponível no hiperdocumento, por meio de estruturas onde o leitor-interpretador ativa processos dinâmicos de interação.

Cada link de um sistema hipertextual pode conter uma trilha de navegação decorrente de um mapeamento conceitualmente deficiente que tende a dificultar tanto a busca por parte do usuário quanto à recuperação por parte das máquinas de busca na Web. Por isso, a organização semântica de uma base de dados e a (re)apresentação da informação na interface humano-computador exige mecanismos dinâmicos e interativos para aproximar modelos conceituais de modelos mentais. Esta organização se constitui requisito essencial para que, em uma etapa posterior, os usuários ou as ferramentas de busca na Web alcancem maior eficácia no momento do acesso à informação (LIMA, 2004).

Na Web, a recuperação relevante é diretamente afetada pela tarefa do usuário e pela visão lógica dos documentos adotados em um modelo computacional. A tarefa do usuário está relacionada à tradução da necessidade de informação de maneira que o sistema possa entender a requisição do usuário e a visão lógica está relacionada ao modo como um documento é representado de forma a permitir que as buscas sejam possíveis (BAEZA-YATES e RIBEIRO-NETO, 1999). Beppler (2008) destaca alguns dos problemas comuns relacionados a essa dificuldade que se referem à composição inadequada de consultas com palavras-chave, à grande quantidade de informação resultante a partir de uma busca com essas palavras e à falta de interatividade em RI.

O princípio de relevância é um conceito fundamental em qualquer modelo de sistema de RI. A relevância é uma noção intuitiva que se refere ao julgamento da informação recuperada em relação à necessidade informacional e determinada por uma relação comparativa. A relevância pode estar relacionada a características do sistema ou às características do usuário. A relevância orientada ao sistema refere-se a propriedades e mecanismos internos caracterizados pelo resultado da correspondência entre os termos utilizados na consulta e os termos indexados e armazenados. A relevância orientada ao usuário inclui aspectos cognitivos, situacionais e psicológicos e refere-se aos contextos subjetivos do mesmo que são empregados em diferentes leituras (SARACEVIC, 1996a).

Neste ponto, a CI se encontra diante de novos paradigmas que Choo (1998) aborda de forma dualista: a) “sistema” computacional com suas barreiras de palavra-chave e que necessita “entender” contexto semântico; e b) “usuário” que seleciona termos relevantes, compara, interpreta, contextualiza e reserva a informação válida e útil. Na visão de Choo, quanto mais a informação obtida for capaz de conectar-se com as necessidades viscerais e conscientes, mais o indivíduo sentirá que a informação é significativa ou útil. Verifica-se que a organização será mais dinâmica, a partir do momento que tiver um “sistema” mais eficaz. Paralelo e intrínseco a esse “sistema”, deve haver profissionais com competências suficientes no gerenciamento de toda massa documental, dentro e fora de uma organização.

Poucos pesquisadores têm visto a unidade de estudo da CI como disciplina, domínio, ambiente ou coletividade. É possível perceber que as abordagens de pesquisas têm sido abordadas de forma dualista, já que focam “usuário” ou “sistema”, e não os aspectos coletivos ou o contexto social no processamento da informação natural (NASCIMENTO e MARTELETO, 2004). Por isso, este trabalho não delimita fronteiras, assumindo uma visão mais ampla de pesquisas de natureza multidisciplinar.

A CI e a CC têm convergido em diversos aspectos no sentido de encontrar mecanismos com o objetivo de minimizar o problema da explosão informacional na Web (MARCONDES, 2009). Esses mecanismos conduzem a um novo tipo de “modelagem informacional” que deve ser entendido como “instrumento linguístico de comunicação” em que o caráter específico de um modelo conceitual computacional deriva da natureza formal de sua linguagem e dos mecanismos preestabelecidos de seus usos. O sucesso dessas ferramentas depende da disponibilidade de um mecanismo humano de inferência que estipule sua aceitação em harmonia consensual de padrões terminológicos pré-estabelecidos por uma comunidade científica na qual está inserida (GÓMEZ, 2004).

Para Lyytinen (1987), existiriam duas principais abordagens, com diferentes premissas, acerca de qual seria a “matéria-prima” a ser operacionalizada no computador: a)

a "visão do mapeamento do mundo real" supõe um processo de mapeamento que se desenvolve entre o "mundo real" e um "modelo formal" ou esquema conceitual, que seria a "representação do mundo real"; dessa forma, a modelagem se preocuparia mais em como representar e não o que e para que representar; e b) o "desenvolvimento de uma linguagem formal" que considera a modelagem de processos comunicacionais e interativos no qual são formuladas, desenvolvidas e adotadas regras convencionais como instrumento linguístico de objetos informacionais conceituais. Essas regras se aproximam da gramaticalização da linguagem, que codifica conforme normas e usos antes que pela lógica formal.

Segundo Castilho (2007), as estruturas mais gramaticalizadas perdem sua eficácia interacional em sistemas hipertextuais, reinstaurando-se mecanismos que fazem da língua muito mais um conjunto de processos entre autor e leitor a serem construídos de acordo com o contexto de uso das palavras do que um catálogo de metadados de produtos prontos. Aqui está um dos maiores desafios desta pesquisa em tentar transpor para um modelo conceitual computacional as estratégias de associações entre significados conceituais e suas relações ao contexto semântico para a extração de "conceitos científicos" em conexão com "conceitos espontâneos" que pode ser explicitado por uma ontologia na IA.

Existe certa homologia entre o estabelecimento de novas relações entre a linguagem, as novas TIC e as técnicas usadas na CI. A passagem das "tecnologias culturais" às denominadas "tecnologias intelectuais" de suporte digital modificam novamente as formas de inscrição, armazenamento e transmissão da produção social discursiva e, ao mesmo tempo, modificam demandas de intervenção institucional, técnica e profissional no tratamento, circulação e gestão dessa produção (GONZÁLEZ GÓMEZ, 2004).

Segundo Wersig (1993), a partir do surgimento da imprensa e, principalmente, com o desenvolvimento das novas TIC, o conhecimento se massificou e se deslocou do produtor, provocando uma tripla fragmentação do conhecimento produzido, representado e demandado. Essa racionalização ganhou importância destacada devido aos grandes efeitos que o conteúdo de um documento científico produz na sociedade e também porque o conhecimento não é mais explicado por um sistema de crenças, mas como registros de objetos informacionais digitais documentados em um suporte.

De fato, desde sua invenção, a utilização da escrita desempenhou papel primordial ao permitir a cristalização do pensar, da ação e da experiência humana a partir de representações simbólicas e gráficas, codificadas por uma linguagem normalizada e sistematizada em um suporte para formar o conteúdo de um documento textual (AGUIAR, 2008). Segundo Bastos (2005), a organização do conhecimento na CI é um produto social que se realiza a partir da informação, e, ao socializar-se, transforma-se em nova informação. O conhecimento se registra sempre em documentos, como conjunto organizado de dados e

admite usos indiscriminados mediante estruturação e comunicação, formando um sistema de conceitos abertos para seu melhor aproveitamento individual e social.

O conteúdo de um documento segue regras próprias do tipo escolhido. Miranda e Simeão (2002) apontam duas direções complementares e interdependentes que devem ser consideradas em qualquer tipo de documento no sentido de informações registradas: uma voltada para o conteúdo e a outra, para a estrutura do próprio documento. Com base neste apontamento, Miranda e Simeão propõem um esquema de decomposição (Figura 2) em quatro elementos constitutivos do documento: a) tipo; b) conteúdo; c) formato; e d) suporte. A ordem desses elementos pode ser vista de forma orgânica ou sistêmica nos processos do ciclo documentário completo.

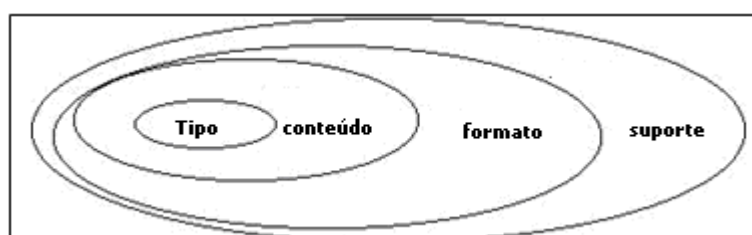


Figura 2 - Elementos constitutivos de um documento.  
**Fonte:** Miranda e Simeão, 2002.

- **Tipo** - O tipo do documento está intrínseca e indissociavelmente relacionado com a sua produção. Tipos de documentos como dissertações e teses têm configurações convencionais consagradas às exigências formais e normas adequadas que estão relacionadas aos modos de produção e uso.
- **Conteúdo** - É a parte substantiva do documento (dado, informação, conhecimento) e está predeterminado pelo seu tipo de acordo com normas e condições de produção.
- **Formato** - está relacionado com o modo de concepção e exposição do conteúdo. O formato molda o conteúdo, tornando-o visível e inteligível. No sentido oposto, diferentes formatos pressupõem diferentes conteúdos e exigem tratamentos técnicos diferenciados. Na prática, é possível a determinação de formatos básicos que servem de modelo e podem ser copiados para determinados tipos de registros.
- **Suporte** - É a parte visível e manipulável do documento, ou o documento propriamente dito no senso comum. É a sua coisificação ou expressão física como produto, mas que compreende todas as características constitutivas já discutidas anteriormente.



Cada ciência se apropria do conteúdo de um documento conforme suas capacidades heurísticas metodológicas, mas a tipificação entra certamente na legitimação do processo de apropriação do conhecimento bem antes do registro deste hiperdocumento em um suporte nos ambientes digitais. Tendo em vista a produção e reprodução dos artefatos culturais, o objeto informação se realiza como “prática informacional”, considerando que toda prática social é “informacional” (MARTELETO, 1992).

Para Marteleto (1992), esta expressão se refere aos mecanismos mediante os quais os significados, símbolos e signos culturais são transmitidos, assimilados ou rejeitados pelas ações e representações dos sujeitos sociais em seus espaços concretos de realização. Wersig (1993) aponta fatores de transformação sociais em que “objetos informacionais” têm a capacidade de alterar o estado de conhecimento do indivíduo, reforçando a possibilidade da informação comportar significações que só se concretizam ao se estabelecerem conexões entre o conhecimento prévio do sujeito e suas relações.

Para estabelecer essas conexões em um sistema de hipertexto, Barreto (1999) nos lembra que todo ato de conhecimento associado ao conteúdo simbólico de uma estrutura de informação textual é uma cerimônia que acontece em mundos diferentes tanto para o emissor quanto para o receptor da informação. Essa cerimônia são atividades construídas (ou reconstruídas) na interação texto-sujeitos e não como algo prévio a essa interação. Na visão de Barreto, a transferência da informação, está condicionada por uma limitação contextual e cognitiva para que possa intervir na vida social e gerar conhecimento com objetos informacionais digitais registradas em um hiperdocumento científico.

Uma estrutura de informação textual possui características de linguagem que admitem uma análise morfológica, permitindo extrair indicações para decisões estratégicas de gestão com intenções de conhecimentos. A estrutura de um texto científico em um hiperdocumento reflete o arranjo utilizado pelo autor para a apresentação das informações como esquemas formais de organização de ideias. Essas características podem ser consideradas como uma “superestrutura”, ou seja, uma organização paradigmática em que o processo de compreensão supõe a transferência das unidades identificadas no texto permitindo monitoramento de leitura e significados conceituais (KOBASHI, 1994).

O texto de um hiperdocumento científico como unidade de linguagem pode ser composto de diversos aspectos funcionais de suas múltiplas facetas tais como: unidade formal e material para dar sentido, coesão, coerência aos significados semânticos e estruturais, além de funções informacionais e comunicativas que necessitam de integração entre elementos, referências, conjunções e relações. A informação desta “superestrutura”, seja ela impressa ou digital, deve ser pensada além dos suportes existindo como qualidade

secundária de um objeto do signo linguístico registrado que regem o funcionamento da “instituição total da sociedade” e da própria dinâmica cultural (MARTELETO, 1992).

González Gómez (2004) sugere um novo olhar sobre a informação, tendo como perspectiva as novas TIC que interagem com contextos de objetos informacionais hipertextuais e que seria próprio das práticas contemporâneas. Neste sentido, a linguagem deve ser concebida como um modelo conceitual e relacional que vai além da visão somente de suporte de um sistema. Embora a linguagem verbal escrita seja um caso particular, ela constitui um “sistema” de signos de espectro tão amplo, que todos os outros sistemas de linguagem podem se repassar de língua. Este estudo tem de ser feito tomando-se um conjunto de enunciados que, independentemente da sua significação, evidenciam as próprias regras de uso e funcionalidades de um corpus<sup>23</sup> teórico como unidade mínima de conhecimento vendo o conceito em suas diferentes representações (CINTRA, 1983).

Um conceito como unidade mínima de conhecimento registrado em um documento científico possui uma grandeza-signo portadora de significado textual dentro de um campo morfossemântico que se refere a um conjunto de lexemas ou outras unidades linguísticas que se encontram ligadas semanticamente. Este tipo de documento possui relatos de experiências em busca de inovações e soluções de problemas como “teoremas-em-ação”. Por isso, é possível pressupor que este tipo de documento possui características linguísticas que permitam a determinados objetos informacionais extrair significados conceituais, transmitindo conhecimentos conceituais como “conceitos-em-ação”.

Neste sentido, um documento científico, seja ele impresso ou digital, não pode ser tratado como um documento qualquer em um suporte. A “superestrutura” deste tipo de documento precisa receber um tratamento mais adequado para socialização da informação, auxiliando usuários-sistema na descoberta de novos conhecimentos. As informações deste tipo de documento podem contribuir para avanços de diferentes áreas do conhecimento que fomentam o fluxo da comunicação científica. Curty e Boccato (2005) confirmam que a produção científica deve ser repassada à comunidade para gerar novos impulsos ao crescimento por meio de descobertas científicas:

“A socialização da informação vem contribuir para o avanço de estudos e pesquisas, indo ao encontro das necessidades do pesquisador científico e da comunidade. Esta última irá beneficiar-se com a aplicação das descobertas do pesquisador no aprimoramento de áreas do conhecimento como a das humanidades, a da saúde e a da tecnologia, entre outras, o que lhe proporcionará bem-estar e melhoria em sua qualidade de vida. A publicação científica é essencial para a pesquisa, que só passará a existir a partir do momento em que é publicada” (CURTY e BOCCATO, 2005, p. 95).

---

<sup>23</sup> *Corpus* é o nome dado ao conjunto de textos digitalizados.

Marcondes (2009) mostra a dimensão que é essencial para a compreensão da importância do documento científico para a sociedade: a publicação deste tipo de hiperdocumento pode ser caracterizada em dois procedimentos regulares, conhecidos como citação e referência. Tais procedimentos destacam diálogos estabelecidos entre o texto atual (nova contribuição) e os textos anteriores (contribuições antigas), revelando que a construção do conhecimento se realiza coletivamente por operações de reconhecimento.

O processo de comunicação científica é revestido pelos aspectos sociais e cognitivos. Os aspectos sociais porque envolvem ações e os aspectos cognitivos porque estabelecem vínculos entre teorias e metodologias geradas no universo científico de um domínio especializado. A reutilização da informação torna evidente que a contribuição científica é uma construção coletiva que representa o conhecimento do estoque teórico e metodológico de uma ciência, como também o reconhecimento de contribuições realizadas em épocas passadas; e demonstra diálogos, desempenhando papéis semânticos e sociais na construção do conhecimento (MARCONDES, 2009).

Considerando que as novas TIC modificaram drasticamente as formas de lidar com a informação, essas mudanças vêm provocando contínuas transformações que afetam também a estrutura do fluxo da comunicação científica e do ciclo de interação entre tecnologia e conhecimento registrado. Miranda e Simeão (2002) destacam a importância de identificar a tipologia do documento porque é o tipo que predispõe a autoria e, conseqüentemente, os demais elementos do ciclo informacional (Figura 3).

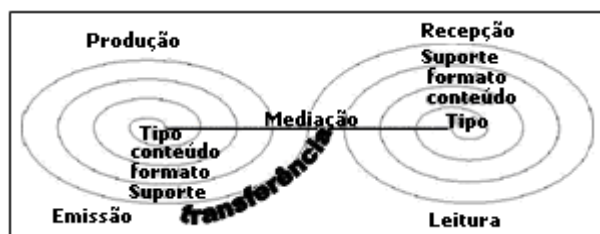


Figura 3 - Interação entre tecnologia e conhecimento registrado.  
**Fonte:** Adaptado de Miranda e Simeão, 2002.

O fluxo da comunicação científica é fomentado por pesquisas e, principalmente, na publicação de documentos como teses e dissertações. A produção do conhecimento científico conformada à tecnologia e aos quatro elementos (tipo-conteúdo-formato-suporte) se dá através da mediação compreendida como absorção das novas ideias, análise, síntese e crítica para a complementaridade do conhecimento acumulado, retornando ao ciclo através de novos documentos e, assim, expandindo a massa documental.

A expansão da massa documental, fenômeno chamado de explosão da informação no conceito de Popper (1975) faz parte deste ciclo de interação entre tecnologia

e conhecimento registrado. Popper apresenta a Teoria dos Três Mundos: a) Mundo 1 é o mundo físico constituído pelo conhecimento relacionado com a biologia, a geologia, a física, etc; b) Mundo 2 é o mundo dos estados mentais constituído pelos conhecimentos subjetivos; e c) Mundo 3 é o mundo do conhecimento objetivo constituído pelo conhecimento registrado.

A CI é uma área de pesquisa típica do Mundo 3, ou seja, de metaciência que estuda o fenômeno dos registros de conhecimentos e trabalha pelo aperfeiçoamento das formas de produção, armazenamento e desenvolveu técnicas para organizar a expansão da massa documental com novos métodos de classificação facetada e rigor científico (MIRANDA, 2005). Martins (2007) alerta que a CI deve se concentrar em organizar o material disponível no Mundo 3 e desenvolver uma teoria sobre a “Filosofia do Conhecimento” e as interações entre os mundos 1 e 2.

Desde suas origens, a CI desenvolveu técnicas de interações entre os mundos 1 e 2, ao gerenciar a estrutura do fluxo da comunicação científica nas práticas documentárias disponíveis no Mundo 3. Por este motivo, as expressões “sociedade da informação” e “sociedade do conhecimento” são, muitas vezes, utilizadas como sinônimas (MARTINS, 2007). Essas técnicas são frutos de um trabalho colaborativo de profissionais com visão multidisciplinar como bibliotecários, arquivistas, documentalistas, professores, analistas de sistemas, engenheiros, entre outros, inseridos no rótulo de “profissionais da informação” na atualidade. Saracevic (1996) identifica três características constituintes da razão de existência e evolução da CI: sua natureza multidisciplinar, sua ligação com as TIC e sua participação ativa e deliberada na evolução da chamada “sociedade da informação”.

Os desafios de organização, representação e tratamento documentário são muito anteriores ao desenvolvimento das tecnologias atuais, tornando extremamente relevante um resgate aos conceitos tradicionalmente usados em bibliotecas, visando identificar possíveis deficiências e melhorias nos processos hoje adotados em modelos de sistemas hipertextuais de uma biblioteca digital. Acredita-se que o entendimento desses processos ajude a automatizar algumas técnicas no MHTX.

### **2.3 Bibliotecas e os desafios da organização da informação**

Uma biblioteca no contexto da universidade é mais do que um espaço de simples acesso físico à informação, criando, assim, uma rede dinâmica para a circulação da informação, visando promover meios para o fluxo da comunicação científica se expandir. Entre os serviços oferecidos pela biblioteca, destaca-se o intercâmbio do seu acervo e o de outras instituições, procurando suprir as necessidades informacionais do meio acadêmico,

além de resguardar a produção intelectual. Esses processos informacionais dependem de pessoas e da tecnologia em seus relacionamentos (BASTOS, 2005).

Há muito tempo, as bibliotecas funcionam “em rede”, dividindo o trabalho de criação de “catálogos coletivos” e procurando garantir o acesso às publicações. Desde a invenção dos computadores, as bibliotecas começaram utilizar seus recursos para melhorar serviços básicos desenvolvendo práticas em torno de coleções das generalidades do conhecimento humano (ALVARENGA, 2001). Cunha (2000) apresenta a evolução das bibliotecas e sua relação com a tecnologia conforme mostra a Figura 4.

Não existe consenso sobre o termo biblioteca digital, diferentes termos podem ser encontrados em documentos científicos: “eletrônica”, “híbrida”, “virtual” ou ainda, “biblioteca sem muros”, “biblioteca sem paredes”, biblioteca do futuro, entre outros. Isso reforça a multiplicidade de vocabulário existente em um sistema de RI. O termo “eletrônica” enfatiza os acervos registrados em meio eletrônico e o “virtual” refere-se ao fato do sistema utilizar tecnologias atuais (KURAMOTO, 2006a, p. 147). Neste trabalho, todos esses termos serão referenciados como sinônimos do termo bibliotecas digitais porque não se pretende discutir as diferenças entre eles, mas enfatizar as semelhanças que notavelmente podem ser observadas em muitos aspectos entre as bibliotecas tradicionais e as bibliotecas digitais.

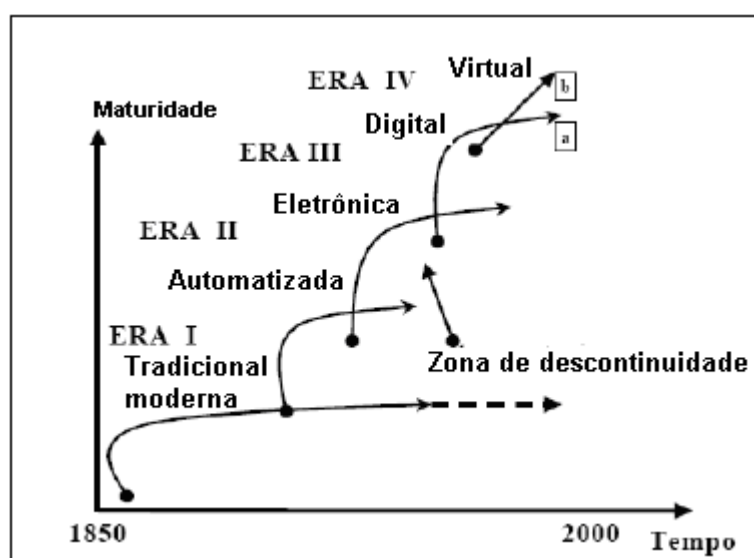


Figura 4 Evolução tecnológica da biblioteca  
Fonte: Cunha, 2000.

As bibliotecas digitais têm as mesmas funções das bibliotecas tradicionais como o gerenciamento e desenvolvimento de coleção, a análise de assunto, a indexação, a catalogação de recursos, a provisão de acesso, a preservação do conteúdo digitalizado, entre outras. A maioria das bibliotecas passa por uma fase de transição híbrida, na qual agregam diferentes tecnologias e fontes de informação tanto em suporte físicos quanto

digitais unindo recursos que não são completamente digitais nem completamente impressos. Esses recursos se apresentam como mais uma forma de acesso e recuperação, proporcionando aos usuários uma ampla divulgação da produção acadêmica e científica desenvolvida na universidade para o crescimento da ciência no país (BASTOS, 2005).

Muitas pessoas desconhecem o fato de terem sido as bibliotecas tradicionais, as primeiras instituições a fazerem uso dos sistemas computacionais de RI, lidando basicamente com a catalogação de documentos científicos e metadados descritivos (ALVARENGA, 2001). A catalogação de recursos vem ao encontro à necessidade de recuperar informação disponível em rede e simplificar a descrição de objetos informacionais para o uso. O MARC (Machine-Readable Cataloging) e o DC (Dublin Core Metadata Element) são exemplos de esquemas de metadados descritivos que ajudam a localizar elementos textuais em um acervo bibliográfico. Esses esquemas são relativamente novos, o DC (Quadro 1) foi criado em 1995 devido à necessidade de equilibrar o aumento dos registros bibliográficos em diversas situações de usos.

**QUADRO 1 – Conjunto de elementos Dublin Core**

<b>Elemento</b>	<b>Descrição</b>
<i>Title</i>	O nome que é dado ao recurso.
<i>Creator</i>	Responsável pelo conteúdo do recurso.
<i>Subject</i>	O assunto sobre o qual o conteúdo do recurso trata.
<i>Description</i>	Uma breve descrição do conteúdo do recurso.
<i>Publisher</i>	Responsável pela publicação do conteúdo do recurso.
<i>Contributor</i>	Outros responsáveis por contribuir com o conteúdo do recurso.
<i>Date</i>	Uma data de um evento no ciclo de vida do recurso.
<i>Type</i>	A natureza ou gênero do conteúdo do recurso.
<i>Format</i>	O formato físico ou digital do recurso.
<i>Identifier</i>	Uma referência única para o recurso dentro de um dado contexto (URI).
<i>Source</i>	Uma referência para um recurso do qual o recurso atual é derivado.
<i>Language</i>	A linguagem do conteúdo do recurso.
<i>Relation</i>	Uma referência para um recurso relacionado ao recurso atual.
<i>Coverage</i>	O escopo do conteúdo do recurso.
<i>Rights</i>	Informações de sobre quem mantém direitos sobre o recurso.

**Fonte:** Dublin Core Metadata Element Set, 2004.

Muitos desses processos ainda são dependentes de mecanismos do século passado. Pereira e Bufrem (2005) destacam que o exercício de produção intelectual no tratamento documentário ainda não foi superado pela tecnologia, especialmente no processo de representação cuja complexidade não pode ser reduzida à intervenção eletrônica. A passagem de um texto original para (re)apresentação no computador é uma operação semântica em que os termos são selecionados em função de ocorrência e interesse para um sistema. Conseqüentemente, muita informação armazenada em bases de dados precisa passar por longos processos de leitura em documentação para extração de conceitos válidos a serem indexados como termos candidatos em um sistema de RI.

Lancaster (1993) alerta que a indexação é um processo subjetivo em vez de objetivo porque duas ou mais pessoas podem divergir a respeito do que trata o assunto de um documento e quais os conceitos ou termos-chave que melhor possam descrever os tópicos selecionados para representação em um sistema. Essa leitura demanda tempo e experiência para condensar o conteúdo em um sistema de significados conceituais terminológicos de um vocabulário comum. Além disso, esses processos precisam re-apresentar um substituto semântico do texto para que este “sistema de conceitos” controle seu uso e viabilize a interface “documento-usuário” (PEREIRA e BUFREM, 2005).

A leitura em documentação tem por objetivo (re)apresentar o conteúdo de um documento em um sistema de conceitos com o uso de linguagens documentárias. O termo linguagens documentárias vem sendo construído com novos instrumentos e técnicas que a CI emprega no aperfeiçoamento do tratamento da informação para a extração de conceitos válidos no processo de indexação (VOGEL, 2009). Assim, quando a linguagem utilizada para (re)apresentar o conteúdo de um hiperdocumento na tela do computador passa por um controle do vocabulário, ela pode receber diferentes nomes tais como metalinguagem, linguagem de indexação, sistema de termos atribuídos, vocabulário controlado, etc.

Lima (1998) destaca que um “sistema” ou “serviço de informação” pode ser denominado biblioteca, centro de documentação, serviços ou sistema de informação, base ou banco de dados, entre outros. Independente de sua denominação, ele tem por função coletar, tratar e disseminar a informação produzida pela sociedade na qual está inserido, garantindo assim, o acesso à cultura por parte de seus membros e possibilitando a sua continuidade. Este “sistema” só poderá transferir informação do seu acervo para o seu “usuário” à medida que a linguagem de representação seja construída a partir do referencial terminológico consensual do domínio que pertence o acervo, independente do suporte físico.

Esta representação é complexa e desafiadora para a organização da informação que se ocupa do estudo dos recursos e dos instrumentos utilizados nos diversos serviços e sistemas de informação para a identificação, extração e descrição da informação registrada (MIRANDA, 2005). A organização da informação no computador, independente de contexto semântico conduziram a vários paradigmas da CI. Lima (1998) afirma que representar o conhecimento acumulado por um domínio é apenas uma parte dos processos do tratamento documentário para que este conhecimento transformado em informação possa ser acessado pelo “usuário” de um “sistema” de RI e efetivar assim a comunicação.

Tanto os processos de organização da informação quanto os de organização do conhecimento produzem dois tipos distintos de representação: a representação da informação, compreendida como o conjunto de atributos que representa determinado objeto

e que é obtido pelos processos de descrição física e de conteúdo documentário, e a representação do conhecimento, que se constitui numa estrutura conceitual representando modelos de mundo (BRÄSCHER e CAFÉ, 2008). Essa multiplicidade de representação do conteúdo documentário gera um (re)trabalho para os profissionais da informação.

Retornando as ideias de Shera, a recuperação deve ser o ponto focal da CI à qual todos os esforços devem ser direcionados. Então, o estudo da informação deve estar baseado na trindade do atomismo, significando a operação tecnológica e do conteúdo documentário, sendo aquilo que é transmitido associado ao contexto de uso, como nos atuais estudos de infometria<sup>24</sup>. Segundo Alvarenga (2003), para que tal recuperação ocorra, torna-se necessário que os profissionais da informação desenvolvam “sistemas representacionais” que estabeleçam a confluência entre a organização cognitiva imposta ao conhecimento pelo seu produtor (representação primária) e a organização conceitual imposta ao documento pelo especialista da informação (representação secundária).

Tudo isso reforça o (re)trabalho desses profissionais em (re)apresentar a informação na tela do computador para, posteriormente, ser recuperada em um sistema de RI. Desde suas origens, a CI tenta aperfeiçoar antigas práticas de organização e representação documentária no computador. A Biblioteconomia está intimamente ligada ao conceito de epistemologia social da CI assim como o estudo da cognição e da “Filosofia do Conhecimento”. Shera (1973) alerta que a Biblioteconomia é uma ciência interdisciplinar e, como tal, deve recorrer aos métodos e ao corpo de conhecimento de outras ciências (MARTINS, 2007). As disputas entre arquivistas, documentalistas e bibliotecários foram fundamentais para o desenvolvimento de novos métodos e técnicas nos sistemas de armazenamento e recuperação de informação (AGUIAR, 2008).

Estudos recentes destacam que a CI se insere no contexto de ciência moderna onde o novo modo de produção e organização da informação envolve diferentes mecanismos geradores de comunicação e conhecimento (TOUTAIN, 2007). Mendonça (2000) analisou alguns pontos em que ocorre a interseção da linguística com a CI: o teórico; o quantitativo (pela visão bibliométrica); o temático (pela representação da informação); o aplicativo (pelos métodos diversos); o ensino (pelas relações curriculares); o tecnológico (pela teoria de sistemas) e o normativo (pelas classificações bibliográficas).

Barreto (2002) apresenta três tempos distintos para o desenvolvimento da CI analisando-a como uma instituição mediadora da relação informação e conhecimento: a)

---

<sup>24</sup> O termo infometria foi proposto em 1987 pela Federação Internacional de Documentação (FID), para designar o conjunto das atividades métricas relativas à informação, cobrindo tanto a bibliometria quanto a cientometria. A infometria abarca as duas primeiras, tendo desenvolvido métodos e ferramentas para mensurar e analisar os aspectos cognitivos das ciências, atribuindo sentido aos dados para melhor uso regional, nacional ou mundial (KOBASHI 2009).



tempo da gerência de informação que vai de 1945 a 1980; b) tempo da relação informação e conhecimento no período de 1980 a 1995; e c) tempo do conhecimento interativo de 1995 até os dias atuais. A Figura 5 mostra a visão de Barreto sobre essas questões.



Figura 5 Sistema de Armazenamento e Recuperação de Informação.  
Fonte: Barreto, 2002.

A visão de Barreto, ao indicar esses três tempos, foi de assinalar as diferentes questões de organização e representação do conhecimento de acordo com os recursos e teorias de cada época. Como a Biblioteconomia se preocupa, essencialmente, com o fluxo interno (seleção, aquisição, catalogação, classificação, indexação, armazenamento, recuperação e disponibilidade para uso de itens de informação), muitos são os problemas que a CI gerencia no tratamento da informação desses fluxos: em uma das extremidades, há a criação da informação e na outra a assimilação da informação pelo receptor, algo que transcende o conceito de uso da informação. A CI depende de romper paradigmas de uma imensa rede científica e interdisciplinar para tratar, armazenar e transmitir a produção social discursiva, documentada e registrada nos sistemas digitais.

O gerenciamento desses fluxos possibilitou o desenvolvimento de muitos instrumentos e técnicas que ajudam na construção de um “sistema de conceitos”. Para desenvolver esses métodos e técnicas, os profissionais da CI se aproximaram de diversas outras áreas, inclusive da Pragmática, numa visão multidisciplinar de ciência moderna, pois reconhece que um campo conceitual de um campo temático possui múltiplas facetas. González Gómez (2004) destaca que a aproximação da Semântica à Pragmática estabelece novas homologias entre as linguagens definidas pelos profissionais de informação e as linguagens daqueles que geram e usam essas informações. Se isso significa, para alguns, que todo modelo seria uma re-apresentação, para outros significa, também, que todas as re-apresentações seriam modelos.

De acordo com Cintra (2002), todas as práticas humanas são tipos de linguagens, já que elas têm a função de demarcar, significar e comunicar. O contexto de um documento possui objetos informacionais com valores semânticos e pragmáticos. Além

disso, uma das funções do vocabulário é representar a informação, o processo de controle de vocabulário pressupõe intervenções na normalização gramatical e o controle semântico dos termos-chave ou controle do significado. Portanto, neste trabalho, adota-se o termo “vocabulário controlado” sempre que se referir a um vocabulário como um produto das práticas documentárias e do processo de controle do significado nos ambientes digitais.

Nos ambientes digitais, Nascimento (2000) resume em cinco os processos do tratamento da informação: a) o processo de seleção que analisa o suporte físico onde está armazenada a informação e a situação do registro deste suporte em relação ao usuário; b) o processo de categorização de assuntos que visa organizar os recursos existentes de forma padronizada por temas e assuntos, seguindo sistemas de codificações específicos ou estabelecidos para a unidade de informação; c) o processo de classificação de documentos que pode compreender o suporte, o formato, o gênero, a espécie, o tipo, a forma, a natureza da informação e o público a que se destina; d) o processo de indexação que atribui índices para expressar o conteúdo de um documento de forma resumida; e e) o processo de catalogação de recursos que utiliza esquemas padronizados que descrevem recursos de quaisquer tipos de objetos para obter a descrição exata desses recursos.

Estas práticas podem ser vistas como re-apresentação em que o profissional que analisa o documento novamente para indexar termos-chave ou extrair conceitos válidos é o mediador do processo de re-significação. Os conteúdos dos documentos, quando estudados, podem integrar diversas áreas do conhecimento, devido às suas diferentes características. Assim, um mesmo documento pode ser descrito em diferentes aspectos, o que acarreta a coexistência de uma multiplicidade de classes. Essas classificações bibliográficas são distribuídas conforme finalidade do sistema. Separados por diferentes pontos de vista, cada documento deve ser sistematizado, com a ordenação de vocabulário que os represente de forma eficiente, sustentado pela garantia literária e que seja reconhecido pelos membros de uma mesma comunidade (NUNES e TÁLAMO, 2009).

A organização da informação em um modelo de biblioteca digital pode ser de forma enumerativa ou em forma de facetas. Em forma enumerativa os objetos são caracterizados pela existência de uma ou mais tabelas que enumeram os assuntos básicos que formam o universo de conhecimento. Em forma de facetas, os objetos são representados em um único conceito ou classe generalizada, tentando sintetizar elementos de um campo conceitual diretamente no contexto semântico para melhor descrevê-los como uma coleção de termos especializados de um domínio e estruturados hierarquicamente.

Devido às práticas documentárias se desenvolverem no universo da linguagem, decorre daí a necessidade de controlar o vocabulário como um recurso normativo para nomear categorias em um sistema de conceitos, de modo que assegure o compartilhamento

e uso da informação mediada por uma linguagem consistente com termos-chave de um domínio. Como o objetivo prioritário a que se propõe a Terminologia é a normalização, estabelecendo as formas de referência e separando as outras variantes para denominar mesmos conceitos, a necessidade de controle de vocabulário é um processo dessas práticas com muitas contribuições dos novos métodos, técnicas e instrumentos da CI.

A Terminologia, ao oferecer um conjunto de termos-chave, tem sido especialmente valorizada para sistematização de bases de dados, arquivos e outros tipos de suportes. O uso de uma terminologia padrão determina igualmente a condição de referência dos termos que integrarão os vocabulários, fazendo com esses vocabulários se remetam às significações estruturadas e registradas no tratamento da informação (DIAS, 2001). Os métodos de análise-síntese de conteúdo balizam-se por critérios como a padronização e a univocidade, qualificadores das principais formas que o representam. Mesmo que os termos candidatos apresentem a mesma forma, a formulação em linguagem documentária funciona como uma metalinguagem e não se confunde com a linguagem do texto, constituindo-se de conceitos, cujos elementos são extraídos do próprio texto (PEREIRA e BUFREM, 2005).

A CI, ao se aproximar da Pragmática e Ciências Cognitivas, pode contribuir na construção de novos modelos conceituais relacionais que melhor se aproximem de modelos mentais como mapeamento de um domínio, representando estruturas de sistemas de “conceitos científicos” como “unidades de conhecimentos registrados”. Descartar a necessidade de linguagens documentárias e de sistemas de classificação revela uma ignorância grande de suas funções nos modelos de sistemas de RI, pois se algum conceito importante escapar na análise de escolha para escolha dos termos candidatos, não pode ser indexado pelo sistema para depois ser recuperado (DIAS, 2001).

Ao observar a dinâmica que o usuário estabelece na relação com um documento, Dias (2004) elaborou um importante estudo sobre as técnicas usadas no tratamento documentário que indicam novos meios que podem ajudar na extração de conceitos válidos no MHTX: a) nas perspectivas dos profissionais de informação, a análise de assunto é feita por profissionais experientes em sistemas clássicos de classificação bibliográfica que ajudam a identificar “conceitos científicos” como termos-chave de um domínio; b) nas perspectivas dos profissionais dos usuários, baseia-se no pressuposto de que oferecem índices para subsidiar ações que visem ao aperfeiçoamento dessas técnicas; e c) na terceira e fundamental perspectiva que deve ser baseada na visão do próprio autor para identificar do que trata o documento que produziu.

O resultado deste estudo sugere padrões de classificação em fundamentais/periféricos e genéricos/específicos. Considerando que a situação ideal da análise de assunto de um domínio do conhecimento deveria ser realizada pelo próprio autor

para identificar e selecionar termos-chave na publicação de um documento, buscam-se novas ferramentas na IA com os padrões de metadados W3C para modelagem informacional. Assim, o novo sistema do MHTX precisa permitir que o recurso computacional possa ser representado e descrito por um “objeto informacional” mapeado em um único mapa. Acredita-se que assim a ontologia possibilitará a compatibilização e interoperabilidade semântica de forma que o computador juntamente com a ferramenta de modelagem possa identificar “conceitos científicos” e relacioná-los com “conceitos espontâneos” durante a leitura de sistemas hipertextuais.

De fato, a análise de assunto realizada nas perspectivas dos profissionais da informação é a parte mais importante do processo de indexação porque condiciona os resultados de estratégia de busca. Como um modelo influencia diretamente no modo de operação do sistema, a eficácia de um sistema de RI está diretamente ligada ao modelo que o mesmo utiliza (DIAS, 2004). Acredita-se que a IA pode contribuir com modelos conceituais por prototipação e princípios de gramaticalização de conceitualização unificando os processos do tratamento documentário com linguagens documentárias em modelos de bibliotecas digitais. A IA utiliza um modelo teórico como fonte da cognição que pode avançar as pesquisas da CI (SARACEVIC, 1996).

## **2.4 Ontologias e seus benefícios no planejamento de sistemas**

O termo ontologia tem suas origens na Filosofia com estudos metafísicos introduzidos pelo filósofo grego Aristóteles que determinou as diretrizes dos sistemas de classificação, dando a distinção entre o aspecto formal e material do ser. Chauí (2000) destaca três critérios para classificação dos saberes utilizados por Aristóteles: a) critério da ausência ou presença da ação humana nos seres investigados; b) critério da imutabilidade ou permanência e da mutabilidade ou movimento dos seres investigados; c) critério da modalidade prática, levando à distinção entre ciências que estudam a práxis e as técnicas.

Esses critérios descrevem a linguagem numa designação lógica como um processo de predicação em que cada item lexical pleno de significado é um predicador. A predicação por designação gera o sentido, que decorre da relação entre o item e seu referente (CASTILHO, 2007). Segundo Brandão (2005), Aristóteles estabeleceu as bases teóricas do raciocínio lógico (Figura 6) e da abordagem analítico-sintética para a combinação de conceitos elementares e compostos na estruturação de categorias.

Ao aplicar critérios de classificação comuns aos seres, Aristóteles contribuiu para o estabelecimento de dez categorias, cujo número não é tão importante como a diferença que deve ser estabelecida entre a primeira e as nove restantes: Essência (substância), Quantidade, Qualidade, Relação, Lugar, Tempo, Posição, Estado, Ação e

Paixão (SCHREINER, 1979). Substância é a categoria mais importante, pressuposta por todas as outras que nada mais são do que suas características ou propriedades (SCHREINER, 1979).

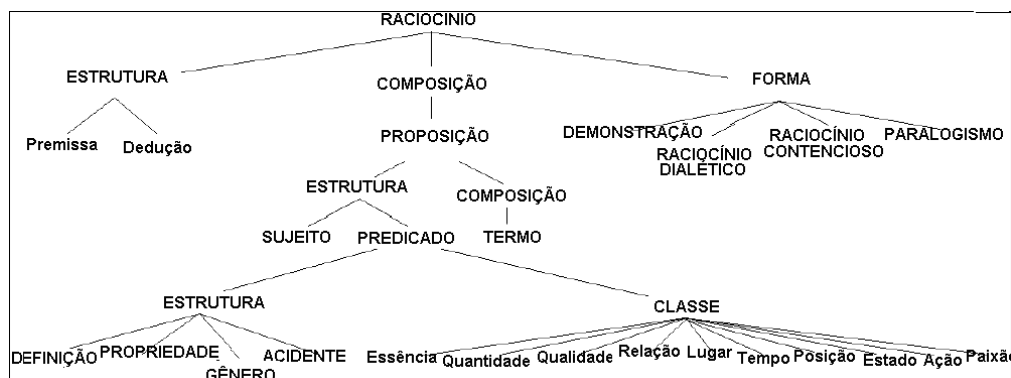


Figura 6 – Categorias de Aristóteles.  
**Fonte:** Brandão, 2005.

O conceito de categorias formais em forma de facetas proposta por Shiyali Ramamrita Ranganathan (1892-1972), Matemático e Bibliotecário, é algo totalmente novo no campo da classificação bibliográfica. Os princípios de Ranganathan com a abordagem analítico-sintético são em oposição ao princípio hierárquico-enumerativo, analisando o assunto em seus elementos constituintes fundamentais, formais e materiais, que são os únicos representados nas tabelas. Ranganathan analisou a síntese destes elementos, através de símbolos de conexão e relação apropriados, de acordo com o conteúdo dos documentos a serem classificados em um modelo de biblioteca.

Lima (2004) sugere a classificação facetada como alternativa aos sistemas enumerativos e utiliza esses princípios na organização semântica na base de dados MHTX para propor a navegação em contexto. A Teoria da Análise Facetada utiliza uma metodologia de organização do conhecimento que permite ao usuário-sistema ver os assuntos de um documento sob diferentes pontos de vista, buscando o mapeamento do conhecimento de uma determinada área específica (LIMA, 2004). Neste trabalho, é importante compreender as propostas dos instrumentos utilizados nos sistemas de classificação da CI com categorias formais tanto na visão de Aristóteles, Ranganathan e Dahlberg para construção de uma ontologia como um sistema de conceitos integrados.

#### **2.4.1 Teoria da Análise facetada e suas relações com os princípios filosóficos de categorias formais nos sistemas de classificação**

As ideias de Ranganathan sobre a Teoria da Análise Facetada podem ser encontradas em cinco obras básicas: *Five Laws of Library Science*, 1931, *Colon Classification*, 1933, *Prolegomena to Library Classification*, 1937, *Philosophy of Book Classification*, 1951, e *Elements of Library Classification*, 1962. Ranganathan não somente

organizou um domínio de um campo especializado, como também provocou grandes impactos na organização do conhecimento, ao promover duas abordagens de holismo: uma direcionada no espaço, representada pela Cultura Brâmane e pela Astrologia, e a outra direcionada no tempo, representada pela Cultura Chinesa (SEPÚLVEDA, 1996).

A classificação facetada, também conhecida como taxonomia facetada ou ainda classificação de dois pontos, é uma nova forma de analisar um domínio de conhecimento com categorias formais exigidas pelo método científico. Campos e Gomes (2005) destacam que a classificação facetada apresenta uma “policotomia ilimitada”. Ranganathan buscou uma alternativa na Árvore Baniana<sup>25</sup> (*Banyan Tree*) para explicar de forma aceitável a sua proposta fazendo uma analogia entre a Árvore de Porfírio em que floresceram as categorias de filósofos como Aristóteles na representação esquemática dicotômica.

Na dicotomia, encontram-se duas divisões no primeiro estágio, duas divisões de cada uma destas divisões são formadas no segundo estágio e assim por diante. No âmbito da representação documentária, a metodologia dicotômica falha porque as categorias de assuntos não fazem parte apenas de um domínio. O rigor do método apresentado por Ranganathan para a ordem dos elementos em seus diversos níveis (classes e subclasses) se justifica porque tal ordem é a base para o sistema de notação conceitual e é este que vai mecanizar o arranjo dos livros ou documentos nas estantes de uma biblioteca (CAMPOS, 2003). Segundo Tristão (2004), este método permite analisar o assunto de um documento, fragmentando-o em suas partes constituintes.

Três planos podem ser destacados no método científico de Ranganathan: a) o plano de ideias que pode ser tomado como análise conceitual, no qual existe a formação do processo do pensar que dá origem às ideias; b) o plano verbal que permite que a linguagem seja a mediadora para a comunicação das ideias e conceitos; e c) o plano notacional que é usado na representação dos conceitos (MOREIRA, 2005). Miranda (2005) considera a espiral do método científico apresentado por Ranganathan como um metaespiral (Figura 7) regido pelas mesmas leis do movimento contínuo e dinâmico de uma ciência, propiciando integração constante do conhecimento, do desenvolvimento de assuntos e a relação com a organização e estruturação da informação.

---

<sup>25</sup> Árvore Baniana é um tipo de figueira indiana que se espalha por uma grande área enviando galhos para o solo, os quais criam raízes formando vários troncos.

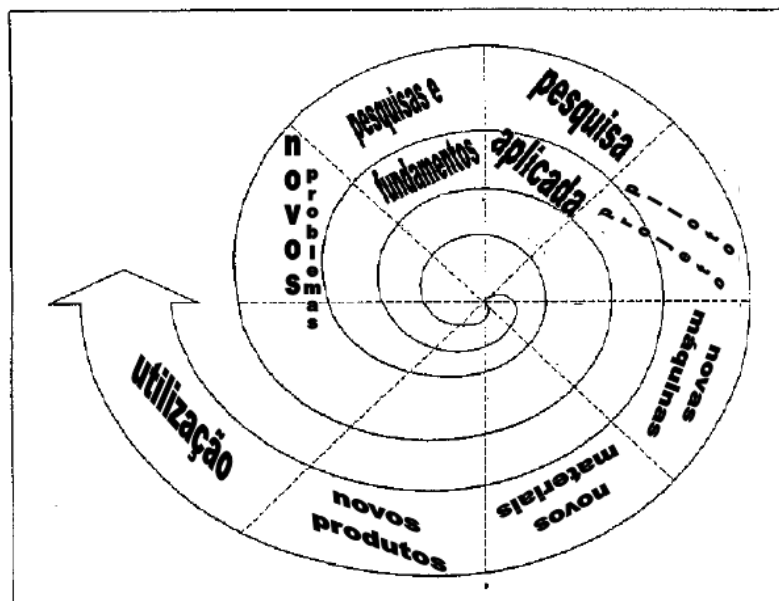


Figura 7 – Espiral do método científico de Ranganathan.

Para Ranganathan, o universo de ideias pode ser dividido em dois tipos: assuntos e isolados. Assuntos complexos são formados por dois ou mais assuntos. Assuntos compostos surgem da combinação entre básicos e isolados. Ranganathan decompõe elementos mais complexos em conceitos básicos e isolados de idéias, procurando examinar cada uma dessas partes, para, posteriormente, uni-las de acordo com as características que serão descritas e representadas por um assunto de um documento.

Primeiramente, Ranganathan apresenta os princípios normativos gerais, depois os cânones para construção das classes, depois, os cânones que ordena os elementos dentro delas e, finalmente, estabelece as categorias fundamentais do seu modelo de classificação como princípio para ordenação da classificação facetada (CAMPOS, 2003). Ranganathan analisou assuntos contextuais com uma abrangência conceitual relacionada a diversos núcleos, dependendo da forma lógica como as unidades de conhecimento se relacionam, ou seja, funcionando como portadoras de significados. As cinco facetas que formam a Teoria da Análise Facetada, denotado pelas iniciais PMEST (*Personality/Entity, Matter, Energy, Space e Time*) correspondem às categorias de nível hierárquico mais abrangente como uma primeira divisão do conhecimento no universo de ideias.

Para Campos (2003), a classificação facetada é muito flexível. Com base nos postulados e princípios de Ranganathan, as facetas são extraídas sem qualquer ideia pré-determinada acerca de seu número ou de suas sequências. Um sistema de conceitos em forma de facetas possibilita estruturar o conhecimento de maneira que os assuntos compostos sinteticamente possam surgir a partir de conceitos "elementares" registrados em um documento. Esses princípios utilizam a técnica analítico-sintética para a identificar e

combinar assuntos: a análise é o processo de decomposição de um tema em seus conceitos básicos e a síntese é a combinação dos conceitos para descrever um assunto.

Sobre as cinco facetas do PMEST, Svenonius (1985) descreve que *Time* é auto-descritiva, e por isso mais fácil de ser reconhecida; *Space* geralmente se manifesta por meio de áreas geográficas, e também não apresenta dificuldades para identificação de suas ocorrências; *Energy* geralmente conota algum tipo de ação; *Matter* abriga manifestações de materiais e propriedades; *Personality/Entity* corresponde ao conceito mais difícil de identificação e, por esse motivo, geralmente os conceitos que não se encaixam em nenhuma das categorias anteriores devem ser alojados nesta categoria (SILVA, 2008).

Para Ranganathan, um termo é a representação verbal de um conceito em uma linguagem natural. Cada termo é representado por um elemento notacional para que o assunto possa ser descrito através de um arranjo ordenado desses elementos. Uma classe é um conjunto de coisas ou ideias que possuem vários atributos, predicados ou qualidades comuns. O agrupamento de isolados com base em suas características comuns formam as facetas. Uma idéia isolada pertencente ao universo selecionado para análise e a partir do momento em que o isolado é inserido no esquema de classificação, passa a ser referenciado como “foco”. Quando um termo genérico é usado para denotar algum componente ele é chamado de faceta. Uma faceta é, portanto, um conjunto de isolados agrupados por possuírem certas características comuns (SILVA, 2008).

Silva (2008) destaca que o modelo simplificado de Spiteri (Quadro 2) propõe uma compilação simplificada da Teoria da Análise Facetada, composta de diversos princípios, agrupados de acordo com o plano de trabalho mental com o qual este modelo simplificado foi construído. É interessante observar os cânones postulados do plano de ideias deste modelo que foram organizados em dois princípios: (1) Princípio para ordem de citação das facetas e (2) Princípio da sucessão relevante e consistente.

Os três planos (ideias, verbal e notacional) foram agrupados para eliminar redundâncias internas na construção de soluções de classificação adequadas aos domínios modelados que contribuiu para a divulgação dos esquemas, muitas vezes adotados intuitivamente nos modelos computacionais. No plano notacional, encontra-se também o princípio da hospitalidade, ou seja, a propriedade representativa da capacidade existente na estrutura classificatória em abrigar novos conceitos oriundos do plano de ideias.

#### **QUADRO 2 – Planos e princípios do modelo simplificado de Spiteri**



Plano	Subgrupo	Princípio	Descrição
Idéias	Princípios para Ordem de Citação das Facetas	Princípio da Sucessão Relevante e Sucessão Consistente	A ordem de citação das facetas deve ser relevante ao sistema de classificação. Não existe uma única ordem aplicável a todos os sistemas. Exemplos: ordem alfabética, espacial, cronológica, quantidade crescente ou decrescente, nível de complexidade, etc.
Verbal	X	Princípio do Contexto	Os termos utilizados não devem ser ambíguos no universo considerado. As classes mais genéricas identificam o contexto de aplicação de um termo.
Verbal	X	Princípio da Aceitação	Os termos utilizados devem ser aceitos por conhecedores do universo de assunto analisado. Não devem ser utilizados termos obsoletos.
Notacional	X	Princípio da Sinonímia	A exclusividade mútua deve ser mantida na notação, ou seja, cada assunto deve ser representado por um e somente um número de classe.
Notacional	X	Princípio da Homonímia	Cada número de classe deve representar somente um assunto.
Notacional	X	Princípio da Hospitalidade	Uma notação deve permitir a representação de novos assuntos.
Notacional	X	Princípio da Ordem de Arquivamento	A notação deve refletir a ordem da tabela utilizada no sistema de classificação.

**Fonte:** Adaptado de Silva, 2008.

Segundo Lima (2004), a classificação em forma de facetas é mais útil na modelagem conceitual de um sistema de hipertexto porque permite ao usuário ver um assunto sob diferentes pontos de vista, do mesmo modo que comungam da não linearidade hipertextual. Todo assunto de um documento é oriundo de um assunto básico e seu componente vem de uma ou mais ideias isoladas, formando assim o assunto composto. Um assunto seria como um corpo de idéias. O universo de isolados não é considerado um assunto sem o seu contexto, mas pode ser moldado para formar um componente de um tópico em uma estrutura de categorias formais de um sistema de conceitos integrados, conforme é a proposta deste trabalho. A Figura 8 ilustra o pensamento de Ranganathan na formação de ideias no cérebro humano.

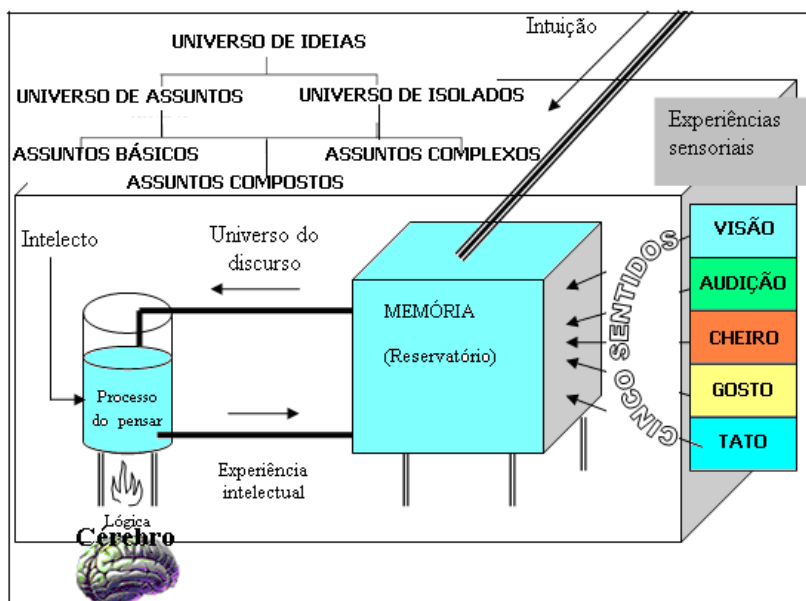


Figura 8 – Compartimento do cérebro: formação da ideias.

Fonte: Adaptado de Lima, 2004.

A lógica do raciocínio humano representa alguma ideia ou complexo de ideias que podem ser moldadas para formar um sistema de conceitos integrados. É interessante observar a presença dos cinco sentidos relacionados ao raciocínio lógico propostos por Ranganathan. Os cinco sentidos são as bases da aprendizagem sob diferentes formas com que o cérebro humano consegue processar um conceito em uma intrincada rede de neurônios na busca por uma informação, utilizando órgãos variados que são: a) visão: o olho capta a imagem e envia para o cérebro, para que este faça o reconhecimento e interpretação; b) audição: o ouvido capta as ondas sonoras e as envia para que o cérebro interprete o som; c) paladar: a superfície da língua capta o sabor dos alimentos e envia as informações ao cérebro, através de milhões de neurônios; d) tato: abaixo da pele humana existem neurônios sensoriais que permite “tocar” objetos externos e através desta ação, estabelecer sentidos e relações; e) olfato: o nariz capta os odores do mundo exterior.

A visão, audição e o tato são os recursos que precisam ser mais bem explorados em um modelo de biblioteca digital interativa não se importando em discutir a presença ou a ausência do autor, mas o texto em si, como uma estrutura livre e com características próprias de existência (BARRETO, 2001). Em um modelo de biblioteca digital interativa e especializada deve haver espaço para publicação de todos os hiperdocumentos científicos produzidos na Universidade e os mais diversos tipos de atividades, sejam elas através de informações escritas, audiovisuais e multimídia.

Acredita-se que o MHTX possui muitas vantagens que possibilitem transformá-lo em um modelo de Biblioteca Digital Interativa da Ciência da Informação doravante

denominada BDICI. A principal vantagem é a navegação em contexto em forma de facetas em que os links mapeados permitem mostrar as relações entre os conceitos de um vocabulário controlado. A ideia principal é que a base de conhecimento deste repositório pode ser representada em um único mapa de “conceitos científicos” com categorias formais.

No MHTX, os conceitos podem ser organizados em uma rede semântica estruturada de forma a permitir uma posterior recombinação com os assuntos tratados nos documentos da base de dados para a navegação em contexto utilizando as classes em forma de facetas. Segundo Campos (2003), os cânones, princípios e categorias fundamentais do método científico proposto por Ranganathan se aplicam tanto à apresentação de uma taxonomia quanto à apresentação de um sistema de conceitos em um tesouro ou ontologia de domínio porque as bases teóricas de ambos são as mesmas.

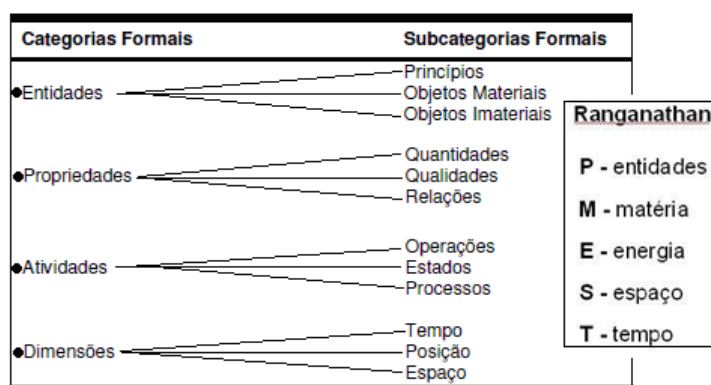
Os conceitos que constituem os assuntos de um documento podem ser estruturados para expressar relações entre eles com o propósito de representação e organização da informação. Uma característica é a qualidade ou atributo escolhido para servir de base à classificação ou à divisão, a partir da qual geralmente formam-se renques e cadeias. Renques são classes formadas a partir de uma única característica de divisão e que são estruturados em fileira de assuntos correlatos. Cadeias são séries de classes, geradas por subdivisões sucessivas que se movem de forma descendente, de um assunto geral para um assunto específico, formando as relações hierárquicas dos assuntos. Os “focos” são correspondem às espécies de um gênero (LIMA, 2004).

Dahlberg (1979) destaca que um assunto só existe relacionando sujeito e predicado com palavras no contexto a partir de elementos físicos de processos sensoriais e permeados pelas experiências cognitivas de cada indivíduo cujo conhecimento pode ser representado pela totalidade do universo de ideias armazenadas em sua memória. A informação existe quando essas ideias são comunicadas de acordo com uma lógica categórica afirmativa que só é verdadeira se encontra adequação entre aquilo que foi escrito no texto de um documento com aquilo que existe na realidade.

Os conceitos, da forma como são representados consecutivamente no enunciado de um assunto de um documento, segue de preferência o modelo de estrutura de uma oração na voz passiva, em que o objeto aparece na primeira posição, o predicado na segunda, e os complementos nas posições seguintes. Geralmente, a quantidade necessária de tais complementos é determinada pela valência de um verbo especial na predicação de uma frase. Isto depende de certas circunstâncias que podem ser mencionadas dando mais concretude à informação registrada em um documento no plano verbal e notacional. Neste sentido, a linguagem verbal funcional como mediadora do processo de comunicação das ideias e representação dos conceitos (DAHLBERG, 1979).

De maneira geral, cada conceito possui um referente sobre o qual pode realizar afirmações passíveis de verificação. Essas afirmações podem ser sumarizadas por um termo-chave que passe a representar o conceito em qualquer processo de comunicação. Dahlberg (1978), para propor a Teoria do Conceito, aplicou quatro princípios de categorias formais da Teoria da Análise Facetada (Quadro 3) devido à necessidade da categorização formal no uso de uma língua natural local ser muito importante na formação dos sistemas e na combinação de objetos, fenômenos, processos, propriedades e relações.

**QUADRO 3 – Princípios do PMEST utilizados por Dahlberg**



Fonte: Adaptado de Silva, 2008.

Nesse sentido, as categorias formais do PMEST que foram utilizadas por Dahlberg para demonstração estão relacionadas aos aspectos multifuncionais da linguagem em suas diferentes possibilidades de usos em um documento científico, ou seja, este tipo de documento possui múltiplas facetas de um vocabulário registrado. É importante destacar que Dahlberg, antes de apresentar a Teoria do Conceito, não concorda com os princípios das normas terminológicas da época que definia 'conceito' como 'unidade de pensamento'. Dahlberg rejeita a definição de 'conceito' como 'unidade de pensamento' porque o pensamento está no quadro cognitivo de um indivíduo e não pode garantir consistência na organização do conhecimento em um modelo de biblioteca (CAMPOS e GOMES, 2002).

As principais normas terminológicas são elaboradas pela International Standard Organization - ISO que, muitas vezes, funcionam como objetos de tradução/adaptação por organismos de normalização de âmbito nacional. Elas são consequência do desenvolvimento do conhecimento científico interdisciplinar e criação de neologismos (RONDEAU, 1984). Campos e Gomes (2002) destacam que durante a reunião da ISO-TC 37, em que se aprovaria a nova versão da norma terminológica, foi possível aprovar a definição de conceito como 'unidade de conhecimento'.

Este trabalho destaca as normas ISO 704 (Terminology work- principles and methods) e ISO 1087 (Terminology work – vocabulary) devido principalmente no que se

refere às relações entre conceitos para sistemas representacionais. A ISO 1087 apresenta a definição de conceito como uma 'unidade de conhecimento' criada por uma combinação única de características. Essas normas estão de acordo com a Teoria do Conceito apresentada por Dahlberg (1978) que define 'termo' como designação de um conceito numa língua especializada, expressa como "unidade de conhecimento" em um documento.

#### **2.4.2 Teoria do Conceito e os princípios de classificação facetada**

Dahlberg (1978) inicia a sua teoria considerando os princípios formais das categorias do PMEST em que um conceito é uma unidade de representação do conhecimento que surge pela síntese rotulada de enunciados verdadeiros sobre objetos reais no sentido de informação registrada em um documento. Para Dahlberg a Teoria da Análise Facetada fornece princípios para estruturar classes de conceitos de um domínio e formar um sistema representacional estabelecendo confluência entre a organização cognitiva imposta pelo autor de um documento no uso de uma língua local.

A ideia principal é que o conceito é uma 'unidade de conhecimento' no sentido de informação, registrada em uma linguagem verbal representacional. Dahlberg (1979) destaca que a existência da Teoria da Análise Facetada é capaz de explicar uma quantidade de fatores anteriormente desconhecidos permitindo avaliar modelos representacionais e ainda saber como esses sistemas podem ser aperfeiçoados. Esta teoria possibilita a identificação de um objeto real no momento da determinação do conceito, ao serem inferidas predicções verdadeiras a respeito de um item de referência.

Com a ajuda do uso da língua natural local, é possível formular enunciados a respeito tanto dos conceitos individuais como dos conceitos gerais e elaborar conceitos relativos aos diversos objetos informacionais. As indicações de tempo e espaço no fim de um enunciado permitem atingir uma maior individualidade possível do objeto informacional. Os objetos individuais, expressos pelos conceitos individuais, também podem se referir a objetos gerais que, de certo modo, prescindem das formas do tempo e do espaço. A esses objetos correspondem os chamados conceitos gerais constituídos por abstração a partir de propriedades comuns atribuídas a um objeto ou classe que pode ser representada por um termo especializado em um domínio.

Exemplificando, se considerar o objeto individual deste documento ECI-UFMG, podem-se formular os seguintes enunciados verdadeiros: é uma Escola relacionada à produção acadêmica da área da Ciência da Informação situada em uma universidade do Estado brasileiro. Se considerar o conceito geral, Escola ou Universidade pode-se formular os seguintes enunciados verdadeiros: é constituída de um grupo de pessoas que trabalham com determinada finalidade comum e está localizada em determinado lugar, etc.

Desde que existem diferentes espécies de objetos informacionais e de conceitos, existem também diferentes espécies de características dos conceitos. Cada enunciado, no verdadeiro sentido de predicação, apresenta um atributo predicável do objeto que, no nível de conceito, se chama característica. Muitas vezes não se trata de um atributo a que corresponde uma característica, mas de uma hierarquia de características, já que o predicado de um enunciado pode tornar-se sujeito de novo enunciado e assim sucessivamente até atingirmos uma característica tão geral que possa ser considerada uma categoria de um conceito.

Segundo Dahlberg, o conhecimento das características dos conceitos facilita a determinação do número de funções que elas exercem em um sistema de classificação que pode ser: a) ordenação classificatória dos conceitos e seus respectivos índices; b) definição dos conceitos; c) formação dos nomes dos conceitos. Na comparação entre as características dos conceitos, aplicam-se também, ao menos em parte, os seguintes tipos de relacionamento semântico entre os conceitos: a) relação hierárquica (implicação); b) relação partitiva; c) relação de oposição (negação); e d) relação funcional (interseção).

Dahlberg distingue as características em dois tipos: a) características simples, características que se referem a uma única propriedade; e b) características complexas, características que dizem respeito a mais de uma característica. Como a ordem das características depende sempre dos objetos cujos conceitos são constituídos pelas mesmas características, Dahlberg sugere especificar essas características em essenciais (necessárias) e acidentais (adicionais ou possíveis).

As características essenciais são indispensáveis para o entendimento do conceito e sua falta pode levar ao entendimento errôneo ou incompleto do conceito. Elas podem ser constitutivas ou consecutivas da essência. As características essenciais consecutivas dependem das características essenciais constitutivas. As características acidentais podem ser gerais ou individualizantes. Essas características dependem de fatores externos e de condições acidentais.

Apesar de não ser fácil determinar as características essenciais dos conceitos, é importante identificar os tipos de relacionamento entre conceitos para determinar a ordem dos conceitos em um sistema de classificação. Sempre que diferentes conceitos possuem características idênticas deve-se admitir a existência de relações entre eles. Este fato tem importância na ordenação dos conceitos e nas suas relações com outros conceitos.

#### **2.4.2.1 Relação hierárquica entre conceitos**

Se dois conceitos diferentes possuem características idênticas e um deles possui uma característica a mais do que o outro, então entre eles se estabelece a relação hierárquica ou relação de gênero e espécie. Pode-se então falar de conceitos mais amplos

ou mais restritos ou, de conceito superior e inferior. O conceito superior é o mais genérico e o inferior é o mais específico. Outro tipo de relação hierárquica é a que existe entre os conceitos específicos do mesmo gênero.

#### **2.4.2.2 Relação partitiva entre conceitos**

A relação partitiva existe entre um todo e suas partes constituindo, também, a relação que existe entre um produto e os seus elementos como, por exemplo, uma árvore (todo) e suas partes (raízes, tronco, galhos, folhas, flores, frutos). Dahlberg destaca que é fácil verificar que as relações hierárquicas e as partitivas aplicam-se principalmente a conceitos que expressam objetos.

#### **2.4.2.3 Relação de oposição entre conceitos**

As relações de oposição podem ser de contradição entre os conceitos, por exemplo, numérico ou não numérico, presente ou ausente, branco ou preto. Essas relações se aplicam principalmente a conceitos que expressam propriedades.

#### **2.4.2.4 Relação funcional entre conceitos**

Para Dahlberg, as relações funcionais aplicam-se, sobretudo a conceitos que expressam processos. A autora exemplifica, questionando sobre as palavras produção, produto, produtor, comprador ao tratar da valência semântica do verbo. Assim, pode-se dizer que entre o conceito do processo "medição" e os conceitos dos complementos mencionados existem relações funcionais, podendo ainda existir outros suplementos adicionais como: as circunstâncias da medição, o tempo e lugar da mesma medição. Tais suplementos podem tornar-se necessários para a individualização do conceito.

A relação funcional entre conceitos permite conhecer o caráter semântico de tais relações tendo por base as chamadas valências semânticas dos verbos, dando atenção aos verbos e respectivos complementos. Essa valência semântica é caracterizada como a soma dos lugares a serem preenchidos de acordo com a ligação de um conceito com outro e pode variar de acordo com o contexto, por exemplo, medição (objeto medido e fins da medição) como instrumento de medição e graus de medição.

#### **2.4.2.5 Elementos para formação de estruturas conceituais**

A Teoria da Análise Facetada permite criar classes no momento da análise de assunto de um documento além de possibilitar processos de análise e síntese na identificação e combinação dos conceitos durante a leitura de documentos textuais (DAHLBERG, 1978a, p. 21). Esta teoria fornece inúmeras possibilidades de combinações de categorias de conceitos que dependem da linguagem local.

No processo de análise de categorias de conceitos, as características dos conceitos expressas nos objetos informacionais textuais são também conceitos que estão

presentes na definição e nas relações entre esses objetos. A combinação desses elementos permite a representação de uma infinidade de assuntos referentes a uma dada área de conhecimento que pode mostrar um mapeamento de “conceitos científicos” de um vocabulário usado em determinado período, estabelecendo conexões e analogias em uma rede semântica representada em um único mapa conceitual.

Muitas destas combinações dependem do uso da linguagem verbal escrita, pois foi este tipo de linguagem que possibilitou o ser humano elaborar enunciados relacionados com determinados objetos informacionais. Os objetos individuais estão ‘aqui e agora’, caracterizados por requerer presença das formas do tempo e espaço. Por exemplo, este documento que você está lendo neste momento possui diversas informações sobre objetos textuais que um usuário-leitor-sistema possa considerar relevante. Esses objetos podem ser pensados como únicos e distintos dos demais de acordo com suas características.

A comparação das características dos conceitos facilita a determinação do número de funções que elas exercem em um sistema de classificação. Dahlberg destaca a facilidade de verificar que os elementos contidos nos conceitos gerais são possíveis de serem reduzidos aos conceitos individuais e ordená-los hierarquicamente. Entretanto, nem sempre é fácil determinar as características essenciais dos conceitos.

Nessa comparação, os tipos de relacionamentos: identidade, implicação, intersecção, disjunção e negação são as relações lógicas possíveis entre dois conceitos diferentes que possuem uma ou duas características em comum conforme mostra o Quadro 4. Para exemplificar, se existe o objeto “plantas” e fenômenos crescimento com processos de arar o solo, existe uma característica comum que é atividade agrícola. Então, tais conceitos possuem relações lógicas entre si que podem combinar suas características.



**QUADRO 4 – Características e relacionamentos entre conceitos**

A. Objetos. Ex.: plantas, produtos, papel, etc.	<b>A + B = crescimento das plantas</b>	
B Fenômenos. Ex.: crescimento, chuva, tráfego, etc.	identidade A (x, x, x) B(x, x, x)	As características são as mesmas;
C Processos. Ex.: imprimir, sintetizar, etc.	implicação A (x, x) B(x, x, x)	O conceito A está contido no conceito B;
D Propriedades. Ex.: cego/cegueira, suave/suavidade,	intersecção A (x, x, o) B(x, o, o)	Os dois conceitos coincidem algum elemento;
E Relações. Ex.: causalidade, necessidade.	disjunção A (x, x, x) B(o, o, o)	Os conceitos se excluem mutuamente. Nenhuma característica em comum;
F Dimensão. Ex.: espaço, tempo, posição, etc.	negação A (x, x, o) B(o, x, o)	O conceito A inclui uma característica cuja negação se encontra em B.

**Fonte:** Adaptado de Dahlberg, 1978.

O agregado das características que constituem o conceito determina a sua intenção. A intenção do conceito é definida como a soma total das suas características. Em alguns casos as características podem estar apresentadas de forma tão geral que pode ser considerada uma categoria. A partir da determinação da intenção, também é possível determinar a totalidade ou o número de conceitos atingidos, determinando sua extensão. A extensão do conceito é entendida como a soma total dos conceitos mais específicos que possui ou a soma dos conceitos para os quais a intenção é verdadeira. A extensão pode ser de um conceito genérico em relação aos específicos ou dos possíveis conceitos individuais.

A extensão e a intenção são maneiras de apreensão e identificação do conceito, e influenciam a elaboração de sua definição. Um conceito é criado pela predicação sobre um objeto de concernência (referente). Qualquer predicação sobre esse referente produz uma característica do conceito. Assim, a soma total das características dessas predicções possíveis irá compor a soma total das características de um conceito. Para Dahlberg, a definição é, de certo modo, uma limitação, ou seja, uma colocação de limites. Trata-se de determinar ou fixar os limites de um conceito ou ideia.

A primeira distinção que se costuma fazer é a separação entre definições nominais e definições reais. Mas existem outros tipos de definições. Dahlberg utiliza três elementos A, B e C representados no triângulo conceitual: a) referente (objeto que se pretende conceituar); b) características (soma dos enunciados verdadeiros sobre o referente); c) forma verbal (síntese das afirmações em forma verbal usada para representar o referente). A Figura 9 apresenta os três passos para a construção de conceitos de objetos informacionais registrados em um documento textual. O item de referência (A) é o ponto de partida para o processo de determinação do conceito, ou seja, a seleção do referente num determinado domínio de conhecimento. A atribuição de características (B) são as asserções

corretas ou predicções verdadeiras acerca do item (A) cuja forma verbal ou termo de referência (C) pode-se chegar através das características à denotação do conceito, representado por um termo.

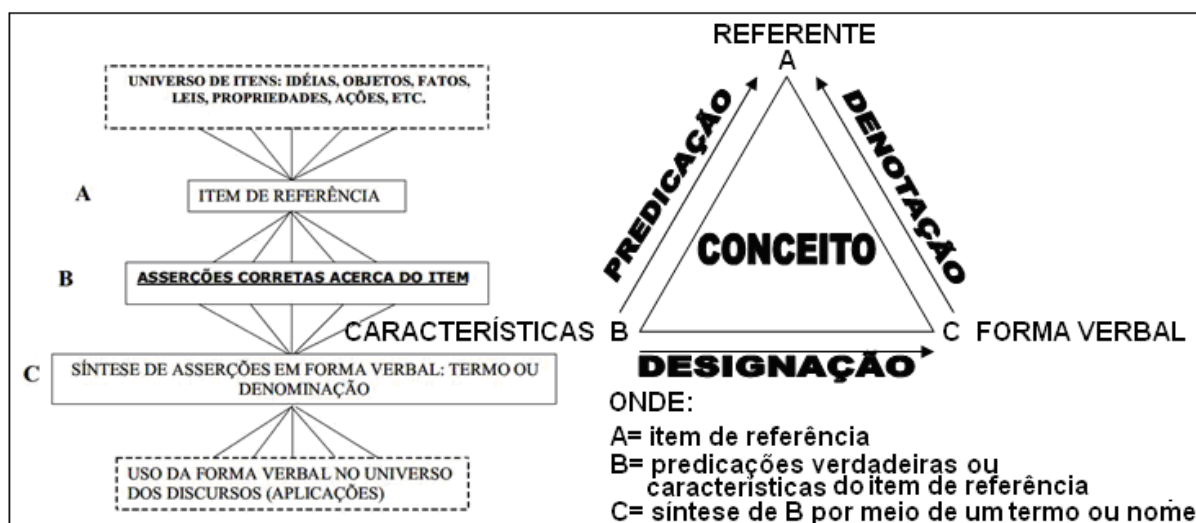


Figura 9 Triângulo conceitual e modelo para a construção dos conceitos.

Fonte: Adaptado de Dahlberg, 1978; Silva, 2008.

Para Dahlberg, ao escrever uma definição pode-se usar símbolos equacionais como na matemática e acrescentar símbolos para expressar a intenção de efetuar uma definição. Fazer uma definição equivale a estabelecer uma "equação de sentido" sendo que, de um lado (à esquerda da base do triângulo conceitual) encontramos aquilo que deve ser definido 'o *definiendum*' e de outro (à direita da base do triângulo conceitual) aquilo pelo qual alguma coisa é definida 'o *definiens*'.

A definição nominal tem por fim a fixação do sentido de uma palavra, enquanto que a definição real procura delimitar a intenção de determinado conceito distinguindo-o de outros com idênticas características. Este tipo de definição relaciona-se com o conhecimento contido na linguagem verbal registrada. Para se equipar a definição de algo ainda não conhecido (o elemento colocado à direita da base do triângulo conceitual) pode-se delimitar ou fixar o conteúdo de um conceito (conteúdo do conceito = intenção, ou conjunto de características ou atributos).

Assim, Dahlberg conclui sua teoria do conceito e define um conceito como a compilação de enunciados verdadeiros sobre determinados objetos reais representado por um símbolo linguístico registrado em um documento. Este símbolo pode ser formado de sinais ou conjunto de sinais cujo signo linguístico é o elemento de representação do objeto informacional que traduz e fixa essa compilação de proposições verdadeiras referente a "unidades de conhecimento registrado".

### **2.4.3 Palavra, termo, conceito, definição e suas relações com os princípios filosóficos de uma ontologia**

O uso de uma palavra em determinado contexto de um documento, não se dá por acaso. Saussure define palavras como signos linguísticos numa visão centrada sobre a ideia de estrutura dicotômica da linguagem, distinguindo o signo em dois elementos: 'significante' e 'significado'. Nessa definição, Saussure mostra que o papel característico da língua frente ao pensamento não é criar um meio fônico material para a expressão das ideias, mas servir de intermediário entre o pensamento e o som, vendo o signo lingüístico como uma entidade psíquica de duas faces (CARVALHO, 2003).

Isso significa que tal sistema lingüístico deve ir além das relações entre signos, tratada pela sintaxe, e considerar as relações entre os signos e as coisas do mundo dentro de um contexto semântico como objetos informacionais registrados. Na CI e na Terminologia, palavras são diferentes de termos. A palavra se refere ao uso da língua em discurso na comunicação e os termos são palavras utilizadas em situações particulares de comunicação. Considera-se que uma característica é a representação de uma propriedade que serve para delimitar um conceito.

De acordo com a ISO 704 (2000), a combinação de um conjunto único de características é representada numa linguagem de especialidade por uma denominação, isto é, um termo ou nome próprio. O conceito é uma representação mental de alguma coisa ou objeto que pode ser percebido em um texto. Então, é por meio do conceito que se estabelece a relação do termo com o objeto/referente (CAMPOS e GOMES, 2004).

Os conceitos são expressos por linguagens (palavras e gramáticas), mas pertencem ao conhecimento extra-lingüístico sobre o mundo. Por isto, a definição de um conceito é determinada pelo ambiente, atividades e cultura das pessoas que falam uma língua. Na palavra em uso, em funcionamento num dado contexto, os conceitos estão sempre relacionados a outros conceitos, formando um "sistema" como um conjunto estruturado, estabelecendo relações com outros objetos, sendo cada conceito determinado por sua posição neste conjunto.

Segundo Bakthin (2003), todo signo linguístico é ideológico, pois tem um significado que remete a algo fora de si mesmo em processo dialógico-discursivo ou numa dada situação enunciativa. Os gêneros discursivos são dinâmicos, por isso trazem consigo diversas vozes e posicionamentos. Essa dinamicidade e pluralidade de vozes são resultantes de relações interpessoais que se estabelecem no próprio texto. A língua é um fato sócio-cultural utilizada por indivíduos. Isso implica dizer que as palavras registradas em um documento são mobilizadas por um discurso historicamente situados num tempo e num

determinado espaço, uma vez que são frutos da interação entre indivíduos e, por conseqüência, trazem em sua constituição apreciações valorativas em relação aos objetos.

Na definição de Saussure, os signos lingüísticos podem ser representados como a união do sentido de uma palavra com sua imagem (acústica) que passam a funcionar como fonemas de representação mental de um objeto da realidade que se articula entre pensamento e som ou significado e significante. A representação gráfica do signo geralmente é feita pelo conhecido Triângulo de Significado apresentado por Ogden e Richards (Figura 10). Para Suonuuti (2001), a definição é uma declaração que descreve um conceito e permite diferenciá-lo de outros conceitos dentro de um sistema. Assim, Suonuuti adicionou o termo 'definição', estendendo o Triângulo de Significado.

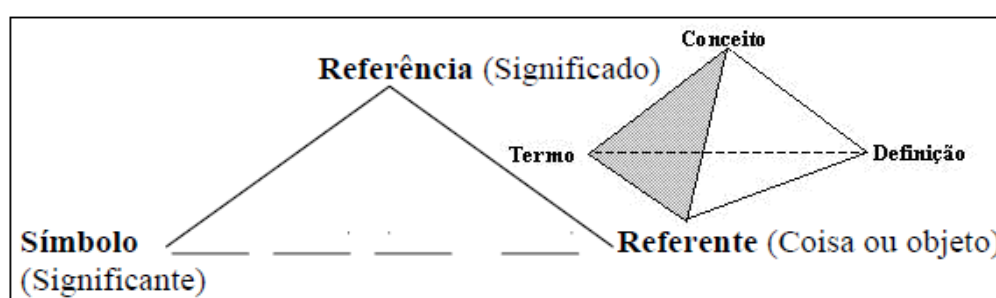


Figura 10 – Triângulo de significado estendido.  
Fonte: Adaptado de Suonuuti, 2001.

Dahlberg (1979) nos lembra que as definições dependem do conhecimento que se tem sobre determinado assunto. Assim como o conhecimento cresce continuamente, aumenta a intenção dos conceitos relativos aos mesmos implicando na criação de novos termos que os representam e também na necessidade de atualização constante desses termos em um modelo conceitual. Para nomear um objeto utilizam-se tantas palavras quantas forem necessárias. A junção das características dos objetos representada na realidade empírica é um elemento de significação do termo.

Segundo Dahlberg (1978) existe uma relação de equivalência entre o termo e sua definição. Os termos das relações de equivalência são chamados de sinônimos. Fazer uma definição equivale estabelecer uma 'equação de sentido' em que fixamos limites de um conceito ou ideia. As relações de equivalência ocorrem entre dois ou mais termos que designam o mesmo conceito (IBICT, 1984 p.22). As características conceituais determinam direta ou indiretamente as relações de equivalência, hierárquicas e associativas.

As relações hierárquicas são aquelas que se definem nas noções subordinadas em um ou vários níveis (ISO 1087). Gomes (1990) destaca que a ordenação dos termos, a partir da relação hierárquica determina uma verticalização: termos superordenados (todas as características) ligados a termos subordinados (pelo menos uma característica a mais).

Essas relações são compostas pelas relações genéricas e partitivas. As relações associativas são as relações que fazem uma conexão entre termos fortemente relacionados que não são termos sinônimos e cuja relação não se enquadra nas relações hierárquicas, como por exemplo: causa/efeito, ausente/presente, preto/branco.

Dahlberg (1978) também destaca a existência da relação funcional que pode ser considerada como um tipo de relação associativa por permitir conhecer o caráter semântico tendo por base as chamadas valências semânticas dos verbos. Os termos nas relações associativas são chamados de conexos e são mais complexas. Cintra (2002) indica a dificuldade encontrada na definição das relações associativas. A maioria dos especialistas primeiramente verifica se a relação é de equivalência ou hierárquica, caso a relação não venha a se enquadrar em nenhuma das anteriores, ela é definida como relação associativa.

Campos e Gomes (2004) questionam a dificuldade de apreender um conceito isoladamente: “como dominar o conceito de 'verniz', sem ter o conceito de 'resina'?” Para dominar ambos os conceitos é necessário dominar as características de cada um dos objetos como “conceitos” a que eles se referem. Assim, se o 'conceito' é uma 'unidade' e o termo denota o conceito, então, pode-se afirmar que o termo é constituído por um grupo de palavras, não existindo, portanto, 'termo composto'.

Todas essas indagações para captar o significado de uma palavra desconhecida ou um termo-chave indexado por um sistema computacional que represente a mesma 'coisa' ou o mesmo objeto informacional são cruciais para se descobrir o significado de estrato ontológico do conceito, até mesmo, para os profissionais da informação. Por isso, acredita-se que este modelo conceitual precisa ser dinâmico, recebendo constantemente atualizações do vocabulário. Há debates e controvérsias, tanto terminológicos quanto substanciais, sobre a natureza do significado das palavras. A semântica é uma área de estudo muito complexa, podendo o significado estar relacionado a questões nebulosas como o conhecimento de mundo e modelos mentais (CHIERCHIA, 2003).

Carvalho (2003) nos lembra que quando se procuram os primeiros princípios e as causas mais elevadas, é evidente que existe necessariamente alguma realidade à qual tais princípios e causas pertencem, em virtude de sua própria natureza. A palavra “conceito” vem da raiz latina *con-cepio*, formada pelo prefixo “*com*”, que significa junto, com o verbo “*cepio*”, que significa “agarrar”, de onde derivam palavras como “catar” e “captar”. Portanto, o significado subentendido na palavra conceito significa agarrar alguma 'coisa' junto com outro objeto referente que pode estar representado em um hiperdocumento científico como uma “unidade de conhecimento registrado”.

Para Carvalho (2003), o conceito é precisamente o instrumento mental que deve permitir captar, ao mesmo tempo, uma palavra ou termo, a ideia ou intenção subentendida e

a coisa real que lhes corresponde em três estratos (ou camadas) do conceito: linguístico, lógico e ontológico. O estrato linguístico corresponde à definição de uma palavra tal como se pode encontrar, por exemplo, num dicionário, tesouro ou ontologia de domínio.

Aos três estratos do conceito correspondem três tipos de julgamentos: a) definição de termo ou definição nominal; b) definição estritamente de uma intenção formal expressa, ou definição de uma ideia pensada enquanto mero objeto de pensamento; e c) definição efetiva ou definição do objeto real. Depois de esclarecido o estrato linguístico, verifica-se como o conceito é pensado e transmitido por um indivíduo numa situação real e concreta denominada conteúdo formal ou lógico do conceito correspondente, distinguindo-o em expresso (ou explícito) e subentendido (ou implícito). Finalmente, deve ser indagado se o conceito aplica-se realmente, adequadamente, sensatamente à ‘coisa’ ou ‘objeto’ representado, ou seja, se a informação é verdadeira (CARVALHO, 2003).

Assim, um conceito corresponde à definição de uma ideia pensada e deve ser julgada ontologicamente e relacionada adequadamente ao objeto informacional digital como “unidade de conhecimento registrado”. Brandão (2005) apresenta uma adaptação do Triângulo de Significado com o Triângulo Conceitual Dahlberg (1978) comparando a relação existente entre objeto, conceito e signo linguístico (Figura 11). A adaptação do Triângulo de Significado ao Triângulo Conceitual está representada na ligação tracejada entre objeto e signo evidenciando o fato de que as relações só podem ser construídas a partir de uma ponte conceitual. Por exemplo, o conceito da palavra “manga” que pode representar diferentes significados: uma fruta, parte de uma blusa, uma pocilga, a parte de vidro que cobre a chama de um lampião e, ainda, uma pequena extensão de terra para cultivo.

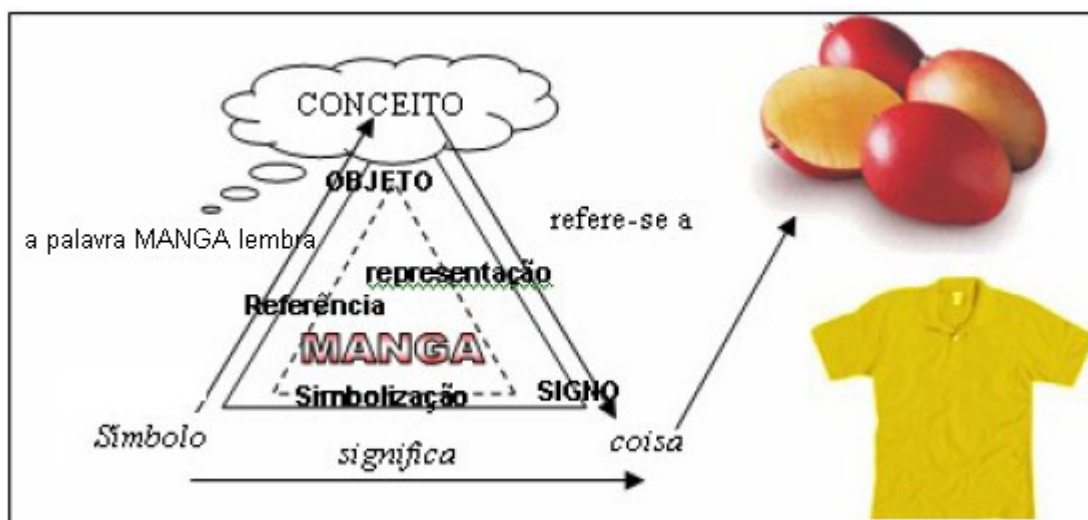


Figura 11 – Adaptação do Triângulo de Significado ao Triângulo Conceitual.  
**Fonte:** Adaptado de Brandão, 2005; Molossi, 2008.

Os objetos (conhecidos como referentes) se comparados ao Triângulo Conceitual também são referenciados por conceitos no Triângulo do Significado (conhecidos como referências ou significados) em que simbolicamente esses objetos são representados por signos. O campo semântico delimita o escopo dos possíveis significados que a palavra “manga” possa ter. Uma possibilidade de definir a palavra “manga” é contando com a ajuda de elementos periféricos no próprio contexto do documento. Se a palavra “manga” for extraída da frase “a manga é doce”, um dos elementos periféricos é a palavra “doce”. A palavra “manga” associada à palavra “doce” passa a ter o significado de fruta. Assim, o campo semântico formado pelas palavras “manga” e “doce” pode determinar o significado extraído de um texto numa estrutura convencional de um texto.

De acordo com Almeida (2006), as palavras escritas no Triângulo de Significado são símbolos que representam objetos, podendo ainda ser dispostos lado a lado para representar símbolos de símbolos. Assim, obtém-se uma metalinguagem em que símbolos representam algo para outros símbolos conforme ilustrado na Figura 12. Almeida (2006) estudando os problemas advindos da prática de reter conhecimento no ambiente corporativo, e de manipular os mecanismos subjacentes à memória organizacional<sup>26</sup> realizou várias adaptações no Triângulo de Significado apresentados na ontologia de Sowa<sup>27</sup>.

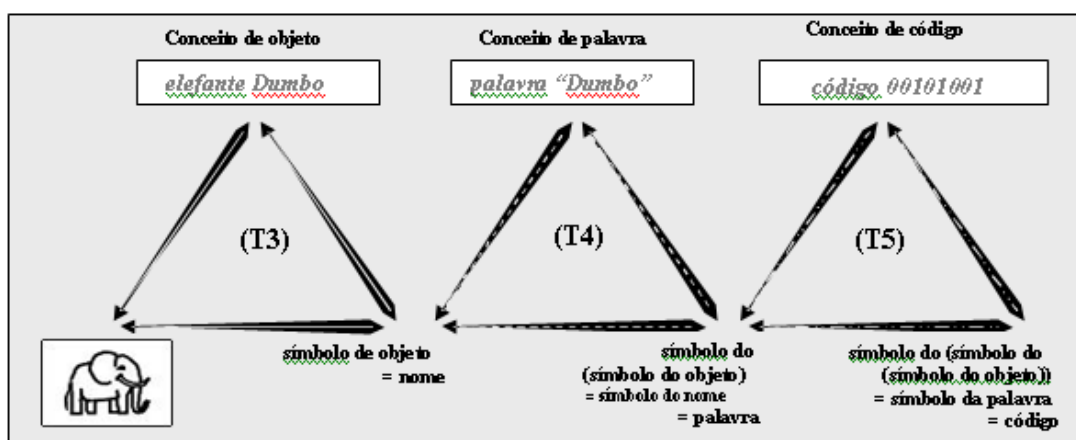


Figura 12 Objeto, nome, símbolo e codificação  
Fonte: adaptado de Almeida, 2006

<sup>26</sup> De acordo com Almeida (2006), o termo memória organizacional tem sido utilizado por diversos autores como uma metáfora para explicar fenômenos verificados no ciclo de vida das organizações. O autor destaca que sua pesquisa busca enriquecer a discussão sobre a memória das organizações. Para isso, o autor desenvolveu um modelo baseado em processos humanos e em processos tecnológicos para representar a memória organizacional e a validação do seu modelo é realizada sobre uma ontologia.

<sup>27</sup> A ontologia de Sowa é composta por categorias básicas e distinções derivadas de fontes variadas da lógica, da filosofia e da inteligência artificial. A ontologia não é baseada em uma estrutura fixa de categorias, mas em um conjunto de distinções a partir das quais uma hierarquia pode ser gerada automaticamente (ALMEIDA, 2006).

O conceito representa o entendimento de que o objeto é denominado por um nome que corresponde ao vértice superior do triângulo (T3); o vértice superior do triângulo (T4) representa o entendimento de que este nome pode ser representado pela palavra escrita “Dumbo”; e o vértice superior do triângulo (T5) representa o entendimento de que esta palavra “Dumbo” pode ser representada por um código binário no computador. Para a obtenção de melhores resultados em um sistema de RI, deve-se considerar a possibilidade do “sistema” trabalhar com o sentido das palavras de acordo com o contexto semântico.

Para Hjelmslev (1973), o sentido de uma palavra não está restrito a ela, mas a toda uma situação de uso de contextualização em funcionamento sócio-semiótico, ou seja, expressão e conteúdo são solidários e um pressupõe o outro de acordo com o contexto de uso. Neste sentido, o significante constitui o plano da expressão e o significado subdivide-se em substância e forma do conteúdo. A informação é um significado que é transmitido através da mensagem inscrita por meio de signos representando ‘objetos informacionais’ com algum significado para uma pessoa, daí o fato de também existir uma estreita relação entre a Semiótica e a CI. Assim, os elementos linguísticos podem, com certa intencionalidade, constituir em duas funções: a documental e a simbólica (TOUTAIN, 2007).

Voltando aos princípios filosóficos de uma ontologia, o “Ser” vive em um mundo de signos culturais. Para Aristóteles, a linguagem é antes de tudo *logos semantikos* (FONSECA, 2009), isto é, expressão significativa em toda sua potência comunicativa, expressiva e transformadora de objetos informacionais representados por um vocabulário multifacetado. O significado de uma palavra é constituído de componentes semânticos mínimos, relativamente constantes, que, em seu conjunto, definem um conceito, distinguem conceitos entre si e estabelecem relações conceituais entre as palavras, no plano da substância e da forma do conteúdo (MARQUES, 2003).

Aristóteles investigou a metáfora dos caracteres da escrita de objetos nos mais variados domínios linguísticos que não são os mesmos para todos, ou seja, no contexto de um documento científico existem múltiplas facetas de representação do vocabulário e suas categorias. Moss (1964) *apud* SILVA (2008) defende que as categorias fundamentais propostas por Ranganathan podem ser consideradas um retorno às categorias filosóficas de Aristóteles. O Quadro 5 apresenta um mapeamento entre esses dois conjuntos.

As categorias formais são as maiores classes de fenômenos, as classes mais gerais que podem ser formadas e empregadas para reunir outras características. Ranganathan esclarece que as cinco categorias do PMEST que ele propôs são usadas estritamente no contexto cujo significado pode ser visto somente nas declarações sobre as facetas de um assunto (sua separação e sequência). Elas são as mais genéricas possíveis e classificam os objetos de acordo com sua natureza conceitual. As categorias do PMEST



correspondem às facetas de nível hierárquico como uma primeira divisão do conhecimento (DAHLBERG, 1979). Então, para determinar as facetas em um único mapa conceitual do MHTX, representado por uma ontologia de domínio deve-se partir de uma estrutura terminológica consensual com categorias formais de campos conceituais como um sistema de conceitos integrados.

**QUADRO 5 – Mapeamento entre as categorias de Aristóteles e Ranganathan**

As 10 Categorias de Aristóteles	As 5 Categorias de Ranganathan
Substância	Matéria / Personalidade
Quantidade	
Qualidade	
Relação	
Lugar	Espaço
Tempo	Tempo
Posição	Espaço
Estado/Posse	
Ação	Energia
Afeição	Personalidade

Fonte: Adaptado de SILVA, 2008.

Resumindo, pode-se dizer que a Teoria da Análise Facetada e Teoria do Conceito abrangem hoje o reconhecimento do conceito como elemento material dos sistemas de classificação e a aplicação de uma teoria analítica de conceitos para a representação do conhecimento em um sistema de conceitos em forma de facetas. Consequentemente, ambas as teorias podem ser utilizadas para: a) a avaliação dos sistemas de classificação existentes; b) a construção de novos sistemas com agrupamentos ou arranjos previsíveis; e c) a formalização de enunciados sobre o conteúdo de documentos; tais enunciados podem ser pesquisados com consistência, quer manualmente quer por computador, a partir de estruturas de sentenças pré-determináveis (DAHLBERG, 1979).

Neste sentido, os elementos que formam um sistema de conceitos são obtidos pelo método científico de análise-síntese com profissionais muito experientes na extração de conceitos válidos para indexação em um sistema de RI. A estruturação desses sistemas em forma de facetas esquematizadas em um modelo relacional pode servir para compatibilização semântica do vocabulário. Isso reforça a importância do MHTX passar a ter um único mapa conceitual representado por uma ontologia de domínio. Este mapa deve representar “conceitos científicos” de um domínio. Considerando as diferentes facetas do vocabulário humano na escrita de um hiperdocumento científico e a ambiguidade no uso de determinadas palavras-chave de um domínio, este mapa precisa explicitar os termos sinônimos para compatibilização semântica de vocabulário.

A compatibilidade semântica de vocabulário está relacionada a capacidade de um sistema aceitar dados da indexação de outro sistema sobre qualquer assunto que seja comum a ambos, podendo ser medida em graus de similaridade entre duas linguagens estabelecendo a distinção entre compatibilidade em plano semântico e no plano lingüístico (SILVA, 2008). Além disso, o planejamento desses sistemas necessita de uma equipe multidisciplinar para aplicar os avanços tecnológicos de interatividade em RI propostos na área de IA. Carvalho (1997) esclarece que toda Filosofia é sistêmica. A razão disso está na natureza mesma dos conceitos que estão sempre vinculados por nexos hierárquicos e lógicos, inescapáveis, cujo conjunto forma mesmo, espontaneamente, um sistema.

Este trabalho não pretende analisar tais questões estritamente filosóficas, mas, verificar suas contribuições como instrumento de classificações e como modelo de relacionamento entre conceitos, visando compatibilização semântica de vocabulário que pode ser reutilizado tanto pelos humanos ou máquinas. A IA considera uma ontologia formal como um artefato de engenharia, constituído por um vocabulário intencional de pressupostos explícitos e lógicos, representando termos de um domínio e relações entre conceitos. Tal vocabulário intencional é descrito mediante uma aceitação, ou seja, um consenso no sentido dos termos-chave que irão compô-lo para especificar uma conceitualização compartilhada em uma comunidade científica.

Gruber (1993) descreve uma ontologia como um conjunto de termos representados por definições de classes são associadas aos nomes de entidades no universo do discurso, funcionando com textos legíveis relacionáveis a outros objetos que descrevem o significado dos nomes e axiomas para interpretação e uso dos termos. As ontologias podem ser genéricas ou apresentarem conceitos sobre um determinado domínio, utilizando um vocabulário a partir de um domínio mais genérico (GUARINO, 1998).

As ontologias não apresentam sempre a mesma estrutura, mas existem alguns componentes que estão presentes em grande parte delas: a) as classes, representativas de conceitos e organizadas hierarquicamente; b) as relações, representativas da interação entre os conceitos de um domínio; c) um conjunto de axiomas que vinculam os elementos da ontologia a condições que devem ser executadas e d) as instâncias, representativas dos objetos pertencentes a uma classe ou atributos que descrevem conceitos próprios ou herdados em uma especialização (ALMEIDA, 2006).

As classes conceituais podem ser organizadas em uma hierarquia, definindo relacionamentos de generalização/especialização. Essas classes representam conceitos e propriedades além de destacar a hierarquia entre elas. As propriedades de uma classe são seus atributos e relacionamentos. As propriedades são modeladas por um conjunto de restrições definindo valores que podem ser assumidos, os quais descrevem ou limitam uma

propriedade. As instâncias são as ocorrências particulares do objeto em relação à classe considerada. Uma instância representa o objeto específico de um conceito, procurando descrevê-lo de forma individualizada, única e concreta, fazendo referência a um objeto real.

Os atributos da generalização são também atributos às entidades especializadas, que geralmente ainda apresentam atributos adicionais. As restrições dos axiomas podem ser utilizadas para uma melhor definição da semântica de classes especializadas derivadas de classes gerais, pois são as restrições, também conhecidas como axiomas, que complementam os relacionamentos de especialização e generalização criados pela taxonomia. Os atributos de instâncias são as propriedades que descrevem as instâncias de um conceito. Os axiomas restringem os conceitos e as constantes são usadas em regras que determinam condições sobre o domínio e possibilitam inferir valores para os atributos. A Figura 13 apresenta um mapa conceitual dos componentes de uma ontologia.

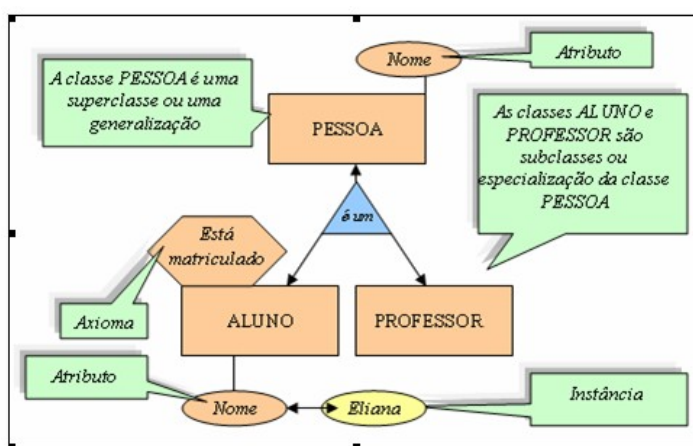


Figura 13 – Modelo conceitual de componentes de uma ontologia  
**Fonte:** adaptado de Oliveira, 2006.

Segundo Noy e McGuinness (2001), é crescente o desenvolvimento de projetos de sistemas que usam ontologias de domínio por vários motivos, destacando o compartilhamento do entendimento comum da estrutura de informação entre pessoas ou entre agentes de *software*<sup>28</sup> e a permissão de reuso. Esses projetos devem ser apresentados como um conjunto de métodos e técnicas de processos automáticos ou semi-automáticos via textos, dicionários, bases de conhecimento, dados semi-estruturados e esquemas relacionais já existentes (GÓMEZ-PÉREZ e MANZANO-MACHO, 2003).

Normalmente, a descrição das técnicas de modelagem utiliza a sintaxe para apresentar sua estrutura, a semântica para determinar os significados e a pragmática para definir a heurística (JACOBSON, 1992). Um modelo conceitual de um sistema

<sup>28</sup> Weiss (1999) define um agente de software como um sistema de computador situado em algum ambiente e capaz de agir de forma autônoma para atingir um objetivo.

computacional pode ser usado como um esquema para permitir a atividade de abstração e explicitação de bases do conhecimento humano além de possibilitar a compatibilização semântica do vocabulário entre agentes de software. O aspecto essencial do raciocínio através de modelos conceituais por prototipação não está só na construção de modelos mais adequados para captar distintos estados de coisas, mas também na habilidade em testar quaisquer conclusões lógicas usando tais modelos.

#### **2.4.4 Uso de ontologias no processamento da linguagem natural**

Uma das áreas mais promissoras para o uso de ontologias na IA com os instrumentos de vocabulário controlado é o Processamento da Linguagem Natural (PLN). Esses aplicam diferentes técnicas de indexação automática, utilizando regras formais segundo a abordagem linguística e a representação de termos-chave adotada pelo sistema. No projeto “Agência de Informação da Embrapa<sup>29</sup>” foram construídas várias ontologias de domínio em que os documentos são catalogados manualmente, selecionando-se termos que podem ser associados automaticamente pelo sistema.

Conteratto (2005) destaca que os sistemas criados para manipular o código linguístico são geralmente modulados em diferentes níveis de processamento: morfológico, sintático, semântico, discursivo e pragmático. Esses sistemas são executados em módulos distintos que se comunicam pela passagem de representações intermediárias do texto sob análise. A Figura 14 mostra cinco bases de conhecimento: o léxico, a gramática, o modelo do domínio, o modelo do usuário e o modelo do discurso. No léxico, cada palavra pode estar associada às suas características morfológicas, sintáticas e semânticas. As outras três bases de conhecimento são modelos que fornecem o contexto para o processamento de cada frase. No modelo de domínio, armazena-se o contexto enciclopédico (dicionários de entidades e relações, etc.). O modelo do usuário pode ser representado com estruturas hierárquicas. O modelo do discurso fornece o contexto textual.

O léxico é um elemento essencial para o tratamento automático de dados linguísticos, sendo a sua análise semântica fator crucial para um efetivo processamento computacional que, não raro, encontra barreiras em questões ligadas a uma representação semântica eficaz e que permita ser representada em linguagem de máquina. Assim, a gramática, através de regras computacionais, determina listas de termos válidos. Essa representação deve ser escolhida de acordo com a representação da gramática, pois essas duas bases de conhecimento interagem durante o processamento do texto além de existir vários formalismos para representar estas informações armazenadas no léxico.

---

<sup>29</sup> <http://www.agencia.cnptia.embrapa.br>

Apenas o fluxo de informação muda, de acordo com a tarefa do sistema: interpretação ou geração. Tem-se o texto como entrada; e uma representação formal como saída utilizando bases de conhecimento para interpretação da linguagem natural. O mapeamento de formalismos e o compartilhamento de conhecimentos com o uso de ontologias oferecem oportunidade para reusar bases de conhecimento, mesmo com adaptações e extensões. A lógica descritiva passa a funcionar como uma linguagem de representação do conhecimento por possuir sintaxe extremamente formal para integração de fontes de dados quando utilizada em modelos conceituais. Assim, a semântica estrutural assume o princípio de que explicitação formal do significado, primeiro depende da determinação de sua sintaxe para depois ser interpretada ao nível semântico.

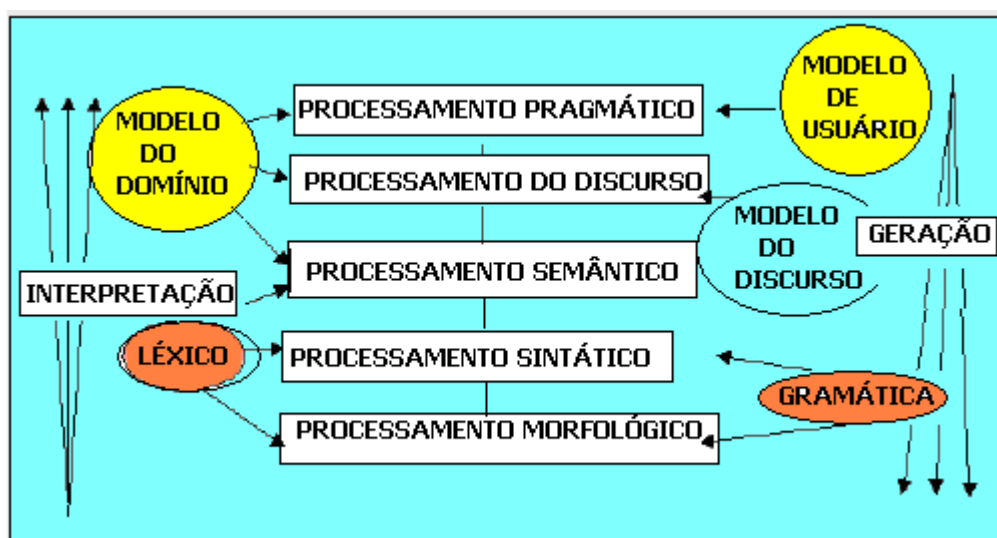


Figura 14 – Arquitetura genérica de um sistema de PLN.  
 Fonte: Conteratto, 2005.

Os processos usados com PLN na Linguística Computacional são práticas de KDT (*Knowledge Discovery from Text*) que aplica técnicas tradicionais de KDD (*Knowledge Discovery in Databases*), e outras novas com uso da lógica descritiva e linguagem orientada a objetos com o uso de anotações semânticas em documentos (LOH, 2001). Muitas técnicas utilizadas com ontologias de domínio e anotações semânticas em documentos eliminam ambiguidades de palavras, ajudando o usuário a interpretar termos errôneos dependendo da construção formada na consulta.

A maioria dos buscadores na Web baseia-se na técnica de recuperação ordenada. Para melhorar as medidas de precisão e revocação, pode-se utilizar técnicas de expansão de consultas por meio de conceitos semanticamente relacionados, além de medidas de distância conceitual. A CI se aproxima da Linguística Computacional para gramaticalização das linguagens documentárias. Assim, instrumentos como taxonomias,

tesauros e ontologias são criados para controlar o vocabulário que estruturados em campos conceituais formam um “sistema de conceitos” hierárquico de um modelo relacional. Dessa forma, uma ontologia de domínio na IA pode ser modelada utilizando os mesmos princípios. Um hiperdocumento científico também pode ser analisado por um sistema “inteligente” de uma base de dados, usando métodos estatísticos de orientação a objetos informacionais digitais para classificar/categorizar e relacionar conceitos integrados sobre um domínio.

Na linguística computacional, as anotações semânticas atribuem rótulos de categorias ao texto. As descrições gramaticais utilizam a categorização das palavras e de outras porções do discurso de acordo com os traços que as posicionem dentro do sistema linguístico, podendo utilizar dicionários, tesauros e ontologias associados aos mecanismos de busca. Bueno (2005) utilizou a metodologia da “engenharia da mente”<sup>30</sup> que foi aplicada para auxiliar os processos de descrever um modelo representado por um vocabulário controlado que possibilitou a recuperação automática mediante a identificação de expressões jurídicas desse vocabulário nos textos (Figura 15). O modelo de Bueno permitiu consultas em linguagem natural e inclusão de novos casos na base de conhecimento. A recuperação foi realizada com *matching*<sup>31</sup> parcial, usando uma medida de similaridade.

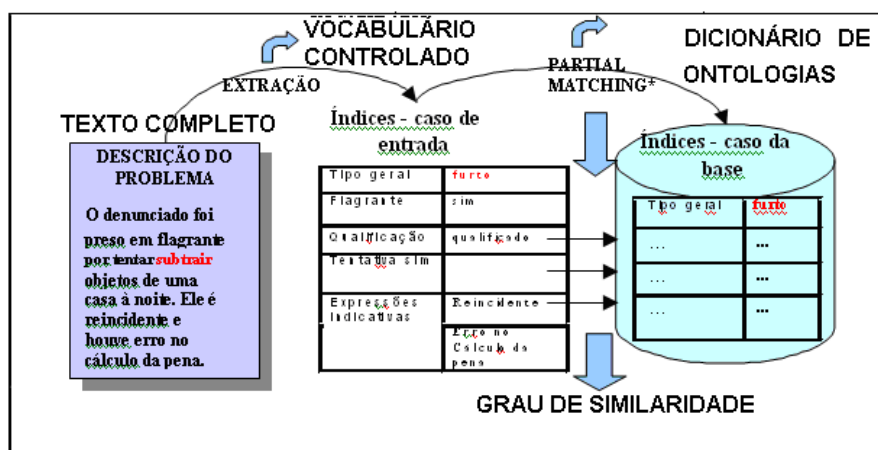


Figura 15 Raciocínio Baseado em Casos.  
Fonte: Bueno, 2005.

Existem diferentes recursos na IA que permitem analisar grupos de textos usando representações gráficas, compararem grafos representando, por exemplo, estruturas sintáticas de textos, combinando conceitos com o uso de mapas conceituais e taxonomias (LOH, 2001). Essas técnicas utilizam a matemática discreta como grafos conceituais e encontram-se entre as técnicas mais recentes de IA. Na teoria dos grafos,

<sup>30</sup> Engenharia da Mente é um processo de sincronização do conhecimento desenvolvido com a finalidade de fixar modelos conceituais em relação ao objetivo e à aplicação do modelo de gestão baseado na IA.

<sup>31</sup> O matching é um algoritmo capaz de comparar dados de uma máquina de inferência com padrões determinando quais regras da base de conhecimento são aplicáveis.

existem programas de tempo polinomial para encontrar um peso máximo de correspondência em um gráfico com base na rápida multiplicação de matrizes. Uma vez que uma regra é disparada, as ações especificadas pela regra são executadas (LUCENA, 2003).

Modelos booleanos estendidos e difusos com uso da lógica descritiva e a lógica *fuzzy*<sup>32</sup> também podem associar termos de uma ontologia com o uso de anotações semânticas em documentos e redes neurais<sup>33</sup>. Assim, o computador consegue agrupar dados para localizar termos-chave e separar uma série de objetos semelhantes a um mesmo assunto que tende a permanecer em um mesmo grupo (cluster), ao associar atributos em comum (LOH, 2001). Essas técnicas são usadas em interfaces que permitem aos usuários entrar em consultas em uma linguagem de sintaxe formal. Os motores removem as palavras de parada (*stop words*)<sup>34</sup> além de normalizar os termos reduzindo-os a seus radicais, processo conhecido como *stemming*<sup>35</sup>.

A indexação para representar documentos pode acontecer em dois níveis (a) em nível de palavra, com equivalência entre os termos de índice e as palavras do texto; e (b) em nível conceitual, onde o mapeamento é realizado entre os conceitos, que as palavras ou as frases carregam, e os índices. Como a indexação visa buscas rápidas, geralmente, usa-se um arquivo invertido<sup>36</sup>. A Figura 16 mostra as etapas do processo de indexação automática são: identificação de termos; remoção de *stop words*; normalização e padronização de vocabulário e seleção de termos relevantes.

---

<sup>32</sup> A lógica fuzzy encontra-se entre as técnicas mais recentes de IA. A maioria dos modelos apresenta algoritmos desenvolvidos em linguagens específicas utilizando principalmente paradigmas de programação lógica e funcional (LOH, 1999).

<sup>33</sup> Uma rede neural compreende um conjunto de nós interligados chamados de neurônios artificiais. Cada neurônio conecta-se e envia informação para os demais da camada seguinte, de acordo com pesos e conexões predefinidos com a capacidade de tratar dados incompletos, produzindo resultados satisfatórios a partir de generalizações (ALMEIDA e DUMONTIER, 1996).

<sup>34</sup> Palavras de parada são palavras muito comuns nos textos como artigos, preposições, conjunções entre outras. Os motores de busca evitam essas palavras porque normalmente não ajudam refinar buscas, demorando o tempo de resposta (TRILLO, 2004).

<sup>35</sup> *Stemming* é um processo para identificar os radicais das palavras e, com isso, diminuir o número de palavras indexadas, as operações mais comuns são a remoção de afixos.

<sup>36</sup> Um arquivo invertido é uma lista de palavras ordenadas ou *strings* (sequência de caracteres) Cada palavra contém um conjunto de ponteiros para os locais onde aparecem que pode ser uma parte do documento ou a posição real e específica da palavra no documento (LOH, 2001).

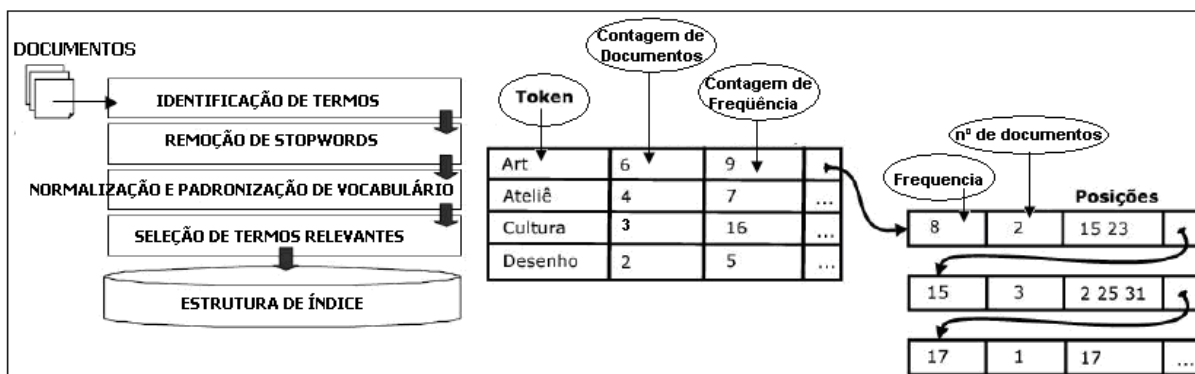


Figura 16 – Indexação automática e índice invertido.

Fonte: Adaptado de TRILLO, 2004 e MORAIS, 2007.

Atualmente, as ontologias também estão sendo vistas como meios de acesso, atuando como mapas conceituais dos tópicos explorados em um serviço de recuperação, revelando as ideias e os conceitos que compõem determinada área do conhecimento. Esses conceitos aparecem representados através de termos-chave, os quais indicam que uma ontologia pode ser modelada representando um sistema de conceitos integrados e indicar relações de equivalência (GOMES, MOTTA, CAMPOS, 2006). No MHTX, Silva (2008) indica o uso de ontologias para a construção de mapas conceituais mais completos sob o ponto de vista semântico e mais preparados para a compatibilização.

A principal vantagem de representação gráfica de redes semânticas com ontologias formais de estruturas taxonômicas é o poder que os mecanismos de inferência têm de derivar as propriedades de objetos mais gerais para os mais específicos (REZENDE, 2003). Esses são os princípios das linguagens de orientação a objetos, como as propriedades de encapsulamento dos atributos, herança e polimorfismo. Uma ontologia pode ter índices múltiplos em uma rede semântica para permitir acesso ao conteúdo estruturado de documento com regiões distintas, explicitando significados de termos.

Um dos benefícios de usar índices múltiplos é que o usuário pode pesquisar em uma região semântica distinta, melhorando o processo de recuperação e evitando que o sistema tenha que pesquisar em todo o índice. Esses índices facilitam a navegação em contexto melhorando o processo de recuperação e evitando que o sistema tenha que pesquisar em todo o índice. Os tópicos (termos-chave) de uma ontologia de domínio são ferramentas que possibilitam uma navegação contextualizada entre os elementos que os compõem, permitindo uma conexão entre o signo representacional e o recurso representado. Esses mapas formam um padrão para representação da informação utilizado para descrever e navegar por objetos informacionais em sistemas digitais (MARTHA, 2005).



Algumas técnicas e ferramentas na IA apresentam documentos e relacionamentos através de agrupamentos visuais, mapas bi ou tridimensionais, conexões gráficas e outras estruturas complexas. As medidas de distância conceitual utilizam os nós para calcular a distância conceitual e medir a similaridade entre os termos da consulta e do documento (LOH, 2001). Um índice de um grafo conceitual é o número total de *matchings*. Este índice é sempre pelo menos um, pois o conjunto vazio de arestas é contado como uma correspondência para essa finalidade, ou seja, este índice é o número de emparelhamentos não vazia. Furgeri (2006) destaca que as redes semânticas têm seu foco nas categorias de objetos e nos relacionamentos entre eles (lado 1 da Figura 18 ). A relação entre conceitos é representada por arcos através de verbos transitivos ou preposições através de um conjunto de nós (conceitos) e um conjunto de arcos (links).

Os nós de uma ontologia são os conceitos e uma aresta dirigida entre dois nós representa uma relação entre os mesmos. O lado 2 da Figura 17 mostra uma visão de uma ontologia extraída e apresentada pelo OntoCover na forma de uma árvore hiperbólica. A visualização de um mapa conceitual hiperbólico reduz o tempo gasto para navegar no computador, fornecendo meios de localizar informações específicas demonstrando como os conceitos estão tão semanticamente estruturados que facilitam a movimentação de um nó para outro. Esta interface conhecida como *fisheye* ou visão “olho-de-peixe” ou “visão mais detalhada” geralmente fornece um esquema que é suficiente para lidar com a orientação de grandes redes, permitindo maior flexibilidade e agilidade na interface (LIMA, 2004).

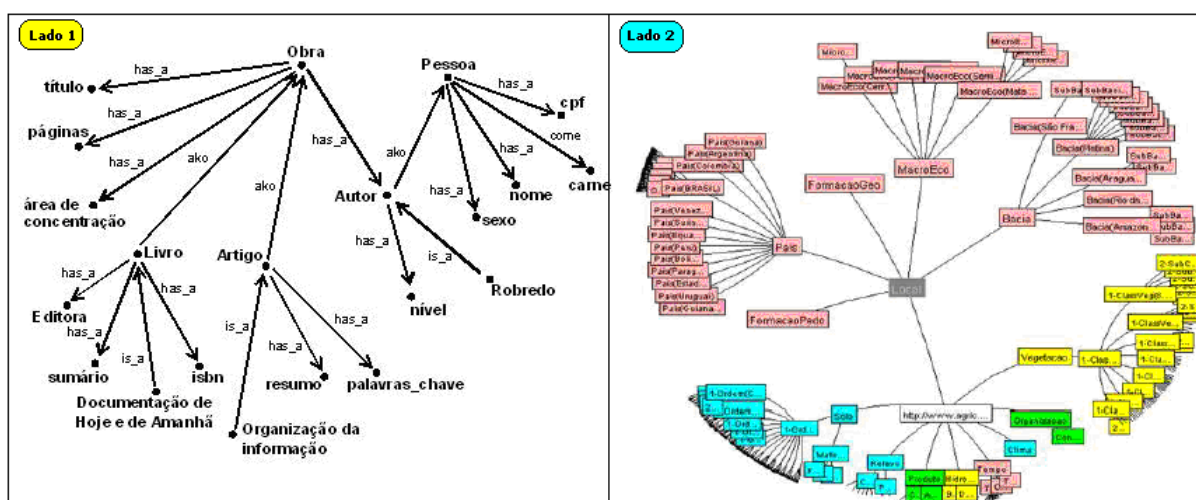


Figura 17 - Modelo de uma rede semântica.

Fonte: Fileto, Medeiros, Pu, Liu e Assad, 2005; Furgeri, 2006.

A distância conceitual entre dois nós é o comprimento do caminho mais curto entre eles no grafo dirigido exibido no Lado 1. Os nomes dos nós representam conceitos e os arcos representam a ligação entre esses os conceitos. Os nós podem ser representados

por substantivos, adjetivos, pronomes ou nomes próprios. Os termos mais comuns usados em arcos são `is_a` (é um), `has_a` (tem um) e `is_part` (é parte de). A seta do arco determina o sentido da leitura, mas podem ser lidas nos dois sentidos. Alguns mecanismos de inferência computacionais podem ser realizados no modelo: Livro é um tipo de Obra; Artigo é um tipo de Obra; Livro tem um ISBN, um sumário e uma Editora; O Autor é um tipo de Pessoa Física que possui CPF, nome, etc. As cores dos nós do Lado 2 distinguem as facetas, isto é, ramos da árvore descrevendo conceitos relativos a diferentes temas ou dimensões, tais como espaço geográfico, tipos de solo, tipos de vegetação, produtos e tempo. No lado direito, a ferramenta permite selecionar termos da ontologia para compor e comparar anotações semânticas, segundo a ordem parcial dos termos em cada ramo da árvore.

As ontologias têm sido utilizadas como uma das possibilidades mais promissoras para garantir a compatibilização e interoperabilidade semântica de aplicações Web por fornecerem uma semântica bem definida e de forma estruturada. Por permitir reuso, outros benefícios podem ser: a verificação/validação automática do modelo conceitual construído, a possibilidade de geração sistemática de infra-estruturas de domínio e a interpretação não ambígua das definições de conceitos e relações (CAMPOS, 2007). Os padrões de metadados digitais W3C permitem que as anotações semânticas em documentos textuais sejam diretamente processáveis por computador utilizando uma ontologia de domínio.

#### **2.4.5 Padrões de metadados digitais propostos pelo W3C com o uso de anotações semânticas e ontologias**

Historicamente, o termo anotação semântica em documentos foi usado para descrever a marcação que era colocada nos textos para instruir os compositores e tipógrafos de como estes deviam ser impressos ou compostos. Os sistemas de hipertextos na Web utilizam a metalinguagem SGML que, em Português significa Linguagem Padrão de Marcações Genéricas. SGML é um padrão internacional ISO 8879 publicado em 1986 para descrever o uso de anotações em documentos digitais com o objetivo de facilitar a descrição de metadados no computador. A SGML possibilitou tanto a HTML (Hypertext Markup Language) que é utilizada para a apresentação quanto a XML (eXtensible Markup Language) que estrutura o conteúdo de um documento, oferecendo mais recursos com ampla expressividade para descrever o formato das informações na Web.

Hoje, a maioria dos sistemas de publicação eletrônica utiliza esses procedimentos para especificar ao programa de computador como deve ser organizado o texto e como o mesmo deve ser formatado. São muitas as vantagens na utilização desses procedimentos, algumas delas podem ser: a) portabilidade - uma vez que através de qualquer documento anotado descritivamente, pode-se trocar informações entre plataformas e sistemas heterogêneos; b) reutilização e reconfiguração do aspecto visual - a anotação

separa conteúdo e forma de um documento possibilitando a associação de diferentes especificações de forma ao mesmo conteúdo; e c) longevidade - a informação possui um formato neutro de fácil processamento (RAMALHO e HENRIQUES, 2002).

A SGML força que cada um dos elementos descritos se ajuste na estrutura lógica do documento, oferecendo portabilidade e permitindo a criação de documentos independentes do tipo de máquina e dos programas usados. Berners-Lee (1999) critica a organização da Web destacando que os sistemas iniciais foram concebidos para facilitar o entendimento humano limitando as operações por parte dos computadores no processamento da informação. Berners-Lee propõe soluções no modo de representação e estruturação da informação como uma extensão da Web para dotá-la de um contexto semântico em que as informações passem a ter significado para os computadores, objetivando um tratamento automático do seu conteúdo.

Assim, a representação da informação deve ser formal para poder ser manipulada por computadores e permitir o seu tratamento automático por agentes de software. Esta proposta ficou conhecida como Web Semântica e, é um projeto ousado, de longo prazo cujo planejamento requer o desenvolvimento de: a) linguagens para expressar meta-informação sobre documentos que sejam processáveis pelo computador; b) terminologias padrões sobre os mais diversos domínios; e c) ferramentas e arquiteturas que usem tais linguagens e terminologias para fornecer suporte à busca, acesso, apresentação e manutenção de informações (BERNERS-LEE; HENDLER; LASSILA, 2001).

A partir dessas propostas, formou-se o grupo W3C com o objetivo de proporcionar estruturas de metadados de padrões digitais para expressar informações que possam ser lidas e interpretadas por máquinas utilizando agentes de softwares “inteligentes”. Desde então, o W3C vem trabalhando arduamente no desenvolvimento de tecnologias avançadas visando à representação estrutural e semântica dos recursos na Web. A XML surgiu para prover um formato de descrição dos dados estruturados que facilitasse as declarações mais precisas do conteúdo e resultados mais significativos de busca através de múltiplas plataformas. Um documento escrito em XML pode conter um conjunto infinito de *tags*, enquanto na linguagem HTML esse conjunto de marcações é limitado. Dessa forma, a linguagem XML possibilita a descrição de informações na Web, uma vez que disponibiliza elementos que possibilitam descrever diversas tags. A estrutura de um documento é definida pelo DTD (Document Type Definitions).

O DTD é um dicionário que ajuda o sistema a representar o documento XML, além de verificar a validade dos dados, ou seja, se este foi escrito de forma sintaticamente correta. Um DTD define as partes de um documento para descrever os elementos que podem ser colocados em seus interiores e se são ou não elementos obrigatórios do

documento. A linguagem XML é considerada um bom formato para a criação de documentos com dados organizados de forma hierárquica. Contudo, o usuário deve ter cuidado com os nomes declarados nos tipos de dados. Estes nomes são conhecidos como *namespace* e devem ser usados como prefixo dos nomes de elementos e de atributos nos documentos XML seguindo algumas regras básicas do SGML.

A XML permite que a própria estrutura do documento seja transformada em um metadado digital (FURGERI, 2006). Os mecanismos de ligação de dados podem ser com o HTML Dinâmico (DHTML) e o XSL (Extensible Stylesheet Language) que adiciona poderes ainda maiores. O que define formalmente quais elementos combinações são permitidas dentro de um documento é conhecido como XML Schema (XMLS). O processo de anotação semântica com o uso de ontologias (Figura 18 – Lado 2) começa pela identificação das entidades que compõem o documento e depois são ligadas as entidades que aparecem no documento com suas descrições na ontologia.

A anotação semântica adiciona ao documento uma camada que descreve o seu conteúdo, permitindo que agentes de softwares possam recuperar informações de forma mais precisa, através da associação do documento com suas descrições semânticas na ontologia. Esses processos se aproximam com os objetivos de modelagem informacional propostos neste trabalho que busca relacionar “conceitos” de um domínio com outros termos sinônimos. O Lado 1 da Figura 18 ilustra como pode ser a apresentação em HTML de um documento XML e o Lado 2 O processo de anotação semântica como uso de ontologias.

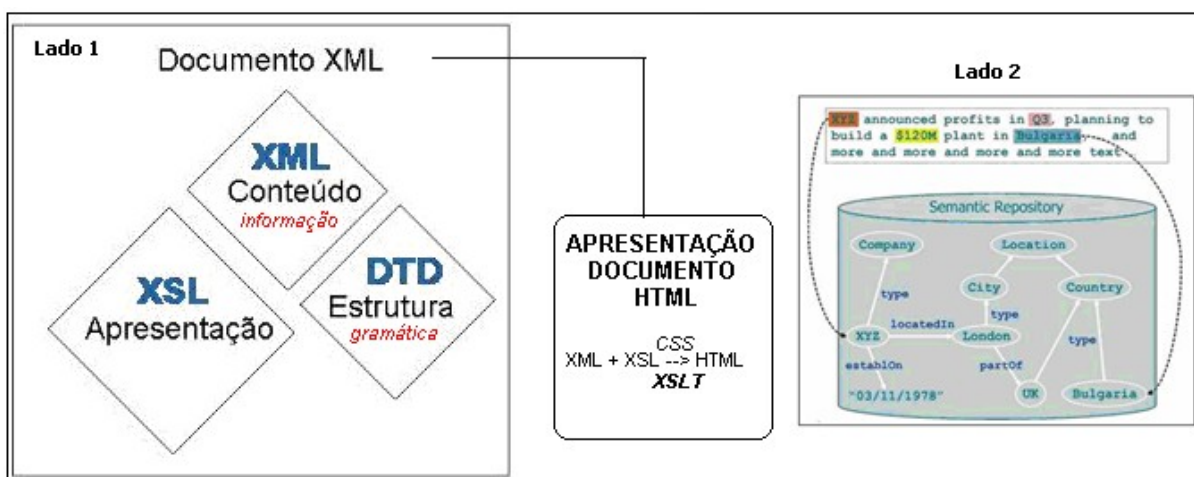


Figura 18 – Apresentação em HTML de um documento XML e exemplo de anotação semântica. **Fonte:** Kiryakov et al., 2003.

Nesta Figura, as entidades presentes no texto como o nome de uma empresa (XYZ) e sua localização (Bulgária) podem ser associadas a termos específicos explicitados por uma ontologia de domínio. A entidade é verbalizada e identificada pela identificação da URI que permite ligação fácil e precisa com a sua descrição no repositório. Para que o

“sistema” interprete essas informações em um hiperdocumento na Web, elas necessitam conter uma anotação semântica, baseada nos termos definidos por uma ou mais ontologias.

É importante destacar que não é necessário possuir uma ontologia completa, definindo todo o domínio de aplicação para anotação semântica. O W3C publicou uma recomendação para definição de namespace identificados por uma referência URI e introduziu o uso da linguagem XSL (Extensible Stylesheet Language) para formatação de documentos XML, substituindo a formatação de documentos XML com o recurso CSSL (Cascading Stylesheet Language). Associado à XSL está o XSL Transform (XSLT) que é o processo de transformação do documento XML num documento HTML. Tudo isso, reforça grandes esforços na padronização de metadados digitais para oferecerem ferramentas, técnicas com mecanismos de inferências computacionais mediados por agentes de software.

Como o processo de anotação semântica é aplicável a qualquer tipo de hipertexto (HTML ou documentos de texto comuns, campos de banco de dados, etc), é possível adicionar ao hiperdocumento científico uma camada que descreve o seu conteúdo. Assim, agentes de software podem recuperar informações de forma mais precisa ao associar termos-chave de um domínio. O tipo de anotação pode ser intrusiva ou não. A anotação intrusiva é aquela gravada no próprio documento. A anotação não-intrusiva utiliza ponteiros usualmente baseados em URI para referenciar as entidades mencionadas no texto podendo ser armazenada sem modificar o documento. O Quadro 6 apresenta uma síntese de algumas ferramentas de anotação semântica com ontologia.

**QUADRO 6 – Síntese de algumas ferramentas para o uso de anotação semântica.**

Nome da Ferramenta	Tipo de Ferramenta	Linguagem da Ontologia	Entrada de Ontologias	Representação das Anotações	Evolução de Ontologia	Formato do Documento	Tipo de Anotação
<b>Ont-O-Mat</b>	Semi-automática	DAML + OIL/RDF	Apenas uma por tarefa	Dentro de páginas da web	Sim	HTML	Intrusivo
<b>MnM</b>	Híbrida	DAML + OIL/RDF	Uma ou mais de uma por tarefa	Dentro de páginas da Web e em uma base de conhecimento	Sim	HTML, texto	Intrusivo
<b>KIM</b>	Automática	RDF/OWL	Ontologia do KIM (Proton)	Dentro de páginas da web	Sim	HTML	Não-intrusivo
<b>Smore</b>	Semi-automática	OWL	Uma ou mais de um por tarefa	Dentro de páginas da web	Sim	HTML, e-mail, imagens, texto	Intrusivo
<b>Annotea</b>	Semi-automática	RDF	Uma ou mais de um por tarefa	Servidores de anotações	Não	HTML	Não-intrusivo

**Fonte:** Eller, 2008.

Algumas dessas ferramentas, de acordo com o processo de anotação semântica em documentos, podem realizar processos automáticos, semi-automáticos ou híbridos sendo importante observar o formato dos documentos suportados e o tipo de anotação, intrusiva ou não-intrusiva. As ferramentas de anotações semânticas têm de lidar com o problema de alinhar os conceitos das ontologias aos termos constantes nos documentos. A maioria das ferramentas automáticas adota um domínio de ontologia pré-selecionado para evitar que a geração automática seja um problema difícil.

Assim, a ontologia pode descrever relações semânticas entre os termos de um domínio e a anotação semântica de um documento descreve o seu conteúdo pela associação dos conceitos descritos na ontologia. Eller (2008) destaca algumas tarefas deste processo: a) análise léxica, para separação dos *tokens* que formam as palavras simples ou compostas do texto; b) extração dos radicais das palavras; c) associação das palavras relevantes identificadas no texto as suas definições semânticas na ontologia; d) armazenamento das anotações.

Os padrões e formalismos de representação de metadados digitais propostos pelo W3C foram construídos sob dois importantes pilares: o Unicode Consortium, padrão para representar caracteres, e o URI (Uniform Resource Indicator) que identifica e localiza um recurso na Web. O esquema de identificadores se distingue em duas grandes categorias de URI: Uniform Resource Locator (URL) e a Uniform Resource Name (URN). O URN se refere ao subconjunto do URI que é preciso permanecer globalmente único e persistente mesmo quando o recurso deixa de existir ou se torna indisponível (SAYÃO, 2007).

Os primeiros níveis desta arquitetura são estruturados por metalinguagens e metadados que foram propostos em 2001 (Figura 19). Em 2005, o W3C sugeriu uma nova arquitetura resumindo a primeira arquitetura em três camadas, reforçando a importância das ontologias para conduzir a interoperabilidade semântica para integração de dados.

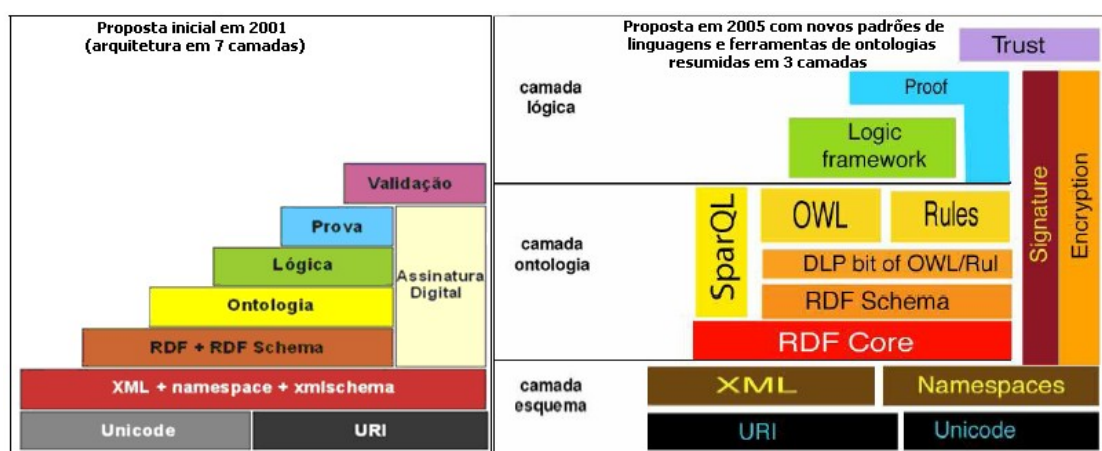


Figura 19 Arquiteturas da Web Semântica  
 Fonte: [http://www.w3.org/2005/Talks/0511-keynote-tbl/#\[17\]](http://www.w3.org/2005/Talks/0511-keynote-tbl/#[17]).

A camada esquema serve para estruturar os dados e definir seus significados seguindo a mesma base proposta em 2001: a) a camada Unicode e URI que garante o uso padronizado dos conjuntos de caracteres e uma forma unívoca para a identificação de recursos através da representação de seu mecanismo primário de acesso, ou seja, sua localização na Internet; e b) a camada XML + namespace + xmlschema que foi criada como uma extensão da linguagem HTML para XML para permitir a representação de outras linguagens de forma padronizada por meio do formato DTD, deixando para as camadas superiores a definição do conteúdo do documento.

A camada Ontologia é a responsável por oferecer a expressividade necessária à representação da estrutura do vocabulário que também ficou conhecida como *Ontology Vocabulary*. O RDF (*Resource Description Framework*) descreve os recursos com um identificador na Web. Para expressar algo sobre os recursos, o modelo de dados de RDF equivale em termos formais às redes semânticas. Os recursos são descritos com a tríplice objetos-atributos-valores semelhantes à tríplice sujeito-verbo-objeto das redes semânticas com a vantagem de esta camada poder ignorar as imposições da estrutura do documento da camada inferior, referindo-se apenas aos dados sobre o conteúdo. O RDFS (*RDF-Schema*) fornece tipos básicos para a criação de esquemas voltados a aplicações específicas para uma padronização de uso do RDF.

As camadas Logic, Proof e Trust são as camadas superiores a camada *Ontology Vocabulary*. A Logic permite a especificação de regras que atuam sobre instâncias e recursos; a Proof executa essas regras e a Trust avalia se a prova executada está correta ou não. Essas camadas definem mecanismos de inferências sobre os dados inferiores que precisam estar semanticamente estruturados para essa execução. Esses mecanismos funcionam como uma ação de raciocínio aplicada com objetivo de derivar uma nova informação a partir das instâncias de uma base de dados.

Várias ferramentas surgiram não só com o objetivo de facilitar a construção de ontologias, como também de manipular, consolidar e traduzir ontologias para reuso e extensão, integrando diferentes grupos de pesquisa. O padrão mais recente de linguagens para ontologias é OWL (*Web Ontology Language*) desenvolvido no âmbito do W3C. A linguagem SPARQL<sup>37</sup> (*SPARQL Protocol and RDF Query Language*) pode ser usada para fazer consultas nos repositórios RDF. Os novos esquemas RDF incluem classe e herdam subclasse para construir hierarquias de propriedades, instância e restrição que passou a

---

<sup>37</sup> SPARQL é uma linguagem de consultas originalmente desenvolvida para acessar dados sobre objetos na Web, representados através de metadados RDF e usa uma sintaxe semelhante a linguagem SQL usada em SGBD (*Sistema de Gerenciamento de Banco de Dados*).

ser uma extensão do RDFS para definir restrições complexas e outras construções que programam características de frames e lógica de descrição.

O RDFS é um sistema de classes extensível e genérico que pode ser utilizado como base para esquemas de um domínio específico. A utilização de regras personalizadas pode ser caracterizada como uma ampliação da expressividade de uma linguagem, pois estende as possibilidades de inferência. Essas regras podem ser usadas para definir a parte de conhecimento deduzido, expressando consultas mais complexas. Para que um mecanismo de inferência computacional tenha “inteligência” e “raciocine” sobre uma determinada informação é preciso que o sistema possua um modelo do domínio sobre o qual poderá executar automaticamente mecanismos artificiais de inferência.

Segundo Fontes (2007), existem dois tipos de inferência na IA: a) inferência imediata quando uma única premissa antecede à conclusão e é gerada a partir da oposição e conversão das preposições; b) inferência mediata quando pelo menos duas premissas antecedem à conclusão e é gerada a partir de raciocínio por analogia, dedução e indução. Esses serviços permitem testar se uma classe é, ou não é, uma subclasse de outra classe, ou seja, as descrições das classes são utilizadas para determinar se existe dentre elas uma relação, podendo oferecer muitos benefícios aos sistemas computacionais de qualquer complexidade. Um exemplo da importância em se utilizar inferência e raciocínio ocorre no âmbito de sistemas de resolução de problemas que requerem sequências de inferência complexas para atingir objetivos específicos (BEPPLER, 2008).

O sistema recebe um dado do domínio como entrada para processá-lo através de expressões lógicas e inferir na saída. Um dado do domínio de entrada pode ser descrito por: a) uma linguagem de representação de conhecimento; b) uma base de conhecimento que contenha informações apropriadas a uma determinada tarefa; e c) uma máquina de inferência capaz de processar o conhecimento armazenado. A abordagem clássica para processar o conhecimento propõe o uso de uma ontologia de domínio explicitamente representada e cujas decisões são tomadas via um raciocínio lógico, casando padrões de metadados e manipulações simbólicas (LUCENA, 2003).

Pesquisas mostram que a Web Semântica não pode acontecer plenamente sem um esforço importante de padronização de vocabulários; tais vocabulários favorecerão o uso de termos interconectados e compartilhados em uma rede semântica estruturada por meio das comunidades de usuários com interesses comuns. A Web Semântica pode contribuir para a melhoria dos índices de revocação e precisão, permitindo aos dispositivos de recuperação evitar os problemas comuns de polissemia e sinonímia, além de considerar as informações em seus contextos de significado (SOUZA e ALVARENGA, 2004). A



sensibilidade ao contexto é uma subárea da Computação Pervasiva<sup>38</sup> sendo ativamente pesquisada, pois possui importância essencial na utilização de ontologias de domínio.

Os padrões de metadados propostos pelo W3C já possibilitam novos serviços Web (Figura 20) que de acordo com Santos e Alves (2009), a Web 3.0, período atual conhecido como terceira geração oferece produtos de uma Web mais organizada e ‘inteligente’. A linha pontilhada entre os dois eixos, vertical e horizontal desta Figura, mostra a evolução pela qual a Web vem passando desde sua criação. A partir da Web 2.0, pode-se claramente observar as contribuições do projeto Web Semântica resultado do uso de plataformas abertas, que têm contribuído principalmente para avanços de aplicações com dispositivos ubíquos sensíveis ao contexto. A Web 2.0 caracterizada pelo surgimento de novos ambientes colaborativos, tais como os *blogs* e *wikis* e principalmente pelas propostas da Web Semântica; a Web 3.0 caracterizada pelos avanços de novas ferramentas de IA e, por fim, o encaminhamento de todas essas aplicações para uma quarta geração destacando-se a contínua atualização das tecnologias e a criação de ambientes informacionais cada vez mais especializados, interativos e colaborativos.

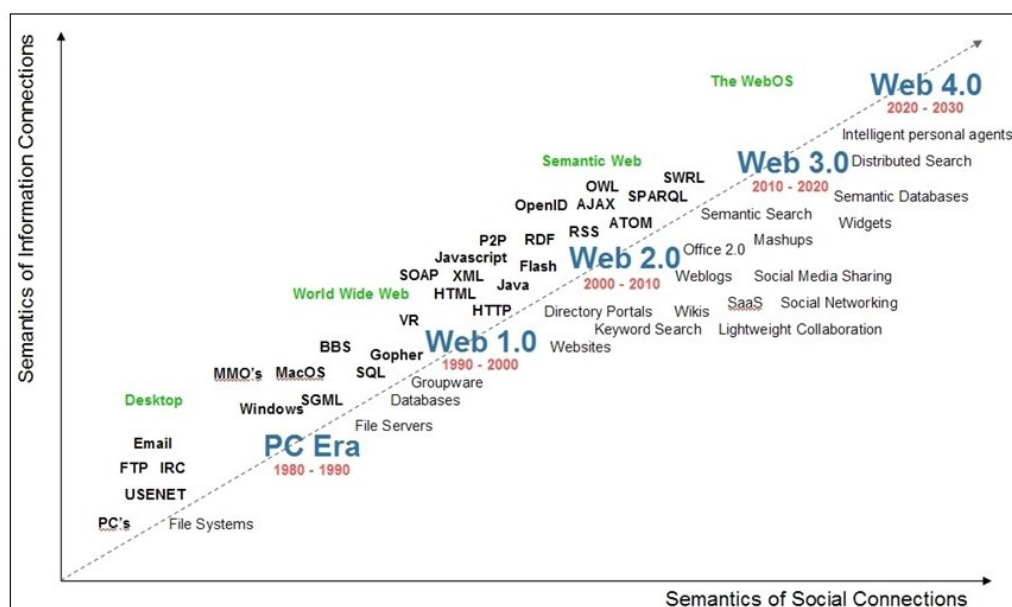


Figura 20 – Evolução da Web.  
Fonte: Santos e Alves, 2009.

Ontologias incluem estruturas que permitem manipular termos de forma eficiente com base em um vocabulário livre de ambiguidades para compartilhar e integrar informações, podendo ainda ser analisadas sob os aspectos sintático, semântico e

<sup>38</sup> A Computação Pervasiva tem origens no final da década de 80, quando Mark Weiser descreveu que o próximo passo da “ubiquidade” dos computadores pessoais e dos dispositivos computacionais móveis seria a integração para permitir aos usuários acessar seus respectivos ambientes computacionais em qualquer lugar ou momento. (NUNES, 2008).

estrutural, usando objetos informacionais de metadados digitais hiperlinkados por ferramentas de anotações semânticas. Segundo Lima (2003), deve-se partir da estrutura existente nas tabelas de subdivisões de áreas do conhecimento abordando toda dimensão teórica da organização e considerando o tratamento para o conteúdo do hiperdocumento textual.

Os links entre os recursos digitais cumprem um papel importante no estabelecimento da funcionalidade entre bibliotecas e arquivos digitais configurados de forma distribuída devido a interoperabilidade entre esses repositórios estar garantida. A condição de permanência dos links por longo prazo é um fator importante para a integridade dos objetos digitais, pois estes recursos são compostos também pelos seus relacionamentos com outros objetos digitais que, por sua vez, se realizam através de links acionáveis (SAYÃO, 2007). Desta forma, a especificação e a resolução de consultas de um modelo de biblioteca digital podem ser beneficiadas pelas relações descritas em uma ontologia de domínio com metadados conceituais integrados, usando padrões W3C.

Sayão (2007) destaca que o sucesso de um sistema de informações depende fortemente da vinculação consistente entre os recursos. Isto se traduz pela estabilidade dos links presentes nos catálogos, nos índices e nas listas que constituem os diversos serviços de descoberta de recursos. Para isso se efetivar, é necessário atribuir um nome que os identifique permanentemente e sem ambiguidades e independente de localização para garantir acesso contínuo ao próprio recurso, assegurando que este seja localizado todas as vezes que seu link referencial for acionado ou clicado por um usuário. Esta é a ideia principal deste trabalho para a modelagem informacional, utilizando uma ontologia de domínio com ferramentas de anotações semânticas com padrões W3C.

#### **2.4.6 Ferramentas, linguagens e modelagem de ontologias**

Um ambiente para desenvolvimento de ontologias deve ter analisadores sintáticos e léxicos, tradutores, editor, navegador, mecanismo de busca e de manutenção automática. Os editores de ontologia são representados em sua maioria por um sistema de frames. Esses frames são apresentados em arranjos de uma hierarquia taxonômica na qual estão relacionados. Uma coleção de frames numa ou mais hierarquias de herança formam uma base de conhecimento. Os *frames* possuem componentes chamados *slots* que descrevem propriedades ou atributos. Os *slots* geralmente possuem outros componentes como nome, valor e restrição que podem ser usados para computar um valor do *slot*. Estes diferentes componentes de um slot são também chamados de facetas.

Este trabalho destaca o Protégé por funcionar em várias plataformas, oferecer suporte de extensões de interface de usuários customizadas (*plugins*) e interagir com vários

padrões de armazenamento como XML, RDF, RDFS, OWL. O Protégé permite diversos formalismos, estratégias de inferência e integra tarefas de aquisição de ontologias num mesmo ambiente. O modelo de aquisição de conhecimento utilizado no Protégé é um modelo abstrato e as ontologias são definidas através de uma interface gráfica, criando-se conceitos do domínio que se deseja representar organizados numa hierarquia de subclasses. O Protégé suporta dois tipos de modelagem de ontologias: Protégé-Frames e Protégé-OWL.

O Protégé-Frames utiliza *frames* onde as classes são organizadas em uma taxonomia, representando os principais conceitos de um domínio, suas respectivas propriedades e relacionamentos (*slots*) e instâncias. O Protégé-OWL possui um conjunto de operadores lógicos que possibilita a definição de conceitos mais complexos. Além disso, permite a utilização de um mecanismo de inferência, o qual pode verificar se as declarações e as definições da ontologia são mutuamente consistentes entre si e reconhecer se os conceitos são adequados a definições. As ontologias OWL têm componentes similares a estrutura do Protégé-Frames, mas a terminologia usada para descrever tais componentes é um pouco diferente da utilizada no Protégé -Frames.

Os indivíduos representam objetos no domínio de interesse e são também conhecidos como instâncias podendo ser referenciados como instâncias de classes. Uma diferença importante é que o Protégé -OWL não usa o UNA (Unique Name Assumption), isto significa que dois nomes diferentes podem remeter ao mesmo indivíduo. Em OWL deve-se declarar explicitamente que os indivíduos são os mesmos, ou diferentes uns dos outros. Por exemplo, linguagem documentária ou linguagem de indexação podem ser referências ao mesmo termo. As classes OWL são conjuntos que contêm os indivíduos e são descritas formalmente (descrições matemáticas) de forma que sejam apresentados os requisitos para a participação na classe. As classes podem ser organizadas em hierarquias superclasse-subclasse, também conhecidas como taxonomias.

Horridge (2005) lista três espécies de ontologias OWL, de acordo com a sub-linguagem utilizada: OWL-Lite, OWL-DL e OWL-Full. A característica principal de cada sub-linguagem é a sua expressividade: a) OWL Lite, usada para casos em que apenas a classificação e restrições simples bastam para atender as necessidades; b) OWL DL, usada para casos em que seja necessária uma expressividade sem perda da completeza computacional e decidibilidade de sistemas de raciocínio; e c) OWL Full, usada para casos em que seja necessária expressividade máxima com toda a liberdade possível, mas sem garantias de computação, como por exemplo, tratar instâncias como classes.

Modelos conceituais por prototipação são delineados e projetados por pessoas experientes sobre determinado domínio do conhecimento para facilitar a compreensão do

funcionamento de sistemas físicos. Para isso, é preciso ter um modelo conceitual de um modelo mental como um mapeamento de um domínio. A modelagem de dados é uma técnica de mapeamento utilizada para análise e representação de uma estrutura conceitual de uma fonte de conhecimento. Existem diferentes abordagens para definir as etapas de trabalho constituindo-se em paradigmas da Engenharia de *Software*. Em geral, especialistas de diferentes áreas do conhecimento trabalham juntos para analisar diferentes aspectos e soluções a serem desenvolvidas por um sistema computacional.

O ciclo de vida com prototipação (Figura 21) prevê a construção de um protótipo logo após a fase de análise de requisitos. Porém, deve-se ter claro que o protótipo não pode ser adotado como produto final, pois sua finalidade é caracterizar certas simplificações e reduções não desejáveis no sistema final (COAD e YOURDON, 1993). Neste processo selecionam-se os objetos construindo-se, assim, esquemas de representação nos quais as 'coisas' são reduzidas a atributos em um modelo conceitual por prototipação e mapeamento de um domínio a ser representado em um sistema.

Na análise de requisitos, captura-se o significado dos objetos de interesse, obtendo-se uma descrição organizada. O processo de aquisição do conhecimento, junto aos especialistas de um domínio, pode ser feito de várias maneiras. O projeto conceitual tem como objetivo principal representar a semântica da informação, independente de considerações de eficiência do sistema. Esses processos utilizam modelos conceituais para facilitar a comunicação, apresentando um vocabulário de fácil entendimento.

O modelo descritivo e o modelo conceitual são criados na fase de definição do problema, sendo a descrição informal elaborada na análise de requisitos transformada em técnicas de modelagem estruturada através de metodologias. O modelo operacional é desenvolvido na fase de projeto na qual são definidas as estruturas de dados e como esses poderão ser representados no modelo físico no qual são desenvolvidos os programas e o armazenamento dos dados para manipulação. O projeto lógico deve conter a especificação detalhada dos elementos do sistema com os procedimentos externos ao computador.

Tudo isso deve ser documentado e submetido ao usuário para aprovação. Segundo Nakanishi et. al. (1993), independente da área de aplicação, todos esses processos podem ser resumidos em três etapas denominadas de definição, desenvolvimento e manutenção conforme mostra a Figura 21. Campos (2004) apresenta quatro princípios fundamentais e seqüenciais que podem ser utilizados no ato de modelar domínios de conhecimento: a) método de raciocínio; b) objeto de representação; c) relação entre objetos; e d) representação gráfica.

O método de raciocínio é o princípio de modelar o conhecimento utilizando o método dedutivo (*top-down*) ou o método indutivo (*bottom-up*). No dedutivo, parte-se do

geral para o específico, isto é, o processo de modelagem se inicia pelo domínio/contexto e os elementos que irão compor a representação. As relações entre os elementos são consideradas em uma etapa posterior. Já no método indutivo ocorre o inverso. O objeto de representação é o princípio de modelar o conhecimento estabelecendo a menor unidade de representação em um dado contexto. Relação entre objetos é o princípio de modelar o conhecimento objetivando verificar as possibilidades de ligação/separação semânticas entre os conceitos de um dado domínio. A representação gráfica é o princípio de modelar o conhecimento elaborando diagramas que expressem relações conceituais (CAMPOS, 2004).

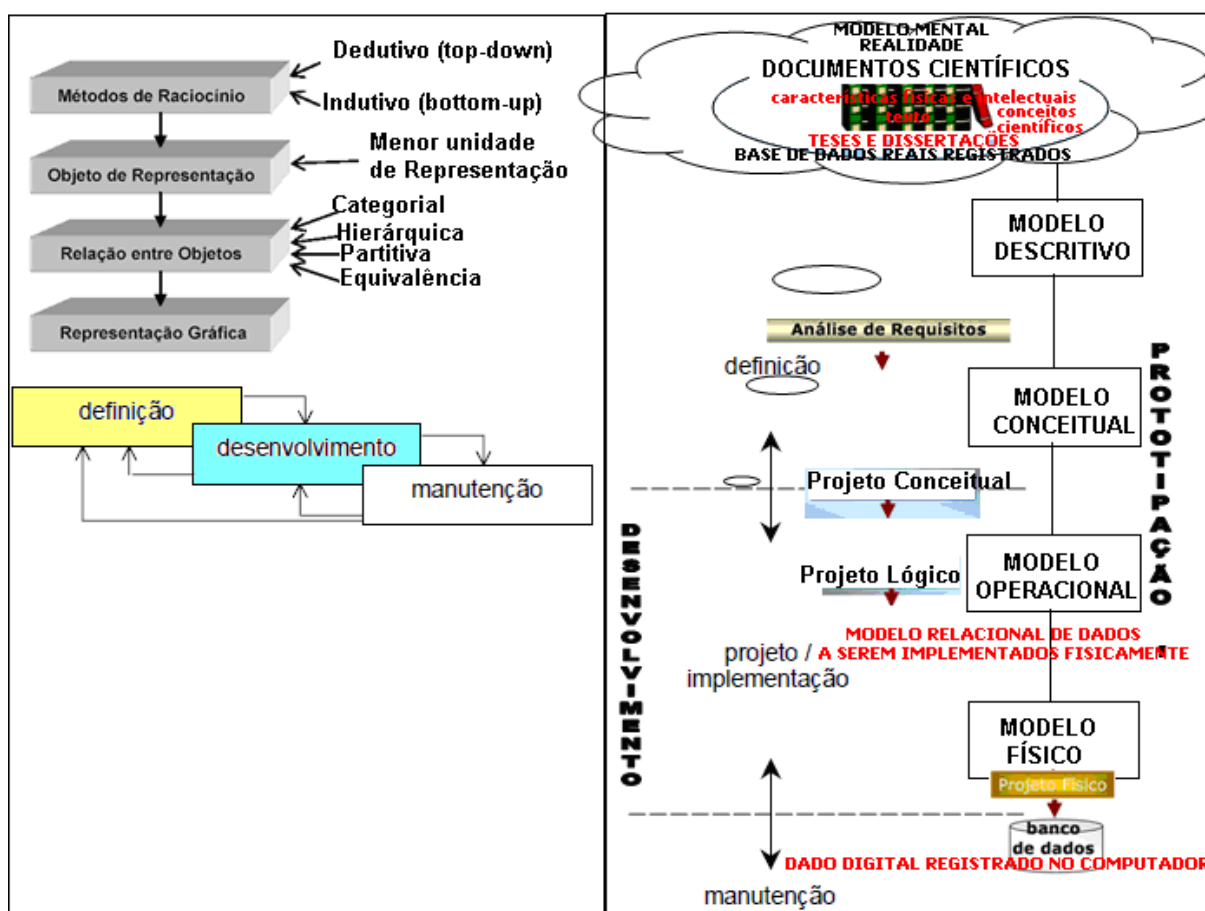


Figura 21 – Resumo das etapas de desenvolvimento de um software  
**Fonte:** Adaptado de Nakanishi, 1993.

O conhecimento da realidade na qual esses sistemas estão inseridos é a primeira etapa do trabalho, a partir da qual serão definidas as responsabilidades do modelo de sistema conceitual por prototipação. Almeida (2003) apresenta uma síntese sobre as metodologias, proporcionando uma visão geral de seu funcionamento, além de seis etapas para construção de ontologias que são descritas a seguir:

- **identificação de propósito e especificação de requisitos:** consiste em identificar a competência da ontologia que está relacionada a representação clara dos propósitos e os usos esperados;
- **captura da ontologia:** consiste em capturar a conceituação do universo de discurso com base na competência da ontologia. Devem-se identificar os conceitos e relações relevantes utilizando um modelo para facilitar a comunicação com os especialistas do domínio acompanhado de um vocabulário de termos. Deve-se, ainda, construir taxonomias, organizando categorias e subcategorias interconectadas do conhecimento do domínio de interesse e prover os axiomas para definir a semântica dos termos, especificando suas definições e restrições sobre sua interpretação;
- **formalização da ontologia:** consiste em escolher um formalismo de representação das diversas categorias de conhecimento da ontologia, ou seja, descrever a ontologia em uma linguagem formal;
- **integração com ontologias existentes:** consiste em analisar e avaliar a necessidade de integrar a ontologia com outras já existentes, visando aproveitar conceituações previamente estabelecidas;
- **avaliação:** consiste em verificar se a ontologia satisfaz os requisitos estabelecidos na especificação. Este passo deve ser realizado seguindo alguns critérios tais como: clareza, coerência, extensibilidade, compromissos ontológicos mínimos. A avaliação é dividida em verificação e validação. A verificação refere-se ao julgamento técnico para garantir a consistência da ontologia e a validação garante que a ontologia corresponde ao sistema de representação proposto; e
- **documentação:** consiste em documentar todo o desenvolvimento da ontologia, incluindo propósitos, requisitos e cenários de motivação, as descrições textuais da conceituação, a ontologia formal e os critérios de projeto adotados.

Essas etapas podem ser agrupadas em um ciclo (Figura 22) proposto por Falbo *et al.* (1998) a fim de contemplar as atividades do processo de análise de domínio orientada a objetos e adaptado para construção de ontologias. As linhas tracejadas indicam que há uma interação constante. As linhas cheias mostram o fluxo principal de trabalho no processo de construção de uma ontologia. A linha envolvendo as etapas de captura e formalização da

ontologia realça a forte interação e, por conseguinte iteração, que ocorre entre essas etapas.

A etapa de captura pode apontar novos requisitos ainda não identificados. Na avaliação, pode-se perceber que os termos são insuficientes para o propósito planejado, impondo um retorno à etapa de captura. Situações semelhantes podem ocorrer na etapa de formalização: incoerências podem ser detectadas, provocando uma revisão das especificações e dos termos definidos na ontologia. Finalmente, quando for necessário integrar uma ontologia com outras existentes, este processo pode ter substancial impacto na definição e formalização dos termos devido a heterogeneidade semântica. Isso é um dos problemas mais difíceis para encontrar a interoperabilidade e a compatibilização do vocabulário entre as múltiplas fontes de informação e classes localizadas em *namespaces*.

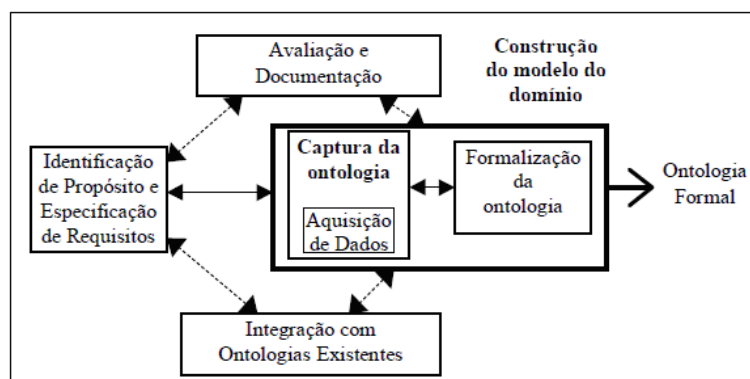


Figura 22 – Etapas processo de construção de ontologias agrupadas em um ciclo.  
**Fonte:** Falbo, 1998.

Na CI, a pesquisa de Silva (2008) traz uma inegável contribuição no sentido de propor princípios metodológicos para construção de ontologias. Silva analisou a metodologia para construção de tesouros e a norma para construção de vocabulários controlados. Neste caso, a formalização não tem o mesmo propósito do que nas ontologias, ou seja, não contempla nenhum formalismo lógico na representação do vocabulário. A formalização abrange os tipos possíveis de exibição (lista alfabética, gráficos) e formato (impressos, eletrônicos) dos vocabulários controlados destinados a seus usuários.

O método de “pensar” sobre um domínio do conhecimento na Teoria do Conceito preocupa-se primariamente com o entendimento do conceito, e em seguida com a classificação deste em um domínio de conhecimento. Este método pode ser considerado híbrido na agregação do método dedutivo e indutivo para modelagem de um sistema de conceitos. No primeiro formam-se categorias para representar contextos, e no segundo analisam-se os conceitos de um contexto no propósito de ordená-los no interior de tais categorias. A relação entre objetos pode ser as relações entre conceitos realizadas através da categorização (agrupamento dos conceitos de acordo com a natureza), pela relação

hierárquica (ordem de dependência entre os conceitos), pela relação partitiva (partes que compõem um determinado conceito) e pelas relações de equivalência (CAMPOS, 2004).

Segundo Campos (2004), o objetivo de se criar uma metodologia para realizar a modelagem não se restringe ao simples fato de criar métodos de modelar, mas principalmente melhorar os processos de recuperação, independentemente da área do conhecimento. Para representar informações em um modelo, tenta-se reproduzir em algum suporte (papel, tela, dispositivo eletrônico, entre outros), através de alguns símbolos ou linguagens, conceitos e significados construídos pela mente humana. Portanto, não existe uma representação perfeita de um objeto, a única representação perfeita é o próprio objeto.

Por isso, este trabalho propõe a modelagem informacional usando dados, ou melhor, “conceitos científicos”, do próprio hiperdocumento com um único mapa estruturado por uma ontologia na IA. Os padrões de metadados W3C oferecem mecanismos de inferências artificiais que lidam com o problema de alinhamentos conceituais que permitem conexões hipertextuais e associá-los a uma ontologia de domínio. A maioria das ferramentas de anotações semânticas que usam padrões de metadados digitais W3C adota um domínio de ontologia pré-selecionado para evitar que a geração automática do termo-chave indexado pelo sistema seja um problema de difícil interoperabilização e incompatibilização para que agentes de software recuperem informações hipertextuais.

Os padrões de metadados propostos pelo W3C deverão ser capazes de resolver a interoperabilidade semântica para poder fazer bom uso das informações disponíveis com a chegada das arquiteturas de sistemas distribuídos. Estes sistemas deverão ser capazes de “entender” o modelo do usuário com os usos e significados e também os modelos por trás das fontes de informação (SHETH, 1999). Além de ontologias, o uso do contexto também é fundamental nesta solução que, segundo Sheth, a nova geração de Sistemas de Informação deve se concentrar, a princípio, em domínios específicos desenvolvendo arquiteturas mais gerais como a proposta deste trabalho.

Os padrões de metadados W3C permitem que a estrutura de um hiperdocumento possa ser diretamente descrita em uma gramática formal usando uma ontologia de domínio como um dispositivo de controle de termos na representação de conceitos-chave. Esses padrões podem estabelecer novas técnicas de modelagem de objetos informacionais digitais usando o “conceito” como um metadado descritivo, administrativo e estrutural. Acredita-se que alguns processos de análise de assunto para a extração de conceitos válidos podem ser automatizados.



## 2.5 Bibliotecas digitais especializadas e a necessidade de padrões de metadados W3C para conduzir interoperabilidade semântica

O desenvolvimento de modelos de bibliotecas digitais especializadas é uma tendência mundial que contribui com novos meios de comunicação científica. Bibliotecas digitais especializadas são unidades pertencentes a instituições governamentais, particulares ou associações formalmente organizadas com o objetivo de fornecer ao usuário a informação relevante de que ele necessita, em um campo específico de determinado assunto (CESARINO, 1980). A produção acadêmica de uma instituição de ensino superior, para que seja citada na literatura, precisa ser facilmente recuperada de modo a contribuir para o aumento da visibilidade, expandindo o acesso aos resultados da investigação e controle acadêmico sobre a publicação científica (BASTOS, 2005).

No Brasil, encontra-se mais de duzentos modelos<sup>39</sup>. Em 2009, foi inaugurada a World Digital Library (WDL), contando com um acervo de mais de mil itens. A WDL é um projeto da Biblioteca do Congresso dos Estados Unidos da América e da UNESCO em parceria com outras instituições de vários países, incluindo o Brasil. Tendências como holismo seriam úteis, ao integrar todas as seções e serviços oferecidos pela biblioteca digital, criando uma conectividade do usuário com o documento existente, não importando o lugar físico onde estão situados (BASTOS, 2005).

Bastos (2005) alerta que há uma organização determinista em modelos de bibliotecas digitais de teses e dissertações que sofrem alterações de acordo com as mudanças que ocorrem na sociedade de um modo geral. As aplicações para a produção de documentos em uma biblioteca digital baseiam-se nos princípios da documentação estruturada, utilizando metadados descritivos na organização de acervos bibliográficos. Metadados descritivos se constituem nas marcações semânticas que irão identificar tanto os elementos descritivos de um objeto digital quanto seus elementos semânticos no ambiente da Web e, em especial, nas bibliotecas digitais especializadas (LOURENÇO, 2005).

Lourenço (2005) comparou o Padrão Brasileiro de Metadados de Teses e Dissertações (MTD-BR) da BDTD<sup>40</sup>-Biblioteca Digital Brasileira de Teses e Dissertações, com o NDLTD-ETD-MS (Networked Digital Library of Theses and Dissertation e An Interoperability Metadata Standard for Theses and Dissertations). Os resultados mostraram que o MTD-BR adotou os mesmos campos do NDLTD-ETD-MS aplicados a quase todos os 15 elementos do Padrão DC, com exceção dos “dc.source” e do “dc.relation” que foram

---

<sup>39</sup> Levantamento realizado pela autora pelas informações <http://www.cg.org.br/gt/gtbv/alfabetica.htm> em junho de 2009.

<sup>40</sup> <http://bdtb.ibict.br>

descartados pelos dois padrões. Isso reforça que os serviços realizados nas bibliotecas digitais são muitos semelhantes aos de uma biblioteca tradicional e podem ser integrados aos padrões de metadados digitais propostos pelo W3C conforme proposto neste trabalho.

A representação documentária descreve os dados ligados à produção editorial dos documentos em bibliotecas digitais que possibilitam a organização e recuperação dos itens segundo suas especificidades sendo necessário que essas bibliotecas acompanhem o avanço tecnológico (LOURENÇO, 2005). Entretanto, esses avanços não acontecem na mesma velocidade em modelos de bibliotecas digitais especializadas, provocando atrasos na disponibilidade dos documentos científicos (MOLOSSI, 2008).

Bibliotecas digitais de teses e dissertações enfrentam os mesmos problemas relacionados a área de RI citados anteriormente em diversas partes deste trabalho. Molossi (2008) comenta sobre as dificuldades na BDTD devido ao aumento de produções científicas disponibilizadas digitalmente em crescimento exponencial. Sobre este assunto, Hélio Kuramoto e Equipe BDTD<sup>41</sup> destacam que algumas das dificuldades apontadas pela comunidade usuária fogem ao controle do IBICT, ao ressaltarem que “as teses e dissertações disseminadas pela BDTD são registradas e depositadas pelas universidades que compõem o consórcio da BDTD. Portanto, a qualidade dos seus metadados e o controle do acesso ao teor das mesmas é de responsabilidade das universidades que as depositam”.

Os resultados de Molossi (2008) indicam que o uso de ontologias com padrões de metadados W3C na BDTD conduz à interoperabilidade semântica. Segundo Bastos (2005), as tendências apontadas para a organização de um modelo de biblioteca digital especializada são as aplicações das teorias baseadas em áreas do conhecimento que permitem determinar a estrutura interna associada ao universo de conceitos, conhecidas como microestrutural e macroestrutural. A microestrutural trata do corpus teórico como unidade mínima de conhecimento, trabalhando assim o conceito e suas diferentes apresentações na CI e na Terminologia. A macroestrutural parte das disciplinas até os âmbitos temáticos mais gerais para explicar a estruturação de sistemas de conceitos como o uso de mapas conceituais, identificando a estrutura interna de um domínio

A microestrutural aborda teorias da linguagem natural que descrevem os resultados de aplicações com vocabulários controlados no processo de RI e propõe que, por meio dos motores ou robôs de busca, é possível o usuário chegar à informação que necessita, sem levar em conta os aspectos linguísticos e semânticos existentes. Todo vocabulário controlado é composto por um conjunto de termos-chave, que representam conceitos como signos lingüísticos, ficando dispostos em uma estrutura como um modelo relacional. Em geral, os instrumentos de controle de vocabulário são apresentados em

---

<sup>41</sup> Informação retirada da página inicial da BDTD

ordem hierárquica e alfabética (macroestrutura e microestrutura). A macroestrutura é o esquema temático global do vocabulário. Apresenta as categorias mais amplas do conhecimento de forma hierárquica. A microestrutura dispõe as unidades do vocabulário em termos descritores em forma alfabética (KOBASHI, 2002; BASTOS, 2005).

Este trabalho propõe avanços no MHTX para integrar aos novos mecanismos utilizados em modelos de plataformas abertas que visam a disseminação dos conteúdos de bibliotecas digitais com uso de ontologias e os padrões de metadados digitais W3C. Na IA, os robôs são programas ordenadores para recorrer de forma automática à estrutura de um sistema de hipertexto de um servidor Web, com o objetivo de alimentar as bases de dados textuais a partir de documentos em qualquer formato de edição eletrônica (TRILLO, 2004).

Senso (2003) cita três grandes linhas de ação desenvolvidas para auxiliar na administração, representação e área de RI: a) a primeira com índices compilados manualmente em base de dados onde os usuários são os criadores das páginas Web, sugerindo sua publicação dentro de algumas categorias, mediante um formulário; b) a segunda com bases de dados rastreadas por robôs ou *spiders* que parte de um esquema clássico de uma interface para a alimentação automática de uma base de dados; c) a terceira que segue o método de indexação distribuída como no método *Harvest*, usando uma arquitetura distribuída de servidores e serviços por toda Web.

Os elementos principais são: *Gatherers*, que consiste em um software instalado em um servidor na Web que, periodicamente, extrai informação relativa aos arquivos disponíveis para a comunidade de usuários da Internet; *Brokers* que recuperam automaticamente a informação extraída por um ou mais *gatherers* e integram em índices sobre os que se podem lançar equações de busca como no protocolo do sistema de metadados denominado SOIF (Summary Object Interchange Format).

A OAI (Open Archives Initiative) é uma tecnologia que permite a interligação provendo acesso a outros repositórios na tentativa de desenvolver e promover padrões de interoperabilidade entre repositórios digitais. O protocolo OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) oferece meios de se compartilhar metadados entre sistemas hipertextuais e servidores distribuídos. São essas tecnologias que permitem interoperabilidade entre repositórios digitais e facilitam o acesso a fontes de informação científica e cultural sem custo adicional além de possibilitar a RI (BASTOS, 2005).

A construção de bases de conhecimento interoperáveis com o uso de padrões de metadados propostos pelo W3C foi facilitada por meio das ontologias desenvolvidas na IA. Os padrões W3C propõem tecnologias avançadas visando a representação estrutural dos recursos na Web e podem documentar procedimentos associados à preservação, especificar direitos de acesso e estabelecer a autenticidade do conteúdo digital. Dentro

dessa perspectiva, o modelo de referência Open Archival Information System - OAIS tem se destacado na comunidade internacional dos profissionais ligados a instituições culturais, já contando com várias iniciativas de implementação (BASTOS, 2005).

A construção de um modelo de biblioteca digital especializada exige grandes investimentos além de uma equipe de visão multidisciplinar. Vicentini (2006) cita os seguintes componentes necessários: a) coleção/conteúdo; b) recursos humanos (equipe multidisciplinar, capacitação); c) padronização de metadados digitais; d) tecnologia (hardware, software, flexibilidade de desenvolvimento, facilidade de gerenciamento da coleção digital, linguagem de programação, utilização de protocolos de regras e convenções entre redes de comunicação para importação e exportação de dados); e) digitalização; f) garantia de direito autoral; e g) preservação do documento com os metadados digitais.

Na atualidade, os modelos de bibliotecas digitais de teses e dissertações necessitam aprofundar seus processos de classificação e categorização (BASTOS, 2005). Nunes e Tálamo (2009) observaram que esses dois processos apresentam limites no tratamento da informação e concluíram que a classificação bibliográfica não contempla amplamente a busca e recuperação da informação desejada. A urgência de um novo sistema de classificação com rigor e métodos científicos é destacada por muitos profissionais da CI como mais um dos paradigmas na organização da informação e representação do conhecimento (MIRANDA, 2005).

Carlan (2010) cita que os esquemas de classificação como CDD (Classificação Decimal de Dewey); CDU (Classificação Decimal Universal), LCC (Library of Congress Classification) têm a função de organizar e localizar, nas estantes, um único item de uma coleção. Assim as bibliotecas digitais necessitam usar sistemas de organização do conhecimento para promover uma visão geral do conteúdo da coleção e a sua recuperação, podendo, também, ser usado para compartilhar informações entre diferentes sistemas.

Schreiner (1979) alerta que a CDU procurou atingir maior flexibilidade através do uso de sinais de relação, divisões comuns e divisões analíticas, mas o resultado, lamentavelmente, foi uma magnificação dos defeitos herdados da CDD. Assim, essas classificações bibliográficas repousam todas elas sobre os cinco predicados aristotélicos. Segundo Schreiner, a relação gênero-espécie, entretanto, continua sendo princípio válido e muito utilizado para subdivisão de uma faceta ou categoria formal nos sistemas modernos. Os sistemas de classificação facetada apresentam contribuições que podem favorecer outras áreas de conhecimento, como Linguística, Filosofia e Lógica (ARAÚJO, 2006).

A diversidade dos sistemas de informação demonstra que eles acarretam em uma série de problemas, como a falta de integração, a redundância e a inconsistência de dados, perda de conteúdo semântico nas traduções, dispêndio de tempo e dinheiro na

transferência de um sistema para outro ou ainda o trabalho de redigitação. Para Freitas (2006), a interoperabilidade e integração plena de informações heterogêneas entre sistemas de informação passam a ser possível quando novos modelos são desenvolvidos a partir de estrutura de dados digitais conhecidos como metabanco de dados ou metadados.

Metadados digitais são dados estruturados para auxiliar na descrição, identificação, compreensão, administração e preservação de conteúdos de documentos digitais, além de facilitar a interoperabilidade de repositórios em modelos de bibliotecas digitais. A especificação e utilização de padrões metadados garantem a existência de um conjunto de informações comuns sobre um determinado tema ou área, com regras claramente estabelecidas e aceitas pela comunidade envolvida.

Padrões de metadados digitais W3C podem ser associados a diversos componentes constituintes de objetos informacionais com anotações semânticas realizadas no próprio hiperdocumento que podem ser associadas em uma ontologia de domínio para armazenamento e preservação de informações digitais de uma base de dados. Os maiores benefícios com a preservação digital distribuída em diferentes locais de armazenamento é o trabalho colaborativo em rede e a dispersão geográfica de servidores para aumentar a solidez contra falhas e minimizar o risco de perda de dados. Esses padrões de metadados digitais podem ser armazenados em locais diferentes, neste caso, é importante um controle cuidadoso para não gerar incompatibilidade de dados (RIVERA DONOSO, 2009).

Segundo Understanding Metadata, documento publicado pela National Information Standards Organization (NISO), existem três tipos de metadados digitais: a) metadados descritivos; b) metadados administrativos; e c) metadados estruturais. As principais funções desses metadados são: mostrar instruções de como interpretar um dado; determinar se o dado é o que se necessita prevenindo certos usos; resumir o significado dos dados; obter informações sobre as condições de uso (direitos do autor); obter informação a respeito do ciclo de vida do dado; oferecer informação relativa ao criador do dado; indicar relações com outros recursos; controlar a gestão da informação; permitir a busca e a recuperação de um dado, além de utilizar a cópia de um dado.

O grupo OCLC/RLG - Working Group on Preservation Metadata observa que, no contexto dos objetos digitais, os metadados estão associados a uma de três categorias funcionais: a) descritivo: facilitar a descoberta e identificação de fonte; b) administrativo: suportar o gerenciamento de fonte dentro de uma coleção; e c) estrutural: juntar os componentes de objetos de informação complexos. Das três categorias, a que tem recebido maior atenção é a categoria descritiva que visa solucionar a pesquisa de fontes eletrônicas.

Contudo, Freitas (2006) alerta que há necessidade de avançar mais nas pesquisas de metadados, além da descoberta de fonte, devido à crescente preocupação

dos desafios colocados pela preservação digital. O modelo proposto neste trabalho pode contribuir para avançar o tratamento documentário no MHTX no sentido de valorizar o “conceito” como um metadado descritivo, administrativo e estrutural, ou seja, considerar as diversas funcionalidades do conceito nos ambientes digitais.

O “conceito” tem sido apontado como mecanismo essencial no desenvolvimento de instrumentos como taxonomias, tesouros e ontologias, podendo se constituir no componente invariável dos processos de bibliotecas tradicionais e digitais. O desenvolvimento de novos sistemas baseados em ontologias visa alcançar interoperabilidade para gerência dos dados e integração dos sistemas existentes. As ontologias na IA tem sido amplamente usada tanto na representação como na recuperação de documentos em sistemas heterogêneos para extração de informações em textos de linguagem natural, troca de informações entre agentes de software “inteligentes”.

Estudar o “conceito” e suas relações é de fundamental importância para a CI uma vez que permite fundamentar a análise de assunto para o processo classificatório em forma de facetas, além de permitir o desenvolvimento de novos modelos de sistemas de classificação bibliográfica com método e rigor científico. Essa interconexão efetiva de diferentes sistemas de computador, base ou bancos de dados em redes tem o objetivo de apoiar sistemas distribuídos para intercâmbio de informações (RIVERA DONOSO, 2009). Outros autores também valorizam o dado (Quadro 7) referindo-se ao termo metadados como uma estrutura descritiva de uma informação necessária para administrar um recurso sobre outro e auxiliar na localização da informação.

**QUADRO 7 – Nível de Gerenciamento do Conhecimento**

Estágio	Recurso a ser administrado	Definição de Metadados
Dados	Valores dos dados	Informação necessária para administrar o recurso dos dados
Informação	Valores dos dados e o contexto da informação	Informação necessária para administrar o recurso da informação
Conhecimento	Valores dos dados, contexto da informação e instruções das regras de negócio	Informação necessária para administrar as regras e políticas de negócio da organização
Sabedoria	Valores de dados, contexto da informação, regras de negócio executáveis, monitoração das regras de negócio e regras e métricas de avaliação	Informação necessária para administrar o comportamento da organização de acordo com suas regras e políticas de negócio.

**Fonte:** FREITAS, 2006.

A relação entre a CI, suas metodologias e práticas e a pesquisa em ontologias é cada vez mais percebida pelos pesquisadores desta área. De acordo com Alvarenga (2001), o “conceito” e a compreensão do que seja conceito é tema crucial, uma vez que pertence à

essência do trabalho de tratamento e organização da informação, compreendendo os processos de análise de assunto, classificação e recuperação da informação. Essa noção de estruturação taxonômica em um sistema de conceitos, por sua vez, concretiza-se, em plano real, fundamentada nas operações de análise-síntese, que se verificam não apenas em plano lógico, mas também num contexto ontológico (CAMPOS, 2007).

Estudos de natureza metodológica para apoiar o levantamento de termos que compõem as unidades de um dado domínio para representação do conhecimento e organização da informação têm sido muito investigados por pesquisadores da CI. Segundo Campos (2004), estudos relacionados à compatibilização de linguagens, desde 1980, já apresentavam princípios que podem ser aplicados em modelagem de ontologias.

A partir das características presentes na definição do “conceito” em um sistema, Campos e outros (2009) vêm desenvolvendo estudos com a intenção de verificar até que ponto o processamento automático traria resultados consistentes para uma análise de potencialidades semânticas utilizando ontologias. Esses estudos visam estabelecer critérios para serem utilizados em uma mostra de termos que possuem coincidência verbal em algumas ontologias selecionadas no campo da Bioinformática.

Os resultados desses estudos apontam para a necessidade de estudos mais aprofundados no que tange as definições do conceito, pois são elas, além das relações conceituais, em ontologias, que evidenciam o conteúdo semântico de um conceito. Naves (2000) destaca que para organizar e representar um ‘conceito’ em um sistema, é necessário trazer à mente o campo do conhecimento que originou o conceito e suas características. Também devem ser consideradas as expectativas dos usuários e as metas do sistema que utilizarão podendo variar os objetivos de acordo com a representação do conhecimento.

Tudo isso demanda tempo, custo e experiência. Naves (2000) detectou que a falta de experiência profissional e do domínio de uma língua natural local interfere negativamente neste trabalho. Os termos de indexação gerados a partir destes procedimentos devem ser associados ao vocabulário do documento, dependendo de um longo processo de leitura, avaliação e citação pelos pares para que os novos conhecimentos possam, enfim, ser incorporados ao acervo de conhecimento público, aceito num determinado campo ou domínio do conhecimento (MARCONDES, 2008). Então, cabe aos profissionais experientes nos processos de tratamento da informação com uso de linguagens documentárias, principalmente nas técnicas de análise de assunto, modelar um único “sistema de conceitos” representado por uma ontologia na base de dados MHTX.

Acredita-se que este mapa conceitual precisa unir processos da CC e da CI com linguagens documentárias. Imagine o tempo e o custo que uma equipe de profissionais pode gastar para realizar o processo de indexação de todos os hiperdocumentos científicos de

uma base de dados para extração de conceitos válidos, representando em cada mapa conceitual de acordo com as propostas iniciais de Lima (2004) no MHTX! Por isso busca-se um novo tipo de modelagem informacional para que o próprio autor no momento da publicação possa utilizar um sistema “inteligente” com um único mapa conceitual representado por uma ontologia de domínio que permita estabelecer relações de equivalência entre conceitos e mecanismos de inferência com anotações semânticas.

Segundo Pereira e Bufrem (2005), quando são criados códigos de classificação como produtos documentários e termos descritores, utilizando linguagens documentárias, se adota a noção de documento impressos ou digitais, passíveis de catalogação de recursos para compreender a representação de conteúdos representativos em um sistema de conceitos. Esta é a ideia principal deste trabalho com o uso de ontologias em modelos de bibliotecas digitais, podendo contribuir como instrumentos modernos da CI e da CC.

Ontologias são ferramentas poderosas para o estabelecimento de um vocabulário consensual na organização de tarefas e que podem auxiliar na realização de consultas específicas em um suporte (ALMEIDA, 2006). O uso de padrões W3C com ontologias e novas ferramentas de anotações semânticas estão possibilitando um tratamento sem ambiguidades do vocabulário e a recuperação de documentos relacionados aos termos expressos na consulta por um usuário-sistema.

Nunes e Fileto (2007) propõem uma arquitetura de um sistema dirigida por semântica em que os documentos se encontram armazenados em uma biblioteca digital especializada utilizando uma ontologia de domínio e anotações semânticas em documentos jurídicos (Figura 23). Nesta arquitetura, o conhecimento do domínio jurídico foi representado por um vocabulário controlado que tornaram possível a recuperação automática da informação, mediante a identificação de expressões jurídicas desse vocabulário nos textos.

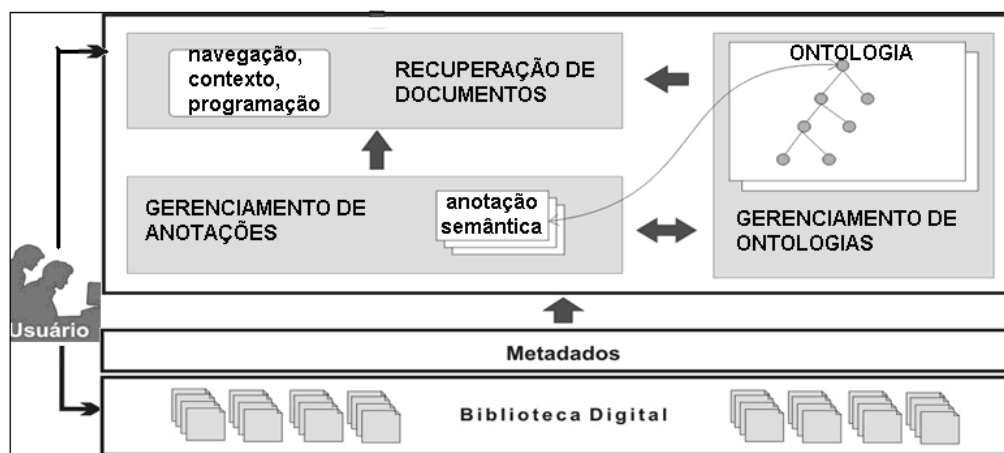


Figura 23 Recuperação de informação dirigida por semântica.  
Fonte: Adaptado de NUNES e FILETO, 2007; MOLOSSI, 2008.



Neste modelo, o módulo de gerenciamento de ontologias é responsável pela gerência da evolução das ontologias. O módulo de recuperação e recomendação de documentos implementa a interface com o usuário que permite a navegação e a busca no conteúdo da biblioteca digital baseada nas ontologias e nas anotações semânticas dos documentos. O módulo de extração e gerenciamento de anotações da arquitetura é o responsável pelo processamento dos documentos inseridos na biblioteca digital, extração de entidades nomeadas, definição das anotações semânticas. Além disso, o módulo de extração mantém e controla o repositório de anotações semânticas, bem como a ligação com seus respectivos documentos para suportar a recuperação relevante de informação.

Neste modelo, a ontologia descreve relações semânticas entre os termos do domínio e a anotação semântica de um documento descreve o seu conteúdo pela associação dos conceitos descritos na ontologia. Esses mecanismos compreendem a disponibilização de um conjunto de coleções estruturadas de informações e regras de inferência que são associados aos agentes de software capazes de percorrer a Web realizando tarefas complexas com base nessas estruturas de conhecimento. A recuperação foi realizada com o *matching* parcial usando uma medida de similaridade para realizar os experimentos necessários à validação dos resultados. Esta arquitetura é a ideal para o MHTX utilizando a ontologia como um único mapa conceitual com conceitos integrados.

Uma estrutura de tópicos conceituais pode fornecer relação de equivalência entre conceitos, usando linguagens documentárias e padrões de metadados W3C. O vocabulário controlado possui limites próprios de unidades de conhecimento de objetos informacionais registrados que se organizam em relações de dependência hierárquica para formar um domínio. Este sistema de conceitos deve representar os campos conceituais, respeitando a cultura da comunidade e os “conceitos científicos” com uma terminologia consensual à qual esta linguagem serve (CAMPOS, 2007). Então, o mapa conceitual do MHTX precisa explicitar o vocabulário, informando significados conceituais de um domínio e respeitando esquemas de metadados descritivo, administrativo e estrutural.

## **2.6 Bibliotecas digitais interativas e as propostas para o MHTX**

Desde o surgimento do hipertexto, a literatura vem enfatizando que um dos principais problemas relativos ao uso dessa nova ferramenta é o direcionamento ineficaz do leitor no momento da busca ou consulta. Partindo dos pressupostos de que essa situação é comumente causada pela falta de critérios consistentes no processo de organização conceitual do hipertexto, Lima (2004) dá os primeiros passos na construção do MHTX – Modelo Hipertextual para Organização de Documentos.

Posteriormente, Lima continuou sua pesquisa no desenvolvimento deste protótipo. Para uma escolha consistente das facetas no MHTX, foram analisados os princípios do plano de ideias das cinco categorias do PMEST oriundos da Teoria da Análise Facetada propostos pelo CRG (Classification Research Group) e comparados com os princípios do modelo simplificado de Spiteri (1998). Baseados em relações de categorias conceituais, os princípios de escolha das facetas e subfacetas do plano das ideias do modelo simplificado de Spiteri foram considerados apenas os seguintes: Diferenciação, Relevância, Verificação, Permanência, Homogeneidade, Exclusividade Mútua e Categorias Fundamentais. Lima esclarece cada um desses princípios:

- Princípio de Diferenciação, proposto por Ranganathan, é baseado em características de divisão nas quais determinadas diferenças ou qualidades comuns distinguem elementos da mesma classe. Por exemplo, o termo “Bibliotecário”, pode ser dividido pela natureza de seu trabalho e depois, pela característica de sua experiência e pelo seu grau de conhecimento.
- Princípio de Relevância, proposto tanto por Ranganathan quanto pelo CRG, ajuda a assegurar que as facetas escolhidas reflitam a proposta, o assunto e o escopo do tema tratado. Por exemplo, em um sistema de classificação construído para a disciplina “Organização do conhecimento”, é sensato que se utilize a faceta Energia (Processos, Atividades, Ações) para dividir os conceitos que representam os processos da organização do conhecimento, como por exemplo, “Processos mentais” e “Indexação”.
- Princípio de Verificação, proposto por Ranganathan e pelo CRG, afirma que é importante escolher facetas que sejam definitivas e que possam ser verificadas. No caso desta pesquisa, em que a terminologia utilizada é a do próprio autor da tese, o princípio de verificação com alto grau de acerto na escolha das características foi assegurado.
- Princípio de Permanência, proposto por Ranganathan e pelo CRG, sugere que as facetas escolhidas devam representar características de divisão com qualidades permanentes ao assunto a ser dividido. O exemplo dado anteriormente também pode ser utilizado como exemplo para exemplificar esse princípio, pois reflete a interpretação de permanência: o indexador será sempre o bibliotecário, o que é uma característica permanente.
- Princípio de Homogeneidade, proposto apenas pelo CRG, afirma que os conteúdos de duas facetas não devem sobrepor-se e cada faceta deve representar somente uma característica de divisão. Por exemplo, ao

dividirmos o conceito “Texto” pelas facetas “Tipologia” e “Estrutura”, os termos que aparecerem em uma faceta não poderão aparecer na outra.

- Princípio de Exclusividade Mútua, proposto por Ranganathan e pelo CRG, afirma que as facetas devem ser mutuamente exclusivas. Muito semelhante ao princípio de homogeneidade, mas de caráter mais limitante, este princípio visa assegurar a exclusividade das classes de um renque, derivando de seu universo imediato uma e somente uma característica.
- Princípio de Categorias Fundamentais, proposto apenas pelo CRG que discorda das categorias fundamentais no sentido de preferir identificá-las com base na referência do contexto do próprio assunto, sugerindo que nenhuma lista pode ser imposta mecanicamente ao assunto.

Quanto ao Princípio de Categorias Fundamentais proposto apenas pelo CRG, Lima (2004) esclarece que nenhuma lista deva ser necessariamente exaustiva ou aplicada a todos os assuntos. Este princípio adotado somente pelo CRG permite aos classificadores moldar categorias para assuntos com termos de vocabulários controlados distintos. Lima utilizou a Sucessão Consistente e Relevante. A Sucessão Consistente é baseada no cânone de consistência de Ranganathan que visa manter a consistência na estrutura de um sistema de classificação. Sua expansão permite que a ordem de citação, tanto das facetas, quanto dos focos, seja mantida consistentemente.

Na Sucessão Relevante, tanto Ranganathan quanto o CRG concordam que a ordem de citação das facetas deve levar em consideração a natureza, o assunto e o objetivo do sistema de classificação. Esta ordem, apresentada aqui, é derivada do Princípio de Sequência Relevante (Helpful Sequence) de Ranganathan e do Princípio de Ordem de Renque (Array) do CRG. Depois de estabelecidas as facetas, foram observadas as seguintes ordens, listadas para construir a ordenação com consistência:

- Ordem cronológica: ordem cronológica crescente, incluindo operações que, necessariamente, são realizadas uma após a outra.
- Ordem geométrica/espacial: ordem para os focos que possuem essa natureza organizacional, segundo arranjos espaciais ou geométricos.
- Ordem do simples para o complexo: ordem crescente de facetas.
- Ordem do complexo para o simples: ordem decrescente de facetas.
- Ordem canônica: ordem tradicional, quando não há nenhum outro princípio.

- Ordem de aumento de quantidade: ordem crescente de quantidade, quando o assunto em um renque de assuntos, ou o isolado em um renque de isolados tiver essa qualidade.
- Ordem de diminuição de quantidade: ordem decrescente de quantidade, quando o assunto em um renque de assuntos, ou o isolado em um renque de isolados tiver essa qualidade.
- Ordem alfabética: ordem alfabética pelos nomes de uso internacional mais corrente, quando nenhuma outra sequência de assunto em um renque de assuntos ou, então, nenhuma sequência de isolados em um renque de isolados for mais útil.

Assim, as relações semânticas propostas para estruturar o mapa conceitual no MHTX foi determinado a partir das relações hierárquicas e associativas da Teoria do Conceito e da Teoria da Análise Facetada. Os princípios para ordem de citação das facetas e focos do plano das ideias determinam como os focos serão organizados dentro de suas respectivas facetas, ou seja, como será a ordenação destes focos no renque. A sequência dos conceitos é construída com critérios firmemente estabelecidos por Ranganathan, mas que transcendem a seu propósito inicial que foi de criar uma tabela de classificação bibliográfica para organizar os livros na estante (LIMA, 2004).

A estrutura do mapa conceitual de interface hiperbólica foi implementada no MHTX utilizando-se o Software Star Tree Studio para compor uma estrutura semântica tridimensional, em que estão representados todos os conceitos relevantes de um documento e suas relações semânticas (Figura 24). Os procedimentos metodológicos foram: a) identificação e leitura do documento; b) análise facetada do assunto: seleção de termos relevantes e categorias (facetas); o reconhecimento das subfacetas; a ordenação das facetas, subfacetas e focos apresentados no mapa conceitual e finalmente, a organização de todos os termos e suas relações; c) criação do mapa conceitual com seus links e suas relações; d) estruturação e criação dos links do sumário expandido para o texto.

Segundo Lima (2004), a interface hiperbólica exibe as seções mais importantes primeiramente. As outras partes podem ser expandidas e lidas quando o usuário arrasta outros *links* com o mouse, ou ainda, quando se apresentam completamente ocultas, por meio de pesquisa direta por conceitos na busca por termos-chave indexados ao modelo hipertextual semanticamente estruturado para navegar. Sua navegação é caracterizada pelas inter-relações existentes entre os conceitos. Assim, a análise facetada implica em apreender e agrupar os conceitos segundo seus traços perceptivos e segundo categorias já estabelecidas anteriormente, envolvendo as etapas de seleção, ordenação e agrupamento.

Na elaboração do mapa conceitual, as relações de semelhanças e diferenças são explicitadas. O mapa permite navegar em conceitos semanticamente relacionados. O acesso ao texto da tese realça a palavra-chave pesquisada e, a partir daí, o usuário poderá navegar dentro do texto, com a opção de ir e vir na estrutura dos links do mapa. A navegação do mapa é caracterizada pelo mecanismo de folheio e pelo mecanismo de pesquisa de conceitos. Na movimentação dessa interface, conceitos integrados aos hiperdocumentos com os nós de uma rede semântica podem ser expandidos quando o usuário arrastar os nós com o mouse ou, ainda, por meio de pesquisa direta pelo nó.

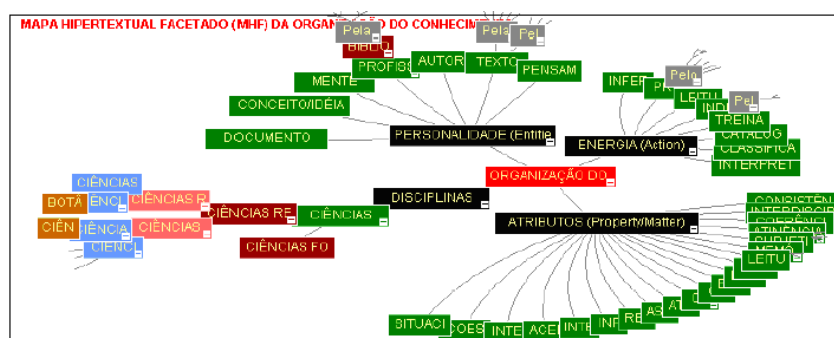


Figura 24 - Primeira tela do mapa conceitual hiperbólico do MHTX.

Os hiperdocumentos científicos produzidos nas universidades merecem um tratamento mais adequado da informação. Bastos (2005) nos lembra que esses documentos devem ser estocados em bases de dados devidamente organizados para posteriormente serem recuperados. Entretanto, nem todos os documentos científicos produzidos nas universidades são estocados com informações registradas e organizadas em modelos de bibliotecas digitais. Ao estocar informações nesses modelos, elas passam a ter uma existência institucional e social para divulgação e acesso ao conhecimento científico.

Para isso, o tratamento envolvido no fluxo da informação científica precisa estar inserido dentro de um contexto para estabelecer vínculos de significados conceituais com usuários-sistema desses modelos. Este trabalho propõe inserir uma ontologia na base de dados do MHTX para auxiliar tanto “usuários” quanto “agentes de software” a estabelecerem esses vínculos entre o mapa e o próprio texto da interface humano-computador. As tecnologias intelectuais, as quais representam formas de expressão simbólica da própria escrita hipertextual, situam-se nos sujeitos através da imaginação e da aprendizagem como códigos compartilhados e textos que circulam impressos ou digitais (LÉVY, 1993).

Albrechtsen e Hjørland (1995) alertam que os princípios para se construir um “sistema” devem ser propostos pelos profissionais da CI e não pelo “usuário”. Esses autores acreditam que trabalhar com a análise de assunto em um documento na perspectiva dos

profissionais da CI corresponde entender a informação por meio dos domínios de conhecimento. Por meio deste conhecimento científico organizado e categorizado, esses profissionais podem fornecer aos “usuários” uma visão mais completa possível sobre as teorias, métodos e abordagens de um determinado assunto para organização de um domínio, demonstrando que o conhecimento pode ser documentado e facetado. Na visão desses autores encontram-se estabelecidos a organização do conhecimento, estrutura, padrões de cooperação, linguagens, formas de comunicação e critérios de relevância para estruturar um modelo de sistema artificial de RI (BASTOS, 2005).

### **2.6.1 Linguagens documentárias e compatibilização do vocabulário**

O texto científico impõe uma seleção cuidadosa de palavras fortemente orientada pela normalização, pois, sem um controle do vocabulário para fins de indexação, as palavras de um texto podem produzir efeitos indesejáveis na recuperação. Lancaster (1993) faz uma importante observação sobre o uso de normas ao promoverem a compatibilidade estrutural dos vocabulários, facilitando a conversão de um vocabulário para outro: a compatibilização de vocabulário controlado pode ocorrer no nível semântico, ligado à capacidade de ambos representarem o mesmo corpo de conhecimento e no nível estrutural que é responsável por garantir a conformidade das relações estabelecidas entre os elementos (SILVA, 2008).

Silva (2008) destaca que os primeiros estudos acerca da compatibilização de linguagens tiveram início na década de 60 e um dos principais resultados alcançados foi a definição de alguns conceitos centrais para a área e suas respectivas aplicações, sobretudo no âmbito da cooperação entre centros de informação. Porém, foi na década de 70 que esses estudos tiveram uma maior produção como o relatório UNISIST, publicado em 1971, que definiu ‘compatibilidade’ como a qualidade dos sistemas cujos produtos podem ser usados de maneira intercambiável sem nenhuma ‘conversão’<sup>42</sup>.

A Teoria do Conceito apresentada por Dahlberg (1978) e seus estudos de compatibilidade e convertibilidade entre linguagens são apontados como principais princípios metodológicos para criação de instrumentos de conversão e/ou desenvolvimento de linguagens compatíveis que viabilizem o acesso a múltiplas bases de dados operando em bases cooperativas. Segundo Campos (2007), essas bases teóricas da CI estão possibilitando identificar níveis de compatibilização lingüística e semântica visando apresentar propostas de harmonização terminológica e determinando diretrizes para as

---

<sup>42</sup> Essa conversão é o processo de transformar registros de informação, baseado na transcrição de códigos de estruturas de dados fazendo com que os mesmos sejam intercambiáveis entre dois ou mais serviços ou sistemas que utilizam diferentes convenções e mídias (DAHLBERG, 1981 *apud* SILVA, 2008).

atividades de integração e compatibilização de linguagens para futuramente possibilitar a viabilização de procedimentos automatizados.

Na CI, Dahlberg (1979; 1981; 1983) propõe um método de compatibilização e conversão de linguagens após verificar os princípios da classificação facetada, avaliando as diferentes facetas de um documento. Segundo Silva (2008), na matriz de compatibilização conceitual apresentada por Dahlberg (Quadro 8) é possível perceber as três fases propostas na compatibilização de linguagens: verbal, conceitual e estrutural.

O método de compatibilização e conversão de linguagens proposto por Dahlberg teve como base a integração de vocabulários e apresenta uma matriz de compatibilidade conceitual formada por três fases: a) coincidência conceitual quando dois conceitos combinam suas características e grau de equivalência; b) correspondência conceitual quando dois conceitos combinam a maior parte de suas características e similaridade; c) correlação conceitual quando dois conceitos são correlacionados através de símbolos matemáticos, estabelecendo uma medida de correlação, quando possuem diferentes níveis de detalhe, ou quando a relação entre eles não é de semelhança (SILVA, 2008).

Este método fornece resultados da análise de assunto sob os pontos de vista: compatibilidade conceitual, semântico e estrutural. A coincidência conceitual acontece quando dois conceitos combinam suas características possuindo uma relação de equivalência conceitual. Essa correspondência conceitual entre dois conceitos que combinam a maior parte de suas características possui uma similaridade sendo que nessa correlação conceitual pode ser estabelecida uma medida de símbolos matemáticos.

**QUADRO 8 – Exemplo de seção da matriz de compatibilização conceitual de vocabulários**

No.	Name	DDC	BBC	UNT	cc
		361			
65	Social Welfare	Social Problems and Social Welfare	Q	R 85/99	3
		361.01	QAE	R 86	
65.1	Soc. Welf. Philos.	Philos. & Theory of	Philos. Of soc. Welf.	Soc. Welfare Philosophy	3
			QAG		
65.2	Soc. Welf. Admin	-	Soc. administration, Soc. welf. administration	R 90 Social welfare administration	2
		> 361.25			
65.3	Welfare Policy	Action within establ. soc. framework (policy)	QAG P	R 87	3
			Policy (in social welfare)	Welfare policy	
		> 361.25	QAH		
65.4	Soc. Welf. Planning	Act. within establ. soc. framework (planning)	Planning for welfare, social planning	R 88	3
			QAT M/Z	Management	2
65.5	Soc. Welf. Econom.	-	of soc. Welf.	R 89	2
				Social welfare economics	2

Fonte: Dahlberg, 1981.

Neste quadro, as duas primeiras colunas correspondem ao sistema escolhido. As três colunas seguintes indicam o nível de equivalência entre termos dos outros sistemas, utilizando-se alguns símbolos: '≠' para indicar divergência de conceitos, '<' para conceitos mais genéricos do que o da segunda coluna, '>' para os mais específicos e '>' para a combinação de conceitos. A última coluna indica o número de coincidências, permitindo calcular o grau de compatibilidade.

Dahlberg (1979) destaca que é através das características de um conceito traduzidas em seus atributos que são realizados os processos de análise-síntese do conceito. Em geral, a quantidade total de complementos de um verbo é limitada à valência "natural" no predicado em questão ("funcionalidade" das características de um "conceito") e maneira necessária de formar e ampliar enunciados para abranger os detalhes necessários à informação sobre determinados objetos informacionais. Mas, só é possível obter essas características formulando enunciados sobre os atributos necessários ou possíveis sobre determinado objeto em um contexto.

A matriz de compatibilidade conceitual é um mapeamento da potencialidade semântica dos instrumentos de vocabulário controlado criados pela CI, consistindo basicamente na extração de conceitos válidos na análise de assunto para traduzir o conteúdo numa linguagem documentária que sirva de intermédio entre usuário-documento (CAMPOS, 2001). Esses instrumentos, ao ligar semanticamente termos que tenham relação entre si, podem ajudar um leitor-interpretador a identificar outros termos-chave que seriam necessários para realizar uma busca completa em um domínio em bibliotecas digitais. Esses instrumentos facilitam o arranjo físico mediante uma linguagem artificial de códigos ou símbolos por reunir fisicamente os documentos em uma base de dados que tratam de assuntos de um mesmo domínio, estruturando um sistema de conceitos.

Os instrumentos de controle do vocabulário desenvolvidos na CI são comumente classificados quanto à coordenação dos conceitos em pré-coordenados e pós-coordenados. Os vocabulários pré-coordenados estabelecem a coordenação dos vários tópicos referentes a um assunto composto no momento da indexação. Os pós-coordenados consistem de entradas (termos indexadores) e de uma saída que permite a comparação das entradas de assunto para determinar coincidências que revelem documentos pertinentes ao assunto composto pesquisado. Assim, a coordenação dos elementos de um assunto composto na busca pode ser transferida, ou seja, pós-coordenação em relação ao momento da indexação, comparando termos formulados pelo usuário na busca de informação (PINTO, 1985).

De acordo com Kobashi (2002), para que um vocabulário controlado seja útil, deve refletir, de um lado, os objetivos do Sistema de Informação para o qual foi elaborado e,



de outro, a linguagem dos usuários. Por essa razão, sua construção é coletiva, requer trabalho integrado, colaborativo, envolvendo tanto os gerenciadores do Sistema de Informação quanto os usuários da informação. Além disso, os instrumentos utilizados para o controle do vocabulário propiciam uma linguagem dinâmica que se desenvolve em consonância com as áreas de conhecimento representadas no Sistema de Informação necessitando, portanto, de atualização periódica dos instrumentos de vocabulário controlado.

Esses instrumentos subsidiar as atividades de indexação e recuperação de conteúdos documentais retifica as dissonâncias terminológicas provocadas pelas multiplicidades de vocabulários envolvidos nos processos de tratamento documentário, pois o vocabulário controlado é representado por um único conceito-chave utilizado como descritor na estruturação semântica de um sistema de conceitos. Devido ao mapeamento de um vocabulário controlado em outro ser considerado uma tarefa intelectual cansativa e demorada, Silva (2008) destaca que alguns algoritmos foram desenvolvidos com o objetivo de automatizar pelo menos uma parte do processo.

Contudo, há necessidade de pesquisas interdisciplinares, envolvendo profissionais da organização e representação do conhecimento de diferentes áreas para integração dos processos de compatibilização. As dificuldades encontradas no processo de compatibilização são originadas, principalmente, na subjetividade da escolha de descritores e relações hierárquicas durante a construção desses vocabulários (SILVA, 2008). Lima (1998) destaca que, a partir do referencial terminológico da área do conhecimento a que pertence o acervo independente do suporte físico de seus documentos, pode-se evitar a diversidade de representações para a mesma informação, o que torna a Terminologia também um instrumento essencial quando há necessidade de compatibilizar vocabulário.

Quando esses instrumentos recebem atualização periódica, eles apresentam muitas vantagens de uso nos atuais sistemas de RI, tais como: controle de termos sinônimos; distinção de homógrafos; minimização de esforço intelectual na atribuição de termos descritores e conseqüente racionalização do trabalho e maior margem de consistência terminológica para indexação; restrição do número de descritores usados pelo sistema, evitando sua sobrecarga; expressão da linguagem no nível conceitual (ideia); aumento da consistência do vocabulário para o mapeamento de um domínio (KOBASHI, 2002). Silva (2008) estudou duas teses da base de dados do MHTX na tentativa de compatibilização de vocabulário controlado, utilizando mapa de tópicos com o objetivo principal de automatização do processo de compatibilização dessas estruturas.

A ideia de Silva era possibilitar o processo de fusão<sup>43</sup> dos mapas de tópicos para promover o inter-relacionamento semântico entre os mapas conceituais e, conseqüentemente, entre os recursos de informação hipertextual. Entretanto, este processo apresentou grande complexidade exigindo um alto nível de interferência manual. O Anexo I apresenta a conversão manual das estruturas facetadas das duas teses em um único mapa de tópicos. Os principais objetivos alcançados foram: (a) a conceitualização detalhada do processo de fusão dos mapas de tópicos, considerando os níveis de compatibilização possíveis e a aplicabilidade dessa tecnologia na integração de estruturas facetadas; (b) a obtenção de uma sequência detalhada de passos que pode ser utilizada para a implementação de mapas de tópicos, a partir de estruturas facetadas.

Estudos futuros são sugeridos por Silva propondo novos modelos de análise de assunto para automatizar processos do tratamento documentário do MHTX utilizando ontologias e associações entre conceitos apresentados no mapa de tópicos da fusão. Por isso, acredita-se neste trabalho que a compatibilização de vocabulário é necessária tanto para o autor (produtor do texto) como para o leitor (usuário de um sistema de uma biblioteca digital). O autor, antes de publicar um documento científico, pode se beneficiar do modelo proposto neste trabalho tendo como referência os termos comuns utilizados em uma comunidade científica (categorizados e classificados em um sistema de conceitos). Assim, os usuários (autor, leitor e co-autor) deste modelo poderão “integrar”, “harmonizar”, “reconciliar”, “concordar” com os termos comuns utilizados no domínio da CI para compartilhar informação.

As palavras “integração”, “harmonização”, “reconciliação” e “concordância” são utilizadas para indicar o mesmo conceito de “compatibilização” apresentado por Maniez (1997) e reforçam as diferentes facetas do vocabulário usado em um documento científico. A compatibilização de vocabulário ocorre quando existe um termo correspondente para cada termo de uma linguagem, de mesmo significado, na outra linguagem, e por isso, é possível converter termos de uma linguagem em outra, sem alterações de significado (SILVA, 2008).

Seria muito interessante, que o mapa proposto neste trabalho possibilitasse a compatibilização semântica das diversas facetas de um domínio para ajudar um leitor-interpretador na busca pelo termo-chave correto. A ontologia servirá de base para explicitar “conceitos científicos” mapeados em um domínio e o novo sistema, mostrando os termos sinônimos das relações de equivalência que poderão ser indicados pelo próprio autor utilizando ferramentas de anotação semântica na modelagem informacional. Para isso, há necessidade de utilizar os padrões de metadados digitais propostos pelo W3C.

---

<sup>43</sup> Fusão foi um termo escolhido pelo autor substituindo o termo em inglês “merge”.

Neste trabalho, acredita-se que o mapa conceitual do MHTX representado como uma ontologia de domínio possa ser modelado por profissionais experientes em linguagens documentárias. Essas estruturas só adquirem significado quando integradas ao contexto semântico e à estrutura cognitiva de um leitor-interpretador diante de uma nova informação (conceito, ideia, proposição) para um usuário-sistema (SAYÃO, 2001; BASTOS, 2005). Por isso, propõe inserir um único mapa de uma ontologia de domínio no MHTX usando a fusão da estrutura apresentada por Silva (2008) para modelagem informacional.

### **2.6.2 Sistemas hipertextuais organizados para interatividade em RI**

A informação esteve cativa em universos simbólicos de importantes livros e enciclopédias científicas por longos anos. Toutain (2007) destaca que entre alforrias e prisões, a informação chegou até a Web com grande parte dos hiperdocumentos científicos completamente liberados em sua língua natural local. Entretanto, muitos modelos de sistemas insistem em continuar operando por uma sublime recapitulação do passado.

A maioria desses modelos ignora o fato de que o usuário pode criar sua própria 'realidade' ao navegar por uma rede semântica e relacionar "conceitos" em seus estoques internos de informação, os quais são usados para compreender informações em diferentes situações de leitura. Saracevic (1997) destaca que os modelos de sistemas de RI consideram que a necessidade de um usuário é estática. Além disso, a maioria desses modelos tem como princípio fundamental a geração de meios de respostas com base na previsão das perguntas mais frequentes a que o sistema estará sujeito. Essa característica influi negativamente por limitar as respostas predeterminadas à busca por palavras-chave antes de receber estímulos diretamente no contexto de leitura.

Para resolver os problemas de busca e de recuperação, muitos modelos foram desenvolvidos, utilizando uma tecnologia correspondente. Entretanto, a julgar pelas reações dos usuários desses modelos, parece que o problema de usabilidade ainda é uma constante, e, pior ainda, parece que muitas soluções resultaram em novos níveis de complexidade e dificuldade para o usuário-leitor-interpretador. O comportamento de um usuário-sistema é dinâmico em vez de estático, sendo afetado durante a busca de informação. A busca diferencia-se da recuperação por tratar-se de uma ação planejada de uso de fontes definidas, implicando em algum conhecimento a priori da informação a ser processada (GONZÁLEZ GÓMEZ, 2004). O julgamento de relevância acompanha cada passo da busca, podendo efetuar-se pelo acesso direto durante a leitura de um hiperdocumento científico e não só a partir de suas representações em esquemas hierárquicos externos ao contexto.

Nos últimos anos, tem havido considerável interesse em pesquisas científicas sobre o comportamento humano no processo de busca de informação (*information seeking*). Esse processo está relacionado com a ação pela qual usuários procuram obter informação a partir de sistema automatizado, incluindo a busca de informações durante a leitura de um hiperdocumento científico em que o autor não está presente para esclarecer a dúvida de determinados conceitos ou termos desconhecidos para um leitor-interpretador. Assim, a busca deve ser traduzida em uma linguagem que ofereça mecanismos de aferir resultados com os objetos informacionais contextualizados e mapeados por um modelo de interface mais dinâmico e interativo para ajudar um usuário-leitor-interpretador com conceitos desconhecidos. Espera-se que a ontologia funcione como um único mapa no MHTX, explicitando “conceitos científicos” com suas relações de equivalência, indicando aos usuários os termos sinônimos de um domínio para a interatividade em RI.

No MHTX, o comportamento de busca e uso de informação podem ser modelados pelo estilo cognitivo do usuário. A navegação em um sistema hipertextual permite também mostrar os tipos de relação semântica existentes entre os conceitos, dando maior flexibilidade ao modelo proposto. A organização semântica de determinada área de conhecimento, para fins de criação de um sistema, teria como produto uma rede semântica, ou seja, demanda a construção de um sistema de conceitos integrados entre si (LIMA, 2004).

Saracevic (1997) apresenta um modelo denominado de Modelo Estratificado de Interação em RI (Figura 25), iniciando com duas suposições: a) usuários interagem com o sistema no intuito de usar a informação; e b) o uso da informação é conectado à cognição e então aplicado de modo situacional. Os usuários e o computador são os principais elementos deste modelo, cada um com variáveis específicas sendo que a interface propicia a interação como um diálogo entre eles. Em relação ao computador, os processos podem ser físicos e simbólicos tais como: a) o nível de engenharia (capacidade e eficiência); b) o nível de processamento está vinculado a interesse específico em algoritmos e modelos que manipulem textos, consultas, interface e outros processos críticos relacionados à interação entre os usuários e o computador; e c) o nível de conteúdo pode concentrar-se nos recursos do próprio hipertexto, modelado com recursos de objetos informacionais digitais com conceitos hiperlinkados.

Beppler (2008) relacionou a busca ao vai e vem do círculo hermenêutico com o uso de ontologias na IA. Os resultados de Beppler (2007) apresentaram uma recuperação mais eficaz que ele denominou de RII (Recuperação de Informação Interativa). A área de *information seeking* tem uma atuação mais abrangente que a área de RI e RII. A condição inicial de busca é contextualizada assim como ocorre com uma falta de informação. A

construção desse processo evolui em estágios de acordo com certas variáveis, tais como, relevância e incerteza. A Figura 25 mostra a relação entre comportamento humano com área *information seeking*, a área de RII e RI na busca de um termo desconhecido.

No modelo de Saracevic (1997), os processos quanto ao usuário podem ser fisiológicos, psicológicos e cognitivos: a) no nível cognitivo, os usuários interagem com textos e com suas representações para interpretação; b) no nível afetivo, os usuários interagem com as suas intenções de seleção de conceitos relacionados, entre outros; e c) no nível situacional, os usuários interagem com uma incerteza ou problema de interpretação ou ambiguidade a qual originou a necessidade de informação. Saracevic deixa claro que a interação é uma sequência de processos que ocorrem em diferentes níveis. Além disso, adaptações podem significar alterações em uma variedade de níveis. Cada nível envolve diferentes elementos e/ou processos específicos. À medida que a interação ocorre, uma série de adaptações dinâmicas acontece em ambos os elementos, usuários e computador.

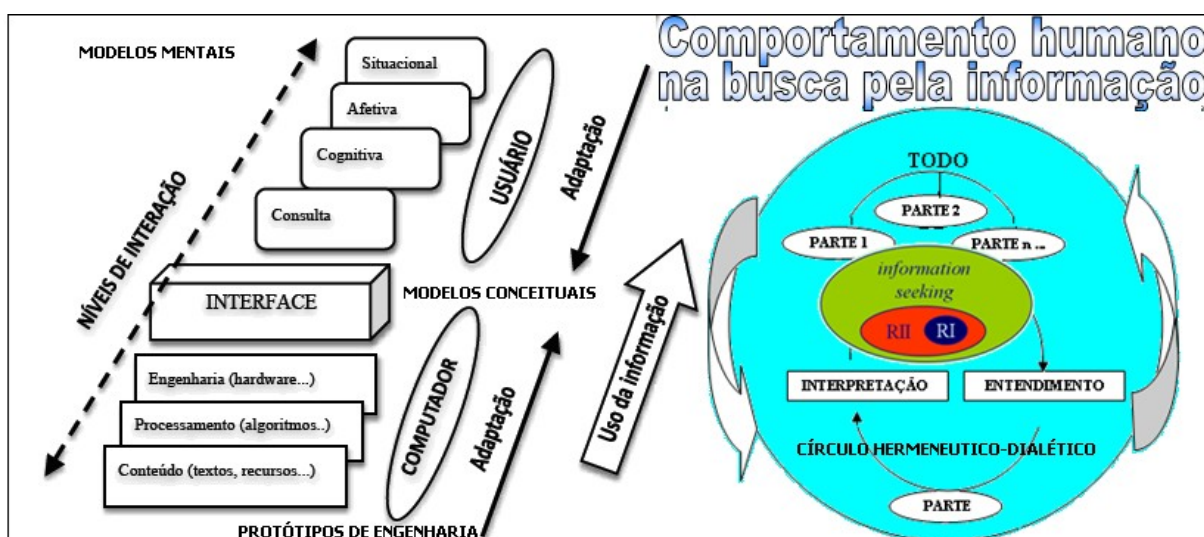


Figura 25– Modelo Estratificado de Interação em RI e comportamento do usuário.

**Fonte:** Adaptado de Saracevic, 1997; Beppler, 2008.

Para Saracevic (1997) não existe um modelo de interação que possa ser aceito universalmente como único e que o papel dos modelos é descrever os elementos essenciais e as relações de um objeto (sistema, processo, entidade, estrutura, ideia, etc.). No computador, a interação envolve processos muitos mais complexos que a interação por parte do usuário. Os mecanismos de inferência computacionais dependem do uso de uma linguagem formal, livre de ambiguidades, que o computador não pode resolver de forma tão simples como fazem as pessoas. Os novos mecanismos de IA são primários, se comparados com as possibilidades humanas, mas significam um avanço para a automatização de tarefas via computador e contexto semântico (ALMEIDA, 2006).

No modelo de Beppler (2008), um leitor-interpretador pode satisfazer sua necessidade de informação ao ler um texto, por exemplo, sem necessariamente utilizar recursos de RI. Neste caso, o próprio contexto semântico pode ser suficiente para esclarecer suas dúvidas. Por isso, um modelo de interatividade em RI deve incorporar três tipos de atividades: (1) física que se refere às ações tomadas; (2) emocional que está associada às experiências vividas; e (3) cognitiva que está relacionada às funções de raciocínio como interpretar, aprender, lembrar. Por meio de ações e interações entre essas atividades em uma interface, um usuário pode passar de uma condição de incerteza para uma condição de solução.

Os atuais modelos de bibliotecas digitais precisam ser (re)pensados sob diferentes perspectivas, considerando a organização semântica da informação fundamental para garantir interatividade em RI. A organização semântica de um sistema dinâmico e interativo de ilimitadas possibilidades merece toda atenção dada por Ranganathan a um leitor “aprendente”. Para Ranganathan, os usuários de uma biblioteca buscam informações sobre um assunto ou seus constituintes, isto é, conceitos e ideias: a) livros são para ser usados; b) a cada leitor o seu livro; c) a cada livro o seu leitor; e d) poupe o tempo do leitor.

Ferreira (1995) destaca que é um mito pensar que o usuário sempre planeja a estratégia de busca ou que faz buscas muito elaboradas. Uma busca geralmente corrobora o princípio do menor esforço cognitivo, o leitor-interpretador ou um usuário “aprendente” deseja que o sistema recupere informação com qualidade e que ele tenha facilidade na utilização da interface hipertextual. Entretanto, a possibilidade de um mesmo termo possuir diferentes significados é um dos principais problemas relacionados ao uso do processamento computacional para compatibilização e interoperabilidade semântica. No MHTX, a proposta é de criar um único mapa com conceitos integrados, utilizando ontologia na IA com linguagens documentárias e padrões W3C.

Os princípios formulados pelo sistema *World Information System for Science and Technology*<sup>44</sup>, de acordo com Fujita (2003), consideram a indexação sob dois pontos de vista distintos: enquanto processo que consiste em descrever e identificar um documento com ajuda de representações dos conceitos contidos em um documento e quanto à sua finalidade, permitindo a busca e o acesso à informação armazenada.

Fujita (2003) aprofundou os estudos sobre extração de conceitos válidos e confirma que as concepções de leitura orientada para o conteúdo do documento e para a demanda devem ser intrínsecas para produzir uma correspondência precisa com o assunto pesquisado em índices. O termo candidato deve se espelhar na configuração das categorias

---

<sup>44</sup> *World Information System for Science and Technology* é um sistema internacional vinculado à UNESCO conhecido pela sigla UNISIST.

fundamentais reconhecidas como importantes para o assunto e dependerá exclusivamente do objetivo para o qual o documento está sendo indexado. Essa extração deve ser realizada em três estágios: a) compreensão do conteúdo do documento; b) identificação dos conceitos que representam este conteúdo; e seleção dos conceitos válidos para recuperação.

A idéia principal deste trabalho é usar “conceitos” extraídos da fusão dos mapas apresentados por Silva (2008) em forma de facetas. A utilização do termo “faceta” não implica apenas numa mudança terminológica, mas uma nova concepção do processo de classificação bibliográfica em forma de um sistema de conceitos. O diferencial da classificação facetada é a utilização de uma estrutura dinâmica que passou a ser considerado nos modernos estudos da CI, o substituto de característica (ARAÚJO, 2006).

Fujita (2003) destaca o funcionamento do sistema PRECIS em forma de facetas que preconizou a "preservação do contexto". Este sistema foi influenciado pelas ideias de Ranganathan que organizava os assuntos em facetas de acordo com uma ordem de citação estruturada em uma entrada de duas linhas e três posições para revelar a estrutura superficial, distinguindo-o em duas partes: a) a parte sintática, formada pela estrutura de entradas e gramática composta de operadores de função atribuídos aos conceitos durante a análise conceitual; e b) a parte semântica, em que os termos são identificados pelos operadores era traduzida por termos de um vocabulário controlado.

Este sistema era baseado em uma análise cuja proposição de identificação de conceitos tinha como base teórica a gramática transformacional, ou seja, é a gramaticalização ou terminologia correta do termo que dá garantia de que o conteúdo será representado por um “conceito verdadeiro”. Segundo Pozo (1998), “conceitos verdadeiros” são “conceitos científicos” que são adquiridos sempre por relação hierárquica com outros conceitos, mediante processos instrucionais. Pozo considera que a cognição humana é essencialmente organizada como uma rede de relações semânticas, na qual conceitos podem ser associados levando a um conhecimento novo pelo agrupamento de entidades.

A tendência de comportamento humano é buscar, cada vez mais, modelos mais dinâmicos e interativos com serviços personalizados e contextualizados (FERREIRA, 1995). Sistemas hipertextuais possibilitam a criação de ambientes onde o utilizador pode experimentar certo grau de autonomia enquanto navega na informação, o que contribui sem dúvida para que se expressem estratégias individuais de aprendizagem, sendo o sujeito responsável pelo seu próprio processo de aprendizagem, seleção e ação de termos-chave desconhecidos. Esses sistemas se compatibilizam com a cognição humana, pois a memória se organiza em estruturas semânticas com conceitos inter-relacionados, permitindo mapeamento de um domínio com objetos informacionais hiperlinkados (JONASSEN, 1988).

Estudos mostram que a interface humano-computador pode ser uma ferramenta muito poderosa na construção de modelos e mapas conceituais, permitindo interatividade às pessoas na comunicação de signos contribuindo para o melhoramento das suas capacidades de entendimento, memorização e tomada de decisão (OLIVEIRA e BARANAUSKAS, 1998). Para Assmann (2000), o que há de novo no uso das novas TIC é a parceria cognitiva que está começando a exercer uma relação com o “aprendente”. O uso do termo “aprendente” é considerado por Assmann como o que melhor expressa a relação cooperativa na interface humano-computador de maneira mais eficiente.

Neste trabalho, acredita-se que o novo sistema do MHTX possa auxiliar o leitor “aprendente” a interpretar “conceitos científicos” ou termos-chave desconhecido em um domínio semanticamente estruturado, fomentando atividades de cooperação entre pesquisadores de uma área específica. A integração de novos serviços de aprendizagem colaborativa assistida por computador, conhecida como CSCL (Computer Supported Collaborative Learning) para que um leitor-interpretador possa criar uma conectividade com o sistema hipertextual para interatividade em RI.

A CSCL cresceu em torno de um vasto leque de investigações na IA e auxiliado por novas ferramentas de agentes de software cresceu em torno do trabalho colaborativo assistido por computador ou CSCW (Computer Supported Collaborative Work). Nesses sistemas, o conhecimento é favorecido pela participação social em ambientes ricos em possibilidades propiciando o crescimento do grupo na troca de informações, inclusive usando mapas conceituais. A aprendizagem colaborativa<sup>45</sup> é uma abordagem promissora para a formação do perfil do profissional contemporâneo, no entanto, ela envolve um conjunto complexo de variáveis que diversas vezes inviabiliza sua utilização (FARIA, 2005).

Mapas conceituais são dinâmicos e estão constantemente mudando termos e relações de um modelo conceitual em ambientes de CSCL. Se a aprendizagem é colaborativa e significativa nesses ambientes, a estrutura cognitiva está constantemente se reorganizando por diferenciação progressiva e reconciliação integrativa e, em consequência, os mapas conceituais traçados por profissionais experientes utilizam termos-chave ou “conceitos científicos” mais adequados na comunicação. A tecnologia e a informação são partes essenciais do processo de aprendizagem, mas não devem ser vistas isoladamente. A interface estabelece um sentido de diálogo em que a tecnologia funciona como a arte, o conselho e a motivação que está por trás de toda a comunicação (WURMAN, 1991).

---

<sup>45</sup> A “aprendizagem colaborativa” pode ser definida como um conjunto de métodos e técnicas de aprendizagem para utilização em grupos estruturados. Segundo Faria (2005), este tipo de aprendizagem utiliza estratégias de desenvolvimento de competências mistas (aprendizagem e desenvolvimento pessoal e social).



O uso de mapas conceituais em bibliotecas digitais também pode ser considerado elementos inovadores de apoio à leitura, compreensão e interpretação. Este uso é bastante recomendado na navegação, pois facilita o processo, permitindo visualização da estrutura dentro da aplicação (SAYÃO, 2001). A visualização de uma ontologia em um grafo ou mapa conceitual no computador fornece meios de localizar informações específicas demonstrando como os conceitos estão estruturados para e facilitando a movimentação de um nó para outro. Assim, agentes de software podem recuperar informações de forma mais precisa ao associar termos de um documento digital publicado na Web.

Moreira (1999) destaca as ideias de Vergnaud sobre o papel do conhecimento prévio como precursor de novos conhecimentos com continuidades e rupturas, se aproximando das propostas de Ausubel com o conhecimento prévio como principal fator, isolado, que influencia a aquisição de novos conhecimentos com a Teoria da Aprendizagem Significativa. Em sua teoria, Ausubel defende a existência de organizadores prévios com diversos tipos de aprendizagens linguísticas para a formação de conceitos. Os aspectos relevantes da estrutura cognitiva que funcionam como uma âncora para a nova informação é chamado de “subsunçores”.

Assim, conceitos e teoremas explícitos em uma ontologia de domínio não constituem mais do que a parte visível de um *iceberg* da conceitualização: sem a parte escondida formada pelos invariantes operatórios essa parte visível não seria nada (MOREIRA, 1999). Isso reforça mais uma vez as “potencialidades da linguagem”, ou melhor, as possibilidades de estruturação de um sistema hipertextual em campos conceituais com conceitos-em-ação e teoremas-em-ação. A Teoria dos campos conceituais propostas por Vergnaud vai de encontro às funcionalidades de um “conceito” nos sistemas hipertextuais de uma biblioteca digital semanticamente organizada, em que o “conceito” pode ser um metadado descritivo, estrutural e administrativo de objetos informacionais digitais registrados.

Uma questão importante desta teoria diz respeito à origem dos “subsunçores” que estabelecem conexões com organizadores prévios, funcionando como uma âncora (um hiperlink) para a descoberta de um novo termo ou conceito desconhecido. A estrutura do mapa conceitual pode orientar o modo de assimilação de novos conceitos para resultar em um modelo de interface para interatividade em RI. Neste sentido, o usuário “aprendente” pode estabelecer ligações a subsunçores relevantes já existentes na estrutura cognitiva.

Interessante destacar que para acelerar o processo de aprendizagem significativa, Ausubel propõe os organizadores prévios interligando conceitos aparentemente não relacionáveis através da abstração. Em sua teoria, os subsunçores podem funcionar com uma estrutura específica ao qual uma nova informação pode se integrar ao cérebro

humano, detentor de uma hierarquia conceitual que armazena experiências prévias do indivíduo. Na interface de um sistema hipertextual, como o autor não está presente para mediar essa ação, como é possível propor esses esquemas de organizadores prévios com “subsunçores” que possam ajudar tanto “usuários” quanto “sistema” no estabelecimento de relações conceituais para interatividade em RI?

### **2.6.3 Funcionalidades da linguagem e potencialidades do “conceito”**

De fato, procurar compreender que tipo de sistema é a linguagem possibilitou avanços importantes com poderosas ferramentas para as elaborações de modelos conceituais mais interativos com a realidade do mundo físico. Estudos sobre a manipulação sistemática dos atributos do uso de palavras têm contribuído para compreensão da memória humana e de seus processos como, por exemplo, o processamento semântico, recuperação, reconhecimento e representação mental de conceitos (JANCZURA, 2005).

Esses estudos procuram desvendar os mistérios do cérebro humano com os aspectos multifuncionais da linguagem e suas relações, explícita ou teórica, contribuindo para verificar que o texto constitui uma unidade mais elevada da linguística podendo se subdividir em unidades menores e também passíveis de categorização e classificação. Muitas dessas contribuições foram proporcionadas, principalmente, pelos estudiosos das Ciências Cognitivas com novas perspectivas multidisciplinares, acreditando que há uma continuidade entre cognição e cultura, pois esta é apreendida socialmente, mas armazenada individualmente. Assim, muitas barreiras foram rompidas entre exterioridade e interioridade, fenômenos mentais e físicos e sociais, modelos conceituais de modelos mentais.

Nesses estudos, os sistemas discursivo, semântico e gramatical são vistos em sua convivência nas expressões linguísticas, excluindo a possibilidade de selecionar um deles como o sistema central da língua. A correlação não biunívoca entre funções e estruturas explica a natural heterogeneidade das línguas. Postula-se que esta natureza não está apenas em dispor de uma estrutura, mas em ancorar essa estrutura na necessidade de cumprir certas funções (CASTILHO, 2007).

O reconhecimento de algum tipo de meio, de alguma forma, ou linguagem, na qual o pensamento ocorre é uma parte importante de quase toda teoria cognitiva contemporânea. Pesquisas recentes sugerem que a linguagem natural humana deve ser um veículo muito rico para poder executar os muitos processos cognitivos tais como percepção, raciocínio, aprendizagem, entre outras coisas. Estas pesquisas indicam que as habilidades e competências humanas formam um sistema natural de ‘inteligências múltiplas’ tais como a linguística, musical, lógico-matemática, espacial, entre outras. Acredita-se que o sistema

nervoso não é um órgão com propósito único e que diferentes neurônios processem tipos de informação em módulos conceituais (GARDNER, 1987).

Uma das visões de mente modular foi proposta por Jerry Fodor, psicolinguísta com ideias claras a respeito de representações e operações cognitivas, considerando que o compromisso de atribuir um sistema representacional a determinados objetos ou organismos deve exigir uma caracterização deste sistema mental e conceitual. Fodor questiona se os sistemas cognitivos envolvem manipulação de símbolos, então estas representações devem existir em algum lugar na mente humana, propondo que ela seja dividida em duas grandes partes: a) percepção ou sistemas de entrada; e b) cognição ou sistemas centrais. As arquiteturas dessas duas partes são bastante diferentes. A percepção humana possui uma série de módulos discretos e independentes, como a audição, visão, o tato e ainda inclui a linguagem entre os sistemas de entrada (GORSKI, 2007).

Fodor argumenta que cada sistema de entrada se baseia em processos cerebrais independentes. Por exemplo, os usados para a audição são totalmente diferentes dos usados para a visão e linguagem, enfatizando o fato destes sistemas operarem muito rapidamente em face de estímulos apropriados. Em contrapartida, os sistemas centrais não possuem uma arquitetura bem definida, e, talvez esta arquitetura sempre permaneça fora do nosso alcance porque é na cognição que reside a inteligência humana para resolução de problemas e imaginação. Essa modularidade é atestada por numerosas evidências, que incluem uma aparente associação com partes específicas do cérebro, os típicos padrões de desenvolvimento das “inteligências múltiplas”.

Os componentes sistêmico-funcionais da linguagem são o experiencial (conteúdo e ideias) e o lógico (relação entre ideias) sendo praticamente impossível separá-los (CASTILHO, 2007). Muitos modelos conceituais na IA foram investigados com contribuições de pesquisas sobre a gramática generativa transformacional na visão de mente-computador para tentar se aproximar de modelos mentais. O ser humano com suas “inteligências múltiplas” possui intuições interpretativas que o levam a resolver problemas de certas ambiguidades do vocabulário humano, usando o contexto semântico. Diferentemente, o computador não possui tais intuições e um dos maiores desafios a que este trabalho se propõe é tentar transportar para um único mapa mecanismos subsunçores inteligentes de interpretação desambiguadora próprios dos seres humanos em um contexto semântico.

No computador, o processo de inferência depende diretamente da expressividade da linguagem porque quanto maior a expressividade oferecida pela linguagem em uma ontologia mais complexa e poderosa a inferência pode ser (BEPPLER, 2008). Esses mecanismos artificiais oferecem um tipo de serviço de raciocínio para aplicações que desejam utilizar processos sobre as definições explicitamente descritas em

uma ontologia de domínio e extrair conhecimento adicional, representado não explicitamente.

Além disso, esses mecanismos empregam um conjunto de regras sobre um conjunto de dados atômicos. Neste trabalho, acredita-se que o “conceito” seja o principal metadado digital deste conjunto de dados atômicos para estender a toda uma classe de objetos, os elementos essenciais, constatados num certo termo de um domínio. Qualquer conceito pode criar efeitos que permitem a interação entre informações velhas e novas, de modo que entre ambas se cria uma continuidade entre texto e contexto possuindo índices múltiplos em uma rede semântica de um mapa ou grafo conceitual de uma ontologia.

Martins (2004) destaca que o ato de relacionar texto e contexto possui sua relevância, pois permite estabelecer uma maior compreensão entre as relações de sentido que compõem o texto em seu conjunto. Tais relações exteriorizam-se de diversas formas, apresentando o conteúdo de um documento, combinando tempos verbais e vocabulário. Isso implica que, em qualquer sistema comunicativo, a partir de inferências do texto para o contexto e vice-versa, os interlocutores fazem previsões sobre os significados negociados e construídos durante processos de interação (SIMÕES, 2006). Essas inferências consideram os processos mentais de que resulta o texto para formar um documento como um todo dotado de unidade própria que ajudam na interpretação de sentido mediante a interação de diversos níveis de conhecimento externos à língua que são exigidos de um leitor.

Ao iniciar a leitura de um texto, os mecanismos humanos de inferência trabalham com certa “nebulosidade” quanto ao significado das palavras. A inferência está baseada na ideia do significado estar relacionado a uma ‘coisa’ ou objeto que pode ou não ser categorizado ou classificado. Segundo Sperber e Wilson (1986), o mecanismo humano de inferência é o processo pelo qual uma suposição é aceita como verdadeira ou provavelmente verdadeira de acordo com a força de suas suposições contextualizadas. Esses mecanismos exigem esforços cognitivos para compreender o significado de uma palavra desconhecida no léxico ou a na busca de um novo conhecimento ou conceito.

Esta forma de inferência é a demonstrativa que, por um processo dedutivo, chega-se ao fato de que a verdade das premissas pode garantir a verdade das conclusões. Esses mecanismos e esforços cognitivos modificam a representação de mundo do leitor-interpretador que, depois de identificar uma palavra ou expressão, passa para o processo de análise-síntese em campos conceituais, o que pode acrescentar outros campos semânticos contextualizados ou mesmo modificá-los em sua estruturação, utilizando um conhecimento prévio. De acordo com Moreira (1999), esse conhecimento é principalmente implícito e o leitor-interpretador tem dificuldades de explicitá-los.

Na ciência, conceitos e teoremas são explícitos e pode-se discutir sua pertinência e sua veracidade, mas esse não é necessariamente o caso dos invariantes operatórios. Esquemas de relacionamento hierárquicos de um domínio têm como ingredientes essenciais aquilo que Vergnaud chama de conceitos-em-ação e teoremas-em-ação que constituem a parte conceitual do conhecimento contido nos esquemas-situação. Moreira (1997) sugere que o modelo conceitual deve contemplar o sistema de crenças de um indivíduo. A interação entre situações e esquemas é a fonte primária das representações simbólicas com significantes conceituais. É nessa interatividade que o novo conhecimento adquire significados e o conhecimento prévio se transforma adquirindo, novos significados. Mas tal interação não é arbitrária, ou seja, o novo conhecimento adquire significados pela interação com conhecimentos prévios especificamente relevantes.

Aqui entram as funcionalidades de um “conceito científico” explícito em uma ontologia de domínio inseridos em um modelo de biblioteca digital interativa, considerando-o como um metadado descritivo, administrativo e estrutural. Além disso, um “conceito científico” integrado ao contexto de um hiperdocumento pode funcionar como “subsunoçores” inteligentes ao explicitar as facetas de um domínio e suas relações de equivalência, considerando as potencialidades e os aspectos multifuncionais da linguagem na visão de categorias formais Aristóteles, Ranganathan e Dahlberg.

Neste trabalho acredita-se que o mapa conceitual da ontologia deve representar uma única estrutura com “conceitos científicos” usando uma terminologia padrão da área modelada para auxiliar um usuário “aprendente” como um mecanismo artificial desambiguador de termos sinônimos. Tarapanoff, Junior e Cormier (2000) também consideram que um dado (Figura 26), a partir de unidades de informação, aliado às novas TIC transforma-se em inteligência competitiva para o perfil do profissional na contemporaneidade e leitores de bibliotecas digitais.

A construção de bases teóricas associada à área de Inteligência é uma atividade recente no Brasil e carece de pesquisas que contemplem seus problemas. Marcial (2007) confirma que, há necessidade, principalmente, de construção de bases ontológicas e epistemológicas da CI com definição de terminologia própria adaptada à língua portuguesa. Segundo Marcial, os motivos da instabilidade de atividades no Brasil associados à área de Inteligência para construção de bases teóricas pode ser fruto da pouca experiência profissional, dos problemas conceituais terminológicos consensuais e de comunicação existentes, devido à falta de pesquisa e de bibliografia disponível em língua portuguesa.

Um conceito como um metadado descritivo, administrativo e estrutural em um modelo relacional pode funcionar como entradas triplas no processamento da linguagem natural: 1) lógica (conjunto de regras dedutivas); 2) enciclopédica (conjunto de suposições);

e 3) lexicais (representações com formas linguísticas) que, ligadas às regras de eliminação, formarão um conjunto finito de premissas para deduzirem um conjunto finito de conclusões não-triviais. O processamento do raciocínio humano inteligente se dá através da análise, seleção, comparação lógica e classificação dessas informações, reagindo com uma resposta conforme os padrões definidos pelo estímulo no sistema natural (WILSON, 1965).

Buscar pela informação verdadeira em um mapa conceitual se faz o primeiro passo para a formulação de estratégias para possibilitar aprendizagem colaborativa e significativa. Acreditando que a estrutura dos estoques do conhecimento, relacionados à demanda de informação dar-se-ia numa situação inversamente proporcional, Barreto (2002) adaptou numa dupla relação piramidal (Figura 26) para analisar o que seria, possivelmente, a demanda e a oferta de informação em sua estrutura básica, relacionadas às necessidades informacionais de um indivíduo.

Na base da pirâmide estão os indivíduos que buscam a satisfação de suas necessidades básicas de alimentação, habitação, vestuário, saúde e educação. Estes indivíduos demandariam, primeiramente, informação utilitária e, de acordo com as necessidades, o indivíduo mover-se-ia da base para o topo, passando de um nível para o outro apenas quando suas necessidades fossem satisfeitas. Assim, tendo resolvido as necessidades anteriores, no topo da pirâmide esses indivíduos demandariam informação que os conduziria à reflexão, criatividade e sucesso profissional e pessoal.

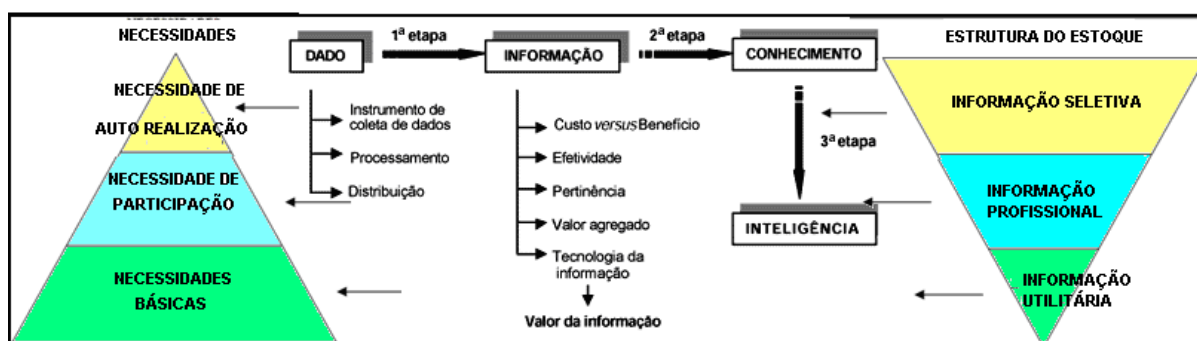


Figura 26 Gestão inteligente e as necessidades informacionais.  
**Fonte:** Adaptado de Tarapanoff, Junior e Cormier, 2000 e Barreto, 2002.

Buscar uma informação relevante é um processo dinâmico que exige métodos e critérios de seleção ou rejeição que variam frequentemente no tempo, dependem dos resultados imediatos e estão relacionados com os hábitos pessoais do indivíduo e com as necessidades informacionais que precisam ser satisfeitas (FERREIRA, 1995). Usuários de uma base de dados fazem buscas mais eficientemente quando estão de posse de um “mapa” com a estrutura esquematizada em árvore para organizar conceitos, ideias e conhecimentos centrados nos objetos mentais cognitivos (SAYÃO, 2001).

De acordo com a Teoria da Relevância apresentada por Sperber e Wilson (1986), qualquer informação representada conceitualmente e disponível para um usuário-sistema, pode ser usada como uma premissa no processo de compreensão inferencial humana e apoiada por mecanismos computacionais de inferência. O contexto cria efeitos de interação entre informações velhas e novas, de modo que entre ambas se cria uma implicação. Essa implicação só é possível porque existe uma continuidade entre texto e contexto e, além disso, a cognição é um fenômeno situado que acontece dentro e fora da mente.

Sperber e Wilson (1986) argumentam que as regras disponíveis dos mecanismos dedutivos humanos são as regras de eliminação ligadas a conceitos, isto é, os menores constituintes das formas lógicas de uma suposição. Geralmente, o autor pretende comunicar somente uma interpretação e, o leitor também, em geral, no seu ambiente cognitivo, escolhe apenas uma das várias interpretações possíveis. O ambiente cognitivo inclui tudo o que está na cabeça do leitor para interpretar e compreender determinado termo-chave desconhecido. Se um dado é relevante, faz a seleção, implicando na integração do sentido selecionado dentro da pré-compreensão do sistema de interatividade, produzindo-se uma nova pré-compreensão como na ideia do círculo hermenêutico-dialético.

Neste sentido o conceito é um componente indispensável para formar o significado de uma palavra em um sistema que represente facetas de “conceitos científicos” de um hiperdocumento. Para Vygotsky (1993), o significado é um critério da “palavra”. Dados isolados são palavras sem significado. O significado pode ser visto como uma generalização ou um conceito e, como atos de pensamento, podemos considerar de acordo com um conjunto de representações simbólicas em contextos e usos.

Segundo Oliveira (1999) existem três dimensões nos conceitos vygotskyanos. A primeira dimensão é a ideia de libertação mediante o processo de abstração e generalização que a linguagem possibilita. A segunda dimensão é a dos conceitos como um sistema complexo de inter-relações e que trata a complexidade da organização conceitual na mente humana. Essas relações entre elementos e estruturas são aplicáveis a todas as espécies de conceitos que Vygotsky distingue em dois: conceitos científicos e conceitos espontâneos. A terceira dimensão é a concepção de conceitos como processos que são construídos mediante negociações de interação com objetos de ação e de conhecimento.

Segundo Pozo, “conceitos espontâneos” são adquiridos e definidos a partir dos objetos, por sua referência, enquanto “conceitos científicos” são adquiridos sempre por relação hierárquica com outros conceitos. Isto faz com que os conceitos científicos cheguem a captar a essência do conceito mediante a análise consciente de suas relações com outros conceitos. O desenvolvimento de conceitos espontâneos, também conhecido como

pseudoconceitos, se dá de forma ascendente (Figura 27), do concreto para o abstrato, ao passo que os conceitos científicos desenvolvem-se de forma descendente, numa pirâmide de conceitos, por caminhos diametralmente opostos (POZO, 1998).

Relembrando Vergnaud, conceitos podem ser apenas relevantes ou irrelevantes e de acordo com Moreira (1999) conceitos-em-ação não podem ser confundidos com teoremas-em-ação. Aqui entra a função do mapa conceitual representado por uma ontologia de domínio no MHTX. Um teorema-em-ação ou um “conceito científico” é uma proposição tida como verdadeira sobre um dado real, ou melhor, um objeto informacional digital registrado em um hiperdocumento científico. O conceito-em-ação é um objeto, um predicado, ou uma categoria de pensamento do leitor-interpretador cujo grau de relevância deve ser definido pelo usuário na seleção de conceitos relevantes.

Os mecanismos humanos de inferência possuem livre acesso às informações conceituais no cérebro para compreensão ou não de um “conceito científico”. A resposta a uma pergunta no sistema natural se dá mediante um estímulo que aciona o mecanismo de processamento e resposta do sistema natural e, após acionado, passa a processar informações armazenadas no cérebro humano. Em um sistema assim, as habilidades inferenciais envolvidas na compreensão verbal são aquelas do tipo que não pode ser provado, pois a comunicação pode falhar até mesmo sob as melhores condições (SPERBER e WILSON, 1986). Esses são os ingredientes necessários para que o modelo conceitual se aproxime de um modelo mental de um usuário para interatividade em RI.



Figura 27 Posicionamento da inteligência na busca por um conceito relevante  
**Fonte:** Adaptado de Puia, 2006; OLIVEIRA, 2006; BEPLER, 2008.

Portanto, cabe ao produtor de qualquer modelo de interatividade em RI, organizar esquemas conceituais de tal forma que um usuário, leitor-interpretador, possa reconhecer a conexão entre os seus constituintes e construir um modelo mental coerente.



Nesse modelo, os conceitos explicitados no mapa e sua estrutura de relações pode ser limitada em um campo conceitual. Neste sentido, um modelo relacional pode ajudar nos mecanismos humanos e artificiais de inferências, expressando relações de equivalência para explicitar significados conceituais em forma de facetas. O que identifica uma relação ou conexão entre conceitos é a existência de algum significado entre eles e isso ocorre no modelo mental do usuário que pode ser convertido diretamente em uma notação.

Há muitos obstáculos na construção de modelos conceituais que se aproximem de modelos mentais. Contudo, um modelo ou mapa conceitual de um domínio não precisa ser nem completo e nem inteiramente acurado para ser útil e funcional. As maiores restrições decorrem da estrutura percebida ou concebida dos significados dos objetos e da necessidade de mantê-los livres de contradições e ambiguidades. Um protótipo é importante para possibilitar a viabilidade do modelo proposto. Pelo seu caráter exploratório, protótipos geralmente não incluem todo o conteúdo do seu potencial de aplicação (LIMA, 2004).

### **3 ESTUDO EXPLORATÓRIO NO MHTX PARA COMPATIBILIZAÇÃO SEMÂNTICA DO VOCABULÁRIO UTILIZANDO ONTOLOGIA**

Para o estudo exploratório proposto, este capítulo apresenta as implementações, conceituais e tecnológicas, visando atingir a meta principal desta pesquisa que busca a compatibilização semântica de vocabulário utilizando uma ontologia na base de dados MHTX. Durante a revisão de literatura, as possibilidades de reuso de uma ontologia de domínio e a procura de soluções para os problemas encontrados no MHTX, resultaram no planejamento de um novo sistema representado por um único mapa conceitual denominado BDICI – Biblioteca Digital Interativa da Ciência da Informação.

Como o prazo para o desenvolvimento de pesquisas científicas é muito curto ainda mais quando se trata de uma proposta de um modelo de inovação tecnológica sendo desenvolvido no tempo limitado de pesquisa de Mestrado, não se pretende, aqui, descrever a implantação ainda incipiente da BDICI. O que se pretende é mostrar a viabilidade do novo sistema proposto para o MHTX, considerando que um dos principais diferenciais desta base de dados está na navegação em contexto em forma de classificação facetada.

A partir dos quatro referenciais utilizados por Lima (2004) para organização semântica estrutural do MHTX, analisou-se a Teoria da Análise Facetada com a Teoria do Conceito proposta por Dahlberg (1978) e a Teoria da Linguagem para representação de um sistema de conceitos, considerando o “conceito” como um metadado descritivo, administrativo e estrutural. A linguagem funciona como um “sistema” mediador que se organiza na forma de categorias e escolhas léxico-gramaticais. Para que essa análise seja bem sucedida é fundamental o conhecimento de gramática normativa e descritiva.

A Teoria dos Mapas Conceituais, estrutura semântica de links hipertextuais e as normas da ABNT foram relacionadas a outras teorias como a Teoria da Aprendizagem Significativa e a Teoria dos Campos Conceituais para a construção do mapa da ontologia de domínio. Esses referenciais teóricos mostram que a gramática normativa, tratando do controle do vocabulário na escrita de um texto científico e dos padrões de uso de uma língua não pode se anular diante da gramática descritiva. O uso de uma terminologia padrão aliada ao contexto semântico propõe a gramática para descrever essas regras de como uma língua é representada em grandezas-signos para a prescrição "correta" do vocabulário.

A Teoria da Hermenêutica foi aplicada aos princípios de metodologia interativa desenvolvida por Oliveira (2001) para os processos de análise-síntese na leitura de “conceitos científicos” e as dificuldades de interpretação que um usuário pode ter. Esses processos cognitivos de interpretação e mecanismos humanos de inferência usando

conceitos integrados em um mapa foram analisados com princípios da Teoria da Relevância como um caso de “relevância mostrada”, ou seja, conceitos-em-ação e teoremas-em-ação que podem ser explicitados por um grafo conceitual de um domínio. Assim, a relevância pode ser relacionada tanto à organização e estruturação do “sistema” na interface com o “usuário” para interatividade em RI.

Todas essas teorias e relações estão de acordo com os princípios de modelagem propostos por Campos (2001 e 2004), considerando a Teoria do Conceito como um método híbrido para extração de conceitos integrados aos documentos de uma base de dados e representados por uma ontologia em um mapa conceitual. O mapeamento do novo sistema de conceitos para a BDICI deve ser realizado por uma equipe multidisciplinar, incluindo profissionais experientes em classificação facetada e linguagens documentárias.

A relação entre conceitos que consideram objetos do próprio hiperdocumento pode ser realizada através da categorização por agrupamento dos conceitos, de acordo com a natureza, pela relação hierárquica (ordem de dependência entre os conceitos), pelas relações de equivalência para o estabelecimento de significados conceituais dos termos sinônimos. Nessas relações, a linguagem deve ser entendida como uma enorme rede de opções inter-relacionadas e de significado paradigmático potencial, reconhecendo seus aspectos multifuncionais para estruturação e organização de um sistema de conceitos.

A organização e o tratamento da informação devem utilizar uma metodologia híbrida para a modelagem informacional, usando o próprio hiperdocumento científico para catalogação e indexação na fonte. Os objetos informacionais digitais do hipertexto servem de fonte primária para o entendimento do “conceito” e os processos de análise-síntese de modelagem conceitual e estruturação do sistema de “conceitos científicos” de acordo com as propostas de Campos (2004). Esta metodologia utiliza um método de agregação híbrida do método dedutivo e indutivo. No primeiro, formam-se categorias para representar contextos, e no segundo, analisam-se os conceitos de um contexto no propósito de ordená-los no interior de tais categorias formais da classificação em forma de facetas.

A ontologia de domínio, associada aos padrões de metadados W3C pode ajudar nesses processos com mecanismos artificiais e ferramentas de anotação semântica para a modelagem informacional. A ênfase desses processos deve analisar o objeto como uma informação registrada, ou melhor, analisando o “conceito” como “unidade de conhecimento registrado” que favorece mapeamento de um domínio com “conceitos científicos”. Neste sentido, a relevância está direcionada apenas ao “sistema computacional” para efeito de indexação. Além disso, há preocupação também com o usuário “aprendente” tanto na busca de termos desconhecidos quanto na recuperação de conceitos-chave, mostrando termos

sinônimos, vislumbrando a navegação em contexto do MHTX com esses conceitos hiperlinkados ao mapa representado por uma ontologia de domínio.

### **3.1 Detalhamento das implementações conceituais e tecnológicas**

O primeiro passo necessário para testar a viabilidade das propostas apresentadas para a BDICI é inserir uma ontologia de domínio na arquitetura do MHTX. O mapa deve mostrar “conceitos científicos” integrados por uma ontologia de domínio construída na IA com os padrões de metadados digitais propostos pelo W3C. Os nós e os links desta ontologia devem ser cuidadosamente avaliados em suas relações léxico-gramaticais para o estabelecimento de conexões dos objetos informacionais registrados.

Essas conexões devem seguir normas e princípios da classificação facetada e por esta razão, trata-se de um caso de “relevância mostrada”, ou melhor, explicitada pelo mapa. Por isso, a ontologia denominada ONTOCI – Ontologia da Ciência da Informação será também uma prototipação que procura utilizar somente informações necessárias para testar sua viabilidade no MHTX. Como não é necessário possuir uma ontologia completa, definindo todo o domínio de aplicação para utilizar os padrões de metadados W3C, buscase simplificar esses passos usando a fusão apresentada por Silva (2008) para verificar um sistema de conceitos em forma de facetadas da área da CI.

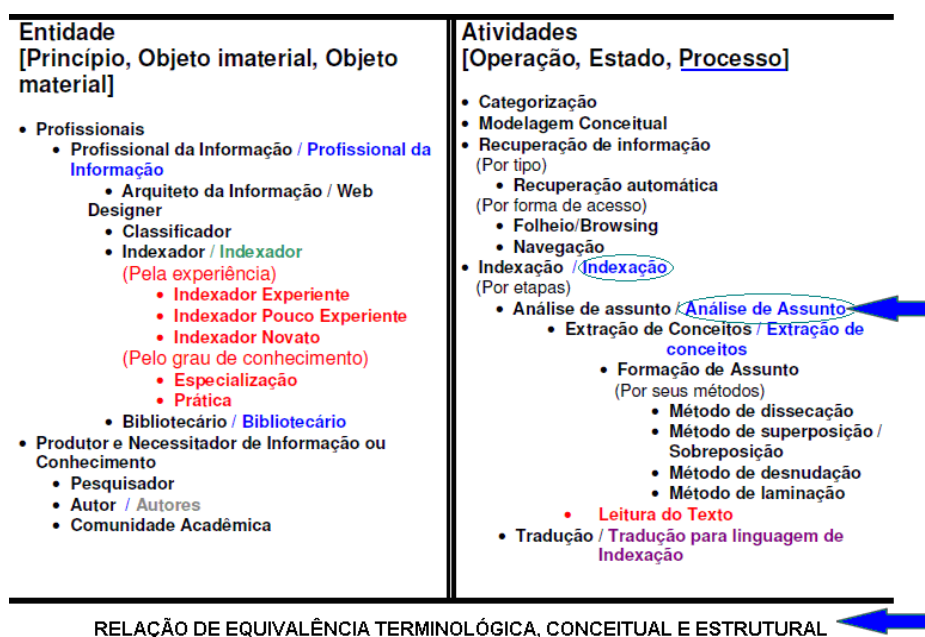
A fusão dos mapas das duas teses estudadas por Silva (2008) representa um cenário ideal para a estrutura taxonômica facetada da ONTOCI porque foi elaborada por profissionais experientes em instrumentos de vocabulário controlado e classificação facetada. Durante o processo de fusão, Silva identificou diferentes tipos de relações entre os elementos das duas estruturas. Essas relações foram classificadas considerando os três níveis de compatibilização: terminológico, conceitual e estrutural. O Quadro 9 mostra uma parte da estrutura unificada e o termo escolhido para representar a ONTOCI.

Segundo Silva (2008), a relação de equivalência acontece quando dois conceitos combinam suas características possuindo uma coincidência conceitual, terminológica e estrutural. Este tipo de relação ocorre entre dois ou mais termos que designam o mesmo conceito, ou seja, são esses termos-chave que identificam a relação de sinonímia entre conceitos, permitindo a compatibilização. A compatibilização de vocabulário ocorre quando existe um termo correspondente para cada termo de uma linguagem, de mesmo significado, na outra linguagem, e por isso, é possível converter termos de uma linguagem em outra, sem alterações de significado.

No Quadro 9, o termo “Análise de Assunto” representa a relação de equivalência terminológica, conceitual e estrutural. Por isso, este termo foi escolhido para verificar a compatibilização do vocabulário utilizando a ONTOCI devido à relação de equivalência

terminológica, conceitual e estrutural do processo de fusão dos mapas conceituais das duas teses já analisados por Silva. Então, a ONTOCI deve mostrar o termo “análise de assunto” no mapa porque ele equivale a uma terminologia consensual que possibilitou a relação de equivalência usada nas duas teses do MHTX. Além disso, o novo mapa deve permitir ao autor do documento estabelecer esses vínculos, indicando outros termos sinônimos, por exemplo, análise documentária, análise temática, análise conceitual, entre outros.

**QUADRO 9 – Esquema resultante da unificação manual das estruturas facetadas das 2 teses**



Fonte: Adaptado de Silva, 2008.

O W3C recomenda que sejam usadas ontologias escritas em OWL para representação de conceitos sobre um determinado domínio de conhecimento e que as anotações semânticas que descreverão a relação entre as instâncias das ontologias aos documentos, sejam escritas em RDF. Isso significa que a ontologia é uma descrição formal dos conceitos e relacionamentos existentes num domínio que se relaciona com vocabulário de uma linguagem específica para conceitualização de classes.

Uma das ferramentas escolhida para construção da ONTOCI foi o Protégé por suportar a modelagem em OWL e interagir com vários padrões de armazenamento como XML e RDFS. A hierarquia das classes em OWL é definida através da propriedade `rdfs:subClassOf`, criando relacionamentos “especialização-generalização”. OWL possui dois construtores próprios que, em conjunto com RDFS, permitem a definição das propriedades. São eles: `owl:ObjectProperty` e `owl:DatatypeProperty`. Indivíduos ou instâncias são a materialização das classes, ou seja, representam objetos e coisas específicas do mundo.

OWL possibilita a definição de diversos tipos de restrições que auxiliam na descrição das propriedades. Em OWL, uma restrição de propriedade é um tipo especial de descrição de classe, que descreve uma classe anônima e define uma classe que será utilizada para satisfazer a restrição, aplicada a todas as instâncias. O padrão escolhido para a construção do protótipo ONTOCI foi o OWL-DL que deve ser usado em projetos que sejam necessários uma expressividade sem perda da completeza computacional e decidibilidade. Dessa forma, garante-se que todas as conclusões lógicas podem ser computadas para terminarem o processamento em um tempo finito (HORRIDGE, 2005).

Com as ontologias do tipo OWL-DL pode ser utilizado um mecanismo de inferência baseado em lógica descritiva para verificar a consistência e para computar automaticamente a hierarquia de classes. As classes são criadas através da *tag* owl:Class, sendo que todas as classes pertencem a uma classe genérica chamada owl:Thing. Uma característica do OWL-DL é que o relacionamento superclasse-subclasse pode ser computado automaticamente por um mecanismo de inferência (HORRIDGE, 2005).

Para definir uma ontologia em OWL-DL, é preciso em primeiro lugar, dizer onde estão as classes primitivas de uma ontologia para que se possam definir novas classes como subclasse. Também é necessário determinar um *namespace* para a nova ontologia. As classes a serem definidas estarão localizadas no *namespace* da primeira definição. A segunda definição serve para que ontologias externas possam referenciar a ontologia sendo definida. As restantes localizam as definições primitivas de OWL, RDF, RDFS e XML.

Nessas classes localizam-se *namespaces* que servem de referência para ontologias externas em que cada tópico representa um único conceito e cada conceito é representado por um único tópico na relação tipo “um-para-um” conforme indicado por Silva (2008). Sob essa perspectiva de relação entre conceitos representados por um único tópico na relação tipo “um-para-um”, os *hyperlinks* indicam focos de conteúdo (tópicos ou temas) que o autor oferece ao leitor. Este decide os conteúdos a serem incorporados no texto e o percurso a ser seguido durante a leitura, produzindo, ele mesmo, as relações de equivalência. Esta é a situação ideal da modelagem informacional.

Como o processo de anotação semântica é aplicável a qualquer tipo de texto, é possível adicionar a um hiperdocumento da base de dados do MHTX, uma camada que descreve o seu conteúdo e associá-la aos “conceitos-chave” de uma ontologia de domínio identificados na URI. A ideia é que o novo sistema do MHTX ofereça aos alunos (autor de uma tese ou dissertação) do PPGCI-UFMG, no momento da publicação oficial (documento aprovado pela banca avaliadora) um cadastro para que eles mesmos possam relacionar os termos sinônimos na ONTOCI.

A visão dos usuários (autor/leitor) na BDICI pode fornecer os índices a serem considerados no processo de elaboração de hipóteses da organização da informação. Por isso, a arquitetura proposta para a BDICI envolve processos muito complexos para a organização e representação do conhecimento. A ONTOCI pode ter índices múltiplos em uma rede de conceitos integrados para permitir a modelagem informacional. Assim, o novo sistema pode oferecer ao autor do hiperdocumento ter acesso ao conteúdo estruturado onde existem regiões semânticas distintas explicitando o significado no contexto. Esses índices facilitam a navegação em contexto melhorando o processo de recuperação e evitando que o sistema tenha que pesquisar em todo o índice.

Os tópicos (termos-chave) de uma ontologia de domínio são ferramentas que possibilitam uma navegação contextualizada entre os elementos que os compõem, permitindo uma conexão entre o signo representacional e o recurso representado. Esses recursos permitem analisar grupos de textos usando representações gráficas, compararem grafos representando, por exemplo, estruturas sintáticas de textos, combinando conceitos em um mapa conceitual.

Considerando todos os esforços do W3C para a organização dos sistemas de hipertexto na Web na tentativa de padronização de vocabulários, muitos desses recursos podem ser aproveitados na arquitetura do MHTX com uso de ferramentas de anotações semânticas que relacionam conceitos de uma ontologia apontando índices sinônimos em um hiperdocumento científico. A ferramenta escolhida para anotação semântica foi Smore (Semantic Markup, Ontology and RDF Editor). Esta ferramenta foi escolhida por ser semi-automática e permitir ao usuário, realizar a anotação semântica de documentos em HTML usando uma ontologia OWL-DL.

A ferramenta Smore utiliza o método intrusivo de anotação, já que as anotações deverão ser salvas no formato RDF no próprio hiperdocumento que está sendo anotado. Além disso, a ferramenta é de simples instalação, apresentando documentação e tutoriais que facilitarão instalação e operação, exigindo apenas uma versão igual ou superior da plataforma Java SE 1.4 instalada no sistema operacional. A Figura 28 mostra a tela inicial da ferramenta Smore, já com as classes da ONTOCI aberta no Lado 1.

O Lado 2 apresenta um editor texto/HTML que permite aos usuários criarem e disponibilizarem páginas Web. Este editor facilita o uso de anotação semântica com a ontologia de domínio, de forma que objetos informacionais digitais podem ser descritos com a tríplice objetos-atributos-valores. Assim, o autor pode selecionar termos sinônimos direto no hiperdocumento que será publicado e arrastá-lo até as classes da ontologia. O sistema deve possibilitar que o autor aponte qual é o termo-chave que estabelece a relação de equivalência tríplice entre sujeito-predicado-objeto, com a vantagem de poder ignorar as

imposições da estrutura do hiperdocumento da camada inferior, referindo-se apenas aos dados sobre o conteúdo.

Primeiramente, a ontologia que será utilizada para o processo de anotação semântica dos termos sinônimos deve ser aberta. No caso da ONTOCI que está salva localmente, basta clicar no menu *File*, escolher a opção “*Load Ontology from Local File*”, selecionar o nome do arquivo da ontologia em OWL, que ela será carregada do lado esquerdo da janela (Lado 1). Depois, no Lado 2, deve-se abrir o hiperdocumento que será anotado. A ONTOCI foi aberta no Lado 1, podendo navegar pela ontologia, observando suas propriedades e instâncias em que são apresentadas as informações do código OWL referentes à classe selecionada na ontologia, na mesma parte da janela da ferramenta Smore, logo abaixo da ONTOCI do mesmo lado (Lado 1).

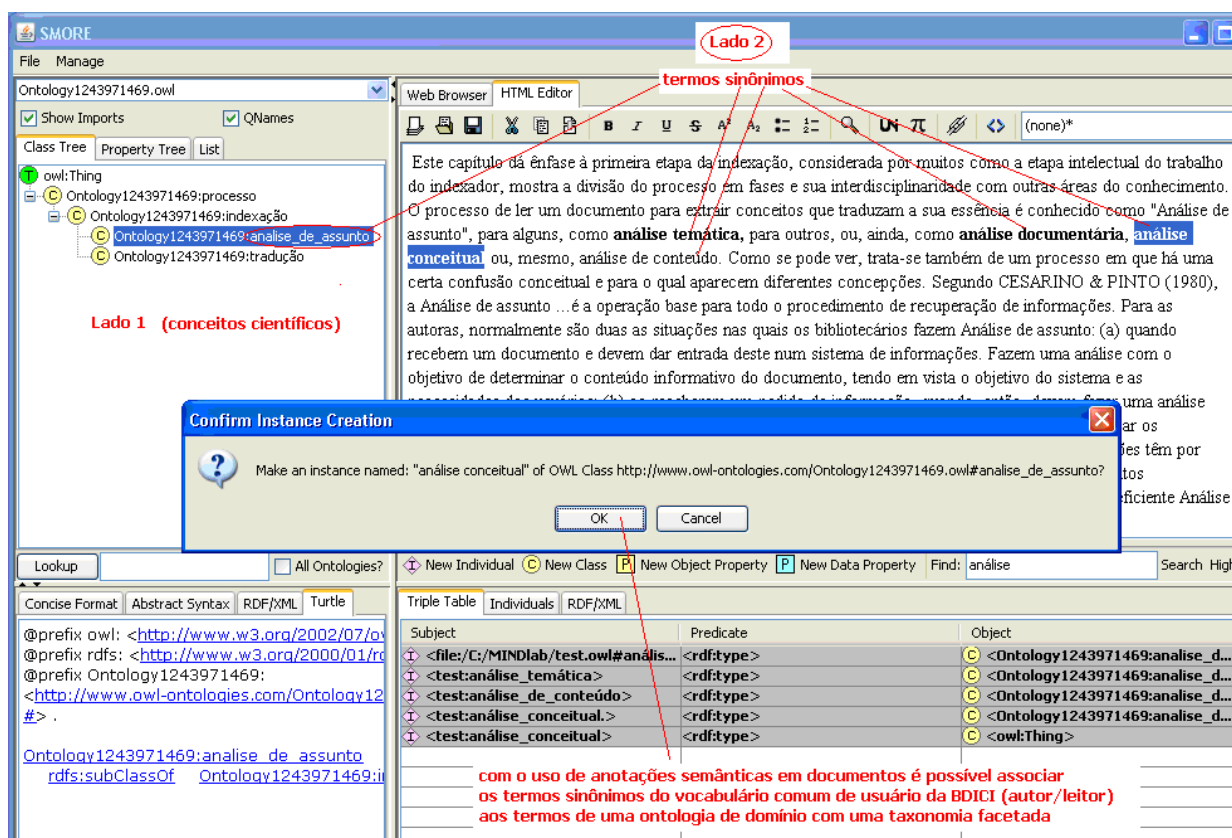


Figura 28 ONTOCI e hiperdocumento aberto com a ferramenta SMORE.

No Lado 2 da janela da ferramenta Smore, o documento em HTML deve ser aberto. Logo após, deve se selecionar a classe deste documento que será associada à ONTOCI, neste caso foi selecionado o conceito-chave “Análise de assunto” no próprio hiperdocumento e, ao clicar com o botão direito do mouse sobre o texto selecionado, aparece um submenu para se escolher qual a opção desejada: a) *create new individual* permite fazer a relação entre conceitos do hiperdocumento com classes da ontologia carregada, determinando que o termo selecionado seja associado a uma determinada



classe da base de conhecimento; b) *create new class*, permite a evolução da ontologia base carregada no início do procedimento, podendo definir novas classes para a ontologia; c) *create new object property*, permite a criação de uma nova propriedade sobre determinado termo do hiperdocumento que referencia uma classe da ontologia, possibilitando a definição das tags *domain* e *range*, que representam respectivamente a classe da ontologia e a faixa de valores que os indivíduos da propriedade definida poderão assumir; e d) *create new data property*, permite a definição de um tipo de dado para determinada propriedade.

Na tentativa de abrir uma das duas teses estudadas por Silva (2008) que estão disponíveis na base de dados do MHTX, a ferramenta Smore apresentou suas limitações sobre a quantidade de número de linhas do documento em HTML. Então, foi convertida uma parte de uma das teses para associar o termo “análise de assunto” a outros termos sinônimos. A Figura 28 (Lado 2) mostra a parte selecionada da tese de Naves (2000) cujo contexto esclarece outros termos sinônimos de “análise de assunto”.

Na ferramenta Smore, as anotações semânticas ficam listadas e codificadas seguindo o padrão de triplas RDF (sujeito – predicado – objeto) na parte abaixo do Lado 2 onde os termos sinônimos ficam nos campos “Subject”, “Predicate” e “Object” respectivamente. O campo “Subject” representa o texto selecionado no hiperdocumento como uma ocorrência de uma classe da ontologia. O campo “Predicate” representa o tipo de relação entre os termos presentes nos campos Subject e Object. Na ferramenta Smore, as relações possíveis são: *type*, *subClassOf*, *domain* e *range*. O campo “Object” representa a classe da ontologia a qual o termo selecionado no hiperdocumento será relacionado. As anotações são geradas automaticamente pela ferramenta Smore e a tripla é definida (Triple Table) com os termos que foram carregados nos campos “Subject”, “Predicate” e “Object”.

Assim, os termos sinônimos que foram selecionados no hiperdocumento aberto na ferramenta Smore, geraram automaticamente os códigos RDF, representando as anotações semânticas com a compatibilização do vocabulário. Esses códigos que podem ser acessadas clicando sobre a guia RDF/XML trazem as informações referentes aos termos do texto que já foram referenciados pela anotação realizada e que, agora, são instâncias da ONTOCI, indicando as relações de equivalência para possibilitar compatibilização semântica. Depois de fazer as relações de equivalência entre os conceitos presentes no hiperdocumento com as classes definidas na ontologia, as triplas RDF ficam definidas automaticamente e o relacionamento efetivado.

Logo após, esses códigos anotados deverão ser salvos no próprio documento em HTML antes de fechá-lo. Para isso, deve-se clicar na guia RDF/XML, selecionar o código RDF gerado pela ferramenta Smore e copiar para a memória do computador. Depois, abrir o código-fonte do documento HTML e colar o código RDF antes da tag `<html>` e salvar as

alterações. Finalmente, o processo de anotação semântica, utilizando a ferramenta Smore pode ser encerrado fechando a janela, pois os metadados (conceitos espontâneos e científicos) descritos em RDF/XML que representam os termos sinônimos que foram anotados e relacionados ao conceito "análise de assunto".

### **3.2 Análise dos Resultados**

Desde a criação inicial do protótipo MHTX, sua estrutura básica sofreu poucas alterações, o que dificulta a avaliação dos resultados apresentados para verificar a compatibilização do vocabulário neste repositório. Por este motivo, utilizou-se novas técnicas de modelagem na implementação conceitual dos atributos dos assuntos, procurando dar continuidade aos estudos de Silva (2008) com a fusão de duas teses da base de dados do MHTX para delimitação de conceitos válidos no processo de indexação.

O processo de modelagem informacional com o uso da ferramenta Smore foi bem sucedido no sentido de usar o mapa da ONTOCI para relacionar os termos sinônimos. Entretanto, esta ferramenta mostrou-se muito limitada, impossibilitando abrir o texto completo para criar relações de equivalência entre conceitos. Então, será necessário o desenvolvimento de ferramentas que atendam a esses critérios no novo sistema do MHTX para a modelagem informacional do modelo proposto.

Mesmo a ferramenta Smore limitando o número de linhas em um hiperdocumento em HTML, o seu uso foi muito simples, proporcionando ao usuário muitas opções de associações dos termos sinônimos de um documento de texto com uma ontologia de domínio em OWL para implementação conceitual e tecnológica do mapa proposto. O objetivo principal do novo sistema do MHTX deve ser facilitar o cadastro para o autor e, principalmente, verificar alguns mecanismos de inferência artificial para verificar a compatibilização semântica do vocabulário, utilizando a ONTOCI.

Os procedimentos realizados com a ferramenta Smore mostram que a indexação de um hiperdocumento científico pode ser feita a partir de descritores elaborados pelos próprios autores para viabilizar a modelagem informacional, apontando para o mapa da ontologia proposta. Neste sentido, a anotação semântica em um hiperdocumento funciona como um esquema específico para geração e uso de metadados digitais, referindo-se a entidades nomeadas na ontologia de domínio, usando os padrões e formalismos de representação propostos pelo W3C.

Este esquema representado por caracteres pode ser identificado e localizado por uma categoria URI com URN de um recurso na Web, referindo-se ao URN como um subconjunto do URI. O URN precisa permanecer globalmente único e persistente mesmo quando o recurso deixa de existir ou se torna indisponível para que a especificação e

utilização desses metadados possam garantir a existência de um conjunto de informações comuns sobre um determinado tema com regras claramente estabelecidas e aceitas pela comunidade (SAYÃO, 2007). Esses padrões de metadados digitais se constituem nas marcações que irão identificar tanto os elementos descritivos de um objeto informacional quanto seus elementos semânticos na Web e, em especial, nas bibliotecas digitais.

Os padrões propostos pelo W3C possibilitam uma melhor representação do conteúdo de um hiperdocumento, possibilitando operar outros padrões de metadados descritivos como o padrão DC em RDF/XML. O Quadro 10 mostra o hiperdocumento da tese de Naves (2000) que foi aberto com a ferramenta Smore no Lado 2 da Figura 29, incluindo as marcações da sintaxe implementada RDF/XML com Dublin Core, em que <dc> representa propriedades, características dos atributos de metadados digitais de um recurso.

#### QUADRO 10 – Exemplo de implementação em RDF/XML com Dublin Core

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/metadata/dublin_core#"
  <rdf:Description about="http://www.dlib.org">
  <dc:Title>Fatores interferentes no processo de análise de assunto:Estudo
    de caso de indexadores</dc:Title>
  <dc:Description>A análise de assunto como uma das etapas mais importantes da
    indexação. </dc:Description> <dc:Publisher>UFMG</dc:Publisher><dc:Date>2000
  </dc:Date> <dc:Subject> <rdf:Bag> <rdf:li>análise de assunto, indexação</rdf:li>
    </rdf:Bag> </dc:Subject>
  <dc:Type> Tese de Doutorado em Ciência da Informação </dc:Type>
  <dc:Format>text/html</dc:Format>
  <dc:Language>en</dc:Language>
  </rdf:Description>
</rdf:RDF>
```

**Fonte:** Elaborado pelo autor.

O novo sistema do MHTX pode compartilhar essas marcações com outros usuários, possibilitando a troca de informação para construção de novos conhecimentos através da discussão, reflexão e tomada de decisões de maneira colaborativa em um modelo de biblioteca digital especializada. Assim, os usuários, autores de um hiperdocumento, ao publicar sua tese ou dissertação no novo sistema podem identificar importantes palavras-chave, fazendo relações de equivalência do conteúdo de documentos com “conceitos científicos” formalmente descritos em uma ontologia.

Para isso, o novo sistema do MHTX deve possibilitar uma nova forma de indexação para a modelagem informacional que deve ser direcionada por um sistema especialista, um software desenvolvido para este fim. O ideal é que o mapa hiperbólico tenha uma única estrutura representada pela ONTOCI. Assim, quando o leitor-interpretador buscar por uma informação deste domínio, o mapa pode mostrar os termos-chave sinônimos que ele poderá fazer. Por exemplo, quando o usuário buscar o termo “análise de

assunto” na árvore, o mapa da ONTOCI poderá indicar quais os termos sinônimos que ele pode consultar sobre este assunto (Figura 29). Além disso, muitos outros benefícios com o reuso da ONTOCI poderão ser alcançados no MHTX.

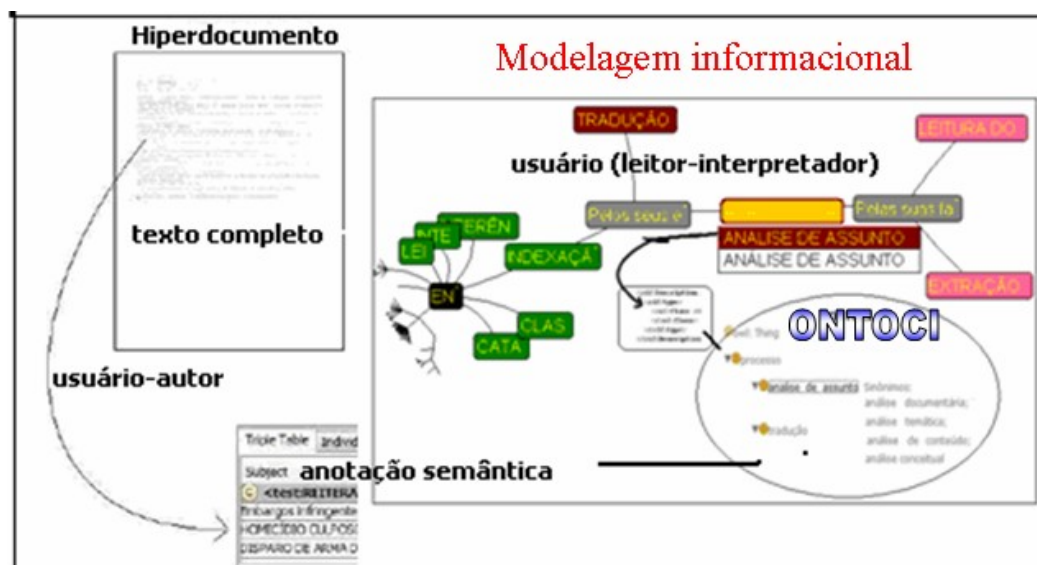


Figura 29 Exemplo do funcionamento da modelagem informacional proposta.

O MHTX pode adotar o sistema de anotação semântica progressiva para que os termos marcados nas futuras teses e dissertações, possam, assim, ser compatibilizados com os termos pré-definidos no mapa da ONTOCI. Assim que o sistema liberar o documento para publicação, ele deve “ler” todos os termos sinônimos que constam no documento para que o próprio autor faça a relação de equivalência entre os “conceitos científicos” indicados no mapa. Por exemplo, se o autor escreveu o termo “análise documentária”, o sistema deve indicar a relação de equivalência com o termo “análise de assunto”. O sistema também pode identificar outros dados no momento da publicação do texto completo tais como nome do aluno(autor), orientador, ano, instituição, entre outros.

A modelagem informacional proposta neste trabalho permitiu usar o conceito como um metadado descritivo, administrativo e estrutural mesmo com as limitações da ferramenta Smore. Tudo isso, confirma a viabilidade de inserir a ONTOCI no MHTX para buscas contextualizadas, além da compatibilização e interoperabilidade semântica. Os padrões W3C analisados favorecem o avanço das técnicas de indexação e recuperação da informação em um sistema de conceitos integrados em uma rede semântica, contribuindo para avanços no tratamento documentário que atualmente é adotado no MHTX.

## 4 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O desenvolvimento de modelos de bibliotecas digitais especializadas é tendência mundial. Os usuários desses sistemas precisam de serviços interativos, contextualizados e dinâmicos. Em função dos benefícios que as ontologias desenvolvidas na IA vêm demonstrando com uso dos padrões de metadados W3C, foi proposto como objetivo principal desta pesquisa verificar a compatibilização semântica do vocabulário, acreditando no aperfeiçoamento do MHTX para transformá-lo em um modelo denominado Biblioteca Digital Interativa da Ciência da Informação (BDICI).

A navegação em contexto semântico em forma de facetas proposta no mapa conceitual do MHTX favorece a inserção de uma Ontologia da Ciência da Informação (ONTOCI) construída na IA como um sistema de conceitos integrados, usando padrões de metadados W3C e linguagens documentárias. Então, a partir dos princípios da classificação facetada que fundamentam a estrutura do mapa conceitual no MHTX, foram analisados os processos de construção de ontologias para verificar a compatibilização semântica do vocabulário. Essa verificação parte dos pressupostos de analisar as potencialidades do “conceito” como um metadado descritivo, administrativo e estrutural para a modelagem de objetos informacionais digitais entre arquivos distribuídos em bibliotecas.

A necessidade de um novo tipo de modelagem de objetos informacionais vem ao encontro das propostas dos padrões de metadados digitais W3C para a catalogação de recursos que podem ser indexados direto na fonte de um hiperdocumento científico. Considerando que um sistema possui compatibilidade semântica quando é capaz de trabalhar em conjunto com outros sistemas, compartilhando informações e com entendimento de um vocabulário comum entre os significados dos seus termos, a proposta da BDICI foi usar a relação de equivalência entre conceitos, explicitados em um único mapa.

Um dos primeiros questionamentos para o estabelecimento de relações é em relação à tipologia convencional de um hiperdocumento científico (dados, informações e conhecimento). Esta é a visão lógica do texto completo que representa o conteúdo de um documento que pode compor o acervo do MHTX com um tratamento mais adequado para aproximar a indexação de conceitos válidos extraídos pelo próprio autor no momento da publicação. Um conceito integrado em um único mapa conforme a proposta deste trabalho para o aperfeiçoamento do MHTX reforça que, para se chegar a estas facetas, é necessário determinar as características dos objetos informacionais digitais, estruturando esses objetos como unidades de conhecimento registrado.

Para a compatibilização semântica de uma estrutura de um modelo conceitual por prototipação, o contexto é um elemento essencial como mecanismo humano de inferência conforme a proposta da BDICI. Esses mecanismos enfocam a natureza sistêmico-funcional e interativa da linguagem, considerando o texto como um complexo de significados léxico-gramaticais que são contextualizados em práticas sociais, situacionais e culturais. A gramática é o mecanismo linguístico que liga essas metafunções a determinadas percepções de objetos informacionais em uma situação de relações significativas, numa forma estrutural unificada, para realizar a troca de ação e informação.

Este sistema encontra-se voltado a análise do discurso, abrangendo o conhecimento e experiências de mundo dos falantes/escritores, enfatizando as relações entre os participantes e seus posicionamentos que expressam as escolhas semânticas, formando o conteúdo de um documento com o seu texto completo. O modo de organizar e estruturar o texto ao mapa é que dá a coerência ao sistema em que as relações devem explorar o significado da oração como troca para interatividade.

Um campo conceitual explicitado no mapa deve se associar aos significados por meio dos padrões de transitividade deste sistema léxico-gramatical. No mapa conceitual do MHTX representado pela ONTOCI, um conceito hiperlinkado passou a funcionar como mecanismo artificial desambiguador de objetos informacionais digitais, facilitando a identificação diretamente na fonte do hiperdocumento. Isto porque a compatibilização e interoperabilidade semântica entre esses repositórios são garantidas pelos padrões de metadados W3C em uma única estrutura.

A “superestrutura” específica de metadados conceituais representada por um esquema hierárquico, semântico e sintático do uso da linguagem verbal como grandezas-signos, associadas aos recursos da IA com uso do vocabulário controlado, possibilitou novos formatos para organização, tratamento e recuperação de objetos informacionais digitais. Assim, a estrutura taxonômica facetada do mapa da ONTOCI cumpriu um papel importante no estabelecimento das funções de preservação de metadados entre bibliotecas e arquivos digitais configurados de forma distribuída. Por isso, foi possível converter termos sinônimos de mesmo significado de acordo com o contexto, sem alterações de significado.

Na BDICI, esses metadados podem ser associados aos padrões de metadados descritivos como o Dublin Core (DC), facilitando também a descoberta de outros objetos para suportar o gerenciamento da estrutura em um único sistema com conceitos integrados, além de juntar os componentes de uma coleção. A preservação desta “superestrutura” de metadados digitais é um fator importante para a integridade dos objetos informacionais registrados, mostrando que estes recursos são compostos também pelos seus relacionamentos com outros objetos para indexação na fonte. Quanto à identificação

permanente com URI e XML/RDF/OWL, ela garante o acesso contínuo ao próprio recurso para assegurar que sua localização se “materialize” todas as vezes que seu link referencial for selecionado por um usuário-sistema.

Consequentemente, tais macro e microestruturas são tão importantes para bibliotecas digitais quanto são os padrões de metadados descritivos propostos com o uso do DC. Esses são os critérios de padronização de metadados digitais W3C que devem ser aplicados a ontologias, visando compatibilização e interoperabilidade semântica. A interface de um modelo conceitual de interatividade em RI exige o diálogo constante entre usuários e computador por meio da linguagem como fenômeno informacional e dinâmico.

A modelagem informacional reforça a necessidade de ter bastante claro o método utilizado para a sua construção das categorias formais em um sistema de conceitos integrados em forma de facetas. Os padrões W3C com uso de ontologias e anotações semânticas no próprio hiperdocumento científico permitem o uso do método híbrido sugerido na compatibilização de vocabulário controlado, utilizando um sistema de conceitos integrados como um modelo relacional. A aplicação sugerida usou uma única estrutura para ajudar a otimizar os mecanismos de indexação no MHTX, utilizando o conceito como um metado descritivo, administrativo e estrutural na modelagem informacional.

No mapa representado pela ONTOCI, um “conceito científico” funciona como entradas triplas no processamento da linguagem natural em um conjunto de regras lógicas dedutivas e relações entre conceitos que foram elaborados por profissionais experientes em instrumentos de vocabulário controlado e sistema de classificação facetada. Isso mostra a existência de dois métodos que conduzem o ato do raciocínio, ou seja, o dedutivo e o indutivo. Esses métodos possibilitam mecanismos de inferência em uma estrutura facetada para resumir o significado dos dados e controlar a gestão da informação, permitindo a busca por um termo sinônimo e a recuperação interativa da informação.

Os resultados deste estudo exploratório confirmam a viabilidade de inserir a ONTOCI como um único mapa conceitual no MHTX para possibilitar a compatibilização semântica do vocabulário, explicitando a relação de equivalência entre conceitos, analisando principalmente os termos sinônimos de um domínio. A modelagem informacional para o compartilhamento do entendimento da terminologia comum entre pessoas e agentes de software e a permissão de reuso da ONTOCI são as principais contribuições desta pesquisa. De acordo com as propostas da BDICI, o mapa indica relações entre conceitos, explorando outros recursos para prevenir certos usos inadequados de um vocabulário em uma consulta, ao utilizar terminologia padronizada que foi definida na ONTOCI.

Os padrões propostos na BDICI favorecem a extração de conceitos válidos representados por uma “superestrutura” de um sistema conceitual em forma de categorias

formais para a interpretação não ambígua das definições terminológicas, estabelecendo relações de equivalência que podem ser utilizadas em qualquer modelo de biblioteca digital especializada. Neste sentido, pode-se concluir que, além dos avanços na organização do MHTX, esta pesquisa também pode contribuir no sentido de propor uma ontologia de domínio como modelo relacional na língua portuguesa para a CI, possibilitando uma representação de “conceitos científicos” desse domínio para reuso em diferentes aplicações.

Uma série de adaptações são necessárias para a construção da BDICI, exigindo dinamismo dos profissionais que organizam e estruturam a informação para constante manutenção e atualização do novo sistema proposto. Na BDICI, a recuperação pode ser facilitada através de sucessivas categorizações, tal como acontece em programas interativos, nos quais o usuário é encaminhado ao assunto desejado. Esses recursos podem funcionar como mais uma forma de interação e recuperação para fomentar atitudes colaborativas em rede, possibilitando ampla divulgação da produção acadêmica e científica desenvolvida não só na ECI, como também em nível nacional de publicação na área da CI.

Tudo isso implica em muito trabalho e desafios futuros e exige investimentos institucionais e governamentais. Os maiores desafios são referentes a novos termos que podem surgir, pois o vocabulário humano é dinâmico e os profissionais que organizam um modelo de biblioteca digital interativa devem ficar atentos ao tratamento da informação mais adequado no momento da publicação. A atualização do mapeamento de uma ontologia deve ser um processo contínuo em qualquer sistema que se propõe a trabalhar com modelos conceituais de interface como um diálogo de interatividade em RI.

Na BDICI, um conceito caracterizado pela palavra no léxico, precisa ser tratado por especialistas de um domínio com instrumentos de vocabulário controlado para que os termos representados na ONTOCI sejam construídos com definições rigorosamente explícitas por um modelo conceitual e relacional. Este mapa deve representar uma terminologia consensual e consistente com padrões propostos pelo W3C para o estabelecimento de relações de equivalência entre os conceitos da ONTOCI, identificando recursos no próprio hiperdocumento científico a ser publicado na BDICI.

Como esses processos podem envolver diferentes aspectos ainda nem conhecidos, os profissionais que organizam a informação em qualquer modelo de biblioteca digital devem estar dispostos a desenvolver pesquisas multidisciplinares que demandam atitudes de ousadia e perspicácia para a criação de novas ferramentas de publicação.



## REFERÊNCIAS

- AGUIAR, F. L. **O controle de vocabulário como dispositivo metodológico para a organização, tratamento e organização da informação arquivística**. Dissertação de Mestrado em Ciência da Informação, PUC- Campinas, 2008. 267p.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília v.32, n.3, p. 7-20, set./dez. 2003.
- ALMEIDA, M. B. Roteiro para construção de uma ontologia bibliográfica através de ferramenta automatizada. **Perspectivas em Ci. Inf.** Belo Horizonte, v.8, n.2, p. 164-179, 2003. Disponível também em: <[http://www.eci.ufmg.br/mba/text/art\\_ontobibliog\\_sub1\\_WEB.pdf](http://www.eci.ufmg.br/mba/text/art_ontobibliog_sub1_WEB.pdf)> Acesso em 2006.
- \_\_\_\_\_. **Um modelo baseado em ontologias para representação da memória organizacional**. Tese de Doutorado em Ciência da Informação – ECI/UFMG. Belo Horizonte. 2006. 345 p.
- ALVARENGA, L. Representação do conhecimento na perspectiva da Ciência da informação em tempo e espaço digitais. **Enc. Bibli. Ci. Inf.**, Florianópolis, n. 15, 1º sem. 2003.
- \_\_\_\_\_. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **Datagramazero**, 2(6), IASI 2001. Disponível em: <[http://www.dgzero.org/dez01/F\\_1\\_art.htm](http://www.dgzero.org/dez01/F_1_art.htm)> Acesso em 2007.
- ALVARENGA, L.; SOUZA, R. R. A Web Semântica e suas contribuições para a ciência da informação. **Ci. Inf.** Brasília, v.33, n.1, jan./abr. 2004
- AMARAL, L. A. M. Planejamento de Sistemas de Informação: metodologias, ferramentas e modelos. **Componente Pedagógica**, Universidade do Minho, 1988a.
- ARAÚJO, C. A. Fundamentos teóricos da classificação. **Enc. Bibli. Ci. Inf.**, Florianópolis, n.22, 2006.
- ASSMANN, H. A metamorfose do aprender na sociedade da informação. **Ci. Inf.** Brasília, v. 29, n. 2, p. 7-15, maio/ago. 2000.
- AZEREDO, J. C. **Gramática Houaiss da Língua Portuguesa**. São Paulo: Publifolha, 2008.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. Addison-Wesley, England, 1999.
- BAKHTIN, M. **Os gêneros do discurso**. In: Estética da criação verbal [1979]. Tradução Paulo Bezerra. 4. ed. São Paulo: Martins Fontes, 2003.
- BARBOSA, M. A. Relações de significação nas unidades lexicais. In: CARVALHO, M.; SILVA, M. E. B. **Anais do 1º Encontro Nacional do GT de Lexicologia, Lexicografia e Terminologia da ANPOLL**. Recife, 1998. p. 19-40.
- BARRETO, A. A. A oferta e a demanda da informação: condições técnicas, econômicas e políticas. **Ci. Inf.** Brasília, v. 28, n. 2, p.168-142, maio/ago 1999.
- \_\_\_\_\_. A Condição da Informação. **Perspectiva**, São Paulo 16(3): 67-74, 2002.
- BASTOS, F. M. **Organização do conhecimento em bibliotecas digitais de teses e dissertações: análise da aplicabilidade das teorias macroestruturais para categorização de áreas de assunto**. Dissertação de Mestrado em Ciência da Informação da UNESP, Marília, São Paulo, 2005, 111p.

BEPPLER, F. D. **Um Modelo para Recuperação e Busca de Informação Baseado em Ontologia e no Círculo Hermenêutico**. Tese de Doutorado em Engenharia da UFSC. Florianópolis, 2008, 135 p.

BERNERS-LEE, T. **Weaving the Web**. London: Orion Business, 1999. Disponível em: <[http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)>. Acesso em 2007.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**. 2001.

BORST, W. N. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. 1997. Tese PhD. Disponível em <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>> Acesso 2008.

BRANDÃO, W. C. **Método para produção de representações conceituais a partir de literatura especializada utilizando a abordagem analítico-sintética**. Dissertação de Mestrado em Ciência da Informação – ECI/UFMG, Belo Horizonte, 2005, 96 p.

BRÄSCHER, M. A Ambigüidade na Recuperação da Informação. **DataGramZero** v.3 n.1, 2002.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, Anais. São Paulo: ANCIB, 2008. Disponível em: <<http://www.enancib2008.com.br>>. Acesso em 2008.

BUCKLAND, M. Information as thing. **Journal of the American Society of Information Science**, v.42, n.5, p.351-360, 1991. Disponível em: <<http://www.sims.berkeley.edu/~buckland/alise00.html>> Acesso em 2009.

\_\_\_\_\_. The landscape of information science. **The American Society for Information Science**. Preprint of article published in "JASIS at 50" special issue of JASIS, 1999. Disponível em: <<http://www.sims.berkeley.edu/~buckland/asis62.html>> Acesso em 2009.

BUENO, T. C. D'A. **Engenharia da Mente: uma metodologia de representação do conhecimento para construção de ontologias em sistemas baseados em conhecimento**. Tese de Doutorado em Engenharia de Produção PPGEP-UFSC, Florianópolis, 2005. 174 p.

BURANARACH, M. **The Foundation for Semantic Interoperability on the World Wide Web**. PhD Thesis (Department of Information Science and Telecommunications School of Information Sciences)-University of Pittsburgh, Pittsburgh, 2001.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, July 1945. Disponível em: <[www.theatlantic.com/doc/194507/bush](http://www.theatlantic.com/doc/194507/bush)>. Acesso em 2006.

CABRÉ, M. T. **La Terminología: teoría, metodología, aplicaciones**. Barcelona: Antártida, 1993.

\_\_\_\_\_. **La Terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos**. Barcelona: IULA: Universitat Pompeu Fabra, 1999.

CAMPOS, M. L. A. **A organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização da autoria**. Tese de Doutorado em Ciência da Informação da UFRJ/ECO/IBTC, Rio de Janeiro, 2001.

CAMPOS, M. L. A.; GOMES, H. E. Organização de Domínios de Conhecimento e os princípios Ranganathanianos. **Perspectivas em Ci. Inf.** Belo Horizonte, v.8, n.2, jul/dez. 2003.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ci. Inf.** Brasília, v.33, n.1, p. 22-32, 2004.

CAMPOS, M. L. A.; GOMES, H. E. Princípios de Organização e Representação do Conhecimento na Construção de Hiperdocumentos. **DataGramZero**, v.6 n.6, dez 2005.

\_\_\_\_\_. Integração de Ontologias: o domínio da bioinformática. RECIIS - Revista de Comunicação, Informação; **Inovação em Saúde**, v. 1, p. 117-121, 2007.

CAMPOS, M. L. A.; GOMES, H. E.; MOTTA, D. F. Elaboração de tesauro documentário: tutorial. Disponível em: <<http://www.conexao.org/bit/tesauro/glossario.htm#tesauro>>. Acesso em 2007.

CAPURRO, R. Epistemologia e Ciência da Informação. In: **ENANCIB**, 5. Anais... Belo Horizonte: ECI/UFMG, 2003. 19p.

CAPURRO, R.; HJØRLAND, B. **The concept of information**. Annual Review of Information Science & Technology, v.37, p.343-411, 2003.

CARLAN, E. **Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação**. Dissertação Mestrado da Ciência da Informação – UNB. Brasília, 2010.

CARVALHO, C. **Para compreender Saussure: Fundamentos e visão crítica**. Petropolis:Vozes, 2003.

CASTILHO, A. T. Fundamentos teóricos da gramática do português culto falado no Brasil: sobre o segundo volume, classes de palavras e as construções gramaticais. **Alfa**, São Paulo, v.51, 2007.

CENDÓN, B. V. Ferramentas de busca na Web. **Ci. Inf**, Brasília, v. 30, n. 1, abr. 2001.

CESARINO, M. A. N.; PINTO, M. C. M. Biblioteconomia. **Ci. Inf**, Brasília, v.8, n1, p.32-43, jan./jun. 1980.

CHAUÍ, M. **Convite à Filosofia**. São Paulo, ed. Ática, 2000.

CHOO, C. W. **The Knowing Organization**. New York: Oxford University Press, 1998.

CINTRA, A. M. M. Elementos de lingüística para estudos de indexação. **Ci. Inf.**, v.12, n.1,p.5-22, 1983.

\_\_\_\_\_. **Para entender as linguagens documentárias**. 2ed. rev. e ampl. São Paulo: Polis, 2002.

COAD, P.; YOURDON, E. **Análise baseada em objetos**. Rio de Janeiro: Editora Campus. 1992.

CONTERATTO, G. B. H. Predicação Secundária: **Uma Contribuição da Lingüística ao Processamento Computacional da Linguagem**. Dissertação de Mestrado em Comunicação da UNISINOS, São Leopoldo, 2005.

CRUZ, F. L. **A necessidade de informação dos projetistas de interfaces de sistemas interativos na web, com foco em usabilidade**.Dissertação de Mestrado em Ciência da Informação da CID/UnB, Brasília, 2008, 101p.

CUNHA, M.B. **Construindo o futuro: a biblioteca universitária brasileira em 2010**. Ci. Inf., v.29, n.1, p.71-89, jan.-abr. 2000.

CURTY, M. G; BOCCATO, V. R. C. O artigo científico como forma de comunicação do conhecimento na área de Ciência da Informação. **Perspect. Ci. inf.**, Belo Horizonte, v.10 n.1, p. 94-107, 2005.

DAHLBERG, I. **Ontical structures and universal classification**. Bangalore: Sarada Ranganthan Endowment, 1978. 64p.

\_\_\_\_\_. A Referent-oriented analytical concept theory of interconcept. **International Classification**, v.5, n.3, p.142-150, 1978.

\_\_\_\_\_. Towards establishment of compatibility between indexing languages. **International Classification**, v.8, n.2, p.88-91, 1981.

\_\_\_\_\_. Terminological definitions: characteristics and demands. In: Problèmes de la définition et de la synonymie en terminologie. Québec, **GIRSTERM**, 1983, p.13-51.

\_\_\_\_\_. Conceptual compatibility of ordering systems. **International Classification**, v.10, n.2, p.5-8, 1983.

\_\_\_\_\_. Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year. **International Classification**, Frankfurt, v. 19, n. 1, p. 1-2, 1992.

\_\_\_\_\_. Knowledge organization: its scope and possibilities. Knowledge Organization, Würzburg, v. 20, n. 4, p. 211-222, 1993.

DIAS, E. W. Contexto Digital e Tratamento da Informação. In **DataGramZero**, 2(5), IASI - Instituto de Adaptação e Inserção na Sociedade da Informação. 2001.

\_\_\_\_\_. Análise de assunto: percepção do usuário quanto ao conteúdo de documentos. **Perspect. Ci. Inf.**, Belo Horizonte, v.9 n.2, p. 146-157, jul./dez. 2004.

DIK, S. C. **Theory of Functional Grammar**. Org.Kees Hengeveld. Berlin: Mouton de Gruyter. v. 2. 1997.

DZIEKANIAK, G. Mapeamento do uso de padrões de metadados por comunidades científicas. **Biblos**, Rio Grande, 20: 229-243, 2006.

ELLER, M. P. **Extração de anotações semânticas de fontes de dados heterogêneas: um estudo de caso com a ferramenta Smore**. Monografia de Bacharelado em Sistemas de Informação - UFSC, Florianópolis, 2008.

ELMASRI, R.; NAVATHE S. B. **Sistemas de Banco de Dados: Fundamentos e Aplicações**. 3.ed. Addison Wesley Longman, 2002.

FALBO, R. A. **Integração de conhecimento em um ambiente de desenvolvimento de software**. Tese de Doutorado em Ciências da COPPE/UFRJ. Rio de Janeiro, 1998.

FARIA, M. P. **Uma biblioteca digital para o fomento de atitudes colaborativas nos usuários de um ambiente de aprendizagem online**. Dissertação de Mestrado em Tecnologia - CEFET-PR, Curitiba, 2005, 131 p.

FARRADANE, J. E. L. Relational Indexing. **Journal of information science**, n.1, p. 267-276, 1980.

FERREIRA, S. M. S. P. Novos paradigmas e novos usuários de informação. **Ci Inf**. Brasília, v.25, n.2, 1995.

FIGUEIREDO, C. X; FRANCISCANI, J. F.; CARDOSO, O. N. P. Recuperação de Informação e Bibliotecas Digitais. **V SECICOM**. Universidade Federal de Lavras, 2003.

FONSECA, R. S. **Aristóteles e a linguagem: estudo e tradução do Peri hermeneias (partes 1-6)**. Tese de Doutorado em Lingüística da Universidade Estadual de Campinas, São Paulo, 2009.

FONTES, C. **Inferência**. 2007. Disp.em: <<http://afilosofia.no.sapo.pt/Inferencia.htm>> Acesso em 2009.

FREIRE, G. H. Ciência da informação: temática, histórias e fundamentos. **Perspect. Ci. Inf.**, Belo Horizonte, v.11 n.1, p. 6-19, jan./abr. 2006.

FREITAS F. et. al. Contribuições de binômio metadados-ontologias para a Ciência da informação – uma análise suscinta. In Proceedings... **CINFORM – Encontro Nacional de Ciência da Informação VI**, Salvador, 2006. Disponível em <<http://dici.ibict.br/archive/00001063/>> Acesso em 2008.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Biblioteconomia e Ci. Inf.**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003

FURGERI, S. **Representação de informação e conhecimento: estudo das diferentes abordagens entre a ciência da informação e a ciência da computação**. Dissertação de Mestrado em Ciência da Informação da PUC-Campinas, São Paulo, 2006, 159p.

GALVÃO, M. C. B. Construção de conceitos no campo da Ciência da Informação. **Ci. Inf.**, Brasília, v. 27, n. 1, p. 46-52, jan./abr. 1998.

GARDNER, H. **The mind's new science**. New York, Basic Books Inc., 1987.

GARDNER, H.; HATCB, T. Multiple intelligences go to school: educational implications of the theory of Multiple Intelligences. **Educational Researcher**, v.18, n.8. p.4-10, 1989.

GERALDI, J. W. **Linguagem e ensino: exercícios de militância e divulgação**. 6. ed. Campinas, SP: Mercado de Letras: Associação de Leitura do Brasil, 2006.

GOMES, H. E. Manual de elaboração de tesouros monolíngües. Ministério da Educação, Ciência e Tecnologia, **Programa Nacional de Bibliotecas de Instituições de Ensino Superior**. Brasília, 1990.

GONZÁLEZ GÓMEZ, M. N. Metodologia de pesquisa no campo da ciência da informação. **DataGramZero**, v. 1, n. 6, dez. 2000.

\_\_\_\_\_. Para uma reflexão epistemológica acerca da ciência da informação. **Perspectivas em Ci. Inf.**, Belo Horizonte, v. 6, n. 1, p. 5-18, jan./jun. 2001.

\_\_\_\_\_. Novas fronteiras tecnológicas das ações de informação: questões e abordagens. **Ci. Inf.**, Brasília, v. 33, n. 1, p. 55-67, jan./abril 2004.

GORSKI, L. **A linguagem como instinto para Steven Pinker**. Dissertação de Mestrado em Filosofia da PUC-Paraná. Curitiba, 2007.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, p. 199-220, 1993.

GUARINO, N. Formal ontology and information systems. **Proceedings of FOIS**, Trento: Italy, p3-15, 1998.

HALLIDAY, M. A.K. **An Introduction to Functional Grammar**. 2ed. London NY: Edward Arnold, 1994.

HJELMSLEV, L. Expressão e conteúdo. In: Prolegômenos a uma teoria da linguagem. Trad. J. Teixeira Coelho Neto. São Paulo: **Perspectiva**, 1973. p. 53-64

HORRIDGE, M. et al. **Um guia prático para a construção de ontologias OWL, plugin Protégé-OWL 3.4**. Trad. SOARES, D. R.; ALMEIDA, M. B. Disponível na Internet <[www.eci.ufmg.br/mba/](http://www.eci.ufmg.br/mba/)>, 2008. 100p. Acesso em nov 2008.

JACOB, E. K. Ontologies and the Semantic Web. **Bulletin of the American Society for Information Science and Technology**, apr./may, p. 19-22, 2003.

JACOBSON, I. **Object-oriented software engineering- A use case driven approach**. Addison Wesley Publishing Company. 1992.

JANCZURA, G. A. **Contexto e normas de associação para palavras: a redução do campo semântico**. Paideia, 2005, 15, 32, 417-425.

JONASSEN, D. H. Designing Structured Hypertext and Structuring Access to Hypertext. Designing Structured Hypertext and Structuring Access to Hypertext. **Educational Technology**, v. 28, Nov. 1988.

KATO, M. **O aprendizado da leitura**. São Paulo: Martins Fontes, 1985. 120p.

KENT, A. **Manual da recuperação mecânica da informação**. São Paulo, Polígono, 1972.

KIRYAKOV, A.; POPOV, B.; OGNANOFF, D. **Semantic Annotation, Indexing, and Retrieval**. 2nd International Semantic Web Conference, Florida, USA. p.484-499, 2003.

KOBASHI, N. Y. **A elaboração de informações documentárias: em busca de uma metodologia**. Tese de Doutorado em Ciências da Comunicação da USP, São Paulo, 1994. 195 f.

KOBASHI, N. Y. Vocabulário controlado USP: desenvolvimento, implantação e gerenciamento. In: **Congresso Internacional de Arquivos, Bibliotecas, Centros de Documentação e Museus**, 2002, São Paulo. Integrar: 1. Congresso Internacional de arquivos, bibliotecas, centros de documentação e museus: textos. São Paulo: Imprensa oficial do Estado, 2002. v. 1. p. 225-235

KOCH, I. Introdução à Lingüística Textual: trajetória e grandes temas. **Coleção Texto e Linguagem**. São Paulo, Martins Fontes, 2004.

KURAMOTO, H. Biblioteca digital brasileira: integrando a ICT brasileira. In: MARCONDES, C. H. (Org). **Bibliotecas digitais: saberes e práticas**. 2ed. Brasília: IBICT, 2006. p. 287-303.

LAKOFF, G. JOHNSON, M. **Metaphors We Live By**, Chicago: The University of Chicago Press. 1980.

LAKOFF, G. **Women, Fire and Dangerous Things: What Categories Reveal About the Mind**, Chicago: TheUniversity of Chicago Press. 1987.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 1993. 347p.

LE COADIC, Yves-François. **A Ciência da Informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LEFFA, V. J. **Aspectos da leitura: uma perspectiva psicolinguística**. Porto Alegre: Sagra, 1996.

LÉVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro: Ed. 34, 1993.

LIMA, G. A. B. A análise facetada na modelagem conceitual de sistemas de hipertexto: uma revisão de literatura. **Perspect. Ci. Inf.**, Belo Horizonte, v. 7, n. 2, p. 189-196, jul./dez. 2002.

\_\_\_\_\_. Interfaces entre a Ciência da Informação e a ciência cognitiva. **Ci. Inf.**, Brasília, v. 32, n.1. p.77-87, 2003.

\_\_\_\_\_. **Mapa hipertextual (MHTX): um modelo para organização hipertextual de documentos**. Tese de Doutorado em Ciência da Informação – ECI/UFMG, Belo Horizonte, 2004.

\_\_\_\_\_. Gerenciamento do conteúdo semântico no protótipo mapa hipertextual MHTX: Perspectivas. **VIII ENANCIB** – Encontro Nacional de Pesquisa em Ciência da Informação. Salvador, 2007.

LIMA, V. M. A. **Terminologia, Comunicação e Representação Documentária**. Dissertação de Mestrado em Ciências da Escola de Comunicações e Artes da USP. São Paulo, 1998, 118 p.

- LOH, S. **Abordagem Baseada em Conceitos para Descoberta de Conhecimento em Textos**. Tese de Doutorado em Ciência da Computação da UFRGS, Porto Alegre, 2001.
- LUCENA, P. **Semantic Agent, uma plataforma para desenvolvimento de agentes inteligentes**. Dissertação de Mestrado em Ciências pelo ICMC da USP, São Paulo, 2003, 151f.
- LUFT, P. C. **Língua e liberdade: por uma nova concepção da língua materna**. 8ed. São Paulo: Ática, 2003.
- LYONS, J. **Semantics**. Cambridge: Cambridge University Press, 1977. v. 1 e 2.
- LYYTINEN, K. Different Perspectives on Information Systems: Problems and Solutions. **ACM Computing Surveys**, v.19(1):5-45, 1987.
- LYYTINEN, K. Two views of information modeling. **Information and Management**, n. 12, p. 9-19, 1987.
- MARCIAL, E. C. **Utilização de modelo multivariado para identificação dos elementos-chave que compõem sistemas de inteligência competitiva**. Dissertação de Mestrado em Ciência da Informação da Universidade de Brasília, Brasília, 2007. 163f.
- MARCONDES, C. H.; KURAMOTO, H.; TOUTAIN, L. B.; SAYÃO, L. (Org.) **Bibliotecas digitais saberes e práticas**. Salvador, BA : **EDUFBA**; Brasília: IBICT, 2005.
- MARCONDES, C. H. Ontologias como novas bases de conhecimento científico. **Perspect. Ci. Inf.**, v.13, n.3, Belo Horizonte, set./dez. 2008.
- \_\_\_\_\_. **Ciência da Informação criadora de conhecimento - Fundamentos evolutivos e culturais da Ciência da Informação**. **Imprensa da Universidade de Coimbra**. Coimbra, v.1, 2009.
- MARCUSCHI, L. A. **Análise da conversação**. São Paulo: Ática. 1986.
- \_\_\_\_\_. Referência e Cognição. O caso da Anáfora sem Antecedente. **Encontro de Linguística em dezembro de 1998**, UFJF, Juiz de Fora.
- \_\_\_\_\_. **Da fala para a escrita – Atividades de retextualização**, São Paulo, Cortez Editora, 2003
- MARQUES, M. H. D. **Aspectos semânticos da linguagem**. In: **Iniciação à semântica**. 6. ed. Rio de Janeiro: Jorge Zahar, p. 58–118, 2003.
- MARTELETO, R. M. **Cultura, educação e campo social: discursos e práticas de informação**. Tese de Doutorado em Comunicação. Rio de Janeiro ECO/UFRJ, 1992, 389f.
- MARTHA, A. S. **Recuperação de informação em campos de texto livre de prontuários eletrônicos do paciente baseada em semelhança semântica e ortográfica**. Dissertação de Mestrado em Ciências da UFSP. São Paulo, 2005, 91 f..
- MARTINS, R. P. **As trocas informacionais e a produção de conhecimento: Um estudo sobre as interações no ambiente das ONGs**. Dissertação de Mestrado em Ciência da Informação – ECI/UFMG, Belo Horizonte, 2007, 129 p.
- MARTINS, S. R. **Geração automática de textos em plataformas de governo eletrônico: um estudo de caso na Plataforma Lattes**. Dissertação de Mestrado em Engenharia de Produção - UFSC, Florianópolis, 2004. 175 f.
- MAUÁ, D. D. **Modelos de tópicos na classificação automática de resenhas de usuário**. Dissertação de Mestrado em Engenharia de Controle e Automação Mecânica da Escola Politécnica da USP, São Paulo, 2009.

MENDONÇA, E. S. A lingüística e a ciência da informação: estudos de uma interseção. **Ci. Inf.**, Brasília, v.29, n.3, p.50-70, set./dez. 2000.

MIRANDA A.; LEITE C.; SUAIDEN E. A biblioteca híbrida na estratégia da inclusão digital na Biblioteca Nacional de Brasília. **Inclusão Social**, Brasília, v. 3, n. 1, p. 17-23, out. 2007.

MIRANDA, A.; SIMEÃO, E. A conceituação de massa documental e o ciclo de interação entre tecnologia e o registro do conhecimento. **DataGamaZero**, v. 3, n. 4, ago. 2002.

MIRANDA, M. L. C. **Organização e representação do conhecimento: fundamentos teórico-metodológicos na busca e recuperação da informação em ambientes virtuais**. Tese de Doutorado em Ciência da Informação IBICT-UFRJ, Rio de Janeiro, 2005, 354 f.

MOLOSSI, S. **Inserção da biblioteca digital de teses e dissertações no contexto da web semântica: construção e uso da ontologia**. Dissertação de Mestrado em Ciência da Informação da UFSC, Florianópolis, 2008. 228 p.

MONTAGUE, R. **Formal philosophy**. 1 ed. NewHaven: YaleUniversityPress, 1974. 375p.

MOOERS, C. N. Zato coding applied to mechanical organization of knowledge. **American Documentation**, v.2, n.1, p.20-32, 1951.

MORAIS, E. A. M. **Contextualização de documentos em domínios representados por ontologias utilizando mineração de textos**. Dissertação de Mestrado em Ciência da Computação pelo Instituto de Informática da UFGO, Goiás, 2007.

MOREIRA, M. A. BUCHWEITZ, B. **Novas estratégias de ensino e aprendizagem: os mapas conceituais e o vê epistemológico**. Lisboa: Plátano Edições Técnicas. 1993.

MOREIRA, M. A. Mapas conceituais como instrumentos para promover a diferenciação conceitual progressiva e a reconciliação integrativa. **Ciência e Cultura**, 32(4): 474-479. 1980.

MOREIRA, M. A. Modelos Mentais. In: Encontro sobre teoria e pesquisa em ensino de ciência – linguagem, cultura e cognição, 1997, Belo Horizonte. **Anais... Belo Horizonte: FAE/UFMG**, 1997.

\_\_\_\_\_. **Aprendizagem significativa**. Brasília: Editora da UnB. 1999.

\_\_\_\_\_. **Teorias de aprendizagem**. São Paulo: Editora Pedagógica e Universitária. 1999.

MOREIRA, M. P. **Ambiente para geração e manutenção semi-automática de tesouros**. Dissertação de Mestrado em Ciência da Informação – ECI/UFMG, Belo Horizonte, 2005, 197 p.

MORIN, E. **Introdução ao Pensamento Complexo**. 3. ed. Lisboa: Instituto Piaget, 2001. 177p.

NAKANISHI, T; CUNHA, J. B. S; Quality and productivity: the control of software system elements. In: The Word Congress on Optimal Design of Structural System: Structural Optimization, Rio de Janeiro, 1993, Proceedings Rio de Janeiro, Herskovits, J. ed., 1993, v. 2, p. 483-490.

NASCIMENTO, D. M; MARTELETO, R. M. A “Informação Construída” nos meandros dos conceitos da Teoria Social de Pierre Bordieu. **DataGramaZero**, v.5 n.5 out/04

NASCIMENTO, R. A recuperação da informação baseada em ontologias e no perfil do usuário. **Revista do Centro Universitário Planalto do Distrito Federal**. Disponível em <<http://www.uniplandf.edu.br/revista/vol4n2.pdf#page=21>>. Acesso em 2009.

NAVES, M. M. **Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores**. Tese de Doutorado em Ciência da Informação – ECI – UFMG. Belo Horizonte, 2000.



NEVES, M. H. M. A gramática e suas interfaces. **Alfa**, São Paulo, 51 (1): 81-98, 2007.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101: A Guide to Creating Your First Ontology**. Knowledge Systems Laboratory – Stanford University, TR KSL-01-05, 2001. Disponível em: <<http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noymcguinness.doc>>. Acesso 2008.

NUNES, A. M.; FILETO, R. Uma arquitetura para recuperação de informação baseada em semântica e sua aplicação no apoio a jurisprudência. **ERBD**, Caxias do Sul, 2007.

NUNES, L.; TÁLAMO, M. F. G. M. Da filosofia da classificação à classificação bibliográfica. **Biblioteconomia e Ci. Inf.**, Campinas, v.7, n. 1, p. 30-48, jul./dez. 2009– ISSN: 1678-765X, artigo 32.

OLIVEIRA, A.; MARCOS, A.; VAASAN, T. Topic maps na visualização de informação no ensino e treino. In: **ata da primeira conferência da associação portuguesa de sistemas de informação**, Universidade do Minho, out. 2000. Disponível em: <[www.dsi.uminho.pt/publicacoes/ficheiros/artigo\\_topic\\_maps\\_final.pdf](http://www.dsi.uminho.pt/publicacoes/ficheiros/artigo_topic_maps_final.pdf)>. Acesso em: nov. 2006.

OLIVEIRA, E. C. **Autoria de documentos para a Web Semântica: um ambiente de produção de conhecimento baseado em ontologias**. Tese de doutorado em Ciência da Informação da Universidade de Brasília – UnB, Brasília, 2006. 207 p.

OLIVEIRA, M. K. **Vygotsky: aprendizado e desenvolvimento, um processo sócio-histórico**. 4.ed. São Paulo: Scipione, 1997.

\_\_\_\_\_. **Investigações cognitivas: conceitos, linguagem e cultura**. Porto Alegre: Artes Médicas, 1999.

OLIVEIRA, M. M. Metodologia interativa: um processo hermenêutico-dialético. **Interfaces**. Brasil/Canadá, Porto Alegre, V.1, n. 1, 2001. 67 -79.

OLIVEIRA, O. L.; BARANAUSKAS, M. C. C. Interface entendida como um espaço de comunicação. **Technical Report**. Disponível em: <<http://www.unicamp.br/~ihc99/lhc99/AtasIHC99/art7.pdf>>

PEREIRA, E. C.; BUFREM, L. S. Princípios de organização e representação de conceitos em linguagens documentárias. **Bibl. Ci. Inf.**, Florianópolis, n. 20, 2º sem. 2005. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/198/302>> Acesso em 2007.

PERINI, M. A. Princípios de lingüística descritiva. **Introdução ao pensamento gramatical**. São Paulo: Parábola, 2006.

PERROTTI, E. A biblioteca interativa. **CFB**, Brasília, ano 1, n. 1, p. 6, abr. 2001.

POPPER, K. R. **Conhecimento objetivo: uma abordagem revolucionária**. Belo Horizonte: Editora Itatiaia; São Paulo, 1975. 394p.

POZO. J. I. **Teorias Cognitivas da Aprendizagem**. Porto Alegre, Artmed, p.191-209, 1998.

\_\_\_\_\_. Aquisição de conhecimento. Porto Alegre: Artmed, 2005.

RAMALHO, J. C.; HENRIQUES, P. R. XML & XSL: da Teoria à Prática, 1ª edição, FCA, 2002.

RAMALHO, R. A. S.; VIDOTTI S. A. B. G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero*, v.8 n.6 dez/2007 a.4.

RANGANATHAN, S. R. **Philosophy of library classification**. New Delhi: Ejnar Munksgaard, 1951.

\_\_\_\_\_. **The five laws of library science**. Bombay, Asia Publishing House, 1963.

\_\_\_\_\_. **Colon Classification**. Bombay: Asia Publishing House, 1963.

\_\_\_\_\_. **Prolegomena to library classification**. Bombay: Ásia Pub. House, 1967.

REZENDE, S. O. **Sistemas Inteligentes: fundamentos e aplicações**. Barueri, SP: Ed. Manole, 2003.

RIVERA DONOSO, M. A. Directrices para la creación de un programa de preservación digital. **Serie Bibliotecología y Gestión de Información**, n. 43, marzo 2009. Disponível em: <[http://eprints.rclis.org/16052/1/Serie\\_N%C2%B043\\_Preservacion\\_digital.pdf](http://eprints.rclis.org/16052/1/Serie_N%C2%B043_Preservacion_digital.pdf)>. Acesso em 2009.

ROBREDO, J. **Da Ciência da Informação Revisitada aos Sistemas Humanos de Informação**. Brasília: Thesaurus, 2003. 245 p.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas**. 4. ed. Brasília: Reprint, 2005. 409 p.

ROBREDO, J. Organização dos documentos ou organização da informação: uma questão de escolha. **DataGramZero**, v.5 n.1 fev. de 2004. Disponível em: <[http://www.dgz.org.br/fev04/F\\_I\\_art.htm](http://www.dgz.org.br/fev04/F_I_art.htm)>. Acesso em 2009.

SANTOS, P. L. V. A. C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGramZero** v.10 n.6 dez/2009.

SARACEVIC, T. Tecnologia da Informação, Sistemas de Informação e Informação como utilidade pública. **Ci. Inf.**, Rio de Janeiro, 3(1): 57-67, 1992.

\_\_\_\_\_. Interdisciplinarity nature of Information Science. **Ci. Inf.** Brasília, v. 24, n. 1, p. 36-41, 1995.

\_\_\_\_\_. Ciência da Informação: origens, evolução e relações. **Perspect. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, 1996.

\_\_\_\_\_. Modeling Interaction in Information Retrieval (IR): a review and proposal. In: **Proceedings of The American Society for Information Science**, v. 33, p. 3-9, 1996a.

\_\_\_\_\_. Relevance Reconsidered. In: Information science: Integration in Perspectives. **Proceedings of The Second Conference on Conceptions of Library and Information Science (COLIS 2)**, p. 201-218, October, 1996b.

\_\_\_\_\_. The Stratified Model of Information Retrieval Interaction: extension and application. **Proceedings of the American Society for Information Science**, v. 34, p. 313-327, 1997.

SAYÃO, L. F. Modelos teóricos em ciência da informação: abstração e método científico. **Ci. Infor.**, Brasília, v. 30, n. 1, p. 82-91, 2001.

\_\_\_\_\_. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes – URN, PURL, DOI, Handle System, CrossRef e OpenURL. **Transinformação**, Campinas, v. 19, n. 1, p. 65-82, jan./abr. 2007.

SCHREINER, H. B. Considerações históricas acerca do valor das classificações bibliográficas. Trabalho apresentado à Conferência Brasileira de Classificação Bibliográfica, Rio de Janeiro, 12-17 set. 1976. **Anais**, v. 1. Rio de Janeiro, IBICT/ABDF, 1979, p. 190-207.

SENSO, J. A.; PINERO, A. R. El concepto de metadato. Algo más que descripción de recursos electrónicos. **Ci. Inf.**; v.32, n.2 2003. Disponível em: <http://www.ibict.br/cionline/viewarticle.php?id=159>>. Acesso em 2009.

SEVERINO, A. J. **Metodologia do trabalho científico: direções para o trabalho didático-científico na universidade**. 4. ed. rev. São Paulo: Cortez & Moraes, 1980. 159p.

SILVA, A. S. A Semântica de DEIXAR. **Uma Contribuição para a Abordagem Cognitiva em Semântica Lexical**. Tese de Doutorado em Filosofia – Uni.Católica Portuguesa, Braga, 1997.

SILVA, D. L. **Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as ciências da informação e da computação**. Dissertação de Mestrado em Ciência da Informação da UFMG. Belo Horizonte, 2008.

SILVA, D. L.; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ci. Inf.**, Brasília, v. 37, n. 3, p. 60-75, set./dez. 2008.

SILVA, G. B. S. **A utilização de mapas de tópicos na compatibilização de conteúdos hipertextuais semanticamente estruturados**. Dissertação de Mestrado em Ciência da Informação – ECI/UFMG. Belo Horizonte. 2008. 144 p.

SIMÕES, D. (org.) Língua Portuguesa. Ensino, pesquisa, pós-graduação e formação docente. Rio de Janeiro: **Dialogarts**, 2006, v.01. p.390.

SIMÕES, D. (org.) **Mundos Semióticos Possíveis**. Rio de Janeiro : **Dialogarts**, 2008, v.1. p.266.

SOERGEL, D. The Art and Architecture Thesaurus: A critical appraisal. **Visual Resources**. v. 10, n.4, p. 369-400, 1995.

SPERBER, D.; WILSON, D. **Relevance. Communication and Cognition**. Oxford: Blackwell, 1986.

SUONUUTI, H. **Guide to Terminology**. 2. ed. Helsinki : Tekniikan Sanastokeskus, 2001. 42 p.

TARAPANOFF, K.; JÚNIOR, R. H. A.; CORMIER, P. M. J. Sociedade da informação e inteligência em unidades de informação. **Ci. Inf.** v. 29 n. 3, Brasília Set./Dez. 2000.

TOUTAIN, L. M. B. B. Para Entender a Ciência da Informação. Salvador : **EDUFBA**, 2007.

TRILLO, C. D. P. Aplicação para a Recuperação de vídeos indexados por conceitos. **Relatório de Estudos de Tópicos em Ciência da Computação**. São Paulo, 2004.

VAN DER LAAN, R. H.; ALVORCEM, R. M. A terminologia expressa no discurso dos especialistas da área da Ciência da Informação: um estudo de caso. **Biblioteconomia e Documentação**, v. 3, n. 2, jul./dez. 2007.

VICENTINI, L. A. **Gestão em bibliotecas digitais**. In: MARCONDES, Carlos H. et al. (Org). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: UFBA; Brasília: IBICT, 2006. p.113-143.

VIEIRA, R.; LIMA, V. L. **Linguística Computacional: princípios e aplicações**. In: **ENIA**, 2001, Fortaleza.

VOGEL Michely Jabala Mamede. A influência da Jean-Claude Gardin e a linha francesa na evolução do conceito de linguagem documentária. **Perspect. Ci. Inf.**, v.14, número especial, p.80-92, 2009.

VYGOTSKY, L. S. **Pensamento e linguagem**. São Paulo: Martins Fontes, 1987.

\_\_\_\_\_. **A formação social da mente**. São Paulo: Martins Fontes, 1991.

\_\_\_\_\_. **A construção do pensamento e da linguagem**. São Paulo: Martins Fontes, 2001.

WERSIG, G. Information science: the study of postmodern knowledge usage. **Information Processing & Management**, v.29, n.2, p.229-239, 1993.

WILSON, J. R. A mente. **Biblioteca Científica Life**, 5, Rio de Janeiro, José Olympio, 1965..

WURMAN, R. S. **Ansiedade de informação: como transformar informação em compreensão**. São Paulo: Cultura, 1991.

## ANEXO I

O quadro 11 mostra a estrutura hierárquica extraída a partir da navegação no mapa de tópicos resultante da fusão apresentada por Silva (2008). Os elementos para os quais foi detectada a poli-hierarquia estão destacados em negrito na própria estrutura.

**QUADRO 11 – Estrutura hierárquica obtida a partir da navegação no mapa de tópicos resultante da fusão automatizada de duas teses no MHTX (continua)**

### Entidade / Personalidade / Princípio / Objeto material / Objeto imaterial

Autores

Disciplinas

- Ciência da Computação
- Ciência da Informação**
- Educação
- Linguística**
- Organização da Informação
- Psicologia Cognitiva**
- Terminologia**

Documento

(Pela forma)

- Tese
- Bibliografia

(Pelo formato)

- Texto**
- (Pela estrutura)
  - Macroestrutura
  - Microestrutura**
  - Superestrutura
- (Pelos seus componentes)
  - Microestrutura

(Por tipo)

- Narrativo**
- Dissertativo
- Descritivo

(Pela natureza do texto)

- Informativo
- Primário
- Secundário
- Narrativo**
- Terciário
- Hipertexto

(Pelo tipo de tecnologia empregada)

- Texto Linear / Documento Impresso
- Documento Automatizado/Documento Eletrônico/Texto Eletrônico
- Sistema Hipertextual / Sistema de Hipertexto
- (Por seus componentes)
  - Link / Elo / Ligação
    - Link Ativo
    - Link Inativo
  - Nós / Nodos

(Princípios)

**(continuação)**

Princípio da metamorfose  
 Princípio de heterogeneidade  
 Princípio de multiplicidade/encaixe  
 Princípio de exterioridade  
 Princípio de tipologia  
 Princípio de mobilidade dos centros  
 Modularidade

(Por tipos)

Hipermídia

**Idéia / Conceito / Pensamento**

Modelo de Representação

Modelo Conceitual

MHTX

Produtor e Necessitador de Informação ou Conhecimento

Autor

Comunidade Acadêmica

Pesquisador

Profissionais

**Profissional da Informação**

**Web Designer / Arquiteto da Informação**

**Bibliotecário**

**(Pela natureza do trabalho)**

**Indexador**

**(Pela experiência)**

**Indexador Experiente**

**Indexador Pouco Experiente**

**Indexador Novato**

**(Pelo grau de conhecimento)**

**Especialização**

**Prática**

**Indexador (...)**

**Classificador**

**Profissional da Informação (...)**

**Sistemas de Armazenamento de Informação**

Base de Dados

Catálogo

Teorias

Teoria da Análise Facetada / Teoria da Classificação Facetada

Categoria

**Estado/Atividade/Ação/Energia/Operação/Processo**

**Qualidade/Atributo/Matéria/Propried./Qtde./Relação**

**Entidade/Personalidade/Princípio/Obj.**

**material/Obj. imaterial (...)**

Espaço

**Tempo / Dimensão / Espaço / Posição (...)**

Faceta

Plano da Idéias

Plano Notacional

Plano Verbal

Subfacetas

(continuação)

**Texto (...)**Ferramentas / Instrumentos  
(do processo de indexação)

Índice

(unidade de representação)

Termo

Linguagem de Marcação

Linguagem HTML

Linguagem SGML

Linguagem XML

(unidade de conhecimento)

**Idéia / Conceito / Pensamento**

(de representação do conceito)

Mapa conceitual

**Estado / Atividade / Ação / Energia / Operação / Processo**

Inferência

Interpretação

Categorização

Modelagem Conceitual

Recuperação da Informação

(Por tipo)

Recuperação Automática

(Por forma de acesso)

Folheio / Browsing

Navegação

Processos Mentais

Abstração

Generalização

Análise

Raciocínio

Indutivo

Dedutivo

Interativo

Bottom-Up

Top-Down

Indexação

(Por seus estágios)

**Análise de Assunto**

Extração de Conceitos

Formação de Assunto

(Por seus métodos)

Método de dissecação

Método de superposição/ Sobreposição

Método de desnudação

Método de laminação

Tradução para Linguagem de Indexação

Tradução

Leitura do Texto

(Por etapas)

**Análise de Assunto (...)****Qualidade / Atributo / Matéria / Propriedade / Quantidade / Relação**

Aceitabilidade

Terminologia

Coesão

(continuação)

**Subjetividade**

Coerência  
 Consistência  
 Intertextualidade  
 Intencionalidade  
 Situacionalidade  
 Informatividade  
 Relevância  
 Assimilação  
 Atitude  
 Estratégia Cognitiva  
     Metacognitiva  
 Compreensão  
 Percepção  
 Conhecimento Prévio  
 Memória  
 (Por tipo)  
     Long Term Memory / LTM  
     Short Term memory / STM  
 Atinência  
 (Por característica)  
     Intrínseca  
     Extrínseca  
 Interdisciplinaridade  
 Sintaxe  
 Semântica  
     Significado

(Propriedade do processo de indexação)

**Subjetividade**

Especificidade  
 (Fatores que interferem na elaboração de hipertexto)  
     Fragmentação Textual  
     Desorientação do Usuário  
 (Fatores constitutivos da unidade textual)  
     Coesão Textual

Disciplina

Ciências  
     Ciências Formais  
     Ciências Reais  
         Ciências Reais Humanas  
             Ciências Humanas  
                 Ciências Cognitivas  
                     **Psicologia Cognitiva**  
                     **Linguística**  
                         Lógica  
                     Ciências Sociais  
                         **Ciência da Informação**  
                             Sociologia  
             Ciências Reais Naturais  
                 Ciências Exatas  
                 Ciências Biológicas  
                     Botânica

**Tempo / Dimensão / Espaço / Posição**

Web  
 Web Semântica

---

**Fonte:** Silva (2008).