Gercina Ângela Borém de O. Lima

# MAPA HIPERTEXTUAL (MHTX):

## Um modelo para organização hipertextual de documentos

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da UFMG como requisito para obtenção do Grau de Doutor em Ciência da Informação

Linha de pesquisa: Organização e Uso da Informação

Orientadora: Profª Dra.Lídia Alvarenga
Co-Orientador: Dr. Hélio Kuramoto

Belo Horizonte
Escola de Ciência da Informação
Universidade Federal de Minas Gerais
2004

*E quando tudo parecia a esmo /*
*E nesses descaminhos me perdia /*
*Encontrei muitas vezes a mim mesmo...*

*Eu temo é uma traição do instinto /*
*Que me liberte, por acaso, um dia /*
*Deste velho e encantado labirinto*

(QUINTANA, Mário. Astrologia. *Baú de Espantos*, 1986)

Dedico este trabalho ao sol, à lua e a estrela que me cercam:
tríade que faz renascer todos os dias
a luminosidade, o mimo e o brilho na minha vida.

À minha mãe, eterno amor
Ao Fausto, amor eterno
À Clara, amor eterno amor

# AGRADECIMENTOS

A Deus, por ter me iluminado e me abençoado para que este sonho nascesse e fosse adiante...

À minha mãe, que embora ausente, me abençoava todos os dias neste período de crescimento intelectual;

Ao Fausto, meu companheiro, amor e amigo, por ter sido meu porto seguro que pacientemente permaneceu ao meu lado durante este período e, também, pela sua valiosa contribuição como experiente pesquisador.

À minha filha Clara, que chegou junto com este projeto e, mesmo pequenina, soube esperar algumas vezes pela sua mamãe.

À minha orientadora Profa. Dra. Lídia Alvarenga, pelo apoio, incentivo, idéias, paciência e cuidados na elaboração desta tese.

Ao meu co-orientador Prof. Dr. Hélio Kuramoto, pela disponibilidade e paciência na sua orientação, pelas suas idéias e sugestões tecnológicas na implementação do protótipo.

Ao Prof. Dr. K S Raghavan, da University of Madras, Índia, pela sua contribuição no ensino da Teoria da Análise Facetada e na orientação da modelagem facetada da tese.

À Profa Dra. Maria Luiza Almeida Campos pelas suas idéias, críticas e sugestões na minha qualificação e por ter acreditado em meu projeto e fazer parte do meu referencial teórico.

Ao Prof. Dr. Ricardo Barbosa por ter me despertado pela integração entre hipertexto e organização do conhecimento.

À minha amiga e colega Madalena Martins Lopes Naves, pelo seu estímulo, amizade e confiança e pela sua presteza em disponibilizar sua tese para ser meu objeto de estudo.

À Profa. Isis Paim, pelo seu apoio, por ter acreditado no meu projeto desde o início e, além de tudo, ter sido para mim uma referência afetuosa e acadêmica todo este período.

A Carmen Pinheiro, por nortear minha vida sempre com conforto, segurança e amor.

A Elza, que foi a dona-de-casa e carinhosamente mãe, quando me ausentava do lar.

A Ana Maria e Theldo, Adriana e Dulce, colegas de doutorado, pelas trocas de experiências e amizade.

Aos meus familiares que sempre tiveram uma palavra de apoio e ternura.

Aos professores e funcionários da ECI que de alguma maneira contribuíram para a realização deste projeto.

Aos meus amigos que compreenderam minha ausência.

**RESUMO**

Este estudo visa a construção de um modelo estruturado semanticamente para auxiliar a organização e representação do conhecimento humano estruturado em hipertextos, baseado nas teorias da análise facetada e do mapa conceitual. O segundo passo neste estudo é a aplicação do modelo semântico para criar um protótipo chamado MAPA HIPERTEXTUAL (**MHTX**): um modelo para organização hipertextual de documentos que deverá ser utilizado para implementar a BTDECI – UFMG (Biblioteca de Teses e Dissertações do Programa de Pós-Graduação da Escola de Ciência da Informação da UFMG). Hipertextos, frutos da tecnologia informacional desenvolvidos nas duas últimas décadas, têm causado uma revolução nos conceitos e procedimentos no tratamento da informação, desde as novas técnicas de entrada de dados até as estruturas de armazenamento, passando pelos processos de recuperação. Entretanto, o hipertexto ainda apresenta problemas devido à grande variabilidade do grau de eficiência no processo de recuperação. Esses problemas estão comumente relacionados com a falta de critérios durante a concepção e os primeiros passos de sua construção. Os autores interessados no desenvolvimento semântico do hipertexto (DUNCAN, 1989; SPITERI, 1997; ELLIS, 2000; CAMPOS, 2001; TUDHOPE, 2002) concordam que a criação de hiperdocumentos mais eficientes pode ser mais facilmente atingida com a aplicação de princípios como (1) análise de assunto, (2) representação do conhecimento, (3) interrelacionamento entre os *links* e (4) conhecimento das necessidades do usuário. Este estudo está voltado para a aplicação de dois princípios, observados sob a luz de dois referenciais teóricos: a Análise Facetada, e o Mapa Conceitual. Foi escolhida uma única unidade de informação, que é uma tese de doutorado sobre o próprio tema da análise assunto (NAVES, 2000), para construir um modelo conceitual anteriormente à implementação tecnológica do seu protótipo. A análise facetada foi utilizada para categorizar os termos indexados, estabelecendo a relação semântica de conceitos e as relações entre eles expressas nos *links*. Finalmente, o mapa conceitual foi aplicado à representação do conhecimento. Para implementar esses processos no protótipo, dois softwares foram utilizados: (1) *Greenstone 2.50*, para criar a base de dados e criar uma estrutura navegacional hierárquica do sumário da tese; (2) *Inxight Star Tree Studio 3.0*, para criar um mapa conceitual tri-dimensional.

## PALAVRAS-CHAVE

Organização do conhecimento, hipertexto, Análise Facetada, Mapa Conceitual, Modelagem Conceitual.

## *ABSTRACT*

This study aims at the construction of a semantically structured conceptual model to help the organization and representation of human knowledge in hypertextual systems, based on the Theory of Facet Analysis and Conceptual Map. A second step in this study is the application of the semantic model to create a prototype called Hypertext Map (*Mapa Hipertextual* - ***MHTX***) which will be used to implement the BTDECI – UFMG (Thesis and Dissertation Library of UFMG´s School of Information Science). Hypertext systems, offsprings of the informational technology developed in the last two decades, have caused a revolution in information treatment concepts and procedures, ranging from new ways of information input, storage structures and retrieval processes. However, hypertext still poses some problems due to the great variability in the degree of efficiency in the retrieval process. These problems are often related to lack of criteria during the conception and initial steps of their construction. Authors interested on the the semantic development of hypertext (DUNCAN, 1989; SPITERI, 1997; ELLIS, 2000; CAMPOS, 2001; TUDHOPE, 2002) agree that the creation of more efficient hyperdocuments can be achieved with the application of principles such as (1) subject analysis, (2) knowledge representation, (3) interrelated links and (4) knowledge of the user´s needs. This study focuses on the application of the two principles, viewed under two specific theoretical references: facet analysis and conceptual map. It was chosen a single knowledge unit, namely a doctoral thesis on the own area of subject analysis (NAVES, 2000), to construct a conceptual model prior to technological implementation of its prototype. Facet analysis was used to categorize the index terms, established the relationship among them expressed by the links. Finally, conceptual mapping was applied to knowledge representation. In order to implement these processes into the prototype, two softwares were used: (1) *Greenstone 2.50* to create the database and to create a hierachical navegational structure map of the thesis´s summary. (2) *Inxight Star Tree Studio 3.0*, to create a three-dimensional conceptual map.



## *KEYWORDS*

knowledge organization, hypertext, faceted analysis, conceptual map, conceptual modeling.

# LISTAS DE FIGURAS

FIGURA 1 - Correspondência dos nós (A, B) e *links* (b) entre a tela do computador e a base de dados hipertextual. . . . . 25
FIGURA 2 - Separação de parte da realidade do Universo do Discurso. . . . . 39
FIGURA 3 - Análise da abstração, determinação dos conceitos e sua estruturação em domínios de conhecimento pelo método *"bottom-up"* . . . . . 40
FIGURA 4 - Síntese da abstração, com seus nós (conceitos) e relacionamentos em rede. . . . . 40
FIGURA 5 - Camadas do *Dexter Reference Model*. . . . . 43
FIGURA 6 - Principais modelos de *design* de hipertextos e suas características 51
FIGURA 7 - Estágios do modelo de processamento cognitivo da informação. . . 66
FIGURA 8 - O sistema de comunicação cognitivo para a CI e a RI. . . . . 68
FIGURA 9 - Compartimento do cérebro: formação da idéias. . . . . 80
FIGURA 10 - Exemplo do Método de Dissecação. . . . . 82
FIGURA 11 - Exemplo do Método da Desnudação. . . . . 83
FIGURA 12 - Estrutura dos mapas conceituais. . . . . 94
FIGURA 13 - Estrutura em teia. . . . . 96
FIGURA 14 - Estrutura hierárquica. . . . . 96
FIGURA 15 - Estrutura *flowchart*. . . . . 97
FIGURA 16 - Estrutura conceitual. . . . . 97
FIGURA 17 - Mapa conceitual paisagem. . . . . 98
FIGURA 18 - Mapa conceitual 3-D. . . . . 98
FIGURA 19 - Mapa conceitual mandala. . . . . 99
FIGURA 20 - Mapa com navegação hiperbólica. . . . . 102
FIGURA 21 - Gravura *Céu e inferno* (1960) da série *Circle Limit-IV* de Escher. . . 102
FIGURA 22 - Exemplo de desnudação. . . . . 112
FIGURA 23 - Lista alfabética de conceitos/termos relevantes da tese. . . . . 113
FIGURA 24 - Aplicação do Princípio de Diferenciação no MHTX. . . . . 115
FIGURA 25 - Aplicação do Princípio de Relevância no MHTX. . . . . 115
FIGURA 26 - Aplicação do Princípio de Homogeneidade no MHTX. . . . . 116
FIGURA 27 - Categorização completa da tese. . . . . 119
FIGURA 28 - Dimensões horizontal (renques) e vertical (cadeias) do MC da tese. . . . . 123
FIGURA 29 - Subdivisões de renque e cadeia na estrutura hipertextual do MC. . 123
FIGURA 30 - Relação de cores para os níveis do MC. . . . . 124
FIGURA 31 - Associação entre níveis e cores na raiz principal do MC. . . . . 125
FIGURA 32 - Tabela de organização dos níveis, cores, *links* e seus

direcionamentos no MHTX  Tabela da organização das cores. . . . 126
FIGURA 33 - Criação de *URL* para o *link* “conceito/idéia” com comandos de nome e endereço . . . . . 127
FIGURA 34 - Navegação dentro do texto da tese acessado através do MC. . . . . 128
FIGURA 35 - Primeira tela do navegador hiperbólico do MC . . . . . 129
FIGURA 36 - Primeira tela do sumário expandido. . . . . 131
FIGURA 37 - Expansão do sumário expandido. . . . . 132
FIGURA 38 - Tela inicial do *Greenstone Digital Library*. . . . . 135
FIGURA 39 - Arquivo contendo os metadados relativos à tese como um todo. . . 137
FIGURA 40 - Tela inicial do *Star Tree Studio* na criação do MC. . . . . 140
FIGURA 41 - Estrutura da base de dados de textos digitais completos de teses e dissertações da Escola de Ciência da Informação da UFMG. . . . . 142
FIGURA 42 - Arquivo com os parâmetros de configuração da base de dados para o protótipo MHTX. . . . . 143
FIGURA 43 - Subdivisões de seções dentro do capítulo 2 da tese em XML. . . . . 144
FIGURA 44 - Pesquisa direta em capítulos ou seções por palavras-chave ou por título. . . . . 145
FIGURA 45 - Resultado de pesquisa direta com o temo “indexação”. . . . . 146
FIGURA 46 - Recuperação do conteúdo sobre o termo “indexação” . . . . . 146
FIGURA 47 - Fluxograma de busca direta e navegação (hierárquica e hiperbólica) na base hipertextual do  MHTX. . . . . 151
FIGURA 48 - Marcação de três conceitos próximos com o mesmo *link* na linguagem *html*. . . . . 153
FIGURA 49 - Exemplos de *tag* inicial e *tag* final. . . . . 153
FIGURA 50 - Exemplos de atributo e valor em uma *tag* inicial. . . . . 154
FIGURA 51 - Excerto da tese ”Fatores interferentes em análise de assunto” como documento *HTML*. . . . . 154
FIGURA 52 - Exemplo da marcação de metadado no título da tese. . . . . 157
FIGURA 53 - Declaração de tipo de documento em XML no documento raiz. . . . 157
FIGURA 54 - Criação de metadados para “título” e “assunto” da tese com valor semântivo em *XML*. . . . . 158
FIGURA 55 - Exemplo de marcação *XML* da tese Fatores interferentes em análise de assunto.  . . . . . 158

# LISTAS DE ABREVIATURAS E SÍMBOLOS

| | |
|---|---|
| ABNT | Associação Brasileira de Normas Técnicas |
| ANCIB | Associação Nacional de Pesquisa em Ciência da Informação e Biblioteconomia |
| BTDECI | Biblioteca de Teses e Dissertações do Programa de Pós-Graduação da Escola de Ciência da Informação da UFMG |
| *CD-ROM* | *Compact Disk Read Only Memory* |
| *CPU* | *Central Processing Unit* |
| *CRG* | *Classification Research Group* |
| *DAML* | *DARPA Agent Markup Language* |
| DERA | Agência de Pesquisa e Avaliação de Defesa |
| *DIM* | *Distributed Information Management Programme* |
| *DTD* | *Document Type Definition* |
| *EORM* | *Enhanced Object-Relationship Model* |
| *EPSRC* | *The Engineering and Physical Sciences Research Council* |
| *HTML* | *Hypertext Markup Language* |
| *HTTP* | *Hypertext Transfer Protocol* |
| *LAN* | *Local Area Network* |
| *LTM* | *Long Term Memory* |
| MC | Mapa Conceitual |
| MHTX | Mapa Hipertextual |
| OO | Orientação a Objetos |
| *OOHDM* | *Object Oriented Hypermedia Design Model* |
| *OIL* | *Ontology Interchange Language* |
| *PMEST* | *Personality, Material, Energy, Space, Time* |
| *RMM* | *Relationship Management Methodology* |
| SE | Sumário Expandido |
| *SGML* | *Standard Generalized Markup Language* |
| *SOHDM* | *Scenario-based object-oriented hypermedia design methodology* |
| *STM* | *Short Term Memory* |
| TAF | Teoria da Análise Facetada |
| UFMG | Universidade Federal de Minas Gerais |

| | |
|---|---|
| *URL* | *Uniform Resource Locator* |
| *WWW* | *World Wide Web* |
| *WSDM* | *Web Site Design Method* |
| *XFML* | *eXchangeable Facet Metadata Language* |
| *XML* | *eXtensible Markup Language* |

# SUMÁRIO

Epígrafe e dedicatória. . . . . ii
Agradecimentos. . . . . iii
Resumo e palavras-chave. . . . . iv
*Abstract* e *keywords*. . . . . v
Listas de figuras. . . . . vi
Lista de abreviaturas e símbolos. . . . . viii

**1 INTRODUÇÃO**. . . . . 1
1.1 Objetivos. . . . . 7
1.2 Método e materiais. . . . . 8
1.3 Apresentação da tese. . . . . 11

**2 FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS**
**2.1 Da tradição oral ao alfabeto**. . . . . 13
**2.2 Do texto ao hipertexto**. . . . . 15
2.2.1 Aspectos cognitivos no hipertexto. . . . . 29
2.2.2 Metodologias para organização de sistemas de hipertexto . . . . . 36
2.2.3 Modelos para estruturação de sistemas de hipertexto . . . . . 41
2.2.4 Conversão de textos em hipertextos. . . . . 53
2.2.4.1 A questão da fragmentação. . . . . 57
2.2.4.2 A alternativa da modularidade. . . . . 58
2.2.4.3 Linguagens de marcação. . . . . 61
**2.3 Da organização à recuperação da informação.** . . . . . 62
2.3.1 Aspectos cognitivos na organização do conhecimento. . . . . 63
2.3.1.1 Categorização. . . . . 64
2.3.1.2 Indexação. . . . . 65
2.3.1.3 Recuperação da informação. . . . . 66
2.3.1.4 A interação homem-máquina. . . . . 70
2.3.2 A Teoria da Análise Facetada (TAF): revisão de literatura e fundamentos.. 72
2.3.2.1 O Modelo de Ranganathan. . . . . 78
2.2.2.2 As contribuições do *CRG*. . . . . 84
2.3.2.3 O Modelo Simplificado de Louise Spiteri. . . . . 86
2.3.3 Mapa conceitual. . . . . 91
2.3.3.1 Mapa conceitual e hipertexto. . . . . 100
2.3.3.2 Aspectos cognitivos no mapa conceitual. . . . . 104

**3 A CONSTRUÇÃO E IMPLEMENTAÇÃO DO MODELO HIPERTEXTUAL**
**3.1 Introdução**. . . . . 106
**3.2 Mapa Hipertextual - MHTX**. . . . . 109
3.2.1 A estrutura conceitual do MHTX . . . . . 109
3.2.1.1 A análise facetada. . . . . 109
3.2.1.2 O Mapa Conceitual (MC). . . . . 121
3.2.1.2.1 A criação dos *links*. . . . . 124
3.2.1.2.2 A navegação no MC . . . . . 127
3.2.2. Estruturação formal do MHTX. . . . . 129

3.2.2.1 O Sumário Expandido (SE). . . . . 130
**3.3 A implementação tecnológica: construindo o protótipo** . . . . . 132
3.3.1 Os *softwares* utilizados para criação do MHTX. . . . . 133
3.3.1.1 Escolha do *software* para criação do banco de dados. . . . . 133
3.3.1.2 Escolha do *software* para criação do MC e do SE. . . . . 137
3.3.2 Criação da base de dados de coleções de teses e dissertações. . . . . 141
3.3.3 Navegação na base de dados do MHTX . . . . . 144
3.3.4. Da tese à hipertese. . . . . 147

**4 CONSIDERAÇÕES FINAIS**. . . . . 160

**5 REFERÊNCIAS** . . . . . 167

**6 ANEXOS**
Anexo 1 - Glossário. . . . . 177
Anexo 2 - Sumário Expandido (SE). . . . . 186
Anexo 3 - Lista de *links* do MHTX: SE e MC com respectivas URLs. . . . . 192
Anexo 4 - Modelo Simplificado para Análise Facetada de Spiteri. . . . . 196
Anexo 5 - Cânones da Teoria da Análise Facetada de Ranganathan. . . . . 197
Anexo 6 - Princípios da Teoria da Análise Facetada do *CRG*. . . . . 199
Anexo 7 – CD ROM com o MHTX. . . . . capa interna posterior

# 1 INTRODUÇÃO

Este capítulo apresenta a justificativa, o problema, os objetivos, os pressupostos, o método, materiais e a apresentação das partes da tese.

Desde o surgimento dos documentos hipertextuais, na última década, a literatura especializada vem enfatizando que um dos principais problemas relativos ao uso dessa nova ferramenta é o direcionamento ineficaz do leitor, no momento da consulta do documento. Essa situação é comumente causada pela falta de critérios claros e consistentes no processo de organização conceitual do hipertexto, especialmente, nas fases de escolha e estruturação das informações relevantes e na criação de *links* a partir das relações existentes entre essas.

Um problema comum em sistemas de hipertextos ou hipermídia[1], relatado na literatura por diversos autores (COKLIN, 1987; EDWARDS, HARDMAN, 1989 p.105; FOSS, 1989 p.407 ; AIGRAIN, LONGUEVILLE, 1992 p.518; ABOUD et al, 1993 p.387; DALAL, QUIBLE, WYATT, 2000, p.608), é a desorientação do usuário na navegação. Freqüentemente, os usuários se esquecem onde estão, como e porque chegaram a determinada parte do texto. Nem sempre o usuário encontra conscientemente a informação desejada, seja em grandes repositórios de texto, como a *Web*, seja em pequenos hiperdocumentos:

> *"É muito fácil criar alguma coisa parecida com uma grande teia de aranha na qual, embora o autor possa gostar disso, o leitor fica rapidamente perdido. Na prática, até os hiperdocumentos tão curtos quanto um capítulo de livro são confusos, quando criados ao acaso. É comum que uma pessoa, que já usou um hiperdocumento diversas vezes, seja surpreendida pela descoberta de algum item útil, que já a iludira anteriormente. Ela fica imaginando que outras pérolas de sabedoria estão ocultas na confusão"* (MARTIN, 1992, p.40).

O trabalho de organização semântica do hipertexto se constitui requisito essencial, visando à criação do sistema de hipertexto, permitindo que, em uma etapa posterior, os usuários ou as ferramentas de busca alcancem maior eficácia no momento do acesso à informação. Uma trilha de navegação desorganizada, decorrente do *design* hipertextual

[1] Hipermídia é uma extensão do paradigma do hipertexto incluindo *links* com diferentes tipos de informações expressos por meio de diversos tipos de mídias: vídeo, áudio, textos, gráficos. Neste trabalho os termos “hipertexto” e “hipermídia”serão utilizados indistintamente.

conceitualmente deficiente de um documento tenderia dificultar a busca e o folheio (*browsing*), por parte do usuário, da mesma forma que dificultaria, ou mesmo, impediria a recuperação por parte das máquinas de busca (*search engines*).

Em 1998, ELLIS, FORD e FURNER (p.44) apontavam a ausência de

> *"... proximidade entre o designer ou o criador (que poderia ser qualquer pessoa) e o usuário em potencial (que poderia ser qualquer um ou todos). . .* [concluindo que] *quanto mais distantes as características e necessidades de informação dos usuários estão do tipo de usuário concebido ou servido pelos que criam ou indexam uma base de dados, mais provável será a existência de problemas no acesso de informações relevantes pelos usuários daquela base de dados."*[2]

Essa preocupação é também compartilhada por ROSENFELD e MORVILLE (1998, p.22), quando suscitam perguntas, tais como, porque existe um alto número de *Web sites* com navegação problemática e porque há grande ineficiência dos *web designers* em prover a recuperação da informação mais facilitada. Ao salientar que essas questões estão relacionadas com o desafio de se organizarem as informações para posterior disponibilização, destacam ainda que ".... *context is king" ,* ou seja, no *design* de sistemas de navegação, a determinação de *links* conceituais dentro do assunto é tarefa de fundamental importância.

A frustração dos usuários na tentativa de fazer uma pesquisa na *Web* é descrita também por BROCOLLI (1998, p.1), ao relatar a grande variabilidade e inconsistência na mesma busca, em que alguns usuários não encontram nada e outros encontram muito mais do que o esperado, porém poucos dos dados recuperados são relevantes.

Na atual fase de transição, os hipertextos vêm sendo criados a partir de textos convencionais, sem qualquer preocupação com a estruturação semântica explícita, assim como com a organização de bibliotecas digitais, ou mesmo dos sistemas de hipertextos é feita, na maioria das vezes, por leigos ou por profissionais da informação

---

[2] *"…There is no closeness at all between designer or creator (which could be anyone) and potential user (which could be anyone or everyone) …the more distant users are, in characteristics and information needs from the types of user conceived of and catered for by those creating or indexing a database, the more likely there are to be problems in accessing relevant information by users from that database.*

que, tais como os usuários potenciais dos documentos, encontram-se distanciados do autor, criador do documento original, tornando ineficaz, moroso e economicamente inviável, na maioria dos casos, o trabalho de organização em ambos contextos.

No âmbito da ciência da informação, antes mesmo do surgimento dos sistemas de hipertextos, o tratamento da informação já se via afetado por esse distanciamento entre o autor do documento, o profissional da informação e o usuário. Pode-se dizer que essa dificuldade, em nossos dias, estende-se ao *designer* do hipertexto como um intermediário, quando o documento já tiver sido criado em forma de texto linear.

A organização semântica de determinada área de conhecimento, para fins de criação de um sistema de hipertextos, teria como produto uma rede semântica ou conceitual e demandaria a implementação de um sistema de conceitos integrados entre si. De maneira ideal, esse processo deveria, de preferência, preceder a fase de implementação do sistema, ocorrendo na fase de construção do hiperdocumento.

Tradicionalmente, a ciência da informação trabalha com documentos impressos, cujas estruturas já estejam definidas. Com a tecnologia, principalmente com a leitura não linear das informações nos sistemas de hipertexto, ganhou destaque a figura do autor de hiperdocumentos, chamado de *Web designer*. Na maioria das vezes, o hipertexto é planejado e estruturado ao acaso, sem considerar a estruturação semântica do documento ou as necessidades dos usuários potenciais. A variedade de conteúdo no mundo atual e, seus interelacionamentos exige um enorme esforço, quando se propõe a construção de uma estrutura analítica de uma página *Web*[3] ou a organização de um CD-ROM, sendo comum a falta de estrutura organizacional que garanta um acesso eficaz à informação.

Nos sistemas de representação e recuperação, um problema recorrente é a identificação de conceitos que reflitam a estrutura do campo de conhecimento enfocado e o entendimento dos mesmos pelo usuário. Quando o hipertexto é criado, seu processo de autoria é, em última análise, um processo de classificação. O autor planeja

[3] O termo *Web* tem sido utilizado como abreviatura de *World Wide Web*, que é também abreviado como *WWW* e, mais raramente, como *W3*.

a estrutura global do hipertexto, seleciona símbolos apropriados (i.e., palavras, ícones) e cria *links* eletrônicos para representá-los. Conceitualmente, esse processo apresenta grande similaridade com o processo de classificação de documentos, no qual o classificador determina o conteúdo, seleciona termos apropriados e cria pontos de acesso. Como na recuperação de documentos convencionais, a eficiência dos sistemas de recuperação em hipertextos depende da qualidade dos *links* inseridos na base de dados. Em princípio, grupos de *links* construídos aleatoriamente seriam pouco eficientes na busca de informação em sistemas hipertextuais.

Segundo SCHIPER[4], citado por LUCAS (1995, p.70), o hipertexto está estruturado na lógica do conhecimento e em princípios de raciocínio psicológico. Porém, permanecem questões sobre a dinâmica do significado dos textos, no qual ocorrem diferentes leituras e ligações em interpretação, compreensão, utilização e aplicabilidade da informação no sistema de hipertexto.

As pesquisas no campo da recuperação automática de documentos na *Web* não privilegiam o momento da estruturação, mas, a recuperação. Partem do pressuposto de que os hipertextos sejam construídos de forma livre e que, a partir daí, devam ser criados mecanismos automáticos para a sua recuperação.

Autores como BORDONARO e LYNCH (1995); COURTOIS et al. (1995), MARCHIONINI, BARLOW e HILL (1994) e WEINBERG (1996) relatam que muitos dos *softwares* usados para indexação/classificação, principalmente na *WWW*, resumem-se à aplicação de algoritmos de indexação automática. Nesse processo, são extraídas palavras de títulos ou de textos completos. Algumas vezes, atribuem-se pesos às palavras, de acordo com suas posições, analisadas como significativas. Esses autores concluem que a classificação automática de algoritmos, utilizada para fazer essa indexação tem apresentado uma performance inconsistente. OUTING (1998) e BROCOLLI (1998) afirmam que, embora alguns *softwares* tenham sido desenvolvidos para fazer o trabalho do indexador, o elemento humano será sempre imprescindível. De

[4] SCHIPER, J. Discourse in hypertext systems, knowledge organizing structures and information strategies: some philosophical remarks. In: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION CONFERENCE. *Proceedings....* Copenhagen, 1994. p.142-6.

fato, parece que a criatividade decisória do indexador bem treinado é insubstituível, pois, a maioria dos índices freqüentemente requer entradas de palavras que não aparecem no documento nem no *site*. Não raro, essas palavras-chave possuem poder de síntese que nenhuma palavra constante do texto apresenta. O autor exemplifica com uma pesquisa em um *site* sobre animais. Muitos hiperdocumentos que incluem o assunto "comida" ou "saúde canina", dentre outros, poderiam estar categorizados sob o cabeçalho "nutrição", termo que poderia não aparecer no corpo do documento. Assim, situações como essa, em que o computador sozinho não é competente para lidar com conceitos implícitos, carecem da interferência humana, exigindo maior aprofundamento, poder de decisão e criatividade.

Apesar da popularização de produtos decorrentes da utilização de documentos na forma hipertextual, - como *sites*, CD ROMs, bibliotecas digitais etc. -, o perfil dos construtores de hiperdocumentos ainda não é regulamentado, sendo na realidade desconhecido. Aparentemente, a maioria desses profissionais não é composta de indivíduos provenientes da área de ciência da informação. Ao contrário, muitos parecem ser leigos que aprendem a manipular *softwares* de construção de hipertexto, ou ainda, são profissionais da área de ciência da computação que não dispõem de informação específica para desenvolver uma estruturação semântica da informação.

Tanto a revisão de literatura desta tese quanto tendências apontadas por fontes do seu referencial teórico permitem afirmar que existem problemas organizacionais bastante recorrentes na elaboração de hiperdocumentos. Os modelos atuais de construção ainda se mostram deficientes, pois carecem de ferramentas que possibilitem apresentar apropriadamente as relações entre conceitos de naturezas diversas, necessários à sua estrutura.

Segundo LANCASTER (1994), *" em trinta anos, passamos da virtual inexistência de base de dados em formato eletrônico à realidade de milhares delas".* Hoje, a informação demandada pelo usuário poderá estar em qualquer formato, e ser localizada em qualquer parte do mundo. LUCAS (1995, p.69) lembra, entretanto, que as novas tecnologias de informação tornam-se inúteis sem os meios de localizar, filtrar, organizar e resumir seus produtos. É nesse quadro que o hipertexto, com suas características

especiais de armazenamento e recuperação da informação, tem se revelado como solução tecnológica com grandes possibilidades de atender à demanda por informação.

Nesse sentido pode-se pressupor que se a informação for disposta de maneira organizada e lógica, será mais provável e mais fácil localizar e recuperar o que se procura no hipertexto. O ideal seria que a organização semântica do texto, que pode ser denominada de estrutura conceitual, refletisse as relações semânticas entre os assuntos que o compõem.

Fica clara a importância da modelagem conceitual, no momento da construção do hiperdocumento, propiciando uma recuperação mais eficiente. Entretanto, dada a enorme quantidade de documentos já disponibilizados na *Web*, sua dispersão e a pluralidade dos produtos e produtores, a construção regular e precisa de hipertextos estruturados semanticamente, nesse ambiente, seria uma utopia.

Um segmento significativo da literatura aponta para a relevância das características analítico-sintéticas da teoria da análise facetada (TAF) de *Ranganathan* como possível solução na modelagem conceitual de sistemas de hipertexto. Justifica-se, portanto, a importância de estudá-la com vistas à possível criação da modelagem conceitual como instrumento componente de um protótipo que visa auxiliar autores de hiperdocumentos e profissionais da informação na transformação do texto linear em hipertexto.

A escolha de um documento acadêmico científico, característico da conclusão de cursos de pós-graduação *stricto sensu* (tese ou dissertação), como unidade de trabalho, justifica-se por se tratar de um tipo de documento com informações muito específicas e pouco divulgadas que, até o presente momento, são produzidas linearmente. Dificilmente o doutorando teria tempo e habilidade para, dentro da realidade da pós-graduação brasileira, integrar dois processos criativos ao mesmo tempo, conciliando as tarefas de autor temático e autor tecnológico, embora isso pareça ser o ideal, conforme sugere CAMPOS (2000, p. 44).

Os pressupostos que justificaram o desenvolvimento desta tese foram:

a) um sistema de hipertexto não alcança níveis desejáveis de eficácia no processo de acesso e recuperação das informações por parte dos usuários, se os hiperdocumentos que o compõem não se apresentem estruturados conceitualmente;
b) a teoria de análise facetada (TAF), baseada em Aristóleles, proposta por Ranganathan, usada pelo *Classification Research Group* - *CRG* ( London, décadas de 50 e 60) e simplificada por Louise Spiteri, tem potencial para compor a estrutura lógica de um hipertexto;
c) os princípios de construção de mapas conceituais, proveniente da área da educação se adequam ao contexto do modelo proposto;
d) as recomendações para elaboração de metodologias de estruturação de hipertexto de Maria Luiz a Campos provêm um caminho que atende à área de ciência da informação.

## 1.1 Objetivos

Esta tese teve por objetivo geral a construção de um modelo hipertextual para teses e dissertações, visando-se a organização desses documentos acadêmicos, modelo passível de implementação, tanto na instância de produção primária (pelos alunos de pós-graduação) como no momento da organização de sistemas de hipertextos (pelos profissionais da informação) ou bibliotecas digitais.

Especificamente, pretende-se (1) propor um modelo conceitual, baseado na conjugação de uma estrutura facetada e de um elemento da estrutura formal do documento; (2) demonstrar a aplicabilidade da teoria da análise facetada na modelagem conceitual; (3) implementar um protótipo tecnológico, tendo como base o modelo conceitual.

O modelo de mapa hipertextual proposto tem potencialmente as seguintes funcionalidades: (1) facilitar a navegabilidade em hiperdocumentos e auxiliar, em última análise, o processo de recuperação do seu conteúdo semântico em textos completos de bibliotecas digitais; (2) possibilitar aos usuários uma navegação organizada utilizando-se de um mapa conceitual que represente os objetos e as relações

semânticas da informação contida em cada documento; (3) facilitar o acesso à informação em diferentes formas de navegação e com diferentes graus de interatividade para o usuário final do sistema; (4) auxiliar aos autores das teses e dissertações numa nova estruturação dos conteúdos desses documentos, por meio da conversão da forma tradicional de comunicação para a forma hipertextual, sem perda semântica, utilizando-se de uma estrutura lógica e coerente; (5) valorizar a simplificação na representação de conceitos e, dessa forma, auxiliar os autores de hiperdocumentos na organização, visando à navegação em textos completos.

Finalmente, acredita-se que este estudo possa contribuir para o desenvolvimento da área de Ciência da Informação e sua integração com áreas correlatas. Baseada na minha experiência como professora de *Organização da Informação para Sistemas de Hipertextos* na Escola de Ciência da Informação da UFMG, percebo a importância e necessidade de pesquisas sobre organização semântica de textos, envolvendo uma estrutura conceitual, a partir de uma análise cognitiva, seguida da correspondente aplicação tecnológica. Assim, pretende-se também com este estudo, estimular frentes de pesquisa que integrem a ciência da informação com suas áreas potenciais de ação interdisciplinar, tais como a ciência da computação, a ciência cognitiva e a lingüística.

## 1.2 – Método e materiais

Os trabalhos que investigam a teoria da análise facetada **TAF**, como possível solução, na modelagem conceitual de sistemas de hipertexto, ainda não apresentam resultados consolidados, como será mostrado na revisão de literatura, o que justifica a crescente demanda por pesquisas de base nessa interface. Por outro lado, a urgência para o acesso mais facilitado, o controle maior e a melhor disponibilização da informação, em um mundo cada vez mais tecnológico, aponta para a necessidade de pesquisas aplicadas, que busquem soluções para os problemas gerados por grande produção, acúmulo e volatilidade da informação, especialmente no meio acadêmico.

Esse estudo envolveu procedimentos de natureza teórica, analítica, descritiva, exploratória e aplicada. A ampliação de sua abrangência e validação por meio de testes

experimentais adicionais com o protótipo criado, tanto quanto à abrangência (outras áreas do conhecimento), quanto à qualidade (eficiência de recuperação), quanto à quantidade (volume de documentos tratados), como quanto à sua generalização, estendendo sua aplicação a diversos níveis, permanecem como perspectivas para etapas posteriores, dentro de sua vocação de natureza dedutiva e de aplicação do referido protótipo, mas são ações que ultrapassam o escopo da presente tese.

O desenvolvimento desta pesquisa implicou em duas implementações: uma conceitual e outra tecnológica. Na implementação conceitual, foram utilizados os referenciais teóricos para a categorização e modelagem da terminologia da tese escolhida. Do campo da ciência da informação, utilizaram-se três referenciais: (1) a TAF, compilada por Louise SPITERI (1998) em um modelo simplificado para análise facetada (*A simplified model for facet analysis: Ranganathan 101*)[5], a partir dos princípios apresentados por S. R. RANGANATHAN (1965) e pelo grupo de pesquisadores do *CRG* (1985) para a facetação da tese; (2) a proposta de estrutura de organização do conteúdo semântico, elaborada segundo os dois primeiros níveis (e seus seis requisitos) para elaboração de modelos conceituais propostos por CAMPOS (2000); (3) uso complementar de uma estrutura formal expandida, de acesso ao conteúdo textual, o Sumário Expandido (SE), de acordo com as normas técnicas da ABNT (Associação Brasileira de Normas Técnicas) sobre informação e documentação, a NBR 6024/2003 (antiga NBR 6027/1989) – Sumário/Apresentação e NBR 14724/2002-Trabalhos Acadêmicos/Apresentação. Adotou-se também a "síntese de abstração", requisito sete do nível três das recomendações de CAMPOS (2000), para adaptar, do campo da educação, a proposta do mapa conceitual de Joseph D. NOVAK (2002) que, por sua vez, é baseada na teoria da aprendizagem significativa de David AUSUBEL (1963, 1968, 1978), da área de psicologia educacional, que trata da representação do conhecimento na forma gráfica.

Na implementação tecnológica, foram utilizados recursos teóricos dos campos da ciência da computação, lingüística e ciência cognitiva, considerando o hipertexto como

[5] *A simplified model for facet analysis: Ranganathan 101.* O número *101* refere-se à prática, nos Estados Unidos, de nomear disciplinas de graduação introdutórias de algum assunto.

forma de escrita nova e dinâmica, como tecnologia de organização do conhecimento que integra textos, gráficos, sons e imagens, de maneira associativa e com grande potencial cognitivo, como forma de assegurar ao usuário a recuperação de conteúdos inseridos em seus contextos. Para a implementação do protótipo Modelo Hipertextual para Organização de Documentos - MHTX" foram utilizados os *softwares Greenstone Digital Lybrary* e o *Star Tree Studio.*

O universo de desta pesquisa referiu-se aos documentos acadêmicos defendidos ao final dos cursos de mestrado e doutorado. Para obtenção do grau acadêmico, dos alunos dos cursos de pós-graduação é exigida a apresentação de trabalho final denominado "tese" para o curso de doutorado, e na maioria das vezes, a "dissertação" para os cursos de mestrado.

A ABNT (Associação Brasileira de Normas Técnicas) na NBR 14724 (2002, p.2) procura diferenciar os dois tipos de trabalho final de cursos de pós-graduação *strictu sensu*:

> *"Dissertação é o documento que representa o resultado de um trabalho experimental ou exposição de um estudo científico retrospectivo, de um tema único e bem delimitado em sua extensão, com o objetivo de reunir, analisar e interpretar informações. Deve evidenciar o conhecimento de literatura existente sobre o assunto e a capacidade de sistematização do candidato. É feito sob orientação de um orientador (doutor), visando a obtenção do título de mestre".*
>
> *"A tese é um documento que representa o resultado de um trabalho experimental ou exposição de um estudo científico de tema único e bem delimitado. Deve ser elaborado com base em investigação original, constituindo-se em real contribuição para a especialidade em questão. É feito sob a coordenação de um orientador (doutor) e visa a obtenção do título de doutor, ou similar*".

Observa-se que as definições acima são insuficientes para abarcar, hoje, os tipos de tese e dissertações das diversas áreas do conhecimento. Poderíamos acrescentar que esses documentos são considerados objetos complexos, de natureza argumentativa elaborados com o intuito de expor, e discutir metodicamente, os resultados de uma pesquisa. São compostos de elementos estruturais que geralmente incluem tema,

problema, hipótese, revisão de literatura, metodologia, procedimentos, reflexões sob a forma de descrição ou discussão de resultados, perspectivas, conclusão, bibliografia e anexos. Embora esses dois documentos acadêmicos sigam projetos de pesquisa e tenham estruturas e planos de trabalhos muito semelhantes, a tese distingue-se da dissertação pelo maior significado de sua contribuição para o avanço científico da área em que o trabalho se inscreve.

Para a implementação do protótipo tecnológico desenvolvido nesse estudo, o documento escolhido foi a tese de doutorado de Madalena Martins Lopes Naves, Professora da Escola de Ciência da Informação da UFMG (ECI/UFMG), intitulada *Fatores interferentes em análise de assunto: estudo de casa de indexadores*, inserida na linha de pesquisa Tratamento da Informação (hoje denominada Organização e Uso da Informação), defendida em 2000, no Programa de Pós-Graduação da Ciência da Informação da UFMG, indicada ao prêmio Teses em Destaque da ANCIB 2000 (Associação Nacional de Pesquisa em Ciência da Informação e Biblioteconomia), tendo ficado em segundo lugar. A tese foi selecionada como objeto de pesquisa pela aproximação facilitada entre o "autor tecnológico" e o "autor temático" do documento, uma vez a autora do presente estudo e a Dra. Naves são especialistas na área de tratamento da informação e trabalharam ambas no mesmo programa. Além disso, o levantamento de termos foi facilitado, tendo em vista que a tese explicita, de maneira clara e abrangente, os principais conceitos do processo de indexação e atividades relacionadas.

## 1.3 – Apresentação da tese

A presente tese subdivide-se em cinco seções. No *Capítulo 1 – Introdução*, que é a presente seção, onde incluem-se a justificativa, objetivos, método e material envolvidos na pesquisa e as partes da tese. O *Capítulo 2 – Fundamentos históricos, teóricos e metodológicos* apresenta os fundamentos históricos, teóricos e metodológicos da organização do conhecimento, explorando a teoria da classificação e da análise facetada, o hipertexto e os aspectos cognitivos que permeiam esses conceitos. Além disso, é apresentada, nessa seção, uma revisão de literatura relacionada à pesquisa na

confluência análise facetada/hipertexto; um levantamento de metodologias e modelos para sistemas de hipertexto. No *Capítulo 3 – Construção do Modelo Hipertextual* descrevem-se os passos de construção e implementação do protótipo chamado Modelo para organização hipertextual de documentos *(**MHTX**)*, discutem-se questões operacionais. No *Capítulo 4 - Considerações Finais*, sumariam-se e apresentam-se considerações inferidas ao longo do estudo, algumas conclusões preliminares e perspectivas para a continuação deste projeto e de outras iniciativas correlatas. Após as *Referências*, são apresentados seis anexos. O *Anexo 1* constitui-se de um glossário para auxiliar a consulta sobre termos, cujos significados são ambíguos ou ainda não estão consolidados na área de Ciência da Informação ou, mesmo, na língua portuguesa. O *Anexo 2* inclui a apresentação do Modelo Simplificado para Análise Facetada de Louise Spiteri. O Anexo 3 é a apresentação dos Cânones da Teoria da Análise Facetada de Ranganathan. *O Anexo 4* é a apresentação dos Princípios da Teoria da Análise Facetada do *CRG*. Do *Anexo 5* consta a apresentação da Lista de *links* do **MHTX** e Sumário Expandido com respectivas URLs utilizados na construção e implementação do protótipo **MHTX.** O *Anexo 6* é um CD-ROM, contendo a tese *Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores*, implementada, segundo o protótipo do **MHTX.**

# 2 FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS

Este capítulo tem por objetivo apresentar os fundamentos teóricos e metodológicos que embasaram a presente tese de doutorado, dividindo-se em duas partes. A primeira parte trata da evolução histórica do conhecimento e de seus suportes, dos primórdios dos registros dos pensamentos filosóficos e científicos aos dias atuais, ressaltando-se o conhecimento acadêmico materializado em teses e dissertações, objeto de trabalho dessa tese. A segunda parte apresenta técnicas de organização do conhecimento, usadas na construção do instrumento tecnológico, parte essencial deste trabalho. Dentre essas, foram abordados os conceitos de hipertexto e dos aspectos cognitivos envolvidos nessa forma de escrita; metodologias e modelos disponíveis para estruturação de sistemas de hipertexto; a organização do conhecimento com a finalidade de recuperação, destacando-se especificamente técnicas de análise facetada, o modelo de Ranghanathan, as contribuições do *Classification Research Group - CRG*, o modelo simplificado de Louise Spiteri, envolvendo essas duas abordagens anteriormente citadas. Também fazendo parte do construto teórico desta tese, encontra-se o conceito de mapa conceitual, seguindo-se esclarecimentos sobre a conexão entre este e o hipertexto, assim como tópicos relativos à construção desse instrumento de estruturação semântica. Os fundamentos, princípios e subsídios que tornaram possível a construção do Modelo Hipertextual para Organização de Documentos - **MHTX** emergiram da assimilação dos conhecimentos constantes do presente capítulo.

## 2.1 Da tradição oral ao alfabeto

Os membros das sociedades orais possuíam apenas os recursos de sua memória para, ao longo do tempo, reter e transmitir as representações que lhes eram convenientes. Para isso, utilizavam recursos como a dramatização, personalização e artifícios narrativos diversos, a fim de que as representações tivessem mais chances de sobreviver em ambiente baseado quase que unicamente na memória humana. As mensagens eram sempre recebidas no tempo e lugar em que eram emitidas. Tanto o emissor quanto o receptor compartilhavam um universo de significados semelhante e todos compartilhavam do mesmo universo semântico, e do mesmo contexto. Nessas culturas, não existia modo sistematizado algum de armazenar as representações para futura reutilização. A transmissão do conhecimento, no

transcorrer do tempo, exigia um contínuo recomeço, uma renovação suscetível de visíveis alterações e perdas, de geração para geração. A história era feita a partir da capacidade de memorização dos membros do grupo social e de suas preferências. Havia, portanto, registros "incertos" e parcial da realidade, fortemente filtrada pelo sujeito da ação. A mediação deste sujeito, nesse tipo de comunicação, era de fundamental importância para a continuidade histórica do conhecimento, pois não havia a escrita.

A escrita foi um dos mais importantes desenvolvimentos técnicos do ser humano. Assim como a fala foi o principal instrumento utilizado no tempo da oralidade primária, diversos tipos de sistemas de sinais gráficos, incluindo o alfabeto, tornaram-se os instrumentos principais da escrita. Os primeiros sinais utilizados pelos seres humanos para se expressarem aparecem na pictografia, forma primitiva que McGARRY (1999, p.72) compara com o início das expressões gráficas utilizadas pelas crianças: desenhos de figuras, rabiscos, imagens toscas ou marcas com supostos poderes de talismãs mágicos ou, mesmo, símbolos de posse. Os sinais pictográficos continuam versáteis em suas funções modernas. Por exemplo, figuras que podem expressar como operar uma máquina complexa a partir da seqüência de idéias e instruções que, além de serem mais eficientes que a linguagem alfabética, muitas vezes transpõem barreiras lingüísticas. De forma semelhante, operam os pictogramas vistos nas telas de computadores, inseridos em sistemas de hipertextos na forma de ícones.

As formas de escrita utilizando ideogramas e equivalência de fonemas, começaram a surgir por volta de 3000 a.C., na Mesopotâmia. Nessa mesma época, no Egito, já eram utilizados papiros e tintas artesanais para representação de signos na comunicação escrita. A escrita era feita em peles de animais, cerâmicas e papiros. Somente no Séc. III a.C. é que surge o pergaminho, como opção de suporte. O livro surge com a conseqüente reunião de vários pergaminhos ou papiros.

No início do Séc. III a.C., foi criada a Biblioteca do Museu de Alexandria por Ptolomeu Filadelfo, tendo como objetivo reunir em um só lugar todo o conhecimento da humanidade. A partir desse momento, o livro tornou-se sinônimo do saber, passando a representar autoridade, poder e prestígio.

Segundo McGARRY (1999, p.72), o alfabeto tem sido considerado por alguns autores como uma das maiores invenções do homem e cita DIRINGIR (1968)[5]: *"Historicamente, foi a última grande forma de escrita a surgir, e a mais altamente desenvolvida, a mais conveniente, e o sistema de escrita mais facilmente adaptável jamais inventado*". O alfabeto mais antigo que se conhece é o semita setentrional, composto de 22 signos, e que foi desenvolvido por volta de 1700 a.C. na Palestina e na Síria. Outros alfabetos, como o hebraico, o árabe e o fenício basearam-se nesse modelo. O fenício serviu de modelo para os gregos, que acrescentaram vogais às consoantes, modelo que foi modificado pelos etruscos por volta de 800 a.C. e do qual vieram as letras do antigo alfabeto romano e de todos os alfabetos ocidentais. A invenção do alfabeto não somente permitiu a humanidade comunicar idéias por símbolos visuais através da escrita, mas também propiciou a criação de registros permanentes destes signos e, assim, a criação da memória externa à mente humana.

## 2.2 Do texto ao hipertexto

A escrita veio permitir a atualização do conhecimento acumulado, de fatos presenciados ou relatos de pessoas que viveram em épocas ou lugares diferentes (DIAS 1999, p.267). Com ela, o discurso pôde desvincular-se da situação particular em que foi produzido, não mais precisando da presença do sujeito social para a reprodução de uma experiência particular. Se o registro escrito do discurso "fala por si só", sofre por outro lado, a interferência de quem o "consulta". Conceitualmente, o tempo e o espaço tornam-se diferenciados daquilo que representam no universo da oralidade, no qual o contexto é vivenciado *in loco*. Com a escrita, há uma autonomia do texto em relação à tradição oral. O tempo não é mais o da circularidade, mas sim da linearidade, linearidade que se traduz no transcorrer da história. Qualquer discurso torna-se possível de ser apreendido, analisado e interpretado fora de seu contexto de produção. Por outro lado, a escrita cria a figura do leitor, para o qual a realidade passa por um filtro muito mais refinado e atento do que ocorre com o ouvinte no tempo da oralidade. Segundo GUINCHAT (1994, p.41), *"documento é um objeto que fornece um dado ou uma informação. É o suporte material do saber e da memória da humanidade".*

[5] DIRINGIR, David. *A history of the alphabet*. 3ed. ed. London: Utchinson, 1968,p.14

O papel foi introduzido no mundo ocidental em meados do século XII e difundiu-se na Europa entre os séculos XIII e XV com a necessidade do aumento da quantidade de novos originais e cópias de livros. À medida que crescia o aumento de exemplares disponíveis, aumentava também o estímulo ao registro do conhecimento, o que influenciou muito o clima intelectual na Idade Média. Ampliou-se o número de editores e possíveis leitores. A partir daí, a leitura e a interpretação do texto adquiriram caráter mais individualizado, com particularidades e exigências diversas. As obras começaram a incluir representações gráficas mais precisas e específicas, como tabelas, gráficos e desenhos, incorporando inventos e artifícios anteriores à tipografia. Formas mais sofisticadas de leitura não linear começaram a surgir, como as notas de rodapé, sumários, por exemplo, que enviavam o leitor a outras partes do mesmo livro ou a outras obras relacionadas como índices, bibliografias.

A partir daí, a comunicação escrita e o modo de transmissão dos textos sofreram profundas mudanças. A escrita em suportes artesanais foi a tecnologia de comunicação mais avançada, desde o quarto milênio a.C. até a invenção da imprensa e da tipografia na Mogúncia (c.1450), quando Johann Gutenberg compôs, com tipos móveis, o texto da Bíblia, o primeiro livro a passar do manuscrito para o papel impresso (McGARRY, 1999, p.73).

Segundo KOBASHI (1994, p.112), o texto, em sentido mais amplo designa uma unidade de comunicação organizada sintagmaticamente e dotada de coesão e coerência. O texto pode ser classificado, quanto à sua estrutura interna, como descritivo, narrativo, dissertativo ou, quanto à sua finalidade, como texto técnico, científico, didático, jornalístico, jurídico, político, de lazer e etc. A grande variedade de tipos de documentos pode ser diferenciada pelas características físicas e intelectuais: livros de leitura corrente, teses, dissertações, livros de referência, publicações periódicas e impressos diversos.

A tese acadêmica, que foi o tipo de documento escolhido para a aplicação metodológica proposta neste trabalho, é caracterizada como texto técnico-científico, de natureza argumentativa, elaborado com o intuito de expor metodicamente os resultados da observação de um fenômeno, dotado de elementos estruturais tais como o tema, o problema, a hipótese, a metodologia, os resultados e a conclusão.

O poder de preservação do pensamento registrado cresceu enormemente com o surgimento dos primeiros computadores, na Inglaterra e nos Estados Unidos em 1945, permitindo grande velocidade na comunicação e a simulação do conteúdo através da demonstração visual. Essa revolução não se deu apenas no processo de transmissão da mensagem, mas também no modo de recepção e interpretação que passou a ocorrer através da mobilidade e direcionalidade das relações de sentido. Hoje, na realidade tanto das nossas relações interpessoais quanto das relações homem/máquina, pode-se constatar a crescente dependência do meio eletrônico. O surgimento dos computadores pessoais, nos anos 70, inaugurou nova fase para a área de informática.

A informática provocou muito mais do que uma revolução nas formas e nos métodos de geração, armazenamento, processamento e transmissão da informação. Pode-se dizer que o desenvolvimento das tecnologias relacionadas com o computador compara-se à revolução causada com a invenção dos tipos móveis de Gutenberg. A mudança do texto impresso para o texto eletrônico criou uma grande mudança na maneira como armazenamos e acessamos a informação. Os textos eletrônicos, que podem ter as mesmas características dos textos lineares bem como dos não-lineares, são armazenados ou disponibilizados em *CPUs*, disquetes, *CD-ROM*, *LANs*, ou redes digitais. A possibilidade de explorar o texto de maneira interativa introduziu o conhecimento por simulação.

O hipertexto, fruto da tecnologia informacional das duas últimas décadas, é a forma texto que estimulou e tornou possível o desenvolvimento do universo digital. O hipertexto, como nova tecnologia, também desterritorializa o texto, tornando-o sem fronteiras, o que remete, de certa forma, à transmissão da informação na tradição oral, em que o elemento da aleatoriedade é inerente à comunicação. Porém, o texto eletrônico acelera o tempo e concatena contextos, encadeando e justapondo diferentes documentos, compartilhando o mesmo espaço de produção e, às vezes, alterando sua compreensão.

A idéia básica do hipertexto de organizar documentos em trechos, e combiná-los conforme as necessidades de compreensão e organização, é anterior ao final da Segunda Grande Guerra, quando Vannevar BUSH (1945), no artigo *As we may think* previu um sistema muito parecido com o hipertexto. Com o *Memex*, que serviria

como suplemento para a memória dos usuários de um banco de dados de literatura científica, ele propôs a organização associativa que tentava imitar a estrutura da memória humana. Seu sistema teria a capacidade para armazenar um grande volume de informações e, por meio de determinado mecanismo, estabeleceria associações entre diferentes pontos da base de dados, propiciando a sua consulta com rapidez e flexibilidade.

Hoje, o hipertexto permite ir muito além do formato uniforme de uma página. Os escritos informatizados variam e se adaptam de acordo com o leitor. As imagens simuladas na tela do computador funcionam como uma extensão da imaginação. Documentos eletrônicos compostos de fragmentos de textos ligados entre si ou com outros documentos, permitem a leitura não seqüencial, mais adequada à flexibilidade do raciocínio humano.

Segundo MARQUES (1995, p.89), os sistemas de hipertexto consistem em abordagens de estruturação e manipulação de textos, caracterizadas pela não-linearidade textual. A divisão padronizada das publicações na linguagem tradicional em capítulos, seções e parágrafos continua sendo o ponto de partida para a estruturação das partes do hipertexto. Entretanto, apenas essa divisão tradicional do texto não atende ao leitor, mesmo que forneça todos os pontos de acesso temáticos do documento ou o conjunto de documentos. A necessidade de criação de índices remissivos, complementando os sumários, atesta essa limitação. Assim, o hipertexto, surge como solução que permite flexibilidade de movimento de um ponto a outro no mesmo documento, ou de um ponto a outro em diferentes documentos. Seu modelo informacional é o da rede estruturada em conjuntos de nodos, e *links* entre os nodos.

Além de Bush, Ted NELSON (1965) e Douglas ENGELBART (1968) foram outros pesquisadores que compartilharam a mesma visão de que seria estabelecida nova forma de interação homem-computador. Theodor Nelson foi quem criou o termo hipertexto, definindo-o como *"uma combinação de texto em linguagem natural com a capacidade do computador de fazer pesquisa interativa e exibição dinâmica de um texto não linear"* (CONKLIN, 1987, p.17). O sistema *Xanadu*, desenvolvido por ele e sua equipe de pesquisadores, possuía um gerenciador de armazenamento que permitia a criação de ligações entre os tópicos correlatos e mantinha informações

sobre a origem, as variações e as interconexões dos textos. Na década de 1960, Nelson começou a estruturar alguns fundamentos do hipertexto e apontou as limitações da informática em face da nova ferramenta.

Paralelamente aos estudos de Nelson, Douglas Engelbart, do Instituto de Pesquisa Stanford, EUA, estudou a utilização do computador como instrumento representacional e ampliador do intelecto humano (MARQUES,1995, p.88). Foram apresentados novos conceitos relativos à interação homem-máquina para a construção de sistemas mais dinâmicos e flexíveis. Juntamente com outros pesquisadores, implementou o sistema *On Line System* (OLS), conhecido mundialmente como sendo a primeira implementação que seria conhecida como um hipertexto, estruturado por uma rede de *links* que interligavam blocos pequenos de informação. Esse sistema foi o primeiro a introduzir noções como edição estruturada, uso do *mouse*, correio eletrônico, janelas múltiplas e tela compartilhada. Em 1969, BUSH , publicou o artigo *Evolution of an information society* (republicado em 1987),[6] concluindo que o *Memex* podia finalmente, ser viabilizado, porém a um alto custo, exigindo também manipulação de alta tecnologia e uma enorme dose de paciência dos usuários.

Dois acontecimentos contribuíram para o aumento do interesse em relação ao hipertexto. O primeiro foi o lançamento do *Guide* (1985) e do *HyperCard* (1987), sistemas comerciais de hipertexto para microcomputadores da *Apple*, os quais conseguiram divulgar o conceito de hipertexto para milhões de pessoas. O segundo foi a realização da conferência *Hypertext'87*, inteiramente dedicada ao tema, na Universidade da Carolina do Norte, em Chapel Hill, EUA. Após cinco décadas da publicação do trabalho de Bush, eram cada vez maiores os números de sistemas de hipertextos comerciais para computadores de portes diversos, com aplicações em educação, medicina, publicações eletrônicas, ciência da computação, biblioteconomia e outras áreas do conhecimento.

Encontram-se na literatura várias definições de hipertexto. Alguns autores não fazem distinção entre hipertexto e hipermídia. CONKLIN (1987,p17), por exemplo, define hipertexto como

---

[6] BUSH, Vannevar. MEMEX revisited. In:CALKELL, A. E. (Editor) *Evolution of an information society*. London: ASLIB, 1987. p.179-191. Publicado originalmente em *Science is not enough*. New York: Apollo Editions, 1969. p.75-101.

> *"Janelas na tela* [que] *são associadas a objetos na base de dados e ligações* [que] *são estabelecidas entre estes objetos, tanto graficamente, na forma de marcas rotuladas, como na base de dados, na forma de ponteiros."*

Já SHNEIDERMAN E KEARSLEY dizem que hipertexto é

> "*Uma rede de nós e ligações entre documentos, onde documentos são os nós e as ligações são referências cruzadas. As redes podem ter a forma de hierarquias, embora geralmente as associações entre os nós sejam mais complexas. E quando os documentos são de natureza (tipo) multimeios, o termo hipermídia é freqüentemente usado*."

Segundo o próprio VILAN FILHO (1992, p.61),

> *Os sistemas de hipertextos são conjuntos de programas de computador (suporte lógico),* [com] *suas tabelas e dados de controle necessários para a operação de um sistema construído para operar com hiperdocumentos, segundo a filosofia de hipertexto. Em geral, esses sistemas são compostos de: (a) um subsistema de autoria, e (b) um subsistema de navegação*."

Uma definição mais resumida é dada por MARQUES (1995,p.85):

> *"Os hipertextos são documentos eletrônicos que permitem uma leitura não seqüencial, mais de acordo com o raciocínio humano* [...] *Hipermídia é o hipertexto enriquecido com outros meios, como imagem e som."*

CONKLIN (1987, p.33) afirma que o ambiente hipertexto resulta basicamente da união de três elementos. (1) ***Um componente de banco de dados****,* em que objetos na tela são associados a objetos armazenados em um banco de dados. As conexões entre estes objetos podem ser definidas não apenas graficamente, mas também por ponteiros em uma base de dados. (2) ***Um esquema de representação****,* no qual a informação textual informal é misturada com processos e operações mais formais e mecanizados. Para ser qualificado como "hipertexto", o sistema deve proporcionar um suporte eletrônico constante para suas inter-relações. (3) ***Uma interface com o usuário***, geralmente janelas representando a correspondência "um-para-um" com os nodos na base referencial de dados. A interface gráfica, com o auxílio de *browsers* e diagramas, auxilia o usuário a navegar através de uma grande quantidade de informações, ativando as ligações e acessando o conteúdo dos nodos.

Os sistemas de hipertexto traduzem uma abordagem de estruturação e manipulação de textos caracterizada pela não linearidade. Seus componentes básicos são:

a) Nó (ou Nodo) : Em cada uma das unidades de informação na base hipertexto, os nós correspondem a uma ou mais exibições de tela. Essas unidades de informação em um hiperdocumento podem conter diferentes tipos de dados, textos, figuras, fotos, sons e são conectadas por *links* a uma variedade de estruturas. Geralmente, o nó descreve um único conceito ou tópico de modo que pode ser caracterizado como auto-contido, não dependendo da leitura prévia de outros nós.

b) *Link* : Conceito mais importante do hipertexto, o *link* é também denominado elo, ligação, vínculo, âncora ou botão, de acordo com o sistema de hipertexto. *Links* são marcas que conectam um nodo com outro. A ativação desses *links* implica abertura de nova janela, contendo o documento referenciado. Assim, é possível deslocar-se fácil e rapidamente, de um ponto a outro, no conteúdo das páginas. Esses *links* podem ser representados por palavras ou frases em destaque (negrito, itálico, cores), ou ainda, por figuras ou ícones.

FREI e STIEGER (1995, p.3) dividem os *links* em dois tipos: *links* referenciais (ou organizacionais) e *links* semânticos (ou de conteúdo). A finalidade do *link* referencial é tornar a leitura do documento mais confortável. Já a função do *link* semântico é apontar para um detalhamento da informação ou uma informação similar, ou ainda, uma informação adicional em tópico específico. O experimento conduzido por BARON, TAGUE-SUTCLIFFE e KINNUCAN (1996, p.897) procurou avaliar qual tipo de *link* tinha melhor performance como apontador no momento da busca ou no folheio. Para isso, ampliaram as categorias de *link* descritas por Frei e Stieger. Além de classificar os descritores da estrutura superficial do documento e de conteúdo, dividiram os *links* de conteúdo em três categorias: *links* semânticos (diferentes significados das palavras e suas associações), *links* retóricos (cuja raiz está na análise do discurso) e *links* pragmáticos (cujo objetivo é levar o usuário à informação específica, mostrando uma associação prática entre a fonte e o destino dos nodos). A partir daí, os autores avaliaram a eficiência dos *links* em três situações: (1) somente *links* organizacionais, (2) *links* organizacionais + *links* sem conteúdo semântico, retóricos e pragmáticos, (3) *links* organizacionais + *links* com conteúdo semântico, retórico e pragmático. O resultado estatístico indicou que, no folheio, não

houve diferença entre as situações criadas, porém, os *links* do terceiro caso, que tinham rótulo de conteúdo, apresentaram desempenho significantemente melhor na busca.

De acordo com os nós conceituais que estão sendo ligados, RADA (1991, p.14) classifica o hipertexto, conforme as tipologias microtexto e macrotexto. Microtexto é um neologismo que significa um hipertexto de pequeno volume, com ligações explícitas entre seus componentes. Aparentemente, os microtextos são muito apropriados para folheio. Uma rede de microtextos pode ser vista como uma rede semântica e, por isso, ser tomada como modelos da memória humana. Através dos recursos computacionais, evidencia-se o caminho ou os caminhos em que o texto foi internamente construído, e que podem variar de acordo com pontos de vista lógicos. Macrotexto é definido como um sistema hipertextual em larga escala, composto de vários documentos, e cuja ênfase está em *links* que os conectam entre si, ao invés de conectar internamente apenas um documento.

A estrutura do hipertexto determina e descreve o sistema de ligações ou relacionamentos entre os nodos ou unidades de informação, sendo um fator decisivo na facilidade de criação, uso e atualização do hiperdocumento. LEVY (1993, p.25), que considera o hipertexto como "*uma metáfora cálida para todas as esferas da realidade em que significados estejam em jogo*", aponta seis princípios básicos de sua lógica, que permite visão panorâmica, a qual organiza, resume e amplia a idéia de rede que se pretende construir:

- **Princípio de metamorfose**: processo de constante construção e renegociação de sentidos, que se dá nos hipertextos O princípio de metamorfose explicita a idéia de que a rede de significados que constitui o conhecimento está em permanente transformação.

- **Princípio de heterogeneidade**: a variedade e diferença entre as informações organizadas em determinada seção do hipertexto, como os nós e as conexões que se estabelecem entre as diversas partes da rede hipertextual. Existem, na memória, informações cognitivas que passam por vários tipos de associações entre estes elementos.

- **Princípio de multiplicidade e de encaixe das escalas**: organização "fractal" do hipertexto, ou seja, qualquer nó ou conexão, quando analisado, pode revelar-se como sendo composto por toda a rede, e assim por adiante, indefinidamente, ao longo da escala dos graus de precisão.

- **Princípio de exterioridade**: a rede não possui unidade orgânica, nem um "motor interno". Seu crescimento, diminuição ou recomposição permanente dependem de ação exterior indeterminada. O princípio de exterioridade pretende caracterizar a permanente abertura da rede hipertextual e do conhecimento em construção. O fluxo constante de elétrons, de dados digitais, que mudam constantemente são incorporados e trocados a todo momento.

- **Princípio de topologia:** nos hipertextos, tudo funciona por proximidade, por vizinhança. Neles, o curso dos acontecimentos obedece a uma topologia, na qual os caminhos seguem a idéia de proximidade entre significados.

- **Princípio de mobilidade dos centros:** a rede não tem centro fixo, mas possui diversos centros que são como pontas luminosas perpetuamente móveis, saltando de um nó a outro, trazendo ao redor de si uma ramificação infinita de pequenas raízes, como finas linhas brancas esboçando por um instante um mapa de informações. A característica de "acentrismo" (que seria melhor descrito como "multicentrismo" ou "centro móvel") da rede hipertextual é especialmente relevante, e pode ser imediatamente transferida para a constituição do conhecimento em sentido amplo.

As seis características da rede hipertextual propostas por Levy constituem uma teia básica a partir da qual pode ser tecida a idéia de rede para a representação do conhecimento.

O hipertexto é tanto uma ferramenta para o autor quanto para o leitor. Para o autor funciona como ferramenta que ajuda a organizar, estruturar e armazenar informações. Para o leitor, ele permite, através da navegação, a recuperação da informação organizada. Neste trabalho, considerou-se para estudo, a primeira abordagem, a do autor, apresentada por Pierre LEVY (1993), na qual o hipertexto é utilizado como ferramenta de organização da informação através da sua

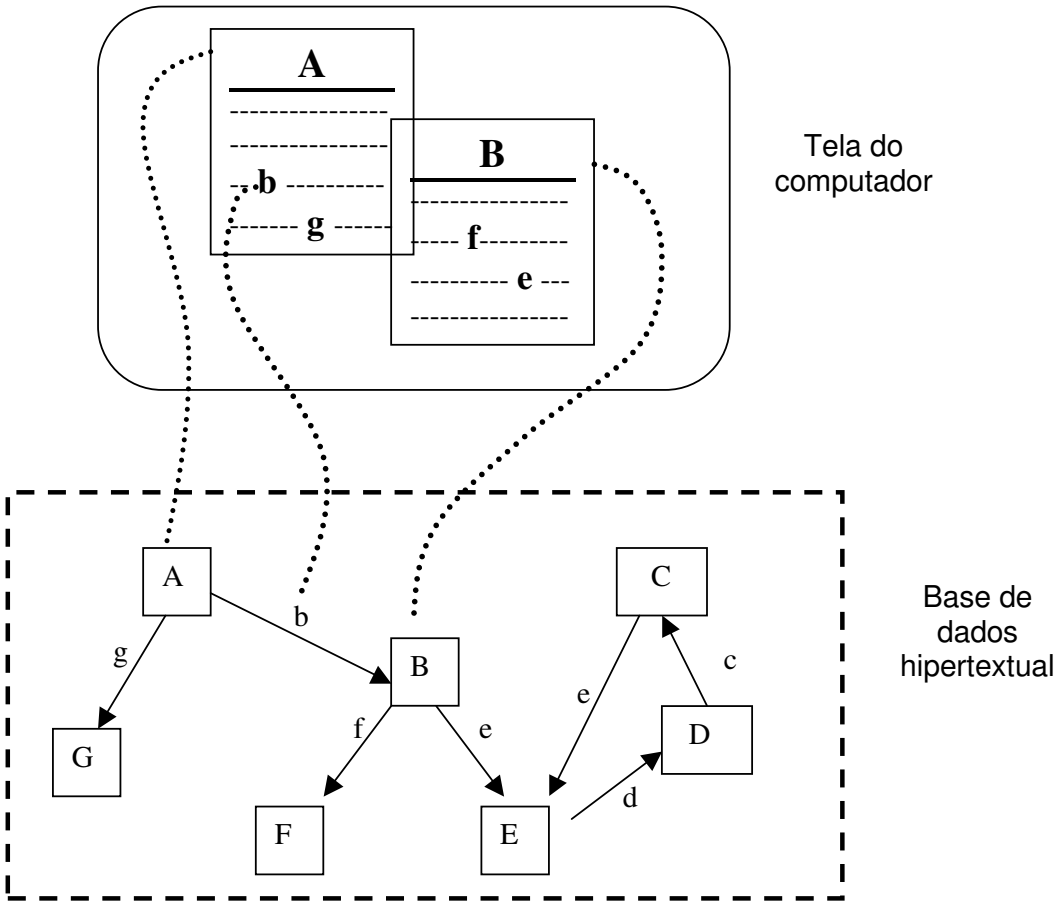
característica de escrita não-seqüencial e da associação dessas informações através de nós, representados por conceitos, possibilitando assim a construção de uma rede navegacional em documentos completos sem perda significativa do conteúdo semântico após a fragmentação.

Os aplicativos de hipertextos são projetados para efetuar a navegação através de um espaço de informações. Para isto, o projeto de navegação de tais aplicativos constitui uma etapa crucial no empreendimento de desenvolvimento. A navegação é uma metáfora utilizada para descrever como os usuários se movimentam por documentos hipertextuais, desenhando um percurso em uma rede que pode ser simples, mas também tão complicada quanto possível. Isto porque cada nó pode, por sua vez, conter uma rede interativa. Assim, navegar significa direcionar uma mudança de foco ou movimentação em relação ao conhecimento disponível no hiperdocumento, por meio da utilização de características abstratas, estruturas ou outras ferramentas para orientação geral (WOODHEAD, 1990, p.4). A estrutura navegacional é construída sobre nós e *links*. Como já foi descrito, os nós contém as informações básicas nas aplicações hipertextuais, e sua estrutura depende do conteúdo semântico que representa os interesses particulares dos usuários aos quais a aplicação se destina. Os nós podem ser acessados em diferentes contextos e seus atributos e *links* podem mudar de um contexto a outro. A realização navegacional das relações ocorre através dos *links*. A estrutura do hipertexto determina e descreve o sistema de ligações ou relacionamentos entre os nós ou unidades de informação, sendo fator decisivo e facilitador de criação, uso e atualização do hiperdocumento.

Para VILAN FILHO (1992, p.14), a estruturação da informação em sistemas de hipertexto é similar à de uma rede semântica, enquanto modelo de representação do conhecimento. Os nós representam conceitos e idéias, e as ligações representam as relações entre eles. As relações devem refletir a estrutura organizacional do assunto relacionado ou, em outras palavras, a rede semântica de uma área do conhecimento (FIG.1). Se a informação for disposta de maneira organizada, será fácil para o usuário recuperar o que estiver procurando. Por outro lado, se a estrutura apresentar uma trilha de navegação desorganizada ou um *design* cognitivamente pobre, o folheio ou a pesquisa mais direcionada por parte do usuário torna-se difícil ou

ineficiente. Uma boa estrutura de hipertexto deve refletir a estrutura organizacional do assunto relacionado ou a rede semântica de sua temática.

FIGURA 1 - Correspondência dos nós (A, B) e *links* (b) entre a tela do computador e a base de dados hipertextual.




FONTE:CONKLIN, p.19, 1987.

SHNEIDERMAN e KEARSLEY (1989, p.10) enumeram alguns mecanismos de navegação:

- **Folheio (*browsing*)**: Através deste mecanismo, os usuários do hipertexto podem perpassar nós e ligações com a naturalidade de quem folheia um livro à procura de algo interessante. Essa manipulação direta possibilita um modo amigável de interação, mas só funciona para ligações (*links*) pré-definidas e requer um processo de tomada de decisão.
- **Pesquisa**: Muitos sistemas de hipertexto possibilitam ao usuário fazer pesquisas por meio do próprio texto que constitui o documento de modo a descobrir nós com informações específicas, o que é geralmente feito com pesquisa por cadeias de caracteres em todo o documento. Este tipo de

pesquisa é útil quando o usuário não está familiarizado com a estrutura do documento ou não se adapta bem a outros tipos de navegação. É considerado por alguns especialistas como um método lento e ineficiente, especialmente se o documento for muito grande.

- **Filtros**: Os filtros são mecanismos de seletividade fornecidos por muitos sistemas, que podem ser implementados mediante palavras-chave ou atributos de nós ou ligações. Complementar à pesquisa, a filtragem visa restringir a quantidade de informações exibidas.
- **Índices**: Embora as ligações sejam o meio principal para conexão de informações em um sistema de hipertexto, a indexação do conteúdo permite disponibilizar a informação alfabeticamente ou pesquisar por termos específicos. Pode-se também indexar hierarquicamente como num sumário que coincida com as estruturas dos nós do hiperdocumento. A grande limitação dos índices resulta da necessidade do autor de hipertexto ter que indexar cada palavra ou frase a ser encontrada pelo usuário.

A navegação pode ser influenciada tanto pela estrutura espacial da informação quanto pela habilidade do usuário. Do ponto de vista da interação homem-computador, a navegação é o resultado da interação entre os elementos do sistema e as necessidades dos usuários, o que caracteriza a interação com o sistema. Esta interação, que ocorre através da interface entre o sistema e o usuário, pode ser física, perceptiva e cognitiva. Assim, a navegação no hipertexto é uma ação que pressupõe aspectos mecânicos, cognitivos e tecnológicos em um só processo.

Como qualquer aplicação interativa no computador, o hipertexto inclui, como uma de suas aplicações, uma interface que permite ao usuário selecionar um nó, ler seu conteúdo e mover-se, a partir de um *link*, para outro nó. Assim, um dos mais importantes componentes do hipertexto é a sua interface, que LEVY (1993, p.181) define como

> *"como uma superfície de contato, de tradução, de articulação entre dois espaços, duas espécies, duas ordens de realidade diferentes: de um código para outro, do analógico para o digital, do mecânico para o humano[...] Tudo aquilo que é tradução, transformação, passagem é da ordem da interface".*

Já para DIAS (1999, p.1), o termo interface designa o elemento discreto e tangível

através do qual o utilizador acessa a informação e sua manipulação num sistema informático:

> "[. . .] *uma superfície de contato com a informação* [. . .] *também um envelope para o conteúdo, procurando-se adequar esta superfície aos fatores humanos envolvidos no processo de contato e regras de organização da informação*, *segundo o modelo internacional de comunicação".*

Para LARA (2003, p.28), as interfaces permitem a visualização do conteúdo, determinam o tipo de interação que se estabelece entre as pessoas e a informação, direcionando sua escolha e o acesso ao conteúdo. Na verdade, as interfaces criadas a partir de linguagens de programação, como *JAVA, HTML, VRML, SGML, FLASH e Javascript*, são mais que dispositivos lógicos organizados, para que se possa visualizar mais facilmente o conteúdo de um hipertexto. Elas exprimem uma forma de relacionamento com a informação, com as idéias, os saberes, os desejos e as aspirações de pessoas e grupos. Suscitam diferentes imagens do mundo e do imaginário e são, ao mesmo tempo, reflexos dos múltiplos agenciamentos do mundo real. No entanto, para que a interface seja eficiente, é necessário que o autor disponha de informações sobre várias áreas do conhecimento, um contexto muito mais amplo e complexo que as próprias linguagens de programação. O autor (*designer)* do hipertexto é responsável por organizar o espaço das funções cognitivas: coleta informações, as armazena na memória, avalia, prevê, decide e concebe o espaço virtual. Pode-se dizer então que a interface é uma forma de representação do modelo organizacional da informação, como forma de visualização do conteúdo e o meio que permite o acesso a esse mesmo conteúdo.

FOWLER, BRADELEY e FOWLER (2002, p.12) afirmam que, em redes relativamente pequenas para pesquisas e resumos de documentos, é possível mostrar todo o sistema da interface. Em redes maiores com todos os termos e todos os documentos na base, somente parte da rede pode ser mostrada, tornando necessária a utilização de mecanismos que permitam ao usuário folhear ou navegar em toda a rede. Esses autores salientam que é sempre útil relacionar a parte mostrada com todo o conteúdo da rede para permitir ao usuário a orientação dentro de toda a estrutura. Dificuldades de orientação e navegação são comumente encontradas em hipertextos e outras grandes estruturas de informação espacial. Os autores sugerem dois mecanismos de navegação e orientação que podem auxiliar

nessa questão: a interface *overview diagram* (diagrama de visão geral ou mapa da rede) e a interface *fisheye views* (visão “olho-de-peixe” ou visão mais detalhada).

No *overview diagram*, o documento é mostrado em uma seqüência linear de textos e figuras, fornecendo ao usuário um meio para navegação, orientação e conhecimento sobre a organização total da rede. Neste tipo de interface, um pequeno número de nós é selecionado para mostrar ao usuário informações sobre a organização completa da rede. Além disso, os nós fornecem pontos de acesso para atravessar a rede, como ponto de referência para ajudar o usuário a conhecer que parte está sendo vista naquele momento.

A interface *fisheye* fornece um esquema utilizado em mapas hiperbólicos (veja definição à frente no item 2.3.3.1), capaz de lidar com diversos fatores importantes na navegação e na orientação em grandes redes de informação. Na movimentação dessa interface, os componentes diminuem e aumentam de tamanho exponencialmente (saem e entram no foco), demostrando grande flexibilidade e agilidade na tela. A interface *fisheye* permite também reduzir o tempo gasto para navegar através do documento uma vez que exibe primeiro as seções mais importantes. As outras partes podem ser expandidas e lidas se o usuário arrastar os diversos *links* com o mouse, ou ainda, quando apresentam-se completamente ocultas, por meio de pesquisa direta por conceitos.

As interfaces *overview* e *fisheye* também podem ser combinadas na navegação. Através de pesquisa direta, um termo pode ser selecionado para servir de foco, centrando na seleção do nó e ativando-se a interface *fisheye* mostrada em uma janela maior. A localização deste sub-gráfico da interface *fisheye* em relação ao diagrama *overview* é indicada pelo sombreamento de qualquer nó de informação. Se nenhum termo do diagrama overview for mostrado na visão detalhada, o termo na trilha próxima do centro da interface *fisheye* estará sombreado.

Até aqui, apresentou-se um panorama sobre o desenvolvimento da escrita nos sistemas de organização do conhecimento, desde a era da tradição oral até o advento do hipertexto. A seguir, serão abordados os aspectos cognitivos utilizados na criação de hiperdocumentos.

### 2.2.1 Aspectos cognitivos no hipertexto

Os aspectos cognitivos têm sido estudados como importantes elementos em dois problemas básicos que se apresentam na navegação hipertextual: imprecisão e excesso na percepção da informação pelo usuário. A seguir, serão analisados os fenômenos que podem influenciar a "leitura" de hipertextos, bem como as variáveis que causam impacto na performance da navegação em um documento.
Segundo FIDERIO (1988, p.238 ) "...*o hipertexto imita a capacidade do cérebro em armazenar e recuperar informações através de ligações referenciais para um acesso rápido e intuitivo* [...]".CARLSON (1991, p.134) completa

> "*os aspectos virtuais do hipertexto imitam o cérebro humano particularmente na qualidade associativa da memória. Como uma ferramenta, o hipertexto não molda somente a realidade externa, mas o interior da mente, proporcionando novas maneiras de "ver" e "sentir" o ambiente de informação*".

Segundo KOCH (2002, p.67), o desafio básico da ciência cognitiva é tentar entender e esclarecer como o homem representa mentalmente o mundo geral e específico que o cerca e que estruturas da mente possibilitam as atividades cognitivas. Considerando-se que a cognição apresenta-se sob a forma de representações (conhecimentos estabilizados na memória e suas interpretações) e processamento das informações (processos voltados para a compreensão e a ação), é necessário distinguir, dentro do processo cognitivo, o que é provisório e o que é permanente.

Tem-se postulado que as informações podem ser retidas em dois tipos de memória: a memória de curta duração, ou STM (*Short Term Memory*), de capacidade limitada e a memória de longa duração, ou LTM (*Long Term Memory*), na qual os conhecimentos podem ser representados de forma mais permanente. Para a leitura de um texto, as estratégias cognitivas são acionadas em cálculos mentais, e assumem a função de facilitar o processamento textual, quer em termos de produção quer em temos de compreensão. Para que haja compreensão entre os interlocutores, é necessário que os aspectos estruturais e processuais da cognição sejam organizados e ativados.
Os processos cognitivos são atividades mentais como o pensamento, a imaginação, a lembrança e a solução de problemas (ALLEN, 1991, p.13). Como ocorre em outras atividades humanas, são desempenhadas por indivíduos que têm diferentes níveis

diferentes de habilidade em raciocínio lógico e em memória visual, o que pode afetar o desempenho na recuperação da informação. A cognição humana é essencialmente organizada como uma rede semântica, na qual os conceitos são ligados pelas associações. O sistema de hipertexto tenta explorar esta natureza básica da cognição. Os aspectos cognitivos humanos possuem características que reforçam o hipertexto como ferramenta computacional que apóia o pensamento e a comunicação. Tanto a leitura quanto a escrita enfatizam muito a natureza não linear do pensamento, como um processo natural do ser humano. A influência do hipertexto, com sua dinâmica e seus aspectos multimediáticos é tal que, a partir dele, as representações de tipo cartográfico têm se tornado cada vez mais importantes nas tecnologias intelectuais de suporte informático. Estudos da psicologia cognitiva mostram que o fato de a memória humana compreender e reter melhor as informações organizadas espacialmente em diagramas e em mapas conceituais manipuláveis favorece o domínio mais rápido e fácil das informações, em relação aos meios áudio-visuais tradicionais. Para LEVY (1993, p.40),

> *"a memória humana é estruturada de tal forma que nós compreendemos e retemos bem melhor tudo aquilo que esteja organizado de acordo com relações espaciais. Lembremos que o domínio de uma área qualquer do saber implica, quase sempre, a posse de uma rica representação esquemática"*.

Uma dos principais pressupostos para navegar bem em um hiperdocumento é entendê-lo. Na ciência cognitiva, a compreensão é geralmente caracterizada como a construção de um modelo mental que represente seus objetos e suas relações semânticas descritas em um texto. (THÜRING, M; HANNEMANN, J. HAAKE, J.M., 1995, p. 57). Essa idéia de redes de memória associativa teve início com BUSH (1945, p.243), quando ele tentou criar um sistema de informações que operava associativamente com a mente: "*Com um item em seu poder, ela* [a mente] *aceita instantaneamente o próximo que é sugerido pela associação de pensamentos, de acordo com alguma intrincada teia de trilhas mantidas pelas células do cérebro...".* Isto nos remete à Teoria da Aprendizagem Associativa, desenvolvida posteriormente por AUSUBEL (1963; 1968; 1978), a qual tem como idéia fundamental a aprendizagem feita por assimilação de novos conceitos. Conforme as concepções desses referenciais, a aprendizagem é a organização das estruturas do conhecimento, que é uma representação da organização de idéias em nossa memória semântica.

Como se pode observar, o modelo da mente humana foi aplicado ao computador desde o começo do desenvolvimento da informática. Segundo DIAS (1994, p.2) a mente era observada como uma rede complexa de conexões internas que poderia ser, em parte, formalizada no *design* de sistemas computacionais. Desta abordagem originaram os modelos de formalização da representação do conhecimento, cujos desenvolvimentos em inteligência artificial (IA) procuram reproduzir a atividade mental do homem em tarefas como a compreensão da linguagem, a aprendizagem e o raciocínio. Este processo de transformar a máquina num simulador dos processos mentais presente nas ciências da computação está profundamente associado às ciências cognitivas, ao formalismo de representação e da atividade de processamento da informação, estabelecendo por outro lado, os limites dos modelos de construção da representação do conhecimento. A tecnologia hipertexto é um modelo não seqüencial de organização da informação orientado para a expansão da atividade de pensamento e a representação do conhecimento e do trabalho colaborativo.

Seguindo o princípio do associacionismo na concepção da máquina *Memex,* BUSH (1945) propõe um conceito de rede para ligar unidades de informação. Para construir um hipertexto eficiente a partir de sua proposta, é preciso combinar habilidades representacionais e conhecimentos técnicos para criar uma rede semântica com alto grau de conectividade, gerando referências cruzadas em todas as suas unidades de informação.

Dois problemas clássicos relacionados aos aspectos cognitivos no hipertexto, freqüentemente citados, são a desorientação do usuário e o transbordamento cognitivo. A desorientação do usuário, ligada principalmente à imprecisão cognitiva, foi tratada humoristicamente por EDWARDS e HARDMAN (1989) no título do artigo *Lost in hyperspace* ("*Perdidos no hiperespaço*"), e se contrapõe à necessidade do usuário saber onde ele se encontra na rede, de onde vem e para onde vai. No texto tradicional, é mais difícil perder-se por causa da existência de mecanismos de referência nas extremidades do documento: o sumário no início e, no final, índices com palavras-chave e números de páginas. Porém, a facilidade de navegação entre os nós induz o usuário a percorrer vários caminhos (ou trilhas) ao mesmo tempo, fazendo com que, em um dado momento, o usuário sinta dificuldade em reconhecer onde ele se encontra e que caminho tomar para chegar aonde precisa. A

desorientação pode manifestar-se com as sensações de confusão e frustração, resultando em grandes períodos de tempo perdidos e uma baixa eficiência na utilização do sistema. Esta tendência é diretamente proporcional ao tamanho e à complexidade do hipertexto.

Um importante e recente campo de pesquisa atual concentra-se na tentativa de desenvolver ferramentas cada vez mais poderosas para a visualização e a organização de estruturas do hipertexto. Deve-se observar que o usuário não pode trabalhar com um nível de complexidade maior que o limite de sua cognição visual, de sua capacidade de mudar o foco de sua visão ou de suprimir detalhes durante o processo de navegação. Segundo CHAIBEN (2002, p.10) muitas estruturas conceituais podem ser representadas melhor e de forma mais eficaz pela exibição gráfica na tela, sugerindo que o problema da desorientação poderia ser corrigido através de algum tipo de mapa.

O transbordamento cognitivo (*cognitive overhead*), ligado principalmente ao excesso de cognição, pode ser tanto um problema de transbordamento mental do autor em criar termos e manter seus relacionamentos com os nós e os *links* na trilha correta, quanto dos usuários na hora de tomar decisões sobre quais *links* seguir e quais devem ser abandonados, devido ao grande número de opções que encontra. O ato de parar em um nó, ao encontrar a informação procurada, pode distrair o usuário, tornando a navegação dispersa, principalmente se houver grande número de nós e *links* a percorrer. Outro problema apresentado pelo transbordamento cognitivo é a dificuldade do usuário em adaptar-se à sobrecarga mental derivada do grande número de operações requeridas em curto espaço de tempo. A necessidade de acompanhar as ligações acarreta uma carga cognitiva adicional, podendo significar que parte da capacidade de processamento das informações seja desviada para as tarefas de tomada de decisão (o que não deixa de ser uma tarefa inerente ao hipertexto), capacidade que poderia concentrar-se mais no material que está sendo pesquisado ou estudado.

Segundo GLEN e CHIGNELL (1992), existem vários fenômenos que podem influenciar o processo de “leitura” em hipertextos. Alguns desses fenômenos que mais se destacam são:

- **Recuperação da memória**: estudos sugerem que a otimização da eficiência em estruturas de recuperação conduz à capacidade máxima de armazenamento de cinco ou sete elementos. A capacidade humana de armazenagem vai depender do nível de familiarização do usuário com relação ao assunto do conteúdo do sistema de recuperação da informação hipertextual.
- **Estímulos à memória de trabalho:** que ocorrem através da forma de memória associativa, pelo estímulo ou predisposição do processo de seleção de um *link* e o uso de estratégias de memorização por meio de associações semânticas.
- **A eficiência da sugestão de recuperação:** o contexto fornecido por um grupo de *links* de um nó pode auxiliar ou impedir a recuperação da informação, dependendo da sugestão de novas informações ou sinais.
- **Decisões de folheio (*browsing*):** os seres humanos algumas vezes utilizam heurísticas cognitivas nas tomadas de decisão complexas mesmo que, em alguns casos, elas possam produzir resultados errôneos. Uma das decisões mais prováveis no folheio parece obedecer à similaridade heurística, na qual o *link* é escolhido com base na sua similaridade com o conceito-alvo. Entre os mecanismos de navegação, o folheio é aquele que exige mais elaboração cognitiva.
- **Estratégias metacognitivas:** desempenhos relativamente pobres na realização de tarefas de memória são geralmente atribuídos a deficiências das estratégias metacognitivas, como as técnicas de repetição (*rehearsal*) e técnicas de memorização (técnicas mnemônicas).

Diferentes estudos identificaram diversas variáveis do processo cognitivo dos usuários que causam impacto na performance de navegação em hiperdocumentos: características individuais como estilo cognitivo ou de aprendizagem (ANDRIS, 1996; KORTHAUER e KOUBEK, 1994), conhecimento prévio (FORD e CHEN, 2000; McDONALD e STEVENSON, 1998; LAWLESS e KULIKOWICH, 1998), nível de interesse (LAWLESS e KULIKOWICH, 1998) e sexo (BEASLEY e VILA, 1992). As demandas das tarefas também influenciam, como por exemplo, se é solicitado ao usuário pesquisar sobre algum material específico (CHEN e RADA, 1996), ou se os usuários são instruídos para estudar em detalhe ou apenas reler o hiperdocumento ao repetir o acesso (BEASLEY e WAUGH, 1997). Características do documento

como, por exemplo, a estrutura inerente do material, dicas (*cues*) de estrutura do documento, e a estrutura dos *links* também podem causar impacto na performance (KORTHAUER e KOUBEK, 1994, p.96). Finalmente, os tipos de ferramentas de navegação que acompanham o documento podem influenciar a eficiência da navegação (BOECHLER e DAWSON, 2002; McDONALD e STEVENSON, 1998; WENGER e PAYNE, 1994).

Segundo DALAL, QUIBLE e WYATT (2000, p.608), uma das principais causas do *design de* hipertexto cognitivamente pobre é a falta de regras no desenvolvimento de aplicativos de hipermídia (resultante da falta de princípios de organização), que sejam aceitos internacionalmente para todos os espaços de informação eletrônica multi-dimensional. Os autores afirmam que isto não quer dizer que não existam regras ou princípios de *design* de hipertexto. Ao contrário, existe um grande número de guias de construção de *Web design*. Porém, a maioria deles são baseados no universo restrito dos próprios *designers* e no senso comum, não se fundamentando em referenciais reconhecidos da ciência da informação e nem contando com validação experimental. THÜRING, HANNEMAN e HAAKE (1995, p.57) lamentam que *"as questões cognitivas e do processamento humano da informação ainda continuem amplamente negligenciados e exerçam pouca influência no design de hipermídia".*

Para melhorar a leitura em hiperdocumentos, não é suficiente concentrar-se somente nos aspectos técnicos da navegação. Há que se considerar a compreensão do usuário, o que implica na construção de interfaces com o objetivo de reduzir o esforço mental do mesmo, buscando coerência no trabalho mental requerido para atividades de navegação e orientação. Para HANNEMANN, THÜRING e HAAKE (1993, p.5), a facilidade da compreensão do conteúdo de hiperdocumentos pode ser atingida se três questões forem respondidas: (1) como inserir uma estrutura coerente no documento?; (2) como transmitir esta estrutura para o leitor?; (3) como dar suporte à navegação e à orientação do leitor?

Para evitar a desorientação do usuário e diminuir suas dificuldades de compreensão, uma solução adequada poderia permitir ao usuário criar uma representação mental coerente do hiperdocumento que está sendo consultado. De acordo com DALAL, QUIBLE e WYATT (200, p.609), o modelo mental nesse contexto é a representação

mental de objetos e as relações semânticas no hiperdocumento, que o usuário constrói quando está processando nova informação. Se o documento consistir de um texto somente - uma tese, por exemplo - o modelo mental é representado pelo assunto tratado pelo texto. Para caracterizar o modelo mental, os autores do hipertexto fazem referência a GLENBERG e LANGSTON (1992),[7] que propõem as seguintes características de um modelo mental:

1) Constituir uma representação mental que faz uso da memória de trabalho (ativa), em particular a memória de "rascunho espaço-visual" (*visuo-spatial scratchpad*) e que, por isso, tem capacidade limitada.
2) Consistir de elementos representacionais do texto e figuras arranjadas em um meio espacial da memória de "rascunho espaço-visual", elementos que devem apontar para informações proposionais e perceptivas na memória de longa duração.
3) Refletir o entendimento atualizado do leitor sobre o texto, sendo o modelo o resultado da atualização do progresso da leitura. Esta atualização é aperfeiçoada por adição, modificação e exclusão do conteúdo e pela localização dos elementos representados no meio espacial do usuário.

Pode-se concluir que, para o usuário construir um modelo mental com rapidez e clareza, são necessários mecanismos de *design* que suportem o processo de formação do modelo mental. Sumariando, deve-se ressaltar (1) a importância da coerência na construção textual do hipertexto, (2) o cuidado de se prover boa orientação e navegação e (3) um *design* da interface que evite o transbordamento cognitivo na construção mental do leitor, que contornoem os problemas de desorientação. A seguir, serão descritos metodologias e modelos para estruturação de sistemas de hipertextos selecionados na literatura.

### 2.2.2 Metodologias para estruturação de sistemas de hipertexto

O *design* dos sistemas de hipertexto pode ser considerado um desafio a seus criadores porque difere muito daqueles *designs* dos sistemas de informação

[7] GLENBERG, A. M. ; LANGSTON, W. E. Comprehension of illustrated text: pictures help to build mental models. *Journal of memory and language*, v.31, n.2, p.129-151.1992

tradicionais. Por um lado, torna-se evidente a dicotomia entre a autoria do texto e a autoria de seu *design*. Por outro, propõe-se lidar, cada vez mais, com a diversidade e quantidade crescentes de unidades de informação. Os primeiros sistemas de hipertextos estruturados em pequenos documentos funcionaram satisfatoriamente. A estruturação de grandes volumes de informação, entretanto, passou a requerer habilidades estruturais e de *design* mais complexas. Parece natural que, somente a partir daí, é que começaram a surgir, na literatura, a apresentação de metodologias e modelos propostos sobre *design* e interfaces para hipertextos.

Um dos primeiros desafios na criação do hipertexto é estruturar o conhecimento de maneira que a visão geral possa ser apresentada ao leitor no nó inicial (ou raiz). Essa visão deve identificar as idéias-chave e amplitude (cobertura), para que os leitores imediatamente formem um modelo mental dos tópicos cobertos. Dependendo da aplicação e do tipo de informação, a forma mais comum de estruturação de documentos é a forma hierárquica. Uma estrutura hierárquica sem *links* cruzados é também chamada de estrutura em árvore. Já as hierarquias com *links* cruzados entre elas constituem uma estrutura de rede. Atualmente, um dos problemas mais recorrente na estruturação de um hipertexto é a fragmentação de unidades do conhecimento (característica inerente ao próprio hipertexto), sem perder o conteúdo semântico da informação.

Existem metodologias e modelos de *design* e construção de sistemas de hipertextos adequados ao momento da autoria (criação) e ao momento da recuperação (folheio e busca), incluindo as fases pré e pós-*Web*, que se destacaram especialmente a partir do final da década de 1980, com o desenvolvimento da Web semântica.

AVISON e FITZGERALD (1995, p.10) definem metodologia de criação de hipertextos como:

> *"uma coleção de procedimentos, técnicas, ferramentas e documentos os quais ajudarão os criadores de sistemas em seu esforço de implementar um novo sistema de informação. As metodologias consistem em fases, que se subdividem em sub-fases, as quais irão guiar os criadores de sistemas na sua escolha de técnicas que podem ser mais apropriadas para cada estágio do projeto e também ajudá-los a planejar, administrar, controlar e avaliar projetos de sistemas de informação."*

KEMP e BUCKNER (1999, p.16) apontam alguns elementos da metodologia de *design* de hipertextos: (a) modelos (conceitual, lógico e físico); (b) requisitos de

análise e especificação; (c) métodos de representação; (d) ferramentas e (e) guia ou documentação para assegurar a consistência do processo de criação.

Embora não existam ainda metodologias de estruturação de hipertextos consolidadas ou internacionalmente adotadas, já se fala em algumas recomendações que orientam a modelagem de hiperdocumentos. SHNEIDERMAN (1989, p.115) menciona "três regras de ouro do hipertexto": (1) a organização de um grande corpo de informações em numerosos fragmentos; (2) a existência de relacionamentos entre os fragmentos, e (3) a necessidade de acesso do usuário a uma pequena fração de cada vez. A convergência destas três características indica que a adaptação de um universo de informações à forma de hipertexto é possível e vantajosa. Inversamente, se o usuário necessita de grandes blocos de informação ou de muitos blocos de informação simultaneamente ou, ainda, se estes não estão relacionados entre si ou se não é possível fragmentar a informação, a forma de hipertexto, provavelmente, não é a mais adequada.

As contribuições de ROSENFELD e MORVILLE (1990) ocorrem em um nível mais simples, com considerações sobre técnicas de *design* gráfico e outros procedimentos, porém com o foco na estrutura da informação. Como eles, MARTIN (1992) também dá orientações sobre a construção eficiente de hipertextos. Discute questões de estilo, objetividade, clareza, estratégias de estruturação da informação e maneiras de ajudar os leitores. Também relaciona a criação de hiperdocumentos com a estruturação de documentos em papel ou livro, dando enfoque para a questão da autoria.

MARCHIONINI (1994, p.87) sugere a utilização de oito algoritmos para criar um hipertexto baseado nos moldes da elaboração de índice de um documento: (1) identificação das principais facetas do tópico; (2) criação de uma lista exaustiva de termos e frases; (3) mapeamento de termos e frases para facetas, revisando-as se necessário; (4) determinação dos termos/conceitos preferidos (*label nodes*); (5) escrita ou introdução dos textos (nodos de informação), como o estabelecimento de referências cruzadas (*links*) para outros nodos durante a criação do hipertexto; (6) revisão do conjunto de textos (nodos) e *links* para outros nodos de acordo com os critérios gramaticais, de estilo, de facilidade de leitura etc.; (7) importação dos

arquivos revisados para o sistema de hipertexto, implementando os *links;* (8) teste e edição do hiperdocumento final.

KEMP e BUCKNER (1999, p.11) descrevem esforços para construir uma metodologia de *design* de hipertextos com vistas ao usuário final na construção de hiperdocumentos de pequeno a médio porte. Os autores fazem uma revisão de literatura das metodologias apresentadas por autores como HORTON (1994)[8], WATERWORTH e CHIGNALL (1993)[9], HARDMAN e SHARRAT (1990)[10], THIMBLEBY (1990)[11]; e DILLON (1989)[12]. Depois, baseados na análise desse material e nas suas próprias experiências, os autores propõem uma síntese, cujos procedimentos metodológicos são (1) checagem dos itens a serem considerados e sua exeqüibilidade; (2) análise do perfil do usuário em potencial; (3) análise das etapas de construção do hipertexto; (4) requisitos para análise e especificações; (5) análise de conteúdo da informação; (6) *design* e estruturação da informação; (7) criação dos *storyboards* e inserção da informação; (8) criação dos *links* e dos acessos; (9) *design* da interface do usuário; (10) teste e avaliação; e (11) manutenção do sistema.

Maria Luiza de Almeida Campos (Universidade Federal Fluminense), na sua tese *A organização de unidades do conhecimento em hiperdocumentos:* um modelo conceitual como um espaço comunicacional para a realização da autoria, propõe sete requisitos para o desenvolvimento de uma metodologia para a elaboração de modelos conceituais em sistemas de hipertextos. A autora organiza estes requisitos em três níveis de compreensão e, em cada nível, sugere procedimentos metodológicos correspondentes (CAMPOS, 2000, p.146-160).

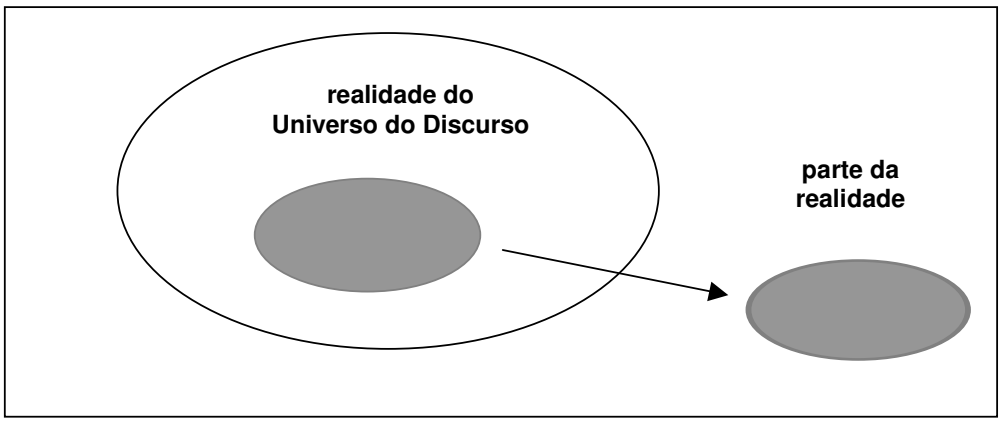
O primeiro nível, definido pela autora como **Nível de entendimento da forma de abordagem do assunto do hiperdocumento**, ocorre na primeira etapa que

---

[8] HORTON, W. *Designing and writing on-line documentation*. 2 ed. New York: John Wiley, 1994.
[9] WATERWORTH, J. A. , CHIGNELL,M.H. e ZHAI, S.M. From icons to interface models: designing hypermedia from the bottom up. *International Journal of Man-Machine Studies*, v. 39, p.453-472, 1993.
[10] HARDMAN, L. e SHARRATT, B. *User-centered hypertext design: the application of HCI design principles and guidelines*. IN: McAleese, R. e Green,C. (Eds.). Hypertext: State of the art.Oxford:Intellect.1990. p.252-259.
[11] THIMBLEBY, H. *User interface design*. Wokingham: Addison-Wesley, 1990.
[12] DILLON, A. *Human factors issues in the design of hypermedia interfaces*. In H. Brown. Proceedings, Hypermedia/Hypertext and object-oriented databases. Uxbridge:Unicom Seminars Ltd. P.43-50.

equivale à *"estruturação de um domínio do conhecimento, que inicia com a separação de certa parte da realidade do universo do discurso"* (CAMPOS, 2000, p.5), como ilustrado na FIG.2. Esse nível tem quatro requisitos que tratam (1) da determinação do domínio do conhecimento, (2) do método de raciocínio utilizado para representação das unidades do conhecimento, (3) do tipo de leitor e (4) da tipologia documental do hiperdocumento.

FIGURA 2 - Separação de parte da realidade do Universo do Discurso

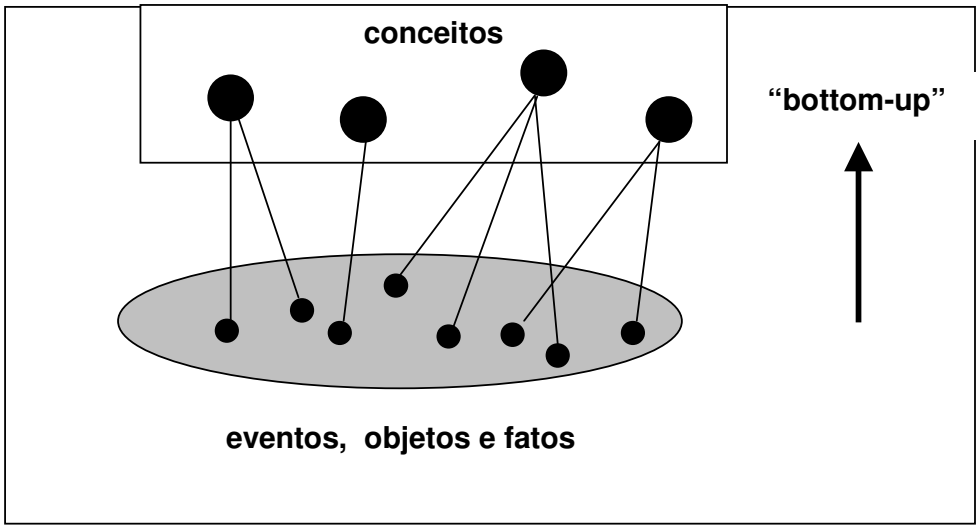



FONTE: Baseado em OESER, [13] citado por CAMPOS (2000, p.5)

O segundo, **Nível da organização das unidades de conhecimento na construção da narrativa do hiperdocumento**, consiste em *"analisar a abstração",* com a determinação dos conceitos e sua estruturação em domínios de conhecimento específicos *"com ênfase nos eventos, objetos ou fatos individuais",* utilizando-se *o* método *"bottom-up"* (CAMPOS, 2000, p.5). A autora apresenta dois requisitos: um que trata da natureza do conteúdo das unidades de conhecimento e outro, que trata do estabelecimento das relações entre elas (FIG.3).

[13] OESER, E. Theoretical principles of terminological knowledge engineering. IN: TKE'93. *Theoretical principles of terminological knowledge engineering*. Frankfurt: Indeks-Velag, 1993, p.34-38.

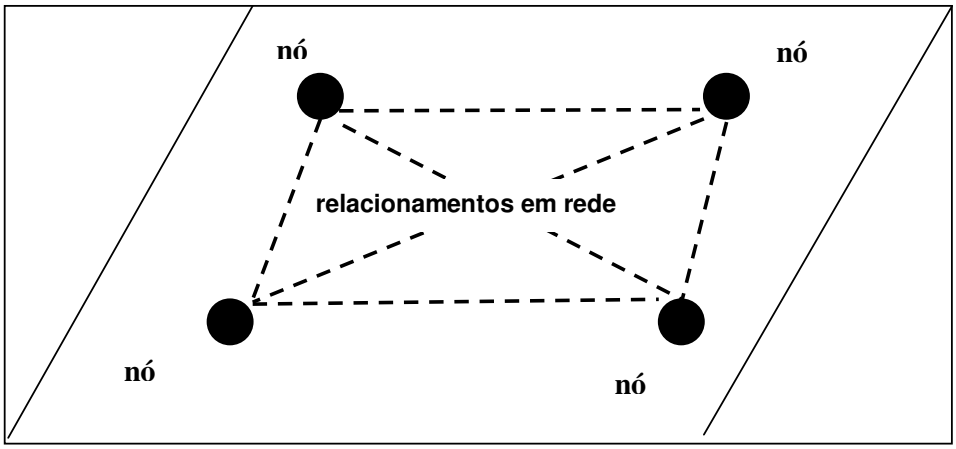
FIGURA 3 - Análise da abstração, determinação dos conceitos e sua estruturação em domínios de conhecimento pelo método *"bottom-up"*




FONTE: Baseado em OESER, citado por CAMPOS (2000, p.6)

No terceiro, **Nível de estabelecimento de um veículo de comunicação e expressão sobre a temática do hiperdocumento**, CAMPOS (2000, p.6) propõe uma "*síntese da abstração*", em que há a elaboração dos nós conceituais e seus relacionamentos em rede, formando um modelo conceitual em um hiperdocumento e que pode ser representado graficamente (FIG.4).

FIGURA 4 - Síntese da abstração, com seus nós (conceitos) e relacionamentos em rede




FONTE: Baseado em OESER, citado por CAMPOS (2000, p.6)

### 2.2.3 Modelos para estruturação de sistemas de hipertexto

Modelos são representações que auxiliam a visualização de um problema, pois sua estrutura baseada na representação de unidades maiores em unidades menores pode clarear o caminho de análise das partes, depois estabelecer uma compreensão por síntese e, finalmente, chegar a uma nova percepção do todo. Assim como os problemas mudam, os modelos, como outros tipos de ferramentas, moldam a consciência daqueles que a utilizam, que por sua vez re-moldam a ferramenta, num processo *ad infinitum*. A elaboração de novos modelos nos leva a níveis mais altos de compreensão sobre o problema de *design*.

RYDER (2003, p.1) descreve "modelo" como a representação que auxilia a visualização de um problema. Sua estrutura, baseada na fragmentação de unidades maiores em unidades menores, busca esclarecer o caminho de análise das partes, depois compreensão por síntese e finalmente, uma percepção do todo. Modelos e mitos nos ajudam a dar sentido ao mundo, independentemente se derivam da imaginação ou de pesquisa sistematizada. Um modelo de *design* instrucional permite estruturar e dar significado a um problema de identificação, permitindo o suposto *designer* negociar sua tarefa com uma imagem da compreensão consciente.

Ainda segundo RYDER (2003, p.1), a avaliação da eficiência do modelo deveria ser feita com base na mediação entre a intenção do *designer* e a tradução das intenções específicas do usuário, como o primeiro compartilha sua carga de informações com o segundo e como efetivamente muda-se o foco do *designer* para o objeto do *design*. A seguir, são apresentados os modelos de *design* de hipertextos/hipermídia pré e pós-*Web* mais conhecidos desde o final da década de 1980.

De acordo com MILLARD (1999, p.8), ao final da década de 1980, pesquisadores de sistemas de hipertextos concluíram que a maior falha dos sistemas de hipertextos até então desenvolvidos, era sua inabilidade de interoperação. Esses sistemas não interagiam satisfatoriamente na troca de informações com o usuário, gerando demandas, que motivaram o surgimento de novas metodologias de *design* de hipertexto.

O primeiro modelo para *design* de hipertextos, chamado de ***Asynchronous Design***, foi desenvolvido por PERLMAN (1989) na Ohio State University (EUA) e surgiu, inicialmente, a partir da proposta de ser um modelo de desenvolvimento para pesquisa em hipertexto com fluxogramas, na tentativa de refletir o processo assincronizado de *design* e avaliação de sistemas de hipermídia. Nesse modelo, o processo de *design* é contínuo, baseando-se em um modelo interativo. Como o próprio nome diz, não precisa seguir nenhuma seqüência, podendo iniciar-se em qualquer um de seus procedimentos.

HALASZ e SCHWARTZ (1990), dos laboratórios norte-americanos Xerox PARC e Tektronics, respectivamente, descrevem o ***Dexter Reference Model****,* desenvolvido com o objetivo de permitir uma comparação mais facilitada entre sistemas, incorporando padrões de interoperabilidade. Este modelo apresenta três camadas - (A) a camada interna dos componentes (conteúdo/estrutura interna e os nós); (B) a camada de armazenamento (base de dados contendo a rede de nós e *links*); (C) a camada de tempo real; e duas interfaces: (1) especificações de apresentação, que operam entre a camada de tempo de execução e a camada de armazenamento e (2) interface âncora, que opera entre as camadas de armazenamento e a camada interna dos componentes. Em resumo, o principal objetivo deste modelo é fornecer as bases principais para comparação de sistemas e, aí, desenvolver padrões de intercâmbio e interoperabilidade de informações, como mostra a FIG.5.

FIGURA 5 - Camadas do *Dexter Reference Model*

FONTE: HALASZ e SCHWARTZ (1994, p.4)

A típica modelagem em camada usada no processo de *design* de um aplicativo inclui (1) o nível conceitual ou estrutural (estrutura e recorte da informação), (2) o nível hipertextual (composição e estrutura da navegação do aplicativo), (3) o nível de apresentação/interface, (4) o nível de personalização (*design* personalizado) e (5) o nível de implementação. No nível conceitual, o campo ou domínio da informação é capturado e modelado usando-se três técnicas de *design* principais (as duas primeiras surgidas nos anos 1990):

- Entidade-relacionamento (E-R): baseada no Modelo de Relacionamento com a Entidade, em que a informação, os objetos e os dados estruturados são descritos pelo significado das entidades e relacionamentos;
- Orientação a objetos (OO): objetos de informação modelados como objetos/classes.
- Ontologia-base: mais recente, esse modelo trata objetos de informação como classes da ontologia[14].

CASANOVA et al. (1991), pesquisadores da Universidade Católica do Rio de Janeiro e do Centro Científico da IBM do Brasil, criaram o ***Nested Context Model – NCM***, composto por três estruturas: a estrutura de definição, a estrutura de apresentação e a estrutura de navegação. Trata modelagem da estrutura do hipertexto em si mesma, sem se preocupar com os aspectos do domínio dos assuntos do hipertexto. Segundo SANTOS et al. (1998, p.114), um documento hipermídia é definido no modelo ***NCM*** a partir dos conceitos usuais de nós e *links*, nos quais os nós são constituídos de fragmentos de informação e os *links* são usados para exprimir todas as possíveis relações entre os diferentes nós que constituem o documento.

No ***NCM***, as características principal são o nó de conteúdo, a lista de *links* e o descritor. Cada *link* define o lugar do conteúdo do nó. Cada nó tem um *link default*, que apresenta o conteúdo total do nó. O descritor contém informações, especificando como o nó deve ser apresentado, no tempo e no espaço. O ***NCM*** distingue duas classes básicas de nós, os nós terminais (ou de conteúdo) e os nós de composição. Os nós de conteúdo contêm um conjunto de dados cuja estrutura interna depende do aplicativo, como os nós tradicionais de hipermídia, podendo ser

[14] The *World Wide Web* Consortium (*W3C*) descreve ontologia como a coleção de termos usados para definir e representar uma área do conhecimento, geralmente chamada de domínio, para ser usada por pessoas, base de dados, e aplicações para compartilhar informações.

especificados em outros formatos de texto, gráfico, áudio, vídeo, etc. Os nós de composição são o ponto central do modelo. Cada um deles contem uma coleção de *links* e nós (de conteúdo ou de composição) interrelacionados.

GARZOTTO, PAOLINI E SCHWABE (1993), pesquisadores da Universidade Católica do Rio de Janeiro e do Instituto Politécnico de Milão, propuseram o ***Hypermedia Design Model* – HDM** para definir estruturas e interações em larga escala em sistemas de hipermídia de leitura somente. Este modelo é adequado para domínios com um alto nível de organização, modularidade e consistência. WOUKEU et al. (2003, p.4) descrevem sucintamente seu funcionamento e os três tipos de *link* que utiliza:

> *"Foca hierarquicamente na descrição dos objetos de informação em termos das entidades formadas pelos seus componentes, contendo as unidades de informação e a estrutura de navegação, independentemente da sua implementação. Sua estrutura de navegação compreende os links de perspectiva entre as unidades, os links estruturais entre os componentes e os links de aplicação entre as entidades, componentes (ou unidades), índice e "visitas monitoradas" (guide tours). Seu design de apresentação consiste de slots (unidades de informações) e moldura (agrupamento de slots)."*

ISAKOWITZ, STOHR e BALASUBRAMANIAN (1995), pesquisadores da New York University e Boston University, criaram o ***Relationship Management Methodology – RMM***, a partir da técnica E-R utilizando, ao mesmo tempo, as abordagens *top-down* e *bottom-up* na construção do aplicativo de hipertexto através de sete fases: (1) *design* E-R, (2) *design* por partes (*slice design*), com agrupamento fracionado das características das entidades como nós/unidades de representação, (3) *design* navegacional (métodos de acesso), (4) *design* de protocolo de conversão (componentes de conversão) (5) *design* da interface do usuário (*layout* da tela), (6) *design* de comportamento em tempo real e (7) construção e teste.

O ***Enhanced Object-Relationship Model – EORM*** (LANGE, 1996) foi o primeiro modelo baseado na metodologia de orientação a objetos cuja principal característica é a representação dos relacionamentos entre objetos (*links*), nos quais as entidades se transformam em nós e as relações se transformam em *links*. É descrito como o processo interativo que possibilita a construção de um protótipo de interface para o usuário no início do processo de *design*.

SCHWABE e ROSSI (1995), da Universidade Católica do Rio de Janeiro, propõem, no ***Object Oriented Hypermedia Design Model – OOHDM,*** um modelo de *design* baseado na orientação a objetos que permite a especificação do aplicativo da hipermídia como forma navegacional em detrimento do modelo conceitual. O conceito das classes navegacionais e o contexto navegacional são introduzidos para descrever a estrutura navegacional. O processo de *design* no ***OOHDM*** compreende quatro tarefas: (1) modelagem conceitual, (2) modelagem navegacional, (3) *design* da interface abstrata e (4) implementação, classificação, agregação e generalização/especialização, que são utilizadas em todo o processo como tentativa de melhorar o poder de abstração e a interação. A modelagem conceitual neste modelo é construída a partir dos princípios de modelagem OO, tendo como principal objeto a definição das classes, subsistemas e relações de acordo com os conceitos do domínio do aplicativo. As classes conceituais são construídas utilizando-se as hierarquias de agregação e de generalização/especialização. O resultado é um esquema conceitual, construído fora do sub-sistema, com classes e relações. Nota-se que o modelo ***OOHDM*** tem como principal preocupação a estrutura ou conteúdo dos objetos, não realizando ainda o passo que seria representar o conteúdo informacional. A modelagem de navegação parte do contexto navegacional, que é induzido pelas classes de navegação como os nós, *links*, índices e guias. Os nós no ***OOHDM*** representam janelas lógicas para as classes conceituais definidas no momento da análise do domínio. O contexto navegacional e as classes levam em consideração o tipo de usuário para o qual se destina o aplicativo e quais tarefas que deverão ser desempenhadas para utilizá-la. Podem ser construídos diferentes modelos navegacionais, utilizando-se o mesmo modelo conceitual, para expressar diferentes visões do mesmo domínio do conhecimento. O modelo de interface é construído pela definição dos objetos que os usuários irão perceber. As classes de interface são definidas como agregação das classes primitivas, tais como textos e ícones. Nesta etapa, são definidas as metáforas da interface e os estilos de interação. A implementação do modelo pode envolver elaboradas arquiteturas, por exemplo, o cliente servidor. Apesar de ser considerado um modelo de difícil compreensão, porque utiliza vocabulário muito técnico, o que é um obstáculo aos profissionais estranhos à área de computação, o ***OOHDM*** tem sido utilizado para grande número de aplicações em *CD-ROM* e em *sites* da *Web*.

DE TROYER e LEUNE (1997), da Tilburg University (Holanda), centram a abordagem do ***Web Site Design Method – WSDM*** no usuário, de forma que seu modelo de navegação se baseia nas necessidades de informações deste e nas suas preferências de aplicativos na Web. O modelo do ***WSDM*** consiste em três fases de procedimentos: modelagem do usuário, modelagem conceitual e implementação. A modelagem conceitual pode consistir na modelagem do objeto – que pode ser baseado em E-R ou OO - e no *design* da navegação. Na fase de modelagem, os usuários da *Web* em potencial e suas atividades, tidos como ponto central, são identificados e classificados. O *design* da navegação consiste em determinar diversos caminhos, um para cada perspectiva, expressando como o usuário de uma perspectiva particular pode navegar através da informação disponível. Cada caminho de navegação consiste em três camadas: contexto, navegação e informação. A camada do contexto é a de nível mais alto no caminho da navegação, começando por este componente. A camada de informação é a de nível mais baixo no caminho da navegação. A camada de navegação conecta as duas camadas externas. Para acessar a informação intermediária, são criados componentes e *links*, como por exemplo, índices. A fase da implementação tem como meta a criação de interface amigável consistente e eficiente para o modelo conceitual.

O ***Scenario-Based Object-Oriented Hypermedia Design Methodology – SOHDM***, criado por LEE, LEE e YOO (1999, p.123), pesquisadores da escola de Administração do Instituto Avançado de Ciências e Tecnologia da Coréia, apresenta outro modelo com abordagem baseada em OO, com foco nos sistemas de hipermídia de processo orientado para dar suporte ao processo organizacional. Apesar de ter muitas similaridades com a ***RMM***, o ***OOHDM*** e o ***EORM***, os criadores dessa metodologia apregoam suas vantagens em relação a seus similares, afirmando que permitem maior flexibilidade para os sistemas de hipermídia porque esses (1) não refletem as necessidades de navegação dos usuários nos seus relacionamentos de informação orientada,; (2) não contemplam suficientemente a integração da hipermídia com a base de dados desenvolvida, visto sua importância para implementação do tipo intranet; e (3) adotam modelos de dados conceituais para capturar as necessidades dos usuários. O modelo ***SOHDM*** consiste em seis fases: (1) análise do domínio, (2) modelagem do objeto, (3) *design* da interface (visão), (4) *design* da navegação, (5) *design* de implementação e (6) construção do aplicativo. Desde os primeiros passos do desenvolvimento do sistema, esta

metodologia identifica os requisitos para as aplicações de hipermídia. Cenários são utilizados com intuito de melhorar a capacidade expressiva de modelagem. Em particular, na interface de visão da OO, são utilizadas unidades de navegação como também a visão lógica do usuário. Os cenários, que determinam o acesso à estrutura de nós, são definidos durante a análise do domínio e servem como base para a modelagem de objetos e para o *design* da navegação. Diferentes interfaces do tipo OO podem ser geradas, oriundas do modelo de objeto de domínio para compor um novo aplicativo. Definidos como mecanismos de "menu", permitem ao usuário acessar outras partes dos documentos de hipermídia. Estas estruturas de acesso, juntas com a visão da orientação a objetos, são chamadas de unidades de navegação. A identificação dos *links* de navegação completa o *design* de navegação. Durante a fase de implementação da interface do usuário, são modelados as páginas e o esquema lógico da base de dados.

O ***Web Modeling Language – WebML, criado por*** CERI, FRATERNALI e BONGIO, 2000) do Instituto Politécnico de Milão, é um modelo dirigido com uma abordagem de *design* baseada em E-R que permite uma especificação conceitual e implementação automática de *Web sites* com grande concentração de dados. O ***WebML*****,** cujos os conceitos são associados à notação gráfica e sintaxe textual do *XML,* possibilita alto nível de descrição através de quatro dimensões ortogonais: (1) dados de conteúdo (modelo estrutural), (2) páginas formadoras (modelo de composição), (3) tipologia dos *links* entre as páginas (modelo navegacional), (4) *layout* e requisitos gráficos para a exibição das páginas (modelo de apresentação), além de características personalizadas para a disponibilização de conteúdo "um-para-um" (modelo personalizado). O ***WebML*** permite uma abordagem de modelo dirigido para o desenvolvimento de *Web sites*, que é o fator principal para a geração da ferramenta original CASE, apropriada para a construção de *sites* complexos, apoiando características avançadas como o acesso a múltiplas máquinas, personalização e gerenciamento diferenciado.

Segundo WOUKEU et al. (2003, p.9) muitos desenvolvimentos da *Web* semântica têm focado em questões relacionadas com a modelagem do conhecimento e publicação do conhecimento (ontologia, bases de conhecimento) e como resultado, tendem a deixar de lado a complexidade dos documentos voltados para os usuários. Porém, as ontologias têm influenciado no desenvolvimento recente de vários

hipertextos, alguns dos quais buscam estabelecer uma ponte entre uma *Web* legível pelo ser humano e uma *Web* semântica processável por computadores. Os hipertextos baseados em ontologias constróem as estruturas de seus componentes baseados em relacionamentos entre objetos virtuais e o mundo real. Abaixo, são descritas três dessas principais aplicações.

O ***Model-Driven Ontology-Based Web Site Managment – OntoWebber*** (YUHUI, DECKER e WIEDERHOLD, 2001), criado na Standford University, USA, resulta de uma abordagem ontológica de criação de *sites* para leitura somente (*ready-only*), permitindo uma integração de fontes de dados heterogêneas para criar portais *Web* com grande concentração de dados. A ontologia de um domínio fornece a terminologia de referência para integração de dados e modelagem do conteúdo, que é o ponto de partida para o *design* de todo o *site*. A modelagem da visão do *site* (*storyboard*) consiste nos modelos de navegação, conteúdo e apresentação. Outras fases incluem a personalização e a manutenção dos modelos. As unidades de navegação ou nós são chamadas de cartões, os quais são mapeados nas classes da ontologia através do modelo de conteúdo. O *design* geral é representado por um meta-esquema baseado no *XML*, por meio da utilização de três recursos: *RDF (Resource Description Framework*), *DAML* (*DARPA Agent Markup Language*) + e *OIL* (*Ontology Inference Layer*). A linguagem *DAML* foi desenvolvida como uma extensão da linguagem *XML* e do padrão *RDF*. O sistema e a metodologia finais dão suporte para a criação de especificações reutilizáveis nos *sites* da *Web*. O ***OntoWebber*** foi utilizado para a criação do *Portal da Comunidade da Web Semântica* como parte do projeto *OntoAgents*, que pretende auxiliar a comunidade de pesquisa no compartilhamento do conhecimento de maneira conveniente e efetiva na *Web* semântica.

Os alemães HANDSCHUH, STAAB e MAEDCHE (2002), da Universidade de Karlsrube, da Ontoprise e do Centro de Tecnologias Informacionais FZI, basearam o ***CREAting Metadata - CREAM*** em ontologia para metadados e criação de documentos hipertextuais. A ferramenta *Ont-O-Mat* possui um sistema de autoria e anotação construído por um editor de documentos e *browser* de ontologia. O *CREAM* suporta a criação de uma *Web* Semântica (de conhecimento) através de anotações tanto durante como depois da autoria (da criação). As anotações são realizadas através do preenchimento de planilha de conhecimento sob o controle do

*browser* de ontologia, tanto pela digitação dos valores quanto pela condução e lançamento “palavra-por-palavra” do editor de documentos. Um aspecto interessante é que os documentos (ou fragmentos de documentos) podem ser construídos pelo processo reverso da anotação: entradas oriundas da ontologia ou da base de conhecimento podem ser utilizadas para criar textos. A maior preocupação com da estrutura *CREAM* é a de gerar conhecimentos que possam ser utilizados nas aplicações da *Web* semântica, a exemplo das buscas, inferência e geração de portais estruturados. Por esse motivo, esta estrutura usa ontologias com o propósito básico de anotação e, por isso, alcança suporte limitado para o conteúdo de autoria como sub-produto da atividade de anotação. Todavia, este suporte é significativo porque, na forma embrionária, a estrutura impõe um esquema íntegro de conhecimento que, de outra forma, se constitui documento textual de forma livre no momento da criação.

O projeto ***Conceptual Open Hypermedia Service Environment - COHSE*** (http://cohse.semanticweb.org), desenvolvido no Departamento de Eletrônica e Ciência da Computação da Universidade de Southhampton (Inglaterra) e financiado pelo EPSRC *Distributed Information Management Programme* (DIM), iniciou suas operações oficialmente em dezembro de 1999, mas foi completado somente em março de 2002. Tem como objetivo pesquisar modelos para melhorar significativamente a qualidade, consistência e amplitude da navegação de documentos da *Web* no momento da autoria (criação de documentos) e no momento da recuperação (folheio dos documentos). Para sua implementação, foram utilizadas três tecnologias: (1) ontológica, para representar um modelo conceitual sofisticado com os termos dos documentos e suas relações; (2) de rede, baseada em *links* de hipermídia abertos, que podem prover grande número de *links* que facilitam no dimensionamento da estrutura e não obstrução da configuração (3) de integração do serviço de ontologia e do serviço de *links* de hipermídia aberto, para formar um sistema de hipertexto conceitual que permita documentos serem ligados através dos metadados que descrevem seus conteúdos.

O ***ONTOPORTAL*** (http://www.ontoportal.org.uk/), também desenvolvido no Departamento de Eletrônica e Ciência da Computação da Universidade de Southhampton (Inglaterra), foi inicialmente financiado pela Agência de Pesquisa e Avaliação de Defesa (*DERA*). A meta-camada semântica dos conceitos da ontologia

e seus relacionamentos podem ser exemplificados ou projetados sobre um recurso *Web* pré-existente e fracamente inter-linkados, para gerar um portal *Web* significativo descrevendo e ligando os recursos e suas relações. A estrutura apresenta quatro modos principais de interação com os usuários: (1) exploração (folheio); (2) captura de conhecimentos (criação ou atualização do conteúdo); (3) discussões entremeadas no folheio; e (4) pesquisa por palavras-chave nos metadados armazenados. Os novos *links* conceituais são deduzidos do sub-domínio da estrutura ontológica para enriquecer a ligação entre os recursos e possibilitar que buscas complexas sejam respondidas simplesmente ao se percorrer *links* ontológicos (*query by linking*).

A apresentação dos modelos acima, sumariados na FIG.6, mostra que, apesar do objetivo comum de produzir hipertextos e *sites* da *Web* bem estruturados e organizados, eles ainda não satisfazem os paradigmas autorais de lidar com as interconexões dos fragmentos semânticos contidos nos textos. As pesquisas mais recentes que tentam, através de ontologias, representar um modelo semântico com os termos dos documentos e suas relações, aparecem como alternativas, mas ainda não apresentaram soluções para as questões da desorientação durante a navegação e do transbordamento cognitivo.

Figura 6 - Principais modelos de *design* de hipertexto e suas características

| Modelos | Processo de design | Estrutura de navegação | Interatividade | Técnica de modelagem |
|---|---|---|---|---|
| *Asynchronous* | 1. Processo assincronizado | Interativa | Somente leitura | Fluxograma |
| *Dexter* | 1. Camada interna de componentes<br>2. Camada de armazenagem<br>3. Camada em tempo real (interface) | Nós<br>*Links* entre as unidades<br>Interface | Somente leitura | interoperabilidade |
| *NCM* | 1. Estrutura de definição<br>2. Estrutura de apresentação<br>3. Estrutura de navegação | Nós<br>*Links* entre as unidades | Somente leitura | Modelagem hipertextual |
| *HDM* | 1. Autoria em larga escala<br>2. Autoria em pequena escala | *Links* entre unidades<br>*Links* entre componentes<br>*Links* entre entidades | Somente leitura | Entidade-Relacionamento |
| *RMM* | 1. Design E-R<br>2. Design por partes<br>3. Design navegacional<br>4. Design de protocolo de conversão<br>5. Design da interface do usuário<br>6. Design de comportamento em tempo real<br>7. Construção e teste | Esquemas<br>Partes<br>Unidades de navegação | Somente leitura | Entidade-Relacionamento |
| *EORM* | 1. Estrutura de classes<br>2. Estrutura de composição<br>3. Estrutura GUI | Classes navegacionais<br>(Unidades de navegação) | Somente leitura | Orientação a objetos |
| *OOHDM* | 1. Modelagem, conceitual<br>2. Modelagem navegacional<br>- esquema das classes de navegação<br>- esquema do contexto de navegação<br>3. Design da interface<br>4. Implementação | Unidades de navegação<br>(Unidades de navegação) | Somente leitura | Orientação a objetos |

**Continua na próxima página)**

| | | | | |
|---|---|---|---|---|
| *WSDM* | 1. Modelagem do usuário<br>2. Modelagem conceitual:<br>- modelagem do objeto<br>- design de navegação<br>3. Implementação do design<br>4.Implementação | Classes navegacionais<br>(Unidades de navegação) | Somente leitura | Entidade-Relacionamento<br>Orientação a objetos |
| *SOHDM* | 1. Análise do domínio<br>2. Modelagem do objeto<br>3. Design da interface<br>4. Design da navegação<br>5. Design da implementação<br>6. Construção do aplicativo | Classes | Somente leitura | Orientação a objetos |
| *WebML* | 1.Modelo estrutural (*design* de dados)<br>2.Modelo Hipertextual:<br>- modelo de composição<br>- modelo de navegação)<br>3. Modelo de apresentação (*layout* e requisitos gráficos)<br>4.Modelo personalizado | Paginas<br>Unidade de conteúdo | Leitura e escrita | Entidade-Relacionamento<br>Orientação a objetos |
| *OntoWebber* | 1.Ontologia de domínio<br>2. Modelagem da visão do site *(storyboard*)<br>- modelos de navegação<br>- Modelos de conteúdo<br>- Modelo de apresentação<br>3. Modelo personalizado<br>4. Modelo de manutenção | Páginas<br>Cartões<br>Unidades de informação | Somente leitura | Ontologia |
| *CREAM* | 1. Autoria<br>2. Anotação | Texto livre | - | Ontologia |
| *COHSE* | 1. Ontologia<br>2. Rede<br>3. Integração dos serviços de ontologia e do serviços e dos serviços de *links* de hipermídia | Documentos ligados através dos metadados de conteúdo | - | Ontologia |
| *OntoPortal* | 1. ontologia<br>2. meta-camada semântica<br>3. relacionamentos | Navegação<br>Captura de conhecimentos<br>Pesquisas por palavras-chave | - | Ontologia |

A seguir, são descritos os passos na construção de um documento hipertextual, suas características, a conversão de texto em hipertexto, a fragmentação hipertextual e as linguagens utilizadas.

### 2.2.4 Conversão de textos em hipertextos

A criação de um hiperdocumento bem estruturado a partir do documento em papel é um processo de conversão que ainda exige um trabalho de re-elaboração do texto. A passagem do "texto" ao "hipertexto", como o acréscimo do prefixo gramatical sugere, é a hiperbolização da forma, dos contornos, dos limites da palavra e do texto. Mas se há mudança pela interrupção, há de se ter também uma continuidade. E esse ultrapassar de fronteiras do suporte impresso para o virtual adquire sentido se forem consideradas as características da textualidade precedente, averiguando a preservação de algumas de suas características, quais limitações são superadas e quais novos domínios são potencializados.

A inquietude para o surgimento de uma nova forma textual já se delineava no início da década de 1970 (LARA, 2003, p.34). Ironicamente, a necessidade de mudança nos conceitos das funções do autor e do leitor, do livro e do texto, foram trazidas à tona por expoentes da teoria crítica literária como FOCAULT (1974), DERRIDA (1975) e BARTHES (1979), cujas posturas filosóficas tratavam do texto tradicional, e não no hipertexto. Para FOCAULT (1974), ". . . o *texto é como uma rede formada por interconexões, na qual as fronteiras não são bem definidas, mas captadas em um sistema de referências e outros livros, outros textos*". Para DERRIDA (1975), ". . . o *texto ideal é composto por unidades que, apesar de separadas, podem se complementar na construção do sentido. Cada unidade pode "quebrar", modificar o contexto e abrir para uma infinidade de novos contextos".* Já BARTHES (1979), descreve o texto ideal como *". . . um composto de blocos de palavras (ou imagens), perpetuamente "não-finalizados", descritos a partir de termos como nós, networks, redes e caminhos*". Permeando a reflexão desses três filósofos, está a idéia de um rizoma que se materializaria na forma de texto hipertextual, na qual não haveria uma hierarquia de início ou fim, nem de acima ou abaixo, nem de antes ou depois. Foucault fala da interrelação, invisível a muitos, entre obras separadas fisicamente. Derrida reconhece a fragmentação inerente ao texto, mas também sua mobilidade e

potencial de transformação como uma entidade viva. Já Barthes aborda o texto como fenômeno em processo, não como matéria acabada.

A definição de texto virtual de BOLTER (1991, p.15), duas décadas depois de Focault, Derrida e Barthes, também confirma a inquietude já presente naqueles autores:

> *". . . texto é um contínuo de parágrafos, dispostos na tela do computador para o leitor ler através de um caminho tradicional. Algumas palavras estão marcadas em negrito; estilo que indica que há uma observação para aquela palavra ou frase, [...] uma segunda janela pode também conter frases em negrito que podem levar o leitor para outros parágrafos. [...] O processo pode continuar indefinidamente [...] formando uma rede que é chamada de hipertexto".*

BOLTER (1991, p.11) nos mostra também que a noção de texto acompanhou o desenvolvimento da tecnologia de transmissão da informação:

> *"Todas as formas de escrita são espaciais* [*. . .*] *Cada tecnologia dá-nos um espaço diferente. Para certas escritas antigas, o espaço* [bidimensional]*, era a superfície interna do rolo contínuo* [de pergaminho]*, que o escriba dividia em colunas. Para a escrita medieval e a impressão na era moderna, o espaço é a superfície branca de uma página, especialmente em volumes encadernados. Para a escrita eletrônica, o espaço é a tela do monitor de um computador, onde o texto é exposto e armazenado em memória eletrônica"*

Segundo KOCH (2002, p.68), se considerarmos a linguagem como atividade interativa que conduz à concepção processual da construção do sentido e que todo texto é constituído por uma proposta de múltiplos sentidos, como querem Focault, Derrida e Barthes, pode-se afirmar que todo texto é um hipertexto. Para que o leitor possa construir um sentido, que nem sempre se dá de maneira linear e seqüencial, é necessário realizar um constante movimento em variadas direções, recorrendo a diversas fontes de informação textuais e extratextuais. Desta forma, o hipertexto é uma forma de estruturação textual que permite ao leitor ser uma espécie de co-autor do texto, oferecendo-lhe a possibilidade de opções entre caminhos diversificados, de modo a permitir diferentes níveis de desenvolvimento e aprofundamento do tema. Esta variabilidade do grau de aprofundamento é notável em textos acadêmicos, a exemplo da tese de doutorado. Povoada de referências, citações, notas de rodapé, quadros, tabelas, exemplos etc., sua concentração de estímulos convoca o leitor a pensar sobre outros textos, consultá-los e, mesmo, suspender sua leitura, em um

processo de alteração de curso para obter outras informações ou fazer anotações. Como o hipertexto oferece uma multiplicidade de caminhos a seguir, compete ao leitor decidir qual o fluxo de informações será incorporado em sua trajetória de leitura.

Nota-se, na literatura, que os teóricos em hipertexto têm dado maior ênfase ao hipertexto enquanto forma de leitura, interessando-se menos pelas suas características de ferramenta de auxílio ao autor e ao texto. No presente estudo, interessa mais este segundo aspecto: os detalhes que permitem criar um texto dinâmico, estruturado semanticamente em seus aspectos cognitivos e seu papel na navegação.

A passagem de documentos originais para a forma hipertextual é um campo em plena fase de descoberta, como novo gênero literário nascido de curto tempo de gestação. No entanto, com o avassalador êxito da Internet, ele tem se tornado forma de expressão cada vez mais explorada e necessária. A conversão de textos originalmente destinados ao suporte impresso para o meio eletrônico envolve uma série de procedimentos. Baseado em trabalhos de GLUSHKO[15](1989) e RINER (1991)[16], BALASUBRAMANIAN (1993) cita seis procedimentos envolvidos na conversão de texto em hipertextos: (1) identificação do potencial de documentos, tradicionais ou não, para beneficiar os leitores, se convertidos para forma hipertextual; (2) determinação e ordenação dos procedimentos de conversão para a forma hipertextual; (3) transformação dos documentos de outros formatos em formatos eletrônicos; (4) identificação e classificação dos tipos de nós e *links*, com vistas à determinação do conteúdo semântico (questão que diz respeito à fragmentação que, assim como a modularidade, será discutida separadamente, mais à frente); (5) construção da estrutura de navegação; (6) realização da conversão e verificação dos resultados.

A abordagem de MARTIN (1992, p.207), mais minuciosa, apresenta os passos a se utilizar no processo de conversão de documentos de papel em hiperdocumentos detalhadamente:

---

[15] GLUSHKO, Robert J. Transforming text into hypertext for a compact disc encyclopedia. In: CONFERENCE IN HUMAN FACTORS IN COMPUTERS SYSTEMS, *Proceedings*…Austin, TX. April,1989. New York: ACM Press, 1989. p. 293-298.

[16] RINER, Rob. Automated conversion, hypertext/hypermedia handbook. New York: McGraw Hill, 1991.

1. Preparação
    a) Conversão do original para um formato processável por computador;
    b) Comparação da estrutura lógica do documento com a estrutura física;
    c) Determinação do que será preservado na estrutura física.

2. Construção da hierarquia
    a) Divisão do documento em uma estrutura hierárquica de nós;
    b) Melhoria da estrutura hierárquica (caso necessário);
    c) Recriação de unidades para melhorar a estrutura hierárquica (caso necessário).

3. Construção do índice
    a) Informatização dos índices do original;
    b) Acréscimo ou melhoria dos índices (caso necessário);
    c) Limpeza do vocabulário do documento (caso necessário).

4. Conversão dos diagramas
    a) Digitalização ou criação de versões computadorizadas dos diagramas;
    b) Acréscimo de novos diagramas (caso necessário);
    c) Melhoria dos diagramas (caso necessário).

5. Construção dos *links* cruzados
    a) Estabelecimento de glossário e sua ligação com o hiperdocumento;
    b) Estabelecimento de unidades de conceitos e suas ligações com o hiperdocumento;
    c) Estabelecimento de *links* para as referências;
    d) Determinação e construção da hierarquia de *links* cruzados (caso necessário);
    e) Exame dos padrões das hierarquias de *links* cruzados.

6. Otimização do hipertexto no auxílio aos seus usuários
    a) Utilização de diagramas para melhorar a estruturação visual;
    b) Construção de rotas de navegação de acordo com necessidades, questões e problemas específicos;
    c) Revisão e crítica extensiva do produto.

A estrutura do documento pode ser vista como o modelo de organização das relações e conexões entre seus módulos. Em primeiro lugar, deve-se analisar os conteúdos do documento e definir uma organização interna adequada. Pararelamente, há que se considerar o ponto de vista e as necessidades do usuário em relação ao assunto documentado, antecipando-se ao mapeamento da informação, de modo que os usuários a reconheçam na estrutura e saibam se situar nela.

Um documento impresso bem organizado geralmente apresenta suas informações em grupos semânticos, configurando uma clara estrutura hierárquica: sumário,

capítulos, seções, sub-seções, parágrafos, sentenças e índices. Existe uma progressão lógica no documento tradicional, mesmo que o leitor não a perceba conscientemente. Essa característica deve ser mantida no formato hipertextual, embora a ordem de consulta possa ser variada. Além disso, se os leitores de documentos tradicionais têm que passar páginas, o leitor de documentos *online* tem que ler o documento através de um monitor, obrigando ao leitor a rolar a página se a informação estiver fora do campo visual.

#### 2.2.4.1 A questão da fragmentação

Um sério problema na conversão de texto em hipertexto refere-se à fragmentação textual, cujas conseqüências podem materializar-se em transbordamento cognitivo, na desorientação do usuário, que resulta de navegação ineficiente, causando ineficácia de desempenho do referido hipertexto.

Segundo HEINONEN (1998, p.1484), a fragmentação é um problema de escolha dos limites dos parágrafos que melhor podem representar o conteúdo semântico fragmentado. Estes fragmentos, ou seções, são conectados através de *links*. Assim, quanto mais *links* existir em um hiperdocumento mais fragmentado ele será.

GLUSHKO (1989, p. 294) aponta outras dificuldades que podem surgir na conversão de documentos devido à fragmentação. Por um lado, a grande quantidade de informação a ser tratada pode ser um empecilho à exeqüibilidade desse processo. Por outro, está o desafio na identificação das unidades de texto que possam constituir-se módulos separados e também servir de referências cruzadas para outras entradas, definindo seu nível de granularidade[17]. Finalmente, a natureza estilística do texto pode impor dificuldades à fragmentação, pois a explicitação de estruturas implícitas pode diminuir ou eliminar a força expressiva do texto original. Nesses casos, o autor do hipertexto se depara com o dilema de reduzir a fragmentação a um mínimo que preserve as idéias e estilo do autor do texto linear, mantendo, ao mesmo tempo, uma manipulação semântica viável.

A fragmentação ainda é problemática na identificação de nós e *links* em hipertextos devido ao desafio de coadunar, ao mesmo tempo, a identificação de unidades de

[17] Nível de detalhamento.

informação que podem ser separadas em módulos e sua utilização com referências cruzadas para outras entradas. Os *links* devem seguir algum modelo para atender às necessidades dos usuários em um determinado contexto. Decidir sobre o nível de granularidade é um problema difícil, na medida em que a melhoria da especificidade possa ser diretamente proporcional ao problema da fragmentação. A fragmentação tende a criar estruturas implícitas entre as partes do documento, podendo gerar um tratamento errôneo do tema. Por isso, é importante achar meios de reduzir a fragmentação das idéias e a perda da estrutura informacional, devido à manipulação da estrutura semântica do documento linear.

Através dos *links* torna-se possível combinar as estruturas lineares e hierárquica da informação. Os *links* permitem fazer conexões entre seções de diferentes níveis ou páginas na hierarquia do texto. A criação de *links* é a tarefa mais importante na conversão do texto linear para o hipertexto. Através deles, torna-se possível combinar a estrutura linear e a estrutura hierárquica da informação, fazendo conexões entre seções de diferentes níveis ou páginas na hierarquia do texto. Porém, HOFMAN e OOSTENDORP (1999, p.129) alertam que, quanto maior o número de *links* da estrutura do hiperdocumento, mais fragmentada ela se torna.

Os recursos que pode constituir-se em uma solução para o problema da fragmentação dos hiperdocumentos são os mapas conceituais, que reproduzem de forma esquemática a estrutura que organiza as páginas *HTML* constituintes do sistema hipertextual do *site*. Normalmente, o mapa funciona simultaneamente como um menu, no sentido de que cada elemento do mapa é um *link* que dá acesso à página correspondente. Além disso, ao se garantirem a coesão e a homogeneidade textual no momento da produção do texto, ou na hora da escrita do hipertexto, garante-se também uma divisão em seções homogêneas, de forma que cada segmento abarque um assunto específico, enquanto segmentos contínuos tratam de diferentes assuntos. Isto ajudará na recuperação de documentos relevantes dentro de grandes bases de dados textuais.

#### 2.2.4.2 A alternativa da modularidade

Modularidade, que será definida aqui como a divisão de documentos completos em pequenos módulos lógicos, é vista como importante conceito para as ferramentas

biblioteconômicas que representam os conceitos de uma área, proporcionando a divisão mais específica dos esquemas dos documentos.

Uma possível solução para a questão da fragmentação está no conceito de modularidade, que é a formação de blocos de informação e a estruturação do documento em módulos, permitindo definir os blocos de informação, os seus conteúdos, as ligações entre eles e a sua organização coerente. O módulo é a unidade básica na construção do sistema. O raciocínio inerente à concepção de documentos multimídia implica nas decisões de como segmentar a informação, de forma que cada módulo seja completo em si e contenha ligações suficientes com outros módulos. Além disso, deve ser possível ao usuário apreender outras informações relacionadas, sem que haja demasia de informações que dificulte a retenção na sua memória de curto-termo.

Cada módulo pode ser composto de um ou mais parágrafos que representam seu conteúdo, mantendo-o independente de outros módulos, sem a necessidade de relacionar entre si, podendo ser representado por um único conceito. Isto permite que o autor do hipertexto use o conceito para representar o conteúdo do módulo, ao invés de mostrar todas as sentenças e períodos que o compõem. Esses módulos de informação formam um documento *online* no todo, permitindo, porém, ser visto parcialmente no monitor. Este texto modular é acessado através de um *link* que pode ser representado pelo conceito que representa seu conteúdo informacional.

A modularidade não implica necessariamente em uma estrutura hierárquica formal. É importante ter um documento principal para que todos os módulos necessários à sua estrutura particular sejam listados em um único lugar. Caso contrário, a identificação de um módulo constituindo-se nível hierárquico intrínseco e não-compartilhado geralmente é contraproducente, porque restringe seu uso e o coloca fora de contexto.

Segundo FISHER (1994, p.19), as formas lineares de seqüência e de estrutura são substituídas, em parte, pelas noções de conteúdo e contexto, respectivamente. Cada módulo de informação é definido pelo seu conteúdo nos nós, mas a sua compreensão depende também das relações que estabelece com outras partes do documento, ou seja, o seu contexto. Em cada módulo de informação, três aspectos

são considerados essenciais: clareza, foco e completude. A clareza assegura que a informação seja facilmente compreendida pelos usuários. O foco implica na identificação e concentração da informação necessária, excluindo-se dados tangenciais que poderiam ser acessados em outros módulos direta ou indiretamente (pelo re-direcionamento do usuário). A completude implica em cada módulo ser completo em si próprio, englobando toda a informação considerada suficiente para que o leitor possa tomar a decisão de prosseguir para outros módulos. Assim, busca-se criar um documento estruturado por níveis, assegurando a coerência em um nível mais alto (*high-level*), a partir do qual se monitoram relações e contextos. Ao mesmo tempo, desenvolve-se módulos em níveis inferiores (*low-level*). Esta interatividade se realiza através dos *links* determinados pelo autor do hipertexto, em níveis de modularidade que caracterizam essa navegação.

A modularidade apresenta benefícios ligados à subdivisão dos textos em unidades menores com conteúdos completos e mais focalizados. Em primeiro lugar, do ponto de vista do usuário, documentos com esquemas modulares pequenos estimulam sua reutilização. Em segundo, a praticidade dos módulos é facilitadora também para a organização do próprio autor do hipertexto e daqueles encarregados de sua manutenção. Em terceiro, o fato de serem esquemáticos os torna mais cognitivos, facilitando sua absorção e retenção na memória.

A maioria das estratégias de estruturação de hiperdocumentos se apóia nas linguagens de marcação de textos como *SGML* (*Standard Generalized Markup Language* ou Linguagem de Marcação Padrão Generalizada*)*, *HTML*(*Hypertext Markup Language* ou Linguagem de Marcação de Hipertexto) e *XML* (*eXtensible Markup Language* ou Linguagem de Marcação Estendida), que serão descritas logo abaixo. Essas linguagens são gerenciadas por uma série de comandos que, uma vez inseridos em um documento tradicional, identificam sua estrutura lógica em títulos, seções e parágrafos. Estes tipos de componentes de um documento servem de guias para o processo de conversão de textos em hipertextos, pois sugerem maneiras de fragmentá-lo e de conectar algumas de suas partes entre si. Estas divisões lógicas em títulos, seções e parágrafos podem ser suficientes para determinar uma boa configuração de um hipertexto, no caso de documentos cuja organização interna é bem definida e rígida, a exemplo de dicionários e catálogos. Todavia, podem ser pouco eficientes em textos livres com subdivisões menos

sugestivas, para os quais os algoritmos de conversão gerariam redes não muito adequadas à estruturação de seu conteúdo semântico. Entre as linguagens de marcação, o *XML* surgiu como recurso diferenciado na autoria de hipertextos, pois permite aos autores a marcação de conteúdo semântico dentro do próprio texto e o acréscimo de comandos, explicitando decisões de fragmentação e interconexão segundo configurações mais convenientes para aquele documento específico.

Segundo SONG e BAYERL (2003, p.4), a combinação da marcação semântica com a estrutura formal pode resultar na abordagem semântica composicional que facilita a extração do significado de documentos dentro de uma área específica. Para isso, todo o texto tem que ser marcado com seus conceitos específicos. O que observa-se, ainda hoje, é que o componente artesanal da especificação semântica de um documento *XML* ainda é necessário para possibilitar ao ser humano ou à máquina interpretar seu conteúdo como um todo.

#### 2.2.4.3 Linguagens de marcação

O termo ”marcação” (*markup*) surgiu antes do advento dos documentos eletrônicos, quando os profissionais da área de editoração recebiam originais em papel e escreviam instruções de marcação para orientar a composição gráfica final na melhor forma ou aparência impressa, como ilustra McGRATH (2003, p.1): *". . .insira uma quebra de parágrafo aqui, coloque esta palavra em negrito, use espaços duplos neste texto. . ."*. Assim, as marcações em hiperdocumentos são quaisquer informações adicionais sinalizadoras acrescentadas ao texto de um documento virtual. O mundo digital apresenta um crescente número linguagens de marcação. A seguir, são apresentadas três linguagens de marcação: *SGML*, *HTML* e *XML*.

A linguagem *SGML* (*Standard Generalized Markup Language*) é um padrão internacional para a representação de conteúdo de documentos, desenvolvido por pesquisadores da IBM. Por isso, exige uma definição rigorosa, de forma que o maior número de programas em sistemas formais possam ser usados no processamento de documentos (GUIMARÃES, 2002, p.3). Sua marcação descreve a estrutura de um documento de forma generalizada, em vez de especificar o processamento a ser feito no mesmo. A *SGML* não impõe qualquer conjunto específico de tipos de elementos a seus usuários. Em vez disso, o padrão possibilita que estes criem seus

próprios de elementos e, por isso, se caracteriza como uma meta-linguagem, ou seja, uma linguagem para definir linguagens de marcação. As linguagens *HTML* e *XML*, descritas a seguir, são exemplos de aplicações baseadas na *SGML* e utilizadas no protótipo construído nesta tese.

A linguagem *HTML* (*Hypertext Markup Language*) é uma linguagem de marcação voltada para a estruturação de documentos e para criação de documentos hipertextos que podem ser interpretados por navegadores como *Netscape Navigator*, *Internet Explorer*, *Mosaic*, etc. Segundo CASTRO (2001, p.12), o sucesso da *HTML* deve-se à sua simplicidade, facilidade de uso e tolerância. Por exemplo, não distingue maiúsculas ou minúsculas ou o uso de aspas simples ou duplas. Porém, sua simplicidade limita seu potencial porque as *tags* da *HTML* (atributos e valores; veja as FIG.54 e FIG.55 no Cap.3 - CONSTRUÇÃO DO MODELO MHTX) são mais adequados para a formatação. Como não fornecem informações sobre o conteúdo de uma página da *Web*, sua reutilização em outro contexto torna-se difícil. Um elemento *HTML* pode ter vários atributos, que devem ser separados por espaços em branco. Já seus valores devem ser sempre delimitados por aspas.

A *XML* é um subconjunto da *SGML* (descritas acima) e permite que uma marcação específica seja criada para especificar idéias e compartilhá-las na rede. Ela tem as virtudes da *SGML* e da *HTML* , mas com a vantagem de não apresentar suas óbvias limitações.

A seguir apresenta-se a segunda fase da fundamentação teórica, quando serão discutidos a teoria da organização do conhecimento, seus aspectos cognitivos e técnicas utilizadas na criação do modelo **MHTX**.

## 2.3 Da organização à recuperação da informação

O ser humano sempre se preocupou, ao longo de toda a história, em desenvolver procedimentos e técnicas com a finalidade de analisar, registrar, guardar e recuperar informações. Atualmente, essa preocupação tem-se tornado, cada vez mais, um grande desafio, já que as inúmeras e profundas mudanças ocorridas no contexto da comunicação mundial e nas áreas de ciência e tecnologia causaram um aumento extraordinário na produção e intercâmbio da informação nas diversas áreas do

conhecimento. A organização do conhecimento é um processo necessário, que se torna mais urgente, à medida que o volume de informação aumenta. Seu objetivo principal é organizar a informação e disponibilizá-la para uma recuperação posterior; a organização e recuperação da informação enfrentam o mesmo desafio visando atender às necessidades dos usuários em potencial. Os princípios e técnicas de tratamento, organização e disseminação da informação, conseqüentemente, têm sofrido profundas transformações. Os instrumentos clássicos utilizados na organização e recuperação da informação, enumerados por ROWLEY (1992, p.7) são o catálogo (*online*, *off-line*, microficha e ficha), a bibliografia, o índice, a base de dados e o arquivo de documentos impressos ou automatizados. Estes instrumentos incluem os esquemas de classificação, como o sistema de Classificação Decimal de *Dewey* (CDD) e o Sistema de Classificação Decimal Universal (CDU), cuja função é organizar os materiais do acervo em nível mais geral. Incluem também os cabeçalhos de assunto, que permitem a organização por assunto em níveis mais detalhados e uma recuperação por conteúdos específicos, como a Lista de Cabeçalho de Assunto da Biblioteca do Congresso Americano (LCSH), a Lista de Cabeçalho de Assunto *Sears,* a Classificação de Dois Pontos de *Ranganathan*, os esquemas desenvolvidos individualmente para coleções especiais e os tesauros.

Esses instrumentos, chamados de linguagens de indexação ou vocabulário controlado, são criados através do processo de indexação e categorização da informação. Por organizarem a informação, podem ser considerados o coração de um sistema de informação. São geralmente preparados por profissionais da área da Ciência da Informação e refletem a garantia literária, o que os tornam instrumentos qualificados para minimizar os problemas na recuperação da informação. Entretanto, um dos problemas das linguagens de indexação é que elas não são atualizadas freqüentemente. LOPES (2002, p.51), após relatos de inúmeros estudos sobre o uso da linguagem controlada e da linguagem natural, conclui que uma alternativa viável seria a combinação das abordagens da linguagem de indexação (vocabulário controlado) com a linguagem natural na organização da informação e na estratégia de busca.

### 2.3.1 Aspectos cognitivos na organização do conhecimento

Da organização à recuperação do conhecimento, a informação passa pelos processos de categorização, indexação e interação homem-máquina até chegar à recuperação da informação, processos que são fortemente influenciados pelos processos cognitivos. A seguir, estes quatro processos, considerados relevantes para esta pesquisa, são discutidos.

#### 2.3.1.1 Categorização

Desde a época de Aristóteles, já havia a preocupação sobre as práticas de nomear, definir e categorizar. Muitos autores consideram os termos categorização e classificação como sinônimos (GARDNER, 1996; JACOB e SHAW, 1991; SMITH e MEDIN, 1981). Com o desenvolvimento de estudos na ciência cognitiva, a visão de como categorizamos sofreu modificações. A categorização passou de um processo cognitivo individual a um processo cultural e social de construção da realidade, que organiza conceitos, parcialmente baseado na psicologia do pensamento. A informação perceptiva é fundamental na definição das extensões de uma categoria, porque a categorização não é feita artificialmente, mas sim levando em conta as informações do mundo a que pertencemos e como respondemos a elas. Na categorização, o reconhecimento das similaridades e diferenças leva à criação de um conhecimento novo, pelo agrupamento de entidades, de acordo com as similaridades e diferenças observadas.

Entre as definições de categorização que caracterizam o processo cognitivo, JACOB e SHAW (1998, p.155) dizem que *"Categorização é um processo cognitivo de dividir as experiências do mundo em grupos de entidades, ou categorias, para construir uma ordem física e social do mundo".* MARKMAN (1989)[18], citado por JACOB e SHAW *(*1998,p.155), descreve a categorização como *" um mecanismo fundamental que simplifica a interação individual com o ambiente: não somente facilitando o armazenamento e recuperação da informação, mas também reduzindo a demanda da memória humana".* Para PIEDADE (1983), este é um processo mental habitual do homem, pois vivemos automaticamente classificando coisas e idéias, a fim de compreender e conhecer. Finalmente, para GARDNER (1996, p.373), "*as*

[18] MARKMAN, Ellen M. *Categorization and naming in children*: problems of induction. Cambridge, MA: MIT Press, 1989. 250p.

*categorias têm uma estrutura interna, centrada em protótipos ou estereótipos, e outros exemplares são definidos como mais ou menos periféricos, dependendo do grau em que eles compartilham características cruciais com o protótipo central"*.

Para organizar o conhecimento de qualquer área, da sua representação à sua recuperação, estudam-se primeiramente os conceitos que compõem esse campo do conhecimento e as relações entre eles. A influência do contexto é tão importante na categorização que, qualquer descrição individual na classificação que o desconsidere, pode mostrar-se muito limitada. Categorias e hierarquias de categorias são a melhor maneira de organizar o conhecimento para recuperação, pelo óbvio motivo de que a informação estruturada é mais fácil de ser recuperada do que uma informação desorganizada. O processo de categorização estaria no bojo das questões de interesses das áreas da ciência da informação e da ciência cognitiva, no que tange à estratégia de se classificarem as coisas, fatos e fenômenos existentes, que são objetos da cognição.

#### 2.3.1.2 Indexação

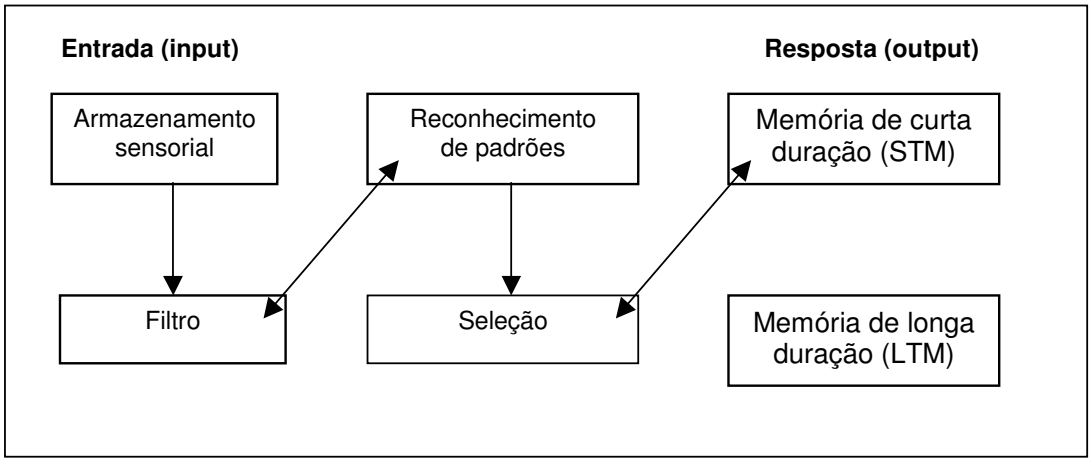
Indexação é o processo intelectual que envolve atividades cognitivas na compreensão do texto e na composição da representação do documento. Como atividade intelectual, é particularmente afeita a teorias e modelos da *Psicologia Cognitiva* e da *Teoria de Soluções de Problemas*. (DAVID et al., 1995, p.49). WELLISH (1995)[19], citado por JACOB & SHAW (1998, p.159), descreve indexação como o ato de indicar ou apontar o conteúdo intelectual de uma coleção. Esta definição mascara a natureza cognitiva do processo de indexação, por enfatizar o produto físico do processo e não a análise de conteúdo.

O processo de indexação possui três etapas: (1) análise do documento e estabelecimento do seu assunto, (2) identificação dos principais conceitos do documento, e (3) tradução destes conceitos em termos de uma linguagem de indexação. A análise cognitiva aparece na análise e na identificação de conteúdo do documento e no processo mental entre estímulo e resposta que processa a informação. Por isto, é importante entender como a estrutura do sistema de informação afeta a interação entre a análise do documento, a representação do

[19] WELLISH, Hans H. *Indexing from A to Z*. New York, NY: Wilson, 1995.569p.

conteúdo e a recuperação da informação. Como a indexação é um processo de categorização, torna-se essencial considerar como a atividade do indexador e a estrutura da linguagem de indexação forçam a formação de categorias. As habilidades intelectuais na indexação poderiam ser harmonizadas de maneira mais eficiente, se as atividades pudessem simular processos cognitivos ou percepções sensoriais. Conforme REED (1992, p.6), para interpretar o conteúdo do documento, o indexador certamente passa pelo processo cognitivo (FIG.7). Após as informações serem filtradas e os padrões de linguagem reconhecidos, o usuário utiliza a memória de curta duração (STM), e também a memória de longa duração (LTM). A informação selecionada passa por dentro da STM, onde ela é processada e armazenada temporariamente ou é transferida, depois, para a LTM, onde é armazenada por períodos de tempo maiores.

FIGURA 7 - Estágios do modelo de processamento cognitivo da informação



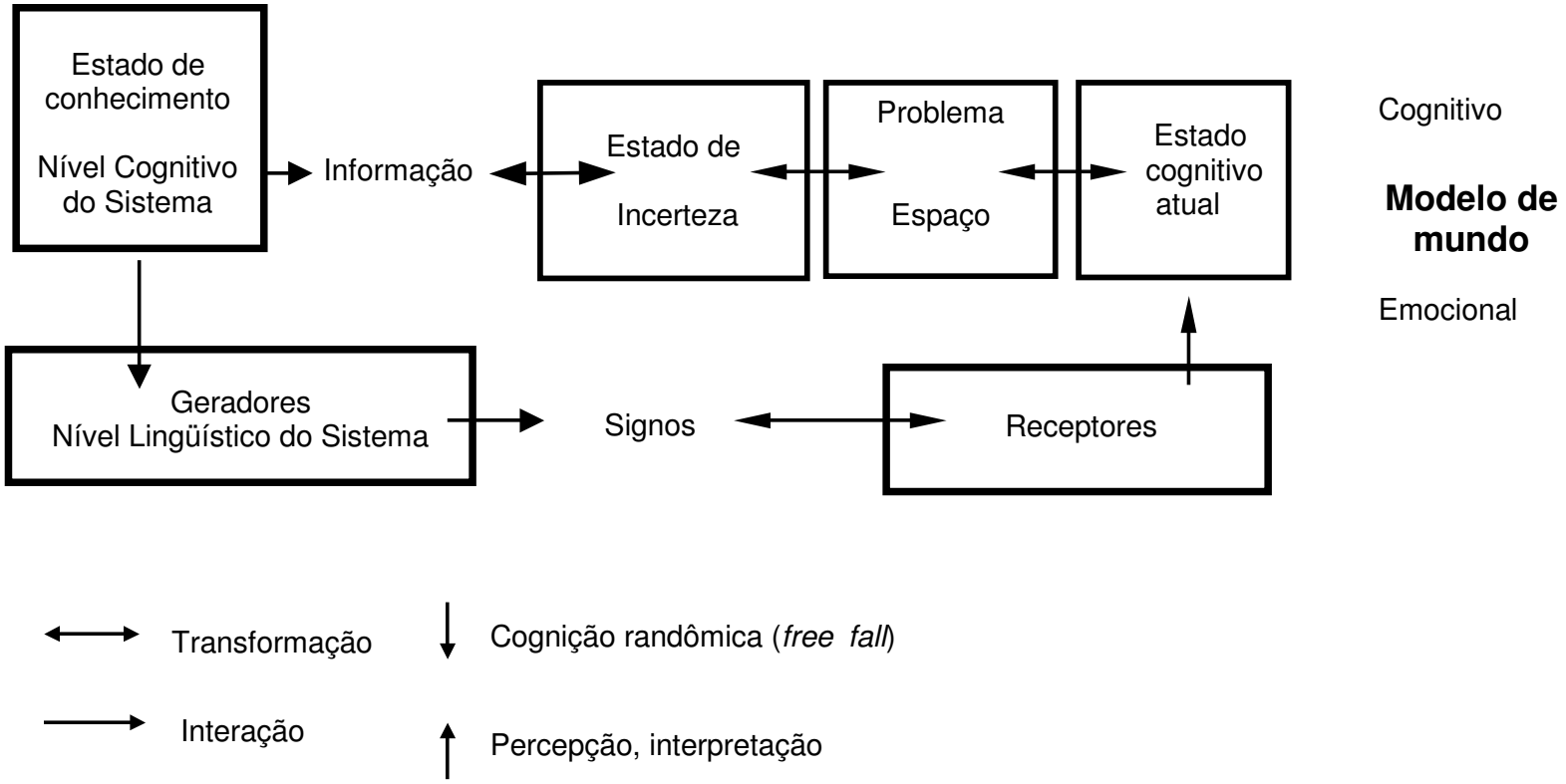

FONTE: REED, 1992, p.5.

Nas duas primeiras etapas, quais sejam (1) análise do documento e estabelecimento do assunto de que trata o item e (2) identificação dos principais conceitos do documento e do processo de indexação; o indexador identifica os conceitos essenciais do documento, buscando compreensão a partir de sua memória e refletindo o processo cognitivo.

#### 2.3.1.3 Recuperação da informação

O conceito de recuperação da informação muitas vezes é tratado como sinônimo de "busca de informação", porque a necessidade de informação é que dispara o processo de busca da informação. INGWERSEN (1982, p.167) propõe a seqüência de 9 etapas para identificar o processo mental da recuperação da informação: (1) a necessidade de informação do usuário, (2) a questão sobre a informação formulada, (3) a negociação usuário-bibliotecário, (4) a formulação da estratégia de busca-análise do tópico, (5) a escolha das ferramentas de busca, (6) a procura na lista alfabética ou sistemática, (7) o julgamento baseado no índice (termos), (8) o julgamento baseado na descrição, resumos e títulos e (9) a avaliação do documento pelo usuário-bibliotecário. Ilustrando a inferência do aspecto cognitivo na RI, o autor desenvolve um fluxograma, mostrado na FIG.8 (INGWERSEN, 1992, p.6), para demonstrar o funcionamento do sistema de comunicação de uma biblioteca do ponto de vista cognitivo. Neste modelo, ele fala de três imagens ou pontos de vista do mundo, onde cada um representa uma parte da estrutura do conhecimento: o usuário, o intermediário (bibliotecário) e o gerador da informação, que devem estar em sintonia para que a RI se realize com sucesso.

FIGURA 8 - O sistema de comunicação cognitivo para a CI e a RI.




FONTE: INGWERSEN,1992, p33; citado por INGWERSEN, 1996, p.6

Para JACOB e SHAW (1998, p.147), a representação na RI envolve um processo complexo de correspondência entre o *campo de modelagem* (*modeling domain*) da linguagem de indexação e o *campo de objetivos múltiplos* (*multiple target domain*). O *campo de modelagem* inclui (1) a estrutura de conhecimento do autor do documento, (2) o conhecimento representado no documento, (3) a estrutura do conhecimento no *campo do discurso* (*discourse domain*) a que o documento pertence, (4) a compreensão do significado da palavra pelo indexador ou pesquisador, tanto do ponto de vista do *campo relevante* (*relevant domain*) quanto do sistema de recuperação. Neste processo, HARBO, INGWERSEN, TIMMERMANN (1977)[20], citados por INGWERSEN (1982, p.169) dizem que

> *"...a tarefa* [de recuperação da informação] *é trazer as estruturas cognitivas dos autores, dos designers de sistemas e dos indexadores de acordo com as estruturas do profissional que lida com a informação e o usuário, para cobrir a necessidade atual".*

A organização da informação, entendida ao nível do armazenamento, e a sua recuperação fazem parte do mesmo processo, de tal forma que uma entrada de dados ineficaz implicará em uma saída de dados também ineficaz. Nesse processo de comunicação, as atividades cognitivas influem tanto na entrada como na saída da informação e dependem da maneira como usamos nossa mente, de como se dá nosso processo de abstração mental.

Nos últimos anos, muitas pesquisas têm sido orientadas para estudar e entender o processo de transferência da informação na recuperação. Muitas dessas pesquisas têm sido relacionadas com a interação entre o sistema de Recuperação da Informação (RI) e o usuário, com o intuito de melhorar a eficácia dos sistemas de recuperação de informação e os métodos de indexação. JACOB e SHAW (1998) observam que a ênfase em sistemas amigáveis para o usuário com interfaces inteligentes caracterizam as tendências de recentes pesquisas cognitivas em Ciência da Informação. A abordagem cognitiva é usada na pesquisa em RI para descobrir características do comportamento do usuário que procura a informação e, então, reprogramar a interface homem-computador para acomodar essas características (BORGMAN; CHIGNELL; VALDEZ, 1989, p.86).

---

[20] HARBO,O. O ., INGWERSEN, P., TIMMERMANN, P. Cognitive process in information storage and retrieval. In: MEY. Marc de.(Ed.) *International Workshop on the Cognitive Viewpoint*. Belgium: University of Ghent, 1977. p. 214-218

O fluxograma da FIG.8 acima, elaborado por INGWERSON (1996, p.6) demonstra como o processo de comunicação interativo ocorre entre o homem e a máquina e como as mensagens enviadas pela máquina ou pelo homem para um indivíduo podem transformar a informação abstrata em senso real. Entretanto, os sinais enviados pelo homem (ou pela máquina) para a máquina, na verdade não se transformam em informação, embora eles sejam percebidos e embutidos nas estruturas cognitivas. Estes sinais permanecem como sinais ou continuam como informação potencial em suas estruturas lingüísticas.

Com o desenvolvimento da informática, o modelo da mente humana começou a ser aplicado ao computador. O cérebro e a mente humanos têm sido comparados ao disco rígido (*hard drive*) e *softwares*, respectivamente: *"O cérebro é somente um computador digital e a mente é somente o programa de computador"* (SEARLE, 1984, p.20).

#### 2.3.1.4 A interação homem-máquina

Na interseção da Ciência da Computação com a Ciência da Informação, tendo como pano de fundo a aplicação da computação e dos computadores na RI, destacam-se as áreas da Inteligência Artificial (IA) e da Interação Homem-computador (IHC). A IA busca reproduzir a atividade mental do homem em tarefas como a compreensão da linguagem, a aprendizagem e o raciocínio. Essas tarefas estão associadas à ciência da informação e à ciência cognitiva nos padrões de representação e nas atividades de processamento da informação, estabelecendo limites nos modelos de construção da representação do conhecimento.

São várias as conceituações de IA. Para MINSKY, (1980)[21], citado por SIQUEIRA e PEREIRA (1989 p.42)

> *" A Inteligência Artificial é a ciência que permite fazer com que as máquinas realizem tarefas que necessitariam de inteligência se elas fossem efetuadas pelos homens"*.

SIQUEIRA e PEREIRA (1989, p.42) afirmam que

> *"IA é a arte de escrever programas capazes de exibir um comportamento inteligente... é um ramo da engenharia do conhecimento e está ligada à*

[21] MINSKY, M. *A framework for representing knowledge*. In: HAUGELAND ed. Mind design: philosophy, psychology, Artificial Inteligence. Bradford Book,1980.

> *ciência cognitiva, mantendo ligações privilegiadas com a lógica, a linguística, a psicologia, a psicolinguística, a biologia e outras ciências"*.

E para GARDNER (1996, p.55) é

> *"... a ciência construída em torno da simulação computacional, é por muitos considerada a disciplina central da ciência cognitiva e a que tem maior probabilidade de excluir, ou tornar supérfluos, outros campos de estudo mais antigos."*

Não só ciência, mas também uma arte, a IA busca, na programação de computadores, simular a inteligência humana. Atualmente, a IA está dividida em IA forte e IA fraca. A IA fraca concentra-se em técnicas de programação, enquanto que a IA forte está mais preocupada com conceitos filosóficos da mente e do conhecimento.

A Ciência da Informação se interessa tanto pela IA fraca como pela IA forte. A primeira, como fonte de inovações nos sistemas de informação, como os sistemas inteligentes, hipertextos, bases de conhecimento, interfaces inteligentes e questões sobre a interação homem-computador. A segunda, como modelo teórico de cognição, no qual a informação desempenha importante papel.

Dentro do processo de organização e recuperação da informação, é muito importante o sistema possuir uma interface amigável. A abordagem cognitiva geralmente é utilizada em pesquisa na RI para descobrir características cognitivas no comportamento do usuário que busque informação, com o objetivo de reformular a interface[22] homem-computador e acomodar essas características.

A essência da abordagem cognitiva na organização do conhecimento é a idéia de que a percepção humana, a cognição e as estruturas de conhecimento são o seu ponto de convergência, relacionando os diversos processos da ciência da informação e permitindo uma estrutura unificada e eficiente.

Em seguida, são descritas as teorias da organização do conhecimento que fundamentam este estudo, como alternativas para organizar o conhecimento em

[22] DIAS (1994, p.1) define interface como *"uma superfície de contato com a informação e também um envelope para o conteúdo, procurando-se adequar esta superfície aos fatores humanos envolvidos no processo de contato e às normas da organização da informação"*.

sistemas de hipertextos. Apresenta-se também uma revisão de literatura sobre a relação entre hipertexto e a análise facetada.

### 2.3.2 A Teoria da Análise Facetada (TAF): revisão de literatura e fundamentos

O objetivo de toda classificação é estabelecer uma ordem ou organização das coisas e dos pensamentos. A classificação facetada pode mostrar um mapeamento do conhecimento científico de um determinado período, ou permitir a descoberta de conexões e analogias entre diferentes campos do conhecimento, facilitando a recuperação da informação (SPEZIALI, 1973, p.462).

Vários autores citam a classificação facetada como um instrumento que pode auxiliar na representação do conteúdo intelectual em sistemas de hipertextos. Um dos primeiros autores a tratar deste assunto foi Elizabeth B. Duncan, em 1989. No seu artigo *Structuring knowledge bases for designers of learning materials* (DUNCAN, 1989, p.20), a autora apresenta um modelo de aplicação direta da análise facetada no *design* de hipertexto, no qual sugere que a análise de facetas tem pontos comuns com a teoria lingüística, entendendo que a combinação entre facetas e análise semântica pode servir de base para o desenvolvimento da tecnologia do hipertexto. As categorias determinadas pela análise das facetas constituem os diferentes pontos de vista do assunto.

A autora enfoca três aspectos para a organização e representação do conhecimento: a) o mapa conceitual, para representar os conceitos de um domínio do conhecimento, b) a análise facetada, representando os diferentes pontos de vista sobre os quais assunto pode ser tratado e c) *links* e seus tipos e a relação entre eles. Em outro trabalho, *A faceted approach to hypertext?*, DUNCAN (1989, p.157) relata uma abordagem da teoria facetada em sistemas de hipertexto, no qual considera possibilidades do hipertexto como uma tecnologia a ser explorada dentro de alguns princípios básicos da ciência da informação, como por exemplo, a análise facetada.

INGWERSEN e WORMELL (1992, p.185) também destacam o potencial do trabalho de Ranganathan no desenvolvimento dos sistemas de recuperação da informação. Concluem que os princípios da organização do conhecimento e a metodologia sobre

análise facetada, por ter uma abordagem estruturada, poderiam ser utilizadas como ponto de partida na modelagem de uma estrutura cognitiva e no trabalho de análise de uma determinada área do conhecimento, antes do processo de elaboração do *design* de um sistema de recuperação da informação.

ELLIS (1996, p.155), comentando os trabalhos de Duncan (citados acima), corrobora sua visão de aplicação das teorias clássicas da ciência da informação no *design* de hipertexto. Discute o uso da análise de facetas na simplificação do processo de reconhecimento de conceitos e suas relações, tanto para a máquina quanto para o usuário.

POLLITT (1997, p.2) sugere que a abordagem facetada para classificação, enquanto alternativa aos sistemas enumerativos da classificação de bibliotecas, seja equivalente à abordagem relacional dos sistemas de bases de dados via computador. O autor descreve o desenvolvimento do sistema *HIBROWSE* (*HIgh resolution Interface for BROWsing and SEarching databases*), que trabalha com os princípios da classificação facetada para melhorar a interação através da pesquisa visual (*view-based).* Esse estudo foi realizado na base de dados EMBASE, publicada pela *Elsevier Science* que, na época, possuía 7 milhões de referências da literatura biomédica.

Aimme GLASSEL (1998), catalogadora da *Internet Scout Project*, escolheu o provocativo título *"Was Ranganathan a Yahoo!?"* para um artigo em que compara o trabalho do pioneiro da classificação facetada com a conhecida ferramenta de busca na Internet. Para a autora, Ranganathan reconheceu que o mundo do conhecimento estava crescendo rapidamente, com o surgimento de novas áreas do conhecimento e novas maneiras de combinar os assuntos. Além disso, nenhuma classificação que tentasse enumerar este enorme e crescente número de assuntos, sem uma capacidade de expansão para permitir novas áreas do conhecimento, teria futuro promissor. Pelo fato da classificação facetada ocupar menos espaço na tabela do que os sistemas enumerativos e, mesmo assim, permitir maior especificidade, fornecendo o acesso a cada faceta de uma notação, é que a autora afirma que a teoria da análise facetada se parece com a ferramenta de busca *Yahoo!.* Esta última ferramenta é conhecida por ser um diretório de assuntos dentro das ferramentas da *Web*, acessível através da navegação hierárquica de categorias, subcategorias e

sub-subcategorias, caracterizando um processo de navegação ramificada, que pode chegar até ao quinto nível categórico ou mais. Cada termo, dentro das categorias mais abrangentes dos *sites* do *Yahoo!*, tem uma palavra com seu próprio significado. Então, quando combinado com outras palavras dentro de uma classe, um novo contexto é criado, gerando um novo significado. Por aplicar análise facetada em uma hierarquia *online*, o *Yahoo!* tem a possibilidade de ter uma classe (*string*) de categorias e subcategorias (facetas). Pela rotação ou permutação destes termos, essa ferramenta pode fornecer acesso a um único recurso, por meio de uma variedade de subcategorias em uma hierarquia maior.

PRISS e JACOB (1999, p.205) descrevem uma outra experiência que utiliza a estrutura facetada no *design* de sistemas de informação, em que foi feita a análise estrutural de três *Web sites* de instituições escolares. As autoras argumentam que a aplicação da abordagem facetada na organização do conhecimento pode assegurar um processo menos ocasional e mais maleável. Para isto, sugerem o uso de um tesauro facetado como base para uma abordagem sistemática na estruturação dos *sites,* onde os elementos mínimos são as facetas constituintes de cada conjunto de termos ou conjunto de conceitos. Nesse trabalho, são descritos três tipos de construção de facetas: (a) por agregação, (b) por composição de termos sem limitação e (c) por composição de termos com limitação.

ELLIS e VASCONCELOS (1999, p.3; 2000, p.97) sugerem a análise facetada como forma de transpor algumas dificuldades associadas com o processo de criação de hipertextos para usuários desconhecidos e, possivelmente, minimizar o problema da distância entre o *designer* e o usuário em potencial. Os autores descrevem e discutem, detalhadamente, o trabalho de Duncan, considerando-o como um bom modelo para estruturação da Internet e da *Web,* mas concluem que a análise facetada não resolve o problema de indexação para o usuário genérico. Porém, a adoção de uma abordagem classificatória *a posteriori* e não *a priori*, ou seja, quando a classificação é derivada indutivamente dos conceitos e dos termos usados em uma área de assunto, com base na garantia literária, pode aliviar alguns problemas de pesquisa na *Web*.

BROUGHTON e LANE (2000, p.143) apresentam breve discussão sobre o uso de sistemas de classificação para a organização da Web. Os dois autores trabalharam na segunda edição da *Bliss Bibliographic Classification* (BC2) e listam as razões

pelas quais as classificações facetadas são melhores: (1) ausência de limites teóricos para especificidade ou exaustividade; (2) facilidade de lidar com assuntos complexos; (3) flexibilidade suficiente para acomodar diferentes necessidades dos usuários. Eles concluem a discussão sobre como e porque ela deve ser aplicada a recursos *online*, e demandam o reconhecimento da importância da classificação e a habilidade de análise de assunto, mesmo quando a pesquisa em texto completo é respondida prontamente pela base de dados.

KOEHLER (2001, p.1), autor do documento *Web site Web Document Management* (http.//www.ou.edu/cas/slis/courses/LIS5990A/slis5990/Catalog/innovation.htm), ressalta a importância da classificação facetada na identificação de características individuais de vários conceitos, combinando-os de maneira sintética, para criar uma estrutura fluída e interativa. Citando STAR (1996), argumenta que a classificação facetada é particularmente importante para a definição da informação e sua recuperação, e levanta as seguintes questões: (a) Os sistemas de classificação têm sido desenvolvidos para a *Web* ?; (b) O que eles têm em comum?; (c) O que seria mais útil: os sistemas de classificação facetados ou os universais?; (d) O típico bibliotecário, classificador padrão, seria útil para *Web*? Embora não responda a nenhuma dessas questões, o autor traz à tona importantes pontos a serem refletidos pela comunidade de profissionais e acadêmicos da área de ciência da informação, especialmente o bibliotecário, cuja atualização e adaptação a um perfil mais moderno é urgente.

Pode-se constatar, nas pesquisas fora do Brasil, que os sistemas de classificação não têm sido desenvolvidos para a *Web*, embora haja consenso de que eles possam ser instrumentos relevantes para a organização e pesquisa nesse ambiente. Os sistemas de classificação facetados são mais úteis para a *Web* por que sua característica analítico-sintética permite ao usuário ver um assunto sob diferentes pontos de vista, do mesmo modo que comungam da não linearidade característica do hipertexto.

No meio acadêmico brasileiro, destacam-se, até o momento, uma tese e uma dissertação dedicadas ao estudo da classificação facetada como possível contribuição para a estruturação de hiperdocumentos. SANTOS (1996), em sua dissertação de mestrado *Engenharia da informação para sistemas hipertexto,*

apresenta os princípios teóricos que regem a construção de sistemas de conceitos na teoria da classificação facetada e a teoria geral da terminologia, como opção na autoria estruturada de hiperdocumentos complexos, em especial na execução da modelagem conceitual dos sistemas de hipertextos. A autora faz um paralelo entre as duas teorias e suas aplicabilidades nos sistemas de hipertexto, apontando para a necessidade de métodos que tracem princípios consistentes para a autoria de hiperdocumentos e conclui que, através de princípios que orientem a modelagem conceitual, elas podem promover a construção de sistemas de hipertextos que expressem cada vez mais o processo cognitivo humano.

CAMPOS (2001), citada anteriormente no item 2.2.2, em sua tese de doutorado apresenta uma preocupação com a autoria no processo de produção de hiperdocumentos. Para ela, a falta de metodologias apropriadas afasta o autor (que desenvolve o documento temático) do analista de sistema (que implementa o hipertexto), ressaltando que os modelos conceituais atualmente construídos e aceitos são deficientes. Através de uma perspectiva interdisciplinar entre a ciência da informação, a ciência da computação e a área de terminologia, a autora estudou critérios para auxiliar a elaboração de metodologias para modelos conceituais voltados para organização de unidades de conhecimento, como os hiperdocumentos. Além de discorrer sobre o objeto do hipertexto, a autora apresenta e compara a teoria da classificação facetada, a teoria da terminologia, a teoria do conceito, recorrendo a métodos de construção de modelos segundo a orientação a objetos e à ontologia. Seu estudo culmina na proposição de uma série de requisitos para a elaboração de metodologias para modelos conceituais de hiperdocumentos, requisitos estes que contribuíram para a construção do modelo ora apresentado.

Pode-se verificar que a literatura já revela muitos esforços no sentido de relacionar análise facetada com a elaboração de hipertexto. Entretanto, nenhum deles apresentou, ainda, uma proposta de solução viável para solucionar o problema da estruturação da informação em sistemas de hiperdocumentos.

A classificação facetada compreende princípios e técnicas para os sistemas de organização e recuperação da informação. Um sistema facetado reconhece muitos aspectos em um único conceito ou classe geral e tenta sintetizar estes aspectos, de maneira a melhor descrevê-los. A facetação mostrou que a relação entre assuntos,

feita através do símbolo de dois pontos (:), pode ser infinita, explicitando que o conhecimento pode ser multidimensional e que suas relações podem tomar rumos diferentes, dependendo de como a síntese entre vários conceitos múltiplos seja realizada (VICKERY, 1980, p.209).

Para elaborar uma classificação facetada, examina-se a literatura do assunto para identificar seus conceitos e termos; estabelece-se suas características e dentre estas, suas facetas. Após investigar e definir a terminologia do assunto, os termos são analisados e distribuídos em facetas. Como já foi dito, faceta é a coleção de termos que apresentam igual relacionamento com o assunto global, refletindo a aplicação de um princípio básico de divisão. As facetas obtidas são inerentes ao assunto. Dentro de cada faceta, os termos que as constituem são suscetíveis a novos agrupamentos, pela aplicação de outras características divisionais, dando origem às subfacetas. Os termos nas subfacetas são mutuamente exclusivos, ou seja, não podem se sobrepor. Estabelecidas as facetas e subfacetas, é importante determinar a ordem de citação em que serão apresentadas no sistema de classificação. Em seguida, ordenam-se todos os elementos em ordem de arquivamento, o que permite colocar o assunto geral antes do específico. Após estas etapas, o sistema está pronto para receber uma notação, que deverá ser flexível para permitir a inclusão de novas classes. Finalmente, compila-se um índice com todos os termos e suas respectivas notações (PIEDADE, 1983, p. 80, BARBOSA,1972, p.76).

Os dois principais referenciais teóricos sobre os processos de análise facetada representam dois momentos cronológicos distintos: a proposta de Ranganathan com o *Prolegomena to Library Classification* de 1933 (republicado em 1967) e os estudos do *Classification Research Group – CRG* (AUSTIN, 1968, 1969; CLASSIFICATION RESEARCH GROUP, 1985; A. C. FOSKETT, 1996; D. J. FOSKETT, 1970, 1971, 1974; MILLS e BROUGHTON, 1977; VICKERY, 1960, 1966, 1975). SPITERI (1989) em seu artigo *"A Simplified Model for facet analysis: Ranganathan 101"* deixa claro que ela não pretendeu suplantar o valioso trabalho de Ranganathan e do *CRG*, e que todos os princípios utilizados em seu trabalho se originaram de estudos anteriores ao seu. Entretanto, o grande mérito no seu trabalho foi realizar uma síntese que poupa não só os alunos, mas também os profissionais da CI, de ter que consultar uma variedade de fontes para lidar com os princípios da análise facetada.

### 2.3.2.1 O Modelo de Ranganathan

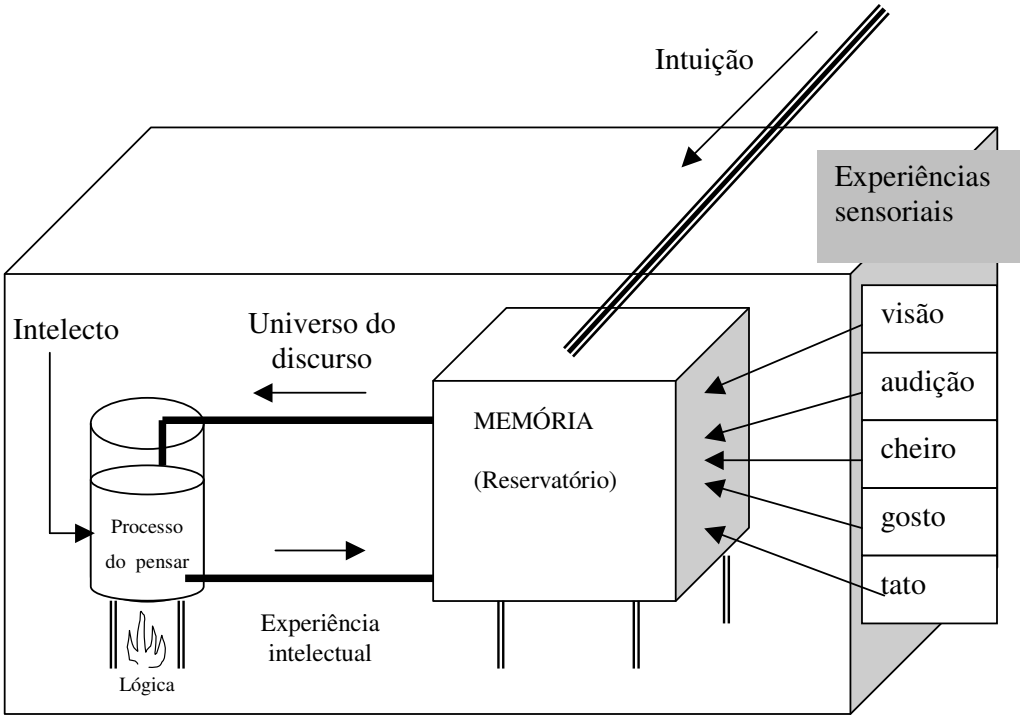
Shiyali Ramamrita Ranganathan (1892-1972), matemático indiano que se tornou bibliotecário, foi um dos estudiosos que mais contribuiu para a teoria da biblioteconomia no século XX, especialmente na área da classificação de assunto. Seu sistema de Classificação de Dois Pontos (*Colon Classification*), publicado em 1933, surgiu de sua insatisfação com os sistemas de Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU). Depois de constatar que a maioria dos assuntos tratados em cinco periódicos diferentes era formada de assuntos compostos, Ranganathan idealizou sua *Classificação de Dois Pontos,* também conhecida como Classificação em Facetas ou Classificação Analítico-Sintética. Baseado nesta evidência, criou um sistema quase matemático, estruturando o conhecimento de maneira que os assuntos compostos sinteticamente surgissem a partir de conceitos "elementares". Ranganathan publicou seis edições desse sistema e faleceu em 1972, quando sua sétima versão estava para ser publicada. Além desse sistema de classificação, Ranganathan publicou *Elements of Library Classifications* (1945; 3ed. 1962) e *Prolegomena to Library Classification* (1937.1957,1967), obras consideradas referenciais pelos teóricos da classificação. Dentre os princípios introduzidos por Ranganathan, o mais conhecido é o princípio de análise de faceta (subdivisões de assuntos em suas partes componentes) e a síntese (recombinação destas partes para se classificar um documento adequadamente). Embora o sistema de Classificação de Dois Pontos não tenha sido aceito mundialmente, o referencial teórico sobre a análise e síntese de facetas proposta por Ranganathan transformou-se em uma base teórica importante para a área de análise de assunto no século XX (RANGANATHAN, 1985, p.86). Uma das grandes contribuições de Ranganathan foi sua ordem de citação, em que divide todo e qualquer assunto em cinco categorias: Personalidade/Entidade (*Personality*/*Entity*), Matéria (*Matter*), Energia (*Energy*), Espaço (*Space*), Tempo (*Time*), mnemonicamente conhecidas pela sigla PMEST.

A origem da classificação facetada se deu com os estudos de Ranganathn e com a Classificação de Dois Pontos (*Colon Classification*), publicado em 1933, mostrando que, em uma determinada área do conhecimento, o número de assuntos específicos que podem ser selecionados numa classificação é infinito. Entre os dois pontos no sistema, pode ser inserido um número infinito de novos dois pontos. Considerou o

conhecimento como um elemento multidimensional: as interligações de cada conceito espalham-se em muitas direções e, freqüentemente, cada assunto consititui-se de uma síntese de vários conceitos múltiplos ligados, o que caracteriza essa abordagem como analítico-sintética. Além disso, Ranganathan explicou os caminhos da divisão, abrangendo a dicotomia difundida por Kant e a Árvore de Porfírio. Examinando o modelo da dicotomia (com dez divisões), concluiu que, em vista do crescimento prolífico e multi-dimensional do conhecimento ao longo da história, esse não seria satisfatório. Sugeriu, então, um padrão para mapear o conhecimento com uma policotomia ilimitada (número ilimitado de divisões das áreas do conhecimento).

Ao discutir a Árvore de Porfírio, com seu eixo único, buscou uma alternativa na Árvore Baniana (*Banyan Tree*) para explicar de forma aceitável a sua proposta de que, em adição ao tronco original, vários troncos secundários podem ser formados de tempos em tempos, idéia que associou a uma árvore de assuntos (RANGHANATAN, 1967, P.368). Esta sua proposta nos remete à discussão de como são construídos os mapas conceituais e como os hiperdocumentos são organizados. Para ilustrar a formação das idéias, Ranganathan nos apresenta um diagrama esquemático (FIG.9).

FIGURA 9 - Compartimento do cérebro: formação da idéias




**Cérebro (Câmara)**

FONTE: RANGANATHAN, 1967, p. 85

Para Ranganathan, a idéia é produto do pensamento. O conhecimento é representado pela totalidade das idéias armazenadas pela memória humana. A informação existe quando as idéias são comunicadas. Todos esses processos são sensoriais, permeados pelas experiências cognitivas de cada indivíduo.

RANGANATHAN (1967, p.82) descreve assunto como:

> *"Um corpo de idéias organizadas ou sistematizadas, cuja extensão e intensão devem ser coerente com o domínio de interesse e confortavelmente ajustadas à competência intelectual e campo especializado de uma pessoa qualquer."*

Todo assunto é oriundo de um assunto básico somente e seu componente vem de uma ou mais idéias isoladas, formando assim o assunto composto. O assunto básico é um assunto sem nenhuma idéia isolada como componente. A idéia isolada (isolado) é alguma idéia ou complexo de idéias, moldada para formar um componente de assunto. Mas, sem seu contexto, ela não é considerada um assunto. Quando combinado com um assunto básico, o isolado é considerado um qualificador

(RANGHAVAN, 1985, p.27). Quando um termo genérico é usado para denotar algum componente, quer seja um assunto básico quer seja um isolado de um assunto composto, tendo ainda, a função de formar renques, termos ou números, ele é chamado de faceta. Grupos de isolados dentro de uma faceta são denominados focos, que são as subdivisões de uma faceta e correspondem às espécies de um gênero (RANGANATHAN, 1967, p.88).

Ranganathan, no *Prolegomena to Library Classification* (1967), define outros termos que devem ser apresentados aqui para um melhor entendimento de sua teoria. Classe *". . .é um conjunto de coisas ou idéias que possuem vários atributos, predicados ou qualidades comuns".* Categorias *". . .são as maiores classes de fenômenos, as classes mais gerais que podem ser formadas*" e que podem ser empregadas para reunir outros conceitos. Característica é a *"qualidade ou atributo escolhido para servir de base à classificação ou à divisão",* 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 formam uma divisão 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 especifico, formando as relações hierárquicas dos assuntos. Termos são as representações verbais dos conceitos em uma linguagem natural. Conceito é qualquer unidade de pensamento de qualquer nível de complexidade.

Antes de abordar os procedimentos da formação das categorias nos planos das idéias (verbal e notacional), utilizando o Modelo Simplificado de Spiteri, será descrita a formação de assunto segundo Ranganathan, uma vez que este processo ocorre antes da categorização.

Ranganathan utiliza a noção de categoria para a análise de assuntos dos conteúdos de documentos e para a organização desses assuntos (**isolados**) em uma estrutura conceitual, quando cria as categorias ou facetas. Isto se dá em dois momentos: na **Formação de Assuntos** e na **Formação de Categorias**.

Para a **Formação de Assuntos**, RANGANATHAN (1967, Parte P, p.351) propõe cinco métodos: Dissecação, Laminação, Desnudação, Reunião/Agregação e Superposição.

**Dissecação** (*Dissecation*): Na Dissecação, propõe-se dividir o universo em partes coordenadas em um mesmo nível. Cada parte, chamada por Ranganathan de "lâmina", pode representar um assunto básico ou um isolado. Esse processo de dissecação pode ocorrer quantas vezes forem necessárias, formando cada um seu próprio universo. Quando organizadas em fileiras, as lâminas formam um renque, como mostra a FIG.10:

FIGURA 10 - Exemplo do Método de Dissecação

| **Universo de assunto básico** | **Universo de isolado – Plantas agrícolas** |
|---|---|
| Botânica | Plantas Sementeiras |
| Agricultura | Plantas Alimentícias |
| Zoologia | Plantas Estimulantes |

**Laminação** (*Lamination*): Na Laminação, propõe-se a superposição de uma faceta com outra, constituindo-se camadas de assuntos básicos e idéias isoladas ou isolados que, quando combinados, formam assuntos compostos. Por exemplo, o assunto composto "Agricultura do milho" é formado pela laminação do assunto básico "Agricultura" e da idéia isolada "Milho."

**Desnudação** (*Desnudation*): Na Desnudação, também chamada de desfolhamento, ocorre uma diminuição progressiva da extensão e um aumento da intensão (profundidade) de um assunto básico ou de uma idéia isolada. Este processo permite a formação de cadeias através de resultados sucessivos da desnudação, permitindo representar o núcleo específico de um assunto básico ou de uma idéia isolada. Por exemplo, considerando-se que "Filosofia" é um assunto básico mais abrangente e que contem "Lógica" como um assunto básico mais específico, pode-se buscar um detalhamento ainda maior pelo processo de desnudação, chegando-se ao assunto básico "Lógica dedutiva" (FIG.11)

FIGURA 11 - Exemplo do Método de Desnudação

♦ Filosofia

♦♦ Lógica

♦♦♦ Lógica dedutiva

**Reunião** (*Loose assemblage*): No processo de Reunião (também chamado de Agregação), ocorre a combinação do assunto básico ou composto com idéias isoladas formando um assunto complexo, ou uma idéia isolada complexa. Ranganathan cita vários exemplos de assuntos complexos, entre eles "Relação geral entre a Ciência Política e a Economia", "Análise estatística para gerentes de ferrovias" e "Influência da Geografia na História". Como exemplos de idéia isolada, cita "Influência do Budismo no Cristianismo" e "Diferença entre vertebrados e invertebrados".

**Superposição** (*Superimposition*): Na superposição (também chamada de Sobreposição), ocorre a conexão de duas ou mais idéias isoladas que pertencem ao mesmo universo de idéias isoladas. A idéia isolada resultante desta superposição é chamada de idéia isolada superposta ou idéia isolada composta. Para dar um exemplo deste processo, Ranganathan nos remete à seção EN3, p.160, do seu *Prolegomena*, onde ilustra com o universo dos professores e suas características. No universo de isolado, o termo "professor" pode ser classificado tanto pela característica "assunto" quanto pela característica "habilidade retórica". Ele considera essas duas características como uma idéia quase isolada, de maneira que os assuntos formados pela reunião destas duas características são idéias superpostas, como por exemplo, "professor de química brilhante", "professor de zoologia medíocre".

Na **Formação de Categorias** (RANGANATHAN, 1967, Parte EA, p. 143), foram avaliadas as características comuns de cada conceito, analisados e definidos os termos isolados acima, conforme suas características comuns, para o procedimento de categorização e a definição das facetas. Segundo RAGHAVAN (1985, p.27), a análise dos conceitos e a relação entre eles encontradas em um determinado assunto servem como base para uma estruturação que expressa o significado e a relação entre eles. Para o autor, a análise de assunto de centenas de documentos tem indicado que os conceitos encontrados no discurso podem ser categorizados em um número ilimitado de categorias. Baseado em suas relações categoriais, os conceitos que constituem o assunto de um documento podem ser hierarquicamente estruturados para expressar as relações entre eles com o propósito de representação e organização da informação.

#### 2.3.2.2 As contribuições do *CRG*

O *Classification Research Group (CRG), de Londres,* começou seus estudos cooperativos em 1952. Em 1955, o *CRG* inlcuia quatorze componentes: D. J. Campell, E. J. Coates, J. E. L. Farradane, D. J. Foskett, G. Jones, J. Milles, T. S. Morgan, B. I. Palmer, O. W. Pendleton, L. G. M. Roberts, B. C. Vickery, A.J. Walford, K. E. Watkins e A. J. Wells. Alguns membros do *CRG* construíram sistemas de classificação facetados para assuntos específicos, como por exemplo, o *London Education Classification*, o *London Classification for Business Studies* e a *Classification for Library and Information Science* (DANIEL e MILLS 1976; FOSKETT, D.J. 1963; VERNON 1979).

O *CRG* utilizou, como base para construção de sistemas de classificação bibliográfica, a Teoria da Análise Facetada a partir das idéias já propostas por Ranganathan no *Prolegomena*, mas modificou alguns aspectos dessa teoria, por considerá-la muito restritiva (AUSTIN, 1969; CLASSIFICATION RESEARCH GROUP, 1985; WILSON, 1972). Os documentos da Teoria da Análise Facetada, segundo o ponto de vista do *CRG,* não se encontram em uma única fonte, mas dispersos em vários trabalhos publicados pelos diferentes membros do Grupo. O *CRG* também não apresenta sua teoria organizada em princípios, como fez Ranganathan. Isto dificulta sua difusão entre profissionais e alunos da Ciência da Informação. Os princípios do *CRG* apresentados por Louise Spiteri foram selecionados a partir de diversas fontes, dentre elas AUSTIN (1968; 1969), CLASSIFICATION RESEARCH GROUP (1985), A. C. FOSKETT (1996), D. J. FOSKETT (1970; 1971; 1974), MILLS e BROUGHTON (1977) e VICKERY (1960; 1966; 1975).

As propostas metodológicas de Ranganathan e do *CRG* possuem muitos pontos comuns. À primeira vista, suas terminologias diferem em alguns pontos e, observadas mais de perto, verifica-se que o conteúdo de alguns conceitos se sobrepõe.

Em relação ao princípio da ordem nos renques, o *CRG* diz que focos não devem ser arranjados necessariamente em algum tipo de ordem. Argumenta que nenhuma é correta *a priori*, mas sugere a seguinte ordem de assunto: simples para complexo;

complexo para simples; espacial/geométrico; cronológico (incluindo as ordens evolucionária e histórica); e alfabética. Esta seqüência, entretanto, corresponde ao princípio da seqüência útil de Ranganathan. Assim, no Modelo Simplificado, Spiteri elas se confundem.

Quanto ao princípio de divisão, o *CRG* afirma que uma faceta deve representar somente uma característica de divisão no domínio, o que, em Ranganathan, corresponde aos princípios de exclusividade e simultaneadade, homogeniedade e de exclusividade mútua.

Quanto ao princípio de relevância, o *CRG* afirma que as facetas devem ser escolhidas pelas suas relevâncias na proposta, no objetivo e no domínio do sistema de classificação, o que corresponde exatamente ao princípio de relevância de Ranganathan. A mesma correspondência acontece com os princípios de permanência e verificação.

Por outro lado, o ponto de vista do *CRG* sobre análise facetada difere significativamente do propósitos de Ranganathan em dois pontos: escolha das categorias fundamentais e ordem de citação. Com relação às categorias fundamentais, o *CRG* afirma que estas devem ser derivadas da natureza dos assuntos a serem classificados e que nem todos os assuntos possuem categorias fundamentais. O *CRG* prefere identificar as categorias fundamentais pelo contexto do próprio assunto, concluindo que nenhuma lista de categorias fundamentais deva ser exaustiva ou imposta mecanicamente aos assuntos. Comparadas ao PMEST de Ranganathan, as categorias fundamentais propostas pelo *CRG*, devido à flexibilidade que permitem, são atraentes aos classificadores, que podem moldá-las à assuntos específicos, permitindo assim, a formação de categorias mais distintas e definidas Assim, como o presente estudo adota a abordagem do *CRG* quanto à escolha das categorias fundamentais, o PMEST de Raganathan não será diretamente obedecido como ordem de citação ( SPITERI, 1998, p.19).

Da mesma forma, em relação ao princípio de ordem de citação, enquanto que Ranganathan sugere que todas as facetas devam ser organizadas em uma única ordem prescritiva, independentemente do assunto, o *CRG* argumenta que a ordem de citação escolhida deve refletir a natureza, o assunto e o domínio do sistema de

classificação. O *CRG* também não concorda que a garantia literária deva ser utilizada para arranjar os focos, devido às sutilezas e à imprevisibilidade das diversas tendências editoriais, o que não era tão problemático na época de Ranganathan.

O *CRG* ampliou as cinco categorias do PMEST de Ranganathan - Personalidade/Entidade (*Personality*/*Entity*), Matéria (*Matter*), Energia (*Energy*), Espaço (*Space*), Tempo (*Time*) - para dez: Tipos de Produto Final (*Final product type*), Partes (*Parts*), Materiais (*Materials*), Propriedades (*Property*), Processos (*Process*), Operações (*Operation*), Agentes (*Agent*), Espaço (*Space*), Tempo (*Time*) e Forma de Apresentação.

No seu modelo simplificado, Spiteri buscou sintetizar os cânones e postulados (aqui tratados sob o mesmo termo "princípios") de Ranganathan e o plano das idéias do *CRG* em apenas dois grupos de princípios: (1) sete princípios do plano das idéias para escolha das facetas e dos focos, e (2) dois princípios para ordem de citação das facetas e dos focos, subdivididos em (a) Princípios das Sucessões Relevantes com oito subidvisões e (b) Princípio da Sucessão Consistente. Spiteri utiliza, ainda, quatro princípios do plano verbal e dois princípios do plano notacional, que não se aplicam ao presente estudo (Anexo 2).

#### 2.3.2.3 O Modelo Simplificado de Louise Spiteri

Louise Spiteri é *Associate Professor* da *School of Library and Information Studies* da Dalhousie University, Halifax, Nova Scotia, EUA. Graduada com o Bachelors of Arts, o Master of Arts (York, Toronto), o Master of Library and Infomation Science MLIS (Western) e o PhD (Toronto), seus interesses de pesquisa incluem os sistemas de análise de assunto, teoria da classificação e construção de tesauros e indexação. Leciona as disciplinas Organização da Informação, Catalogação e Classificação, Indexação e Resumos e Gerenciamento de Documentos. Spiteri, que tem vários artigos publicados nas áreas em que atua, esboçou a integração entre os dois referenciais anteriores, que originou, ao final do século XX, um modelo simplificado dos princípios da análise facetada.

O primeiro objetivo de Louise Spiteri em desenvolver o *Modelo simplificado para análise facetada* foi a construção de uma ferramenta de ensino para introduzir aos alunos um modelo de classificação mais fácil de ser lido e compreendido, que pudesse possibilitar a eles o entendimento de como os sistemas de classificação facetados são elaborados e como funcionam. O *Modelo Simplificado* de Spiteri tem servido para outras funções, como auxiliar bibliotecários na elaboração de sistemas de classificação facetada e de tesauros para recuperação da informação, fornecendo critérios necessários para a avaliação da integridade estrutural desses sistemas.

Como já dito, uma das finalidades dessa tese foi valorizar a simplificação na representação de conceitos e, assim, auxiliar os autores de hiperdocumentos na organização dos mesmos, visando à navegação em textos completos. Por isso, levou-se em consideração a escolha de um modelo mais simples para o procedimento de categorização na análise facetada. A escolha do trabalho de Spiteri como a principal referência teórica nesse estudo se deu pelo fato do *Prolegomena to Library Classification* de Ranganathan ser um manual que exige maior esforço na sua compreensão e pelo modo de como seu conteúdo é apresentado e estruturado, o que dificulta sua consulta. Além disso, os princípios do *CRG*, embora mais atualizados, encontram-se dispersos em artigos e livros publicados em períodos diferentes, o que torna sua consulta menos objetiva e produtiva, especialmente no tratamento de um conjunto maior de teses, que é um objetivo do futuro Banco de Teses e Dissertações da ECI/UFMG. O Modelo Simplificado de Spiteri (apresentado completo no Anexo 6.2) foi utilizado somente na aplicação dos princípios do Plano das idéias para a formação das categorias e para a ordem de citação das facetas, pois foi este o plano escolhido nesta tese. A seguir, descreve-se o *Modelo simplificado para análise facetada* de Spiteri, observando quais são os princípios originados da Teoria de Ranganathan, quais vêm da Teoria do *CRG* e quais pertecem a ambas as teorias.

**Princípios para o Plano das idéias**

**1 . Princípios de escolha das facetas**

**a) Princípio da diferenciação:** Este princípio, proposto por Ranganathan, é baseado em características de divisão determinadas por diferenças ou qualidades comuns que distinguem elementos de uma mesma classe. Por exemplo, “Seres

humanos", pode ser dividido pela característica gênero, em "Feminino" e "Masculino".

**b) Princípio da relevância:** Este princípio, proposto tanto por Ranganathan quanto pelo *CRG*, visa 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 EDUCAÇÃO, é sensato que se utilize a faceta "grau escolar" para dividir os conceitos "primeiro grau" e "segundo grau".

**c) Princípio da verificação:** Este princípio, proposto tanto por Ranganathan quanto pelo *CRG*, visa a escolha de facetas definitivas e que podem ser verificadas. Por exemplo, em "Raças de cães", "Raça" não é uma qualidade que pode caracterizar com certeza um cão, porque existem autores diversos, como veterinários ou criadores de cães, que listam raças de cães de maneiras divergentes.

**d) Princípio da permanência:** Este princípio, proposto por Ranganathan e pelo *CRG*, prevê que as facetas escolhidas devem representar características de divisão com qualidades permanentes ao assunto a ser dividido. No exemplo dado anteriormente, há uma observância do Princípio de permanência porque, por exemplo, um cachorro de raça "Dálmata" será sempre um dálmata.

**e) Princípio da homogeneidade:** Este princípio, proposto apenas pelo *CRG*, afirma que o conteúdo da faceta deve ser homogêneo, não permitindo a superposição de duas facetas, de tal forma que cada faceta deve representar somente uma característica de divisão. Por exemplo, na divisão do conceito "Cães" pelas facetas "tamanho" e "cor", os termos que aparecerem em uma faceta não poderão aparecer na outra.

**f) Princípio da exclusividade mútua**: Este princípio, proposto tanto por Ranganathan quanto pelo *CRG*, afirma que as facetas devem ser mutuamente exclusivas. Muito semelhante ao princípio de homogeneidade, mas de caráter mais limitante, esse princípio estabelece que, para assegurar essa exclusividade, as classes de um renque devem derivar-se de seu universo imediato uma e somente uma característica.

**g) Princípio das categorias fundamentais**: Este princípio, proposto somente pelo *CRG*, afirma que não existem categorias consideradas fundamentais sem as quais qualquer assunto ficaria indefinido, e que todas as categorias devem ser derivadas com base no assunto a ser classificado. Pode haver uma correspondência com o acrônimo PMEST de Ranganathan, mas o *CRG* prefere identificar as categorias fundamentais baseadas na referência do contexto do próprio assunto, sugerindo que nenhuma lista pode ser imposta mecanicamente ao assunto. Além disso, o *CRG* acredita que nenhuma lista de classificação deva ser necessariamente exaustiva ou aplicada a todos os assuntos. Spiteri preferiu, neste princípio, a abordagem do *CRG* na síntese do seu Modelo Simplificado, porque o mesmo permite aos classificadores moldar melhor as categorias para assuntos muito específicos, permitindo a formação de categorias mais distintas e mais bem definidas. Segundo essa pesquisadora, a maioria dos sistemas de classificação facetada e tesauros consultados usam a abordagem do *CRG* na escolha das categorias fundamentais.

**2. Princípios para a ordem de citação das facetas e focos**

Estes princípios coordenam a organização dos focos dentro de suas respectivas facetas e, conseqüentemente, a ordenação destes focos no renque.

**a) Sucessão relevante:** Tanto Ranganathan quanto o *CRG* concordam que a ordem de citação das facetas deve ser relevante quanto à natureza, o assunto e o objetivo do sistema de classificação. O *Modelo Simplificado* sugere que esta ordem seja expandida para garantir que, tanto a ordem de citação das facetas, quanto a dos focos devem ser relevantes ao sistema de classificação. As seguintes ordens de citações são derivadas do Princípio de Seqüência Útil de Ranganathan e do Princípio de Ordem no renque do *CRG*:

- **Ordem cronológica:** segue-se a ordem cronológica, incluindo nesta ordem operações que, necessariamente, sejam realizadas uma após a outra.
- **Ordem geométrica/espacial:** Este princípio sugere que os focos possam ser organizados em uma ordem espacial ou geométrica.
- **Ordem do simples para o complexo:** Este princípio sugere que os focos devem ser arranjados na ordem do simples para o complexo.
- **Ordem do complexo para o simples:** Este princípio sugere que os focos devem ser arranjados na ordem do complexo para o simples.

- **Ordem canônica:** Esse princípio privilegia a ordem tradicional de um assunto, ou isolados em um renque de isolados, quando não há outro princípio a seguir.
- **Aumento de quantidade:** se um assunto em um renque de assunto, ou isolados em um renque de isolados tiver a distinção de quantidade, eles devem serem arranjados de acordo com a quantidade crescente.
- **Diminuição de quantidade:** se um assunto em um renque de assunto, ou isolados em um renque de isolados tiver a distinção de quantidade, eles devem ser arranjados de acordo com a quantidade decrescente.
- **Ordem alfabética:** Quando nenhuma outra seqüência de assunto em um renque de assunto, ou isolados em um renque de isolados for mais útil, aplica-se a ordem alfabética pelos nomes de uso internacional corrente.

**b) Sucessão consistente:** Baseado no Princípio de Sucessão Consistente de Ranganathan, este princípio ajuda a manter a consistência na estrutura de um sistema de classificação. O *CRG* não apresenta um princípio equivalente a este. Apesar disto, ele foi mantido porque ajuda a manter a consistência na estrutura de um sistema de classificação, pois sua expansão permite que a ordem de citação tanto das facetas quanto dos focos seja mantida consistentemente.

**Princípios do plano verbal**

Nem todos os princípios seguintes têm equivalência na teoria de Ranganathan ou na Teoria do *CRG*, porém suas especificidades justificaram sua manutenção**.**

**1. Princípio do contexto:** Este princípio, proposto por Ranganathan, foi mantido porque ele ajuda a assegurar a clareza dos termos utilizados em sistemas de classificação.

**2. Princípio da terminologia usual (uso geral):** Este princípio, proposto por Ranganathan, foi mantido porque ajuda a manter a relevância dos termos usados em um sistema de classificação.

**Princípios do plano Notacional**

**1. Princípio de sinônimo:** Este princípio, proposto por Ranganathan, foi mantido por ajudar a assegurar a exclusividade mútua da notação utilizada no sistema de classificação.

**2. Princípio de homônimo:** Este princípio, proposto por Ranganathan, foi mantido por ser um resultado ou conseqüência do princípio de sinônimo.

**3. Princípio de hospitalidade:** Este princípio, proposto apenas pelo *CRG*, foi mantido porque ajuda assegurar a habilidade do sistema notacional de estar aberto às mudanças e adições de assuntos que precisam ser feitos em sistemas de classificação.

**4. Princípio de ordem de fichamento:** Este princípio, proposto apenas pelo *CRG*, foi mantido para assegurar que a notação refletisse a ordem da tabela usada no sistema de classificação.

Observa-se que o Modelo Simplificado para Análise Facetado pode ser um instrumento de grande valia para os profissionais da área, ajudando-os na criação de tesauros facetados e sistemas de classificação facetados. Sua grande flexibilidade permite moldar linguagens conforme a necessidade tanto dos usuários como dos assuntos a serem organizados e, conseqüentemente, permite melhorar a recuperação da informação em um SRI.

### 2.3.3 Mapa conceitual

Mapa conceitual, termo composto cujo significado podemos inferir a partir de seus termos simples, é uma técnica de organização do conhecimento. Entre os significados do termo “mapa” encontra-se as conotações de representação, lista descritiva ou relação (AURÉLIO, p.1086). Já a palavra "conceitual" é tida como sendo representativa das características gerais de um objeto ou idéia (AURÉLIO, p. 445). Partindo dessas definições, pode-se dizer, então, que mapa conceitual seja uma representação que descreve a relação das idéias do pensamento, relação esta pré-adquirida ao longo de um processo de aprendizagem na construção do conhecimento e que vai se arquivando na memória. O mapa conceitual pode

também estar relacionado com outros termos como “rede semântica”, “estrutura do conhecimento”, “estrutura cognitiva”, “mapa cognitivo”, “mapa mental” ou “mapa da *Web*”. O presente estudo aborda mapa conceitual como uma ferramenta de organização do conhecimento, capaz de representar idéias ou conceitos na forma de um diagrama hierárquico escrito ou gráfico e capaz de indicar as inter-relações entre os conceitos, procurando refletir a organização da estrutura cognitiva do indivíduo sobre um dado assunto.

As definições sobre mapa conceitual vêm esclarecer suas aplicabilidades. MOREIRA (1987, p.10 ) diz que são:

> *"representações gráficas de uma estrutura de conhecimento demonstrada hierarquicamente, apresentando forma e representação condizentes com a maneira como os conceitos são relacionados, diferenciados e organizados".*

SHERRATT e SCHLABACH (1990, p.60) descrevem o mapeamento conceitual de forma semelhante à definição acima:

> *“O mapeamento conceitual envolve a identificação de conceitos ou idéias pertencentes a um assunto, e a descrição das relações existentes entre essas idéias na forma de um desenho esquemático. O objetivo deste mapa é representar a compreensão de um indivíduo sobre um corpo de conhecimento e ilustrar as relações entre as idéias que são significativas para este indivíduo.”*

Para GAINES e SHAW (2002, p.1), mapa conceitual é

> "*uma forma de diagrama especificamente direcionado para fornecer uma linguagem visual parecida com as características da linguagem natural do texto, no sentido de que eles possam estar sujeitos às limitações sintática e semântica, e sua capacidade de representação pode variar de uma forma muito informal a uma forma extremamente formal."*

Nessa última definição, os autores chamam a atenção para as limitações sintática e semântica que podem ocorrer na representação do conhecimento através do mapa conceitual. Apesar de possuir uma estrutura hierárquica e permitir as relações entre termos, quando organizado à luz da linguagem natural, o mapa conceitual pode não representar o conteúdo semântico do termo, cuja linguagem visual seja uma alternativa para a comunicação da informação. Por outro lado, fica clara a flexibilidade do mapa conceitual em lidar formal e informalmente com os conteúdos.

Em outra definição, LANZING (2002, p.1) aproxima a estrutura do mapa conceitual à estrutura de uma rede hipertextual:

> *"Mapeamento conceitual é uma técnica para representação do conhecimento em forma gráfica. Gráficos do conhecimento são rede de conceitos. A rede consiste de nodos* [de informação] *com apontadores e links com arcos ou subdivisões. Os nodos representam os conceitos e os links representam as relações entre conceitos".*

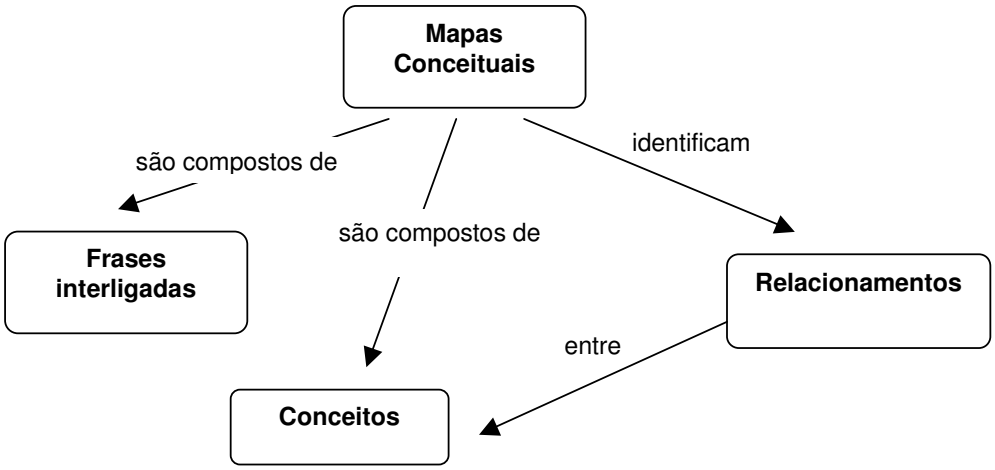
Joseph D. NOVAK (2002, p.1), da *Cornell University* desenvolveu a Teoria do Mapa Conceitual a partir da Teoria da Aprendizagem Significativa, da Psicologia Educacional de David AUSUBEL (1963, 1968, 1978). Novak considera o mapa conceitual como uma forma de organizar e representar o conhecimento. Em 1960, Ausubel aplicou essa técnica em uma pesquisa na área de educação, nos EUA, que tinha como objetivo entender o processo de aprendizagem do conhecimento científico pelas crianças. Esta teoria parte do pressuposto de que o armazenamento de informações ocorre a partir da organização de conceitos de forma hierárquica, tanto nas relações envolvendo os elementos mais genéricos, quanto os mais específicos, apontando similaridades e diferenças e tirando vantagem da seqüência natural dos tópicos do conteúdo a ser trabalhado. A idéia fundamental da teoria de Ausubel é que a aprendizagem é feita por assimilação de novos conceitos. Segundo ele (AUSUBEL, 1963; 1968; 1978), a criança adquire os primeiros conceitos por volta dos 3 anos de idade, quando começa a reconhecer o mundo em volta deles. A partir daí, novos conceitos são adquiridos através da linguagem e processo de aprendizagem, onde interagem perguntas e esclarecimentos sobre relações dos velhos e novos conceitos. A Teoria da Aprendizagem Significativa é baseada no modelo construtivista do processo cognitivo humano, que explica como os conceitos são adquiridos e organizados dentro de um aprendizado cognitivo.

NOVAK (2002, p.2) refere-se a duas características básicas do mapa conceitual. Primeiro, os conceitos são representados de forma hierárquica, com o conceito mais geral no início do mapa e depois os mais específicos, arranjados hierarquicamente. Segundo, a estrutura hierárquica de um campo específico do conhecimento também depende do contexto em qual o conhecimento está sendo considerado. Além disso,

os mapas conceituais apresentam referências cruzadas que permitem verificar como é representada a relação dos conceitos em um domínio do conhecimento.

O diagrama na FIG.12 representa o conceito do mapa conceitual pode ser, em que a técnica usual de subdividir classes hierarquicamente em um número de subclasses, em forma de árvore, que equivale à técnica utilizada por *Ranganathan* de criar subgrupos pelo processo de facetação. Todo conceito em qualquer nível na hierarquia terá sempre um nível mais geral. As relações entre os conceitos no mapa conceitual representam as preposições, que constituem unidades semânticas, por ligarem dois ou mais conceitos.

FIGURA 12 - Estrutura dos mapas conceituais




FONTE: Baseado em TAROUCO, Liame. http//penta2.ufrgs.br/edu/ambvirt/sld058.htm

O mapa conceitual tem sido utilizado em diversas áreas do conhecimento como uma técnica formal ou semi-formal de diagramação. Nas áreas de educação, estudos políticos, lingüística e filosofia da ciência, essa técnica tem sido usada para apresentar visualmente a estrutura do conhecimento e suas formas de argumentação. Na área de administração, AXELROD (1976),[23] citado por GAINES e SHAW (2002, p.2), o utilizou para representar a estrutura do conhecimento que suporta o processo de decisão para análise organizacional de empresas, sistemas sociais e diretrizes de líderes políticos. Na área de inteligência artificial, QUILLIAM

[23] AXELORD, R. *Structure of decision*. Princeton, New Jersey: Princeton University Press. 1976.

(1968)[24], também citada por GAINES e SHAW (2002, p.2), desenvolveu uma forma de mapa conceitual que resultou em uma rede semântica usada extensivamente para representação formal do conhecimento. O mapa conceitual nos permite, pois, entender de forma facilitada e ao primeiro contato visual complexos de informação. Uma das principais funções da mente é interpretar o significado das informações adquiridas e transformá-las em conhecimento, o que se torna mais fácil quando são apresentadas em formato gráfico. O mapa conceitual pode ser um suporte apropriado para a arquitetura de sistemas de hipertexto por possibilitar uma interface atrativa, interativa e fácil de ser utilizada, facilitando a navegação em redes semânticas.

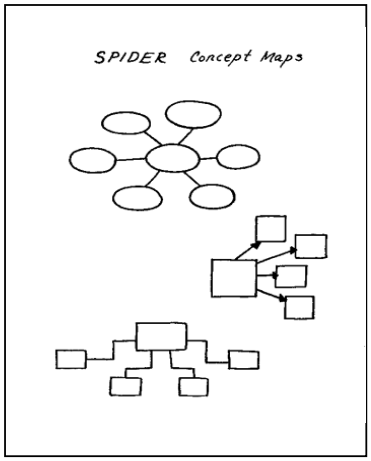
Todo conhecimento é processado e organizado pela interação da memória de curta duração com a memória de longa duração. Existe um limite para o processamento da memória de curta duração que nos permite processar de cinco a nove unidades de informação por vez, sendo que cada conceito pode combinar duas ou três unidades de informação. Assim, podemos processar somente a relação entre dois ou três conceitos de cada vez. O mapa conceitual, com sua característica gráfica, é um instrumento poderoso para permitir a compreensão das relações entre os conceitos e do conhecimento no todo. Para o cientista da informação, que lida com a análise de assunto para estruturação de uma área do conhecimento, o mapa conceitual pode tornar-se instrumento importante para ajudá-lo a entender e a lidar com complexos informacionais.

O *site* do *College of Agricultural Consumer and Environmental Sciences* da *University of Illinois at Urban-Champaign*, na disciplina *The Mind Module*, apresenta o artigo *Kinds of Concept Maps*, em que os mapas conceituais são classificados em quatro tipos de estrutura:

1. Estrutura em Teia, em que o tema central é colocado no meio do mapa (FIG.13)

[24] QUILLIAN, M:R. *Semantic memory* In: MINSKY, M.,Ed. Semantic Information Processing. Cambridge, Massachisetts: MIT PRESS, 1968. p. 216-270.

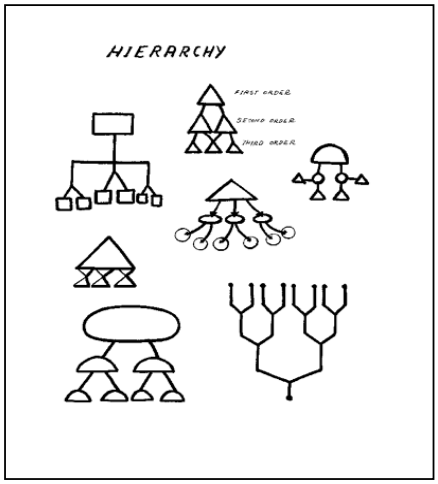
FIGURA 13 - Estruturas em teia




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

2. Estrutura Hierárquica, que apresenta a informação em forma descendente de importância, sendo que a informação mais importante é colocada no início da cadeia hierárquica (FIG.14).

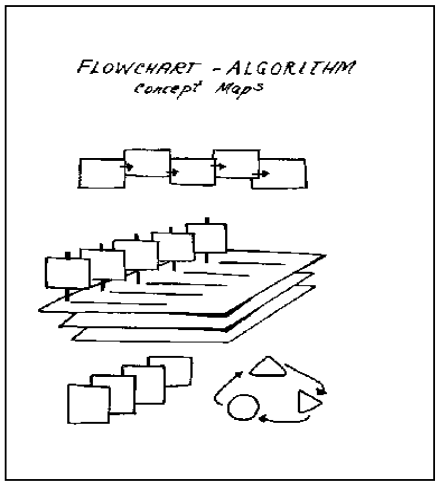
FIGURA 14 - Estruturas hierárquicas




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

3. Estrutura *Flowchart*, que organiza a informação em formato linear, semelhante à estrutura de um livro (FIG.15).

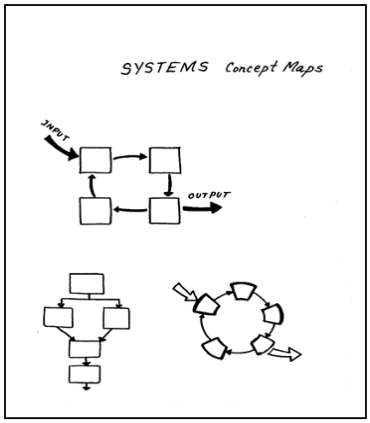
FIGURA 15 - Estruturas *Flowchart*




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

4. Estrutura Conceitual, que organiza as informações no formato parecido com o *flowchart*, mas com a possibilidade de inserção e exclusão de novos conceitos (FIG.16).

FIGURA 16 - Estruturas conceituais

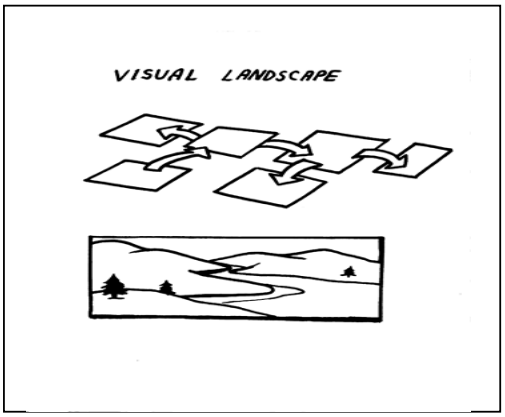



FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

Quanto à forma de apresentação, os mapas conceituais podem classificar-se segundo os seguintes tipos:

1. Mapa Conceitual Paisagem (*Landscape picture*): para situações em que há necessidade de apresentar a informação em contextos panorâmicos (FIG.17);

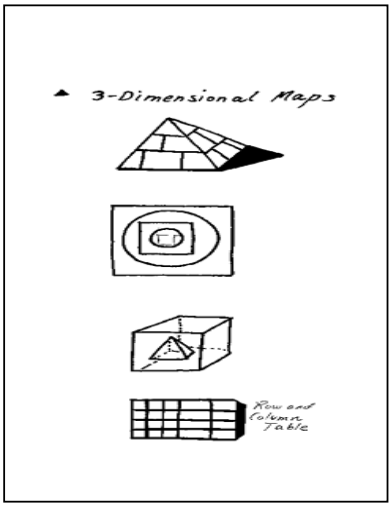
FIGURA 17 – Mapa conceitual paisagem




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

2. Mapa Conceitual 3-D (multidimensional): que utiliza a profundidade (terceira dimensão) para representar relações entre os conceitos que não são atendidas pelos mapas em duas dimensões (FIG.18);

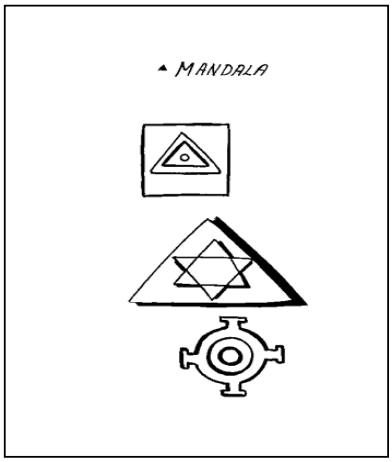
FIG.18 – Mapa conceitual 3-D




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

3. Mapa Conceitual Mandala: que apresenta as informações em formatos geométricos, nos quais sua característica telescópica permite um efeito visual no qual o foco da atenção está no processo de pensamento do usuário (FIG.19)

FIGURA 19 – Mapa conceitual mandala




FONTE: http://classes.aces.uiuc.edu/ACES100/Mind/c-m2.html, 2002, p.1.

Quando se estrutura o conhecimento através do mapa conceitual, é necessário que, a partir de um termo ou uma palavra-chave, a relação existente com os demais assuntos já pré-concebidos seja analisada, interpretada e compreendida e, através desta, sejam construídos novos conhecimentos.

Os mapas conceituais podem ser construídos para diferentes finalidades: (1) para gerar idéias por meio de *brain storming* (“tempestade de idéias”) e a partir da qual são feitas a compilação e análise das informações e estabelecidos relacionamentos para formação de outros conceitos; (2) para desenhar uma estrutura complexa de maneira mais amigável, facilitando a estruturação de textos, documentos, hipertextos e *sites* da *Web*; (3) para estruturar e comunicar idéias, com a apresentação de informações na forma gráfica; (4) para auxiliar no processo de aprendizagem, explicitando graficamente a integração de conhecimentos novos e velhos, por meio de comparação de conhecimentos já existentes com novos conhecimentos que vão sendo agregados a um determinado domínio do conhecimento, e (5) para acessar o entendimento ou diagnosticar um desentendimento, pois através da forma gráfica é possível detectar e comparar idéias antagônicas.

O processo de construção de um mapa conceitual envolve etapas semelhantes às da análise facetada: (a) seleção: escolha do assunto e identificação das palavras-chave ou frases relacionadas; (b) ordenação: organização dos conceitos do mais

abstrato para o mais específico; (c) agrupamento: agrupamento dos conceitos que estão no mesmo nível de abstração e aqueles que estão fortemente interelacionados; (d) arranjo: organização dos conceitos em uma representação diagramática; (e) *link* e preposição: conexão dos conceitos com linhas e nomeação de cada linha com uma proposição.

Entre as vantagens do mapa conceitual, podemos citar: (1) a definição de uma idéia central, através do posicionamento do assunto no centro da página; (2) uma clara indicação da importância relativa de cada idéia; (3) a facilidade para encontrar os *links* entre as idéias-chave; (4) a visão geral de toda informação básica em uma página; (5) em decorrência, uma revocação e uma revisão mais eficientes; (6) a inserção de novas informações sem atrapalhar a estrutura informacional; (7) a facilidade para acessar a informação em diferentes formatos e diferentes pontos de vista; (8) a facilidade de compreensão da complexidade de relações entre as idéias; (9) a facilidade de verificar contradições, paradoxos e falhas no material organizado.

#### 2.3.3.1 Mapa conceitual e hipertexto

O mapa conceitual, enquanto componente da navegação hipertextual, é uma técnica de representar o conhecimento em forma gráfica, construindo redes de conhecimentos constituídas de nodos e *links*, nas quais os nodos representam os conceitos e os *links* representam as relações entre os conceitos. A representação concisa e gráfica do conhecimento por meio do mapa conceitual, com ampla possibilidades de relações, sejam elas hierárquicas ou horizontais, resulta em um ambiente ideal para criar uma estrutura navegacional, onde os usuários possam encontrar a informação com conteúdo semântico, principalmente se esta representação for feita por um especialista da área a ser estruturada. O mapa conceitual funciona como um guia navegacional, ajudando o usuário a caminhar pelos *links* entre os *clusters* dos objetos relacionados. Um hiperdocumento, por exemplo, constitui-se de uma série de documentos que possuem interligações entre si, cuja leitura é feita pelo usuário de forma dinâmica.

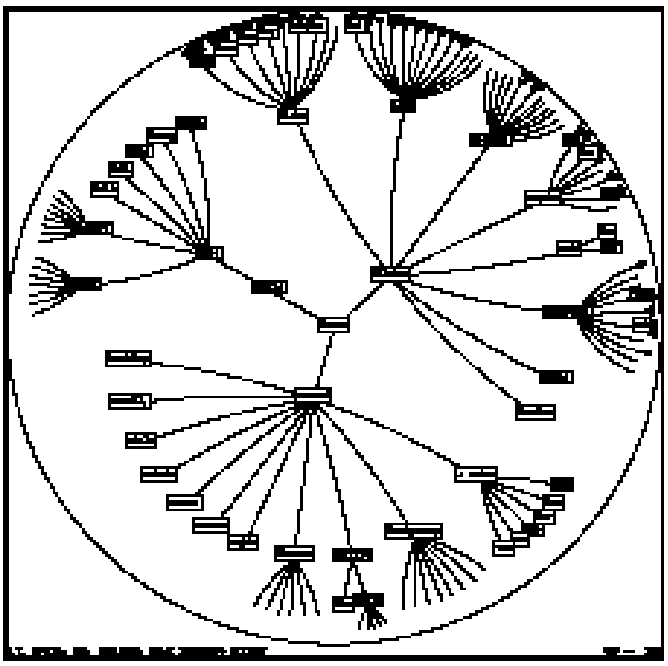
Segundo CHAIBEN (2003, p.14), a maneira mais direta de mapear a estrutura semântica de um hiperdocumento é utilizar o mapa semântico como um *browser* gráfico, ou seja, por meio das listas ou mapas gráficos dos nodos disponíveis no

ambiente hipertexto. O arranjo dos nodos em um *browser* gráfico de acordo com o mapa semântico ilustra os esquemas na estrutura do conhecimento do domínio do hiperdocumento.

Os mapas conceituais permitem uma interface elegante, ágil e amigável para a navegação em sistemas de hipertexto. A navegação orientada por um mapa conceitual, parece ser uma solução para o problema *"lost in cyberspace",* apontado por CONKLIN (1987, p.40) e revisitado por vários autores de hiperdocumentos para os quais o transbordamento cognitivo pela sobrecarga mental e, conseqüentemente, desorientação do usuário derivam do grande número de informações em curto espaço de tempo. Qualquer informação visualizada requer três componentes: organização da informação, representação espacial visual da informação organizada e sua disponibilização com mecanismos de interação.

Um dos tipos de mapas conceituais aplicados ao sistemas de hipertextos mais recentes são os mapas hiperbólicos. O mapa hiperbólico (FIG.20) foi inspirado a partir de uma obra artística de ESCHER (1960), artista conhecido por suas representações tridimensionais em planos bidimensionais, tanto em desenhos quanto em gravuras. Na gravura *Céu e inferno* (*Heaven and hell*) de 1960 (FIG.21), Escher utiliza apenas dois motivos complementares - a pomba e o morcego – para criar, no espaço de um círculo, a perspectiva de uma abóbada celeste, cuja curvatura sugere uma superfície infinita com focos variáveis (LAMPING; RAO; PIROLLI, 1995, p.1).

FIGURA 20 – Mapa com navegação hiperbólica



FONTE: LAMPING, RAO e PIROLLI, 2001, p. 24

FIGURA 21 - Gravura *Céu e inferno* (1960) da série *Circle Limit-IV* de Escher
©1994. Todos os direitos reservados

FONTE: LAMPING, RAO e PIROLLI, 2001, p. 24

Este tipo de mapa hiperbólico possui a interface *Fisheye* (descrita ao final do item 2.1). A partir do pressuposto de que algumas partes do mesmo documento são mais importantes que outras, HORNBAEK e FROKJAER (2001, p.295) observam que a

interface “olho-de-peixe” (*fisheye*) permite reduzir o tempo gasto para navegar através do documento, uma vez que exibe primeiro as seções mais importantes. Inicialmente, as outras partes do documento apresentam-se deformadas ou semi-ocultas abaixo da parte focalizada, mas podem ser expandidas e lidas se o usuário arrastar os diversos *links* com o mouse, ou ainda, quando completamente ocultas, por meio de pesquisa direta pelo conceito, tal como será demonstrado no **MHTX**. Na movimentação da interface “olho-de-peixe”, os componentes diminuem e aumentam de tamanho exponencialmente (saem e entram no foco), demonstrando grande flexibilidade e agilidade na tela.

Todas as seções no mapa hiperbólico podem ser expandidas ou retornar à seção principal do documento. FURNAS (1986, p.16) afirma que a estratégia fundamental da interface “olho-de-peixe” é fornecer um equilíbrio entre o detalhe local e o contexto global. O detalhe local é necessário para as interações com a estrutura geral do mapa. O contexto global é necessário para mostrar ao usuário quais outras partes existem na estrutura e onde elas estão. Uma das aplicações da interface “olho-de-peixe” é a manipulação de grande hierarquias e sua visualização através do mapa hiperbólico.

Outra característica da estratégia do “olho-de-peixe” é a possibilidade de dois focos simultâneos, que pode ser útil quando uma estrutura pequena deve se copiada de um local para outro ou quando se pretende fazer uma autoria (*authoring*) entre duas partes. A interface “olho-de-peixe” também trabalha com o conceito de Grau de Interesse (*DOI*, do inglês *Degree Of Interest*), que é uma função que atribui a cada nó um valor, indicando o grau de interesse do usuário. O DOI relaciona-se com dois parâmetros: (1) o grau de importância *a priori*, diretamente proporcional ao DOI e que considera a relevância do nó na estrutura geral do documento e (2) a distância entre o nó e o foco, inversamente proporcional ao DOI. Somente serão exibidos os nós para cujos valores o DOI ultrapassa um limite de filtragem, chamado de *K*, o qual pode ser configurado pelo próprio usuário. Essa característica não só permite uma personalização do documento, mas também o controle de sua legibilidade (FURNAS, 1986, p.19).

Uma vez que um nó pode estar contido em diferentes nós compostos, o mesmo pode aparecer mais de uma vez no mesmo mapa, desde que a função DOI de cada

um esteja acionada em contexto separados. HORNBAEK e FROKJAER (2001, p.297) explicam detalhadamente as propriedades e programação dos nós na interface “olho-de-peixe”, especialmente a composição de nós compostos, o que extrapola os objetivos desse trabalho. Mas é importante observar que, se o usuário não estiver interessado em ver os componentes dos nós compostos, pode, a partir da função *DOI*, simplesmente ocultá-los, para evitar a poluição visual do mapa, ou exibi-los se necessário.

#### 2.3.3.2 Aspectos cognitivos no mapa conceitual

Uma das principais funções da mente é interpretar o significado das informações adquiridas e transformá-las em conhecimento, o que torna-se mais fácil quando são apresentadas em formato gráfico. Segundo MOREIRA (1993, p.33), a estruturação do conhecimento na mente humana tende a seguir uma estrutura hierárquica a partir das idéias mais abrangentes Um sistema de hipertexto organizado em rede está muito mais próximo dos processos e estruturas do pensamento do que um texto linear. O pensamento desenvolve-se através da mudança e do salto de uma idéia para outra, seguindo um processo de livre associação de idéias, refletindo formas e formatos das ligações orgânicas das próprias células do cérebro.

Sabe-se que as ciências cognitivas têm como objeto de estudo os processos gerais que regem a percepção, o tratamento, o armazenamento, a recuperação e a utilização da informação, bem como as formas como se organizam as representações das atividades no tratamento da informação. A organização conceitual, que está diretamente relacionada à capacidade de aprender, supõe a assimilação de novas informações, sua estocagem e sua acomodação**.**

Para AMORETTI *et al.* (2000, p. 106),

> *... no mapa conceitual, o conhecimento é representado sob a forma de conceitos organizados de forma relacional e modular, em classes e subclasses em que os nós, arquétipos cognitivos ou protótipos integram os actantes e seus atributos. Desta forma, o protótipo é uma representação mental de um exemplar de uma categoria de estímulos que visa a representação de um exemplar genérico a um determinado grupo social.*

A percepção, o raciocínio, a aprendizagem, a linguagem, a comunicação, a organização conceitual e a ação finalizada, são aspectos que fazem parte das ciências cognitivas. O mapa conceitual aparece como uma ferramenta eficaz que pode ser adotada para desenvolver estes estudos dentro das ciências cognitivas. A representação do conhecimento, sob a forma de mapas conceituais, é uma alternativa de estruturar a informação, pois procuram refletir a organização da estrutura cognitiva do indivíduo sobre um dado assunto. Estas estruturas do conhecimento são representações da organização das idéias na nossa memória semântica. Cada estrutura de conhecimento existe como um objeto, uma idéia ou evento e, também, como um grupo de atributos o qual é ligado a outras estruturas do conhecimento. À medida que aprendemos, apreendemos novas estruturas e ligações, adicionando informações às estruturas existentes, ou alterando essas estruturas através do processo de restruturação. Nosso conhecimento reside na memória semântica, a qual EKLUND (1996, p.4) define como uma rede de conceitos interrelacionados.

Para o cientista da informação, que lida com a análise de assunto para estruturação de uma área do conhecimento, o mapa conceitual é um instrumento importante para ajudá-lo a entender e a lidar com um complexo de informações. AMORETTI (2001, p. 51) o relaciona ao esquemas de classificação:

> *". . . os conceitos ficam armazenados na memória de longo prazo, refletindo as experiências vividas pelo sujeito ao longo da sua vida. . . possuem as características básicas dos esquemas, tais como o de necessitarem de um estímulo interno ou externo para serem ativados e conservarem-se na memória a longo termo".*

Observa-se, então, que o mapa conceitual é uma ferramenta apropriada para representar um domínio do conhecimento, auxiliando a externalização das estruturas cognitivas dos autores de hiperdocumentos.

# 3. A CONSTRUÇÃO E IMPLEMENTAÇÃO DO MODELO MHTX

## 3.1 Introdução

Neste capítulo, são apresentadas as etapas de desenvolvimento do modelo hipertextual, sua implementação em um protótipo digital e os processos subjacentes a esse trabalho. O protótipo digital, chamado Modelo Hipertextual para Organização de Documentos (**MHTX**), consiste no mapa semântico chamado de Mapa Conceitual (**MC)**, e no Sumário Expandido (**SE**), instrumento formado pelo sumário da tese, ao qual se agregaram pontos de acesso. O modelo hipertextual foi instalado em uma base de dados, que abriga a tese selecionada para ser objeto desta pesquisa, em formato digital e que, posteriormente abrigará também outras teses e dissertações em textos completos digitalizados, pertencentes ao Programa de Pós-Graduação em Ciência da Informação da UFMG.

Antes de falar de modelagem de dados, é preciso explorar o conceito de modelo, que COUGO (1997, p.7) define como *"a representação abstrata e simplificada de um sistema real, com a qual se pode explicar ou testar o seu comportamento, no todo ou em partes"*. Assim, sob o ponto de vista dessa definição, as representações dos conteúdos da produção intelectual de um programa de Pós-Graduação, que é uma meta desse estudo a longo prazo, deverão atingir uma simplificação em três aspectos: organização, acesso e recuperação, que são processos comumente complexos e insuficientes quando se trata de versões impressas de teses e dissertações. Freqüentemente, após defendidos e armazenados nas bibliotecas tradicionais, esses documentos apresentam sérios problemas especialmente quanto à localização de seu conteúdo, cujas partes tendem a se perder, a se dispersar ou a não relacionar convenientemente entre si.

A situação ideal de modelagem de dados de teses e dissertações se traduziria na simultaneidade entre a utilização de um modelo de representação em domínios específicos do conhecimento, na forma hipertextual, e a própria criação do texto, pois levaria o autor do hipertexto a um nível mais alto de abstração, coadunando a autoria do documento e a autoria de sua forma de apresentação, na primeira etapa da escrita hipertextual. Porém, na prática, observa-se que quase todos os autores intelectuais dos documentos não dominam a linguagem hipertextual no momento da escrita ou, devido a prazos e formatos tradicionais, não se dispõem a fazê-lo.

Quando RYDER (2003, p.1) diz que o modelo conceitual nos ajuda a ter uma compreensão do nosso mundo, nos remete tanto aos procedimentos da análise facetada, com sua característica de abrigar qualquer assunto a partir de diversos pontos de vista, quanto à característica do hipertexto de trabalhar com dinâmica de trilhas que se conectam por associações, tal qual o fluxo dos pensamentos na mente humana.

Na elaboração do **MHTX**, foram seguidas as recomendações de CAMPOS (2000). No primeiro nível (Nível de entendimento da forma de abordagem do assunto), foram definidos os seguintes requisitos:

Requisito 1 - Determinação do domínio do conhecimento: Organização do Conhecimento.

Requisito 2 - Método de raciocínio utilizado para a representação das unidades do conhecimento: Análise Facetada.

Requisito 3: Tipo de leitor: Comunidade acadêmica, pesquisadores e profissionais da Ciência da Informação e áreas correlatas.

Requisito 4: Tipologia documental do hiperdocumento: teses e dissertações.

No segundo nível (Nível da organização das unidades de conhecimento na construção da narrativa), foram definidos osseguintes requisitos:

Requisito 5: Natureza do conteúdo das unidades de conhecimento: Modularidade.

Requisito 6: Estabelecimento das relações entre as unidades de conhecimento: Criação dos *links*.

No terceiro nível (Nível de estabelecimento de um veículo de comunicação e expressão sobre a temática do hiperdocumento), foi definido o seguinte requisito:

Requisito 7: Elaboração de uma representação gráfica para os nós conceituais e seus relacionamentos: Implementação gráfica do **MHTX**.

No processo da criação desse modelo foram consideradas quatro frentes. (1) Procedimentos da **Teoria da Análise Facetada** (**TAF**), que utiliza uma metodologia

analítico-sintética de organização do conhecimento, permitindo ao usuário ver os assuntos de um documento sob diferentes pontos de vista, buscando o mapeamento do conhecimento de uma determinada área, seja ela científica ou não. A **TAF,** como técnica de classificação e indexação, provê suporte na organização do conhecimento *à priori,* tendo como ponto de partida a modelagem de uma estrutura semântica. (2) **Mapas conceituais**, enquanto ferramentas de visualização, facilitam a estruturação de documentos publicados em forma hipertextual, disponibilizando, de uma maneira amigável, o conhecimento a ser recuperado. O potencial de representação e recuperação da informação do mapa conceitual é visto como uma alternativa navegacional para o texto virtual. (3) ***Links*** **hipertextuais**, que permitem mostrar as relações entre os conceitos e, enquanto ferramentas de navegação, permitem também mostrar os tipos de relação semânticas (hierárquica, associativa) existentes entre os conceitos, dando maior flexibilidade ao modelo proposto. (4) Normas técnicas da **Associação Brasileira de Normas Técnicas - ABNT**, que estabelece requisitos para apresentação de sumário de documentos, permitindo uma visão de conjunto e facilidade de localização das seções e outras partes e a que específica os princípios gerais para elaboração de trabalhos acadêmicos.

Um dos primeiros procedimentos desse estudo foi a seleção de uma tese, como material de trabalho. A escolha de um microtexto[25] deve-se à limitação do escopo proposto para o modelo, que é trabalhar no universo de um documento específico, no domínio de um assunto específico. Pretende-se, em estudos posteriores, poder expandir sua aplicação a um conjunto de documento da mesma natureza.

A escolha da tese como amostragem para a conversão de um texto linear em texto não linear se deu em função dos seguintes critérios: (1) a característica das teses e dissertações serem escritas linearmente; (2) a dificuldade do autor temático em trabalhar diretamente com o autor tecnológico ao mesmo tempo (3) a impossibilidade das teses serem escritas concomitantemente na forma linear e não linear e, finalmente; (4) a possibilidade da aplicação deste modelo na BTDECI – UFMG (Biblioteca de Teses e Dissertações do Programa de Pós-Graduação em Ciência da Informação da UFMG.

[25] Microtexto é o texto com ligações explícitas entre seus componentes e está situado no âmbito de um documento.

Como aplicação mais ampla do modelo ora apresentado, cada tese a ser inserida na base de dados terá uma estrutura conceitual que retrate seu conteúdo e uma estrutura formal que ressalte as partes do mesmo (sumário). Apesar de estar utilizando uma tese que possui características próprias quanto à sua tipologia e sua temática, a modelagem conceitual proposta pode ser testada na construção de outros hiperdocumentos, a partir de suas próprias características.

## 3.2 Mapa Hipertextual - MHTX

A combinação de técnicas das abordagens descritas acima serviu como embasamento para a criação de um instrumento interativo, com um potencial organizacional de conteúdo semântico em documentos completos em bases de dados hipertextuais e com a possibilidade de uma recuperação em contexto eficaz. A seguir, descreve-se a construção das estruturas conceitual e formal, tendo em vista a elaboração do **MHTX**.

### 3.2.1 A estrutura conceitual do MHTX

Os procedimentos metodológicos para a implementação conceitual incluíram, seqüencialmente, as seguintes etapas:

1 - Identificação do documento básico de trabalho (a tese);
2 - Leitura da tese;
3 - Análise facetada do assunto da tese: a seleção de seus termos relevantes e categorias (facetas); o reconhecimento das subfacetas; a ordenação das facetas, subfacetas e focos a serem apresentados no mapa conceitual e finalmente, a organização de todos os termos e suas relações;
4 - Criação do Mapa Conceitual (MC) com seus *links* e suas relações;
5 - Estruturação do Sumário Expandido (SE)
6 – Criação dos *links* do sumário para o texto.

#### 3.2.1.1 A análise facetada

Para a facetação da tese escolhida para esse estudo, foi utilizada a técnica da análise facetada, com base em dois processos: (a) <u>análise</u>, que se dá a partir da

identificação dos conceitos relevantes; b) classificação dos conceitos em categorias, em que cada categoria representa uma característica. Ressalte-se que o processo referente à síntese, que ocorre quando cada conceito pertencente a essas categorias é combinado com outro para expressar um assunto composto, não foi implementado na presente tese.

Para a formação das categorias, foram utilizados os princípios normativos do plano das idéias constantes do trabalho *Modelo simplificado para análise de facetas: Ranghanathan 101*, elaborado por SPITERI (1998), no qual a autora discute e sintetiza princípios estabelecidos anteriormente em duas frentes de pesquisa: aqueles inicialmente apresentados por S. R. Ranganathan e aqueles apresentados posteriormente pelo *Classification Research Group-CRG* em Londres (1952).

No plano das idéias, ocorrem diversos procedimentos: (1) processo de definição do assunto; (2) seleção das características que constituem o assunto; (3) seleção de um modelo para o mapeamento da informação sobre os conceitos; (4) agrupamento e divisão destes conceitos conforme suas características comuns e diferentes; (5) organização e o arranjo de grupos e subgrupos. Os princípios que regem esses procedimentos foram utilizados para embasar o processo facetado, objetivando garantir a determinação do conteúdo das bases fronteiriças de assuntos de interesse dos usuários, a seleção apropriada de padrões para a categorização do assunto e a organização destas categorias em forma seqüencial apropriada.

A **TAF,** Teoria da Análise Facetada, é um modelo com natureza metodológica dedutiva, no qual considera-se primeiro o domínio/contexto para depois se selecionarem os termos representativos de uma área. Assim, possui mecanismos de representação para se trabalhar com níveis conceituais na formação de categorias, a partir das quais, os conceitos são ordenados para formarem classes de conceitos.

Na análise da facetação do assunto da tese, procedeu-se ao mapeamento minucioso do documento para identificação correta de sua temática e à análise conceitual, para determinação dos seus assuntos básicos e isolados. Para a indexação, considerou-se o texto no todo, tendo como potencial todos os termos que compunham o documento na linguagem do autor, desde que considerados

relevantes. Por se tratar de uma modelagem conceitual, foram desconsiderados todas as entidades coletivas e nomes de autores citados na tese.

No processo de análise da terminologia da tese levou-se em consideração a noção de categoria para a análise e classificação de assuntos do documento e para a organização desses assuntos (isolados) em uma estrutura conceitual. Para a formação de assunto (análise) utilizou-se *Prolegomena to Library Classification* (RANGANATHAN, 1967, Parte P). Para a formação de categorias (síntese) utilizou-se o Modelo Simplificado de SPITERI (1998), descrito nesta tese no item 2.3.2.2

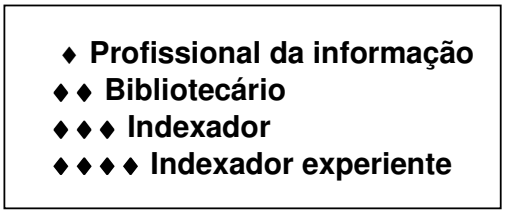
Para a formação de assuntos, RANGANATHAN (1967, p 351) propõe cinco métodos: **dissecação, laminação, desnudação, reunião/agregação e superposição**, os quais foram também discutidos mais recentemente por CAMPOS (2000, p.68). Na construção do modelo conceitual desta tese, apenas dois desses métodos foram utilizados na formação de assuntos: **dissecação** e **desnudação.** Os outros três métodos não foram utilizados por tratarem da formação de assuntos compostos (**laminação**), formação de assuntos complexos (**reunião/agregação**) e da superposição de duas idéias isoladas (**superposição)**, não relacionados ao objeto. A não utilização desses últimos três tipos de formação de assuntos se justifica, um a vez que foi utilizada a linguagem natural do documento. Em outras palavras, os termos levantados permaneceram na linguagem original do autor, não sendo utilizado nenhum vocabulário controlado para a padronização da terminologia.

O método de **dissecação** foi utilizado para a verificação de um único domínio que englobasse a tese e suas partes. A partir deste procedimento, chegou-se à conclusão de que o universo do assunto básico da tese é “organização do conhecimento”.

No método da **desnudação** (ou **desfolhamento**), ocorre a diminuição progressiva da extensão e o aumento da intensão (profundidade) do assunto básico ou de idéia isolada. O resultado sucessivo da desnudação permite a formação de cadeias, que passam a representar um núcleo específico de um assunto básico ou de uma idéia isolada. Por exemplo, a área mais abrangente tomou como assunto básico o termo “Profissional da informação” e o assunto básico mais específico considerou o termo

"Indexador". A desnudação ainda mais detalhada nos levou ao assunto básico "Indexador experiente" (FIG.22).

FIGURA 22 – Exemplo de desnudação




Na aplicação dos métodos de **dissecação** e **desnudação** sugeridos por Ranganthan para a formação de assuntos, foram seguidos alguns passos e, mesmo, estabelecidas outras regras para ajudar a padronização do mapeamento dos conceitos na modelagem facetada:

a) Leitura minuciosa de todas as partes do documento;
b) Identificação dos conceitos que melhor representassem o conteúdo semântico do documento;
c) Seleção dos termos mais relevantes para a área organização do conhecimento, constantes no documento;
d) Seleção dos termos com um conteúdo semântico significativo, que apresentassem uma quantidade maior de informação textual;
e) Seleção dos termos que apesar de não possuir uma grande quantidade de informação textual tivessem relevância na construção da modelagem conceitual para representação do assunto;
f) Representação dos conceitos selecionados, tal como aparecem na linguagem natural do texto, sem a necessidade de uma tradução que utilizasse vocabulário controlado.

A FIG.23 relaciona, em ordem alfabética, os 89 conceitos/termos considerados relevantes na análise conceitual do objeto de pesquisa.

FIGURA 23 - Lista alfabética de conceitos/termos relevantes da tese

| | |
|---|---|
| Abstração | Intencionalidade |
| Aceitabilidade | Interdisciplinaridade |
| Análise | Interpretação |
| Análise de assunto | Intertextualidade |
| Assimilação | Leitura |
| *Atinência* | Leitura |
| *Atinência Extrínseca* | Leitura do texto |
| *Atinência Intrínseca* | Lingüística |
| Atitude | Lógica |
| Autores | *Long-Term memory – LTM* |
| Bibliotecário | Macroestrutura |
| Botânica | Memória |
| Ciência da informação | Mente |
| Ciências | Metacognitiva |
| Ciências Biológicas | Microestrutura |
| Ciências Cognitivas | Percepção |
| Ciências Exatas | Prática |
| Ciências Formais | Processamento *Bottom-up* |
| Ciências Humanas | Processamento Interativo |
| Ciências Reais | Processamento *Top-down* |
| Ciências reais Humanas | Processos mentais |
| Ciências reais Naturais | Profissional da informação |
| Ciências Sociais | Psicologia Cognitiva |
| Coerência | Raciocínio |
| Coesão | Raciocínio Dedutivo |
| Cognitiva | Raciocínio Indutivo |
| Compreensão | Relevância |
| Conceito/Idéia/Pensamento | Semântica |
| Conhecimento prévio | *Short-Term Memory – STM* |
| Consistência | Significado |
| Documento | Sintaxe |
| Emoção | Síntese |
| Especialização | Situacionalidade |
| Extração de conceitos | Sociologia |
| Generalização | Subjetividade |
| Hipertexto | Superestrutura |
| Indexação | Terminologia |
| Indexador | Texto |
| Indexador experiente | Texto Informativo |
| Indexador novato | Texto narrativos |
| Indexador pouco experiente | Texto Primário |
| Inferência | Texto Secundário |
| Informatividade | Tradução para a linguagem de indexação |
| | Treinamento |

Em um segundo momento, no processo de **formação das categorias**, procedeu-se à avaliação e à definição de cada conceito, com análise e definição dos termos isolados acima, conforme suas características comuns. Segundo RAGHAVAN (1985,

p.27) a análise dos conceitos e a relações entre eles encontradas em um determinado assunto, servem como base para uma estrutura que expressa o significado e a relação entre esses elementos. Ainda segundo esse autor, a análise de assunto de centenas de documentos tem indicado que os conceitos encontrados no discurso podem ser categorizados em número ilimitado de categorias.

Baseado em relações categoriais, os conceitos que constituem os assuntos de um documento podem ser estruturados para expressar as relações entre eles com o propósito de representação e organização da informação. As relações semânticas consideradas foram determinadas a partir das relações hierárquicas e relações associativas.

Para uma escolha consistente das categorias ou facetas do objeto em questão, passou-se à análise e à abordagem dada, pelos princípios do plano das idéias para categorização do modelo simplificado. No seu modelo simplificado, SPITERI (1998) agrupou todos os cânones, postulados e princípios do plano das idéias em dois princípios, com subdivisões: (1) Princípio para escolha da faceta e (2) Princípio para ordem de citação das facetas e focos. Ressalta-se que conceito de faceta foi usado neste trabalho, característica mais abrangente e subfaceta, subdivisões de uma faceta.

Dentre os Princípios de escolha das facetas e subfacetas do plano das idéias do Modelo Simplificado para Análise Facetada foram considerados apenas os seguintes: **Diferenciação**, **Relevância**, **Verificação**, **Permanência**, **Homogeneidade**, **Exclusividade Mútua** e **Categorias Fundamentais**, princípios que são oriundos da **TAF** de Ranganathan, dos fundamentos do *CRG* ou de ambos.

O **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 “Bibliotecários”, pode ser dividido pela natureza de seu trabalho e depois, pela característica de sua experiência e pelo seu grau de conhecimento, como mostra a FIG.24. Deve-se observar aqui, que a autora da tese modelada considerou o quesito experiência na profissão como diretamente proporcional ao tempo de trabalho no processo de indexação.

FIGURA 24 - Aplicação do Princípio de Diferenciação no MHTX

**Bibliotecário**

(Pela natureza do seu trabalho)
**Indexador**
(Pela sua experiência)
**Indexador experiente**
**Indexador pouco experiente**
**Indexador novato**
(Pelo seu grau de conhecimento)
**Especialização**
**Prática**

O **Princípio de Relevância,** proposto tanto por Ranganathan quanto pelo *CRG*, ajuda 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", como mostra a FIG.25:

FIGURA 25 - Aplicação do Princípio de Relevância no MHTX

| | |
|---|---|
| **Processos mentais**<br>(Pelo tipo de)<br>**Raciocínio**<br>(Pela abordagem lógica)<br>**Indutivo**<br>**Dedutivo**<br>(Pela abordagem cognitiva)<br>**Bottom-up**<br>**Top-down**<br>**Interativo** | **Indexação**<br>(Por seus estágios)<br>**Análise de assunto**<br>(Pela suas fases ou aspectos)<br>**Leitura do texto**<br>**Extração de conceitos**<br>**Tradução para a linguagem de Indexação** |

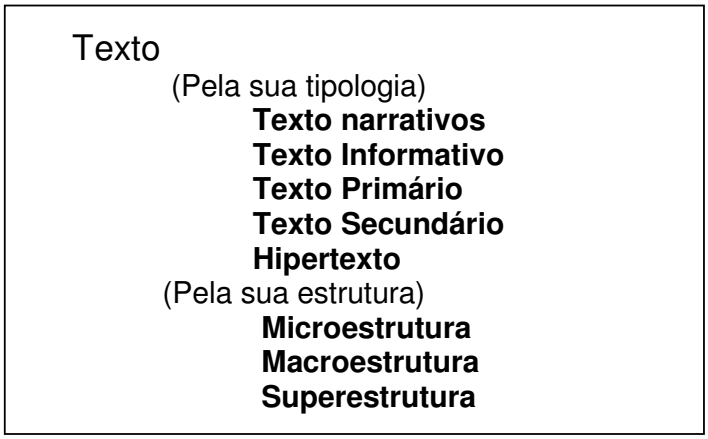
O **Princípio de Verificação,** proposto por Ranganathan e pelo *CRG*, afirma que é importante escolher facetas que sejam definitivas e que possam ser verificadas. No caso desta pesquisa, em que a terminologia utilizada é a do próprio autor da tese, o princípio de verificação com alto grau de acerto na escolha das características foi assegurado.

O **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 este princípio, pois reflete a interpretação de permanência: o indexador será sempre o bibliotecário, o que é uma característica permanente.

O **Princípio de Homogeneidade**, proposto apenas pelo *CRG*, afirma que a faceta deve ser homogênea. 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, como mostra a FIG.26:

FIGURA 26 - Aplicação do Princípio de Homogeneidade no MHTX




O **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.

O **Princípio de Categorias Fundamentais**, proposto apenas pelo *CRG*, afirma que não existem categorias que são fundamentais para o assunto como um todo e que todas as categorias devem ser derivadas com base no assunto a ser classificado. Assim, esse princípio pode parecer contrário ao conjunto básico de categorias fundamentais de Ranganathan representado pela sigla PMEST (Personalidade, Matéria, Energia, Espaço e Tempo). O *CRG* 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. Além disso, o *CRG* acredita que nenhuma lista deva ser necessariamente exaustiva

ou aplicada a todos os assuntos. Louise Spitieri escolheu a abordagem do *CRG* para ser inserida no *Modelo simplificado* considerando que essa abordagem permite aos classificadores moldar melhor as categorias para assuntos específicos, permitindo a formação de categorias mais distintas e mais bem definidas. A autora salienta que a maioria dos sistemas de classificação facetada/tesauros por ela consultados usa a abordagem do *CRG* na escolha das categorias fundamentais.

Os Princípios para ordem de citação das facetas e focos do plano das idéia**s** determinam como os focos serão organizados dentro se suas respectivas facetas, ou seja, como será a ordenação destes focos no renque. Dentre esses princípios, foram utilizados a **Sucessão Relevante** e a **Sucessão Consistente.**

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 Seqüência 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 possuírem essa natureza organizacional, segundo arranjos espaciais ou geométricos.
- **Ordem do simples para o complexo**: ordem para elementos de complexidade crescente.
- **Ordem do complexo para o simples**: ordem para elementos de complexidade decrescente.
- **Ordem canônica**: ordem tradicional mais corrente, quando não há nenhum outro princípio a seguir.
- **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 seqüência de assunto em um renque de assuntos ou, então, nenhuma seqüência de isolados em um renque de isolados for mais útil.

A **Sucessão Consistente,** baseada no cânone de consistência de Ranganathan (*Consistent Sequence*), 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.

Para a determinação de quais categorias seriam utilizadas na tese, separaram-se os assuntos conforme suas características. Geralmente, microdocumentos não abarcam toda a terminologia do assunto, apresentando uma manifestação limitada de facetas. No caso dessa tese, cujo tema é "Análise de assunto", a qual está situada dentro do assunto "Organização do conhecimento", a síntese de sua terminologia gerou conceitos agrupados segundo quatro categorias apenas: Personalidade, Atributo, Energia e Disciplina.

A primeira categoria determinada foi **Personalidade**, que pode ser definida também como Entidade e que é considerada pelos estudiosos, e mesmo por seu criador Ranganathan, como uma das mais difíceis de ser definida. A análise das entidades dentro da organização do conhecimento serviu como referência para esta escolha, o que resultou na seqüência "Bibliotecário", "Autores", "Conceito", "Mente', "Documento" e "Texto".

A segunda categoria, **Atributo,** que pode ser definida também como Matéria ou Propriedade, foi determinada com base na análise dos conceitos que caracterizava ou qualificava os conceitos da categoria Personalidade, descrita acima. Geralmente, os atributos são representados por conceitos que se referem a propriedades abstratas, tais como intencionalidade, emoção, abstração; conceitos que representam propriedades intrínsecas: "coesão", "subjetividade" etc.

A terceira categoria, **Energia,** é manifestada através de ações e operações. Como exemplo desta categoria na tese, temos “processos mentais gerais" (análise, síntese, abstração etc), “processos de indexação”, “análise de assunto”, “treinamento” e “interpretação”. RAGHAVAN (1985, p.29) observa que se deve ter cuidado com a afirmação de que somente situações reveladas por verbos intransitivos possam ser manifestações da categoria Energia.

Embora a tese estudada seja focada na área “organização do conhecimento”, a autora Madalena Naves incluiu, no Capítulo 5, um histórico sobre a área de “ciências” e suas diversas disciplinas e sub-disciplinas, o que justificou a determinação da quarta categoria, **Disciplina**.

Prosseguiu-se, então, com a distribuição dos termos de acordo com o princípio das Categorias fundamentais sugeridos pelo *CRG*. SPITERI (1998, p.19) afirma que *“para categorias fundamentais todas as categorias devem ser derivadas baseadas na natureza do assunto a ser classificado”.* Porém, a **TAF** propõe o esquema de categorias que, nesse caso, é considerado válido para todos os assuntos inseridos em qualquer campo do conhecimento.

Após o agrupamento dos conceitos dentro de cada faceta, a estruturação facetada da tese “*Fatores interferentes no processo de análise de assunto”* gerou a categorização mostrada na FIG.27.

FIGURA 27 – Categorização completa da tese com seus níveis de 1 a 8 no MC

**Classe Geral: ORGANIZAÇÃO DO CONHECIMENTO** (1)

**Personalidade [Entities]** (2)

- **Autores** (3)
- **Profissional da informação** (3)
  - **Bibliotecário** (4)
    (Pela natureza do seu trabalho)
    - **Indexador** (5)
      (Pela experiência)
      - **Indexador experiente** (6)
      - **Indexador pouco experiente** (6)
      - **Indexador novato** (6)

      (Pelo grau de conhecimento)

- **Especialização** (6)
- **Prática** (6)

- **Conceito/Idéia/Pensamento** (3)
- **Documento** (3)
- **Texto** (3)

(Pela natureza do texto)

- **Narrativos** (4)
- **Informativo** (4)
- **Primário** (4)
- **Secundário** (4)
- **Hipertexto** (4)

(Pela estrutura)

- **Microestrutura** (4)
- **Macroestrutura** (4)
- **Superestrutura** (4)

## Atributos [Property/Matter] (2)

- **Coesão** (3)
- **Coerência** (3)
- **Consistência** (3)
- **Intertextualidade** (3)
- **Intencionalidade** (3)
- **Aceitabilidade** (3)
- **Situacionalidade** (3)
- **Informatividade** (3)
- **Relevância** (3)
- **Assimilação** (3)
- **Atitude** (3)
- **Cognitiva** (3)
  - **Metacognitiva** (4)
- **Compreensão** (3)
- **Percepção** (3)
- **Conhecimento prévio** (3)
- **Leitura** (3)
- **Subjetividade** (3)
- **Memória** (3)

(Por tipo)

- ***Long-Term memory – LTM*** (4)
- ***Short-Term Memory – TM*** (4)

- **Atinência** (3)

(Por característica)

- **Intrínseca** (4)
- **Extrínseca** (4)

- **Interdisciplinaridade** (4)
- **Sintaxe** (3)
- **Semântica** (3)
  - **Significado** (4)
- **Terminologia** (3)

**Energia [Process/Action]** (2)

- **Treinamento** (3)
- **Indexação** (3)
  (Por seus estágios)
  - **Análise de assunto** (4)
    - **Leitura do texto** (5)
    - **Extração de conceitos** (5)
  - **Tradução para a linguagem de indexação** (4)
- **Leitura** (3)
- **Processos mentais** (3)
  - **Análise** (4)
  - **Síntese** (4)
  - **Abstração** (4)
  - **Generalização** (4)
  - **Raciocínio** (4)
    - **Indutivo** (5)
    - **Dedutivo** (5)
    - **Bottom-up** (5)
    - **Top-down** (5)
    - **Interativo** (5)
- **Inferência** (3)
- **Interpretação** (3)

**Disciplinas [Disciplines]** (2)

- **Ciências (3)**
  - **Ciências Formais** (4)
  - **Ciências Reais** (4)
    - **Ciências Reais Naturais** (5)
      - **Ciências Exatas** (6)
      - **Ciências Biológicas** (6)
        - **Botânica** (7)
    - **Ciências Reais Humanas** (5)
      - **Ciências Sociais** (6)
        - **Sociologia** (7)
        - **Ciência da informação** (7)
      - **Ciências Humanas** (6)
        - **Lógica** (7)
        - **Lingüística** (7)
        - **Ciências Cognitivas** (7)
          - **Psicologia Cognitiva** (8)

#### 3.2.1.2 O mapa conceitual (MC)

O objetivo do Mapa Conceitual **(MC)** do **MHTX** é possibilitar ao usuário uma visão geral da estrutura semântica do texto escolhido através de sua representação

gráfica, facilitando a navegação semântica em contexto, através de seções e subseções, digitalizadas, constantes da base de dados hipertextual. A estrutura conceitual é composta dos conceitos relevantes da tese em questão e organizada de acordo com a estrutura facetada, com as relações semânticas, hierárquicas e associativas.

A estrutura hipertextual do **MC** foi implementada utilizando-se o *Software Star Tree Studio*, que possui vários aplicativos para interpretar e disponibilizar informações fragmentadas em formato gráfico hiperbólico. O Mapa conceitual compõe-se de uma estrutura semântica tridimensional, onde estão representados todos os conceitos relevantes da tese e suas relações semânticas. Sua navegação é caracterizada pelas inter-relações existentes entre os conceitos.

Assim como na **TAF**, a elaboração de um mapa conceitual implica em apreender a agrupar os conceitos segundo seus traços perceptivos e segundo categorias já estabelecida anteriormente, envolvendo as etapas de seleção, ordenação e agrupamento. Na elaboração do mapa conceitual, as relações de semelhanças e diferenças são explicitadas. As denominações “renque” e “cadeia” são utilizadas na análise facetada para evidenciar as relações hierárquicas gênero-espécie e todo-parte. A navegação através do mapa conceitual permite ao usuário final navegar através dos conceitos semanticamente relacionados, permitindo uma recuperação mais precisa. A dimensão vertical representa a hierarquia entre os conceitos, a seqüência em cadeias, enquanto que na dimensão horizontal, localizam-se os conceitos que apresentem um grau de generalidade idêntico, ou seja, a formação de renques no **MC** (FIG.28). A FIG.29 mostra como exemplos extraídos da tese se apresentam na tela do computador no mapa conceitual.

FIGURA 28 - Dimensões horizontal (renques) e vertical (cadeias) do **MC** da tese

**Renque**

♦Indexação

♦♦Análise de assunto ----------

♦♦Tradução par um vocabulário de assunto

} **relações horizontais/associativas**

**Cadeia**

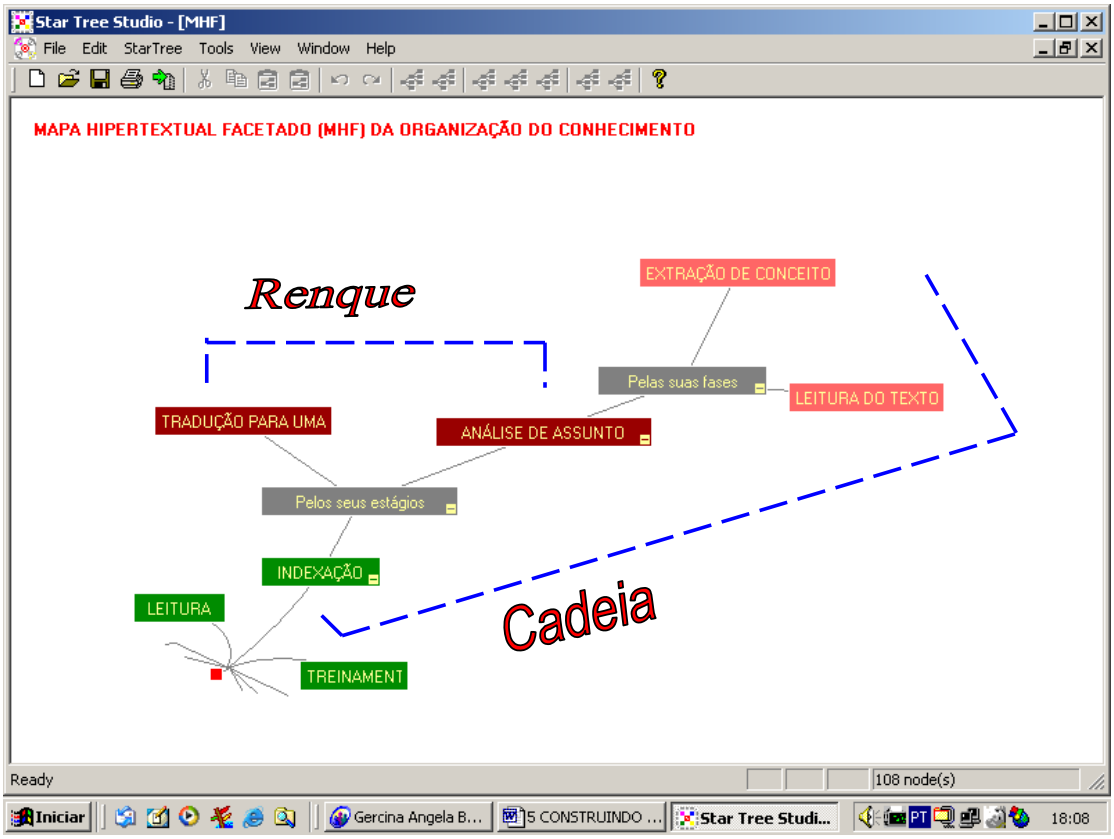
♦Indexação ----------

♦♦Análise de assunto

♦♦♦Extração dos conceitos

} **relações verticais/hierárquicas**

FIGURA 29 - Subdivisões de renque e cadeia na estrutura hipertextual do **MC**.



#### 3.2.1.2.1 A criação dos *links*

Para a estruturação dos *links* conceituais, foi considerada a lista de todos os conceitos inseridos no mapa conceitual, com seus respectivos endereços, marcados no interior do texto da tese. Como cada conceito poderia ter somente um caminho eletrônico, cada conceito teve seu *link* direcionado para a seção ou subseção da tese em que seu conteúdo tivesse maior relevância. O Anexo 5 lista todos os *links* do **MC** e do **SE** com suas respectivas URLs.

Quanto à função, foi utilizado o tipo de *link* relacional. Como o próprio nome sugere, os *links* relacionais, que são a essência do hipertexto, apresentam as informações mais relevantes de forma não linear, auxiliando o usuário a escolher, encurtar ou aumentar seu caminho na procura da informação desejada.

Quanto ao direcionamento do usuário, os *links* utilizados foram classificados como ativos ou inativos. Enquanto que os primeiros remetem a informações retiradas da própria tese, os segundos servem apenas para encaminhar o usuário para outros *links de conteúdo*.

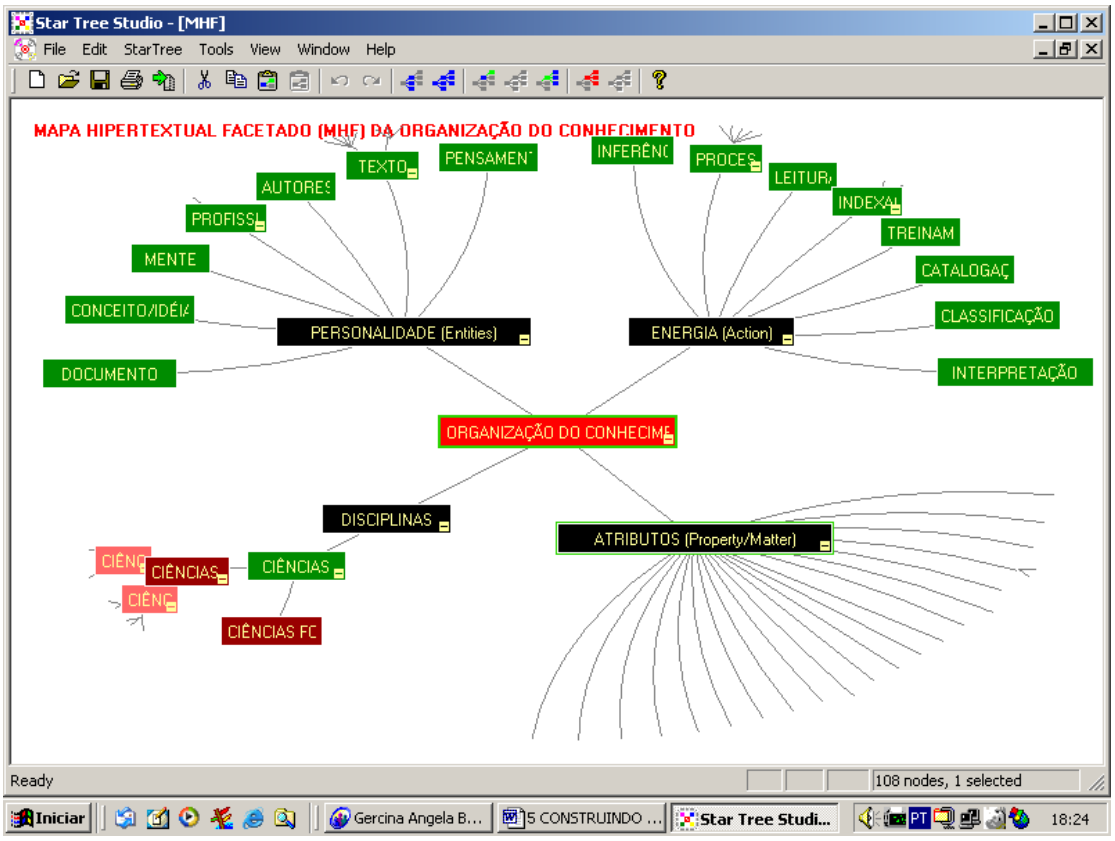
Na estruturação do **MC**, visando facilitar a localização no mapa conceitual pelo usuário, utilizou–se cores diferentes para cada nível. Esta correspondência é apresentada na Fig.30.

FIGURA 30 - Relação de cores para os níveis do **MC**

| **Nível** | **Cor** | ***Link* Ativo/ Inativo** |
|---|---|---|
| 1º nível: Faceta básica | **Vermelho** | Ativo |
| 2º nível: Facetas | **Preto** | Inativo |
| 3º nível: Subfacetas | **Verde** | Ativo |
| 4º nível: Subfacetas | **Vinho** | Ativo |
| 5º nível: Subfacetas | **Rosa** | Ativo |
| 6º nível: Subfacetas | **Azul** | Ativo |
| 7º nível: Subfacetas | **Ocre** | Ativo |
| 8º nível: Subfacetas | **Roxo** | Ativo |
| Características de divisão (podem ocorrem entre quaisquer níveis) | **Cinza** | Inativo |

A FIG.31 mostra o esquema de cores na tela principal do **MC**, a partir da raiz principal ou assunto básico (nível 1, em vermelho), que é circundado pela facetas (nível 2, em preto) e subfacetas (nível 3, em verde; nível 4, em vinho e nível 5, em rosa).

FIGURA 31 – Associação entre níveis e cores na raiz principal do **MC**



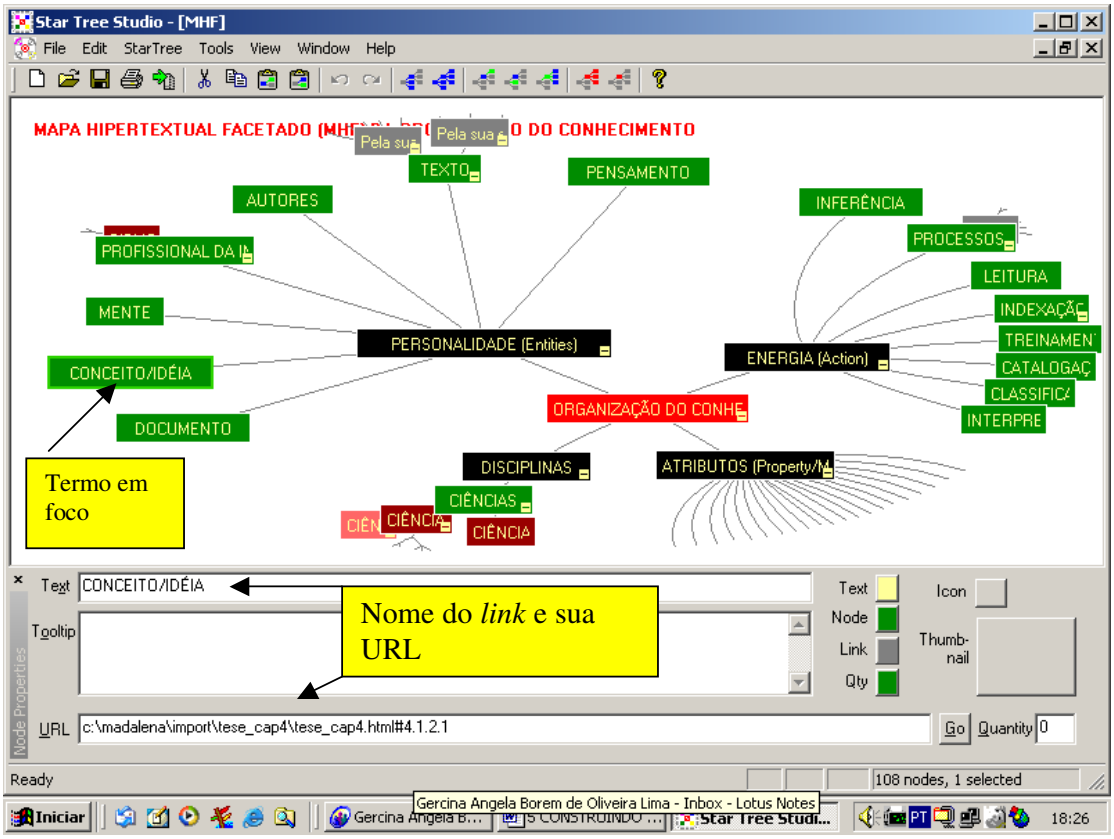

A colocação dos *links* no mapa conceitual partiu do assunto mais abrangente da tese, o assunto básico "Organização do conhecimento" (*link* do 1° nível, vermelho), passando a seguir para as quatro categorias (*links* do 2° nível, pretos), determinadas pela facetação da tese: **Personalidade**, **Atributos** (características) , **Energia** e **Disciplina**. Assim, o *link* do 1° nível inicia a navegação no **MC** e remete o leitor à capa da Tese. No 2° nível, os *links* foram considerados inativos, por incluírem apenas o nome das quatro primeiras facetas e não remeter a nenhuma informação inserida na tese. Nos níveis seguintes, todos os *links* são ativos. Entretanto, as Características de divisão, que podem ocorrer entre quaisquer dos oito níveis, são *links* inativos. No total, o mapa recebeu 108 *links*, sendo 91 ativos e 17 inativos (FIG.32).

FIGURA 32 – Tabela de organização dos níveis, cores, *links* e seus direcionamentos no MHTX

| **Nível** | **Cor** | **Número total de *Links*** (com ou sem direcionamento) | **Direcionamentos para outros *links*** |
|---|---|---|---|
| 1º nível (faceta básica) | vermelho | 1 (ativo) | **4 pretos** |
| 2º nível (facetas) | preto | 4 (inativos) | **37 verdes:**<br>1 (disciplina) + 7 (personalidade) +<br>8 (energia) + 21 (atributos) |
| 3º nível (subfacetas) | verde | 37 (ativos) | **26 na cor vinho:**<br>1 (profiss. da Informação) + 8 (texto) +<br>2 (ciências) +1 (cognitiva) +<br>2 (memória) + 5 (atinência) +<br>2 (indexação) + 5 (processos mentais) |
| 4º nível (subfacetas) | vinho | 26 (ativos) | **11 rosas:**<br>1 (bibliotecário) + 2 (ciências reais) +<br>1 (semântica) + 2 (análise de assunto) + 5 (raciocínio) |
| 5º nível (subfacetas) | rosa | 11 (ativos) | **9 azuis:**<br>5 (indexador) + 2 (ciências reais naturais) + 2 (ciências reais humanas) |
| 6º nível (subfacetas) | azul | 9 (ativos) | **6 na cor ocre:**<br>1 (ciências biológicas) + 2 (ciências sociais) + 3 (ciências humanas) |
| 7º nível (subfacetas) | ocre | 6 (ativos) | **1 roxo** (psicologia cognitiva) |
| 8º nível (subfacetas) | roxo | 1 (ativo) | - - - - - - - - |
| Características de divisão<br><br>(podem ocorrer entre quaisquer níveis) | cinza | 13 (inativos) | **entre o 3º e o 4º níveis:**<br>2 (texto) + 2 (atinência) + 1 (memória) +<br>1 (processos mentais) + 1 (indexação)<br><br>**entre o 4º e o 5º níveis:**<br>1 (bibliotecário) + 1 (análise de assunto) + 2 (raciocínio)<br><br>**entre o 5º e o 6º níveis:**<br>2 (indexador) |
| Total de *Links* | | 108: 91 ativos e 17 inativos | |

Cada *link* conceitual ativo recebeu um endereço único, uma URL que remete ao seu conteúdo no texto da tese. Por exemplo, a URL para o conceito "**Conceito/idéia**" é c:\madalena\import\tese_cap4\tese_cap4.html#4.1.2.1, como mostra a FIG.33.

FIGURA 33 - Criação de URL para o *link* “Conceito/idéia” com comandos de nome e endereço

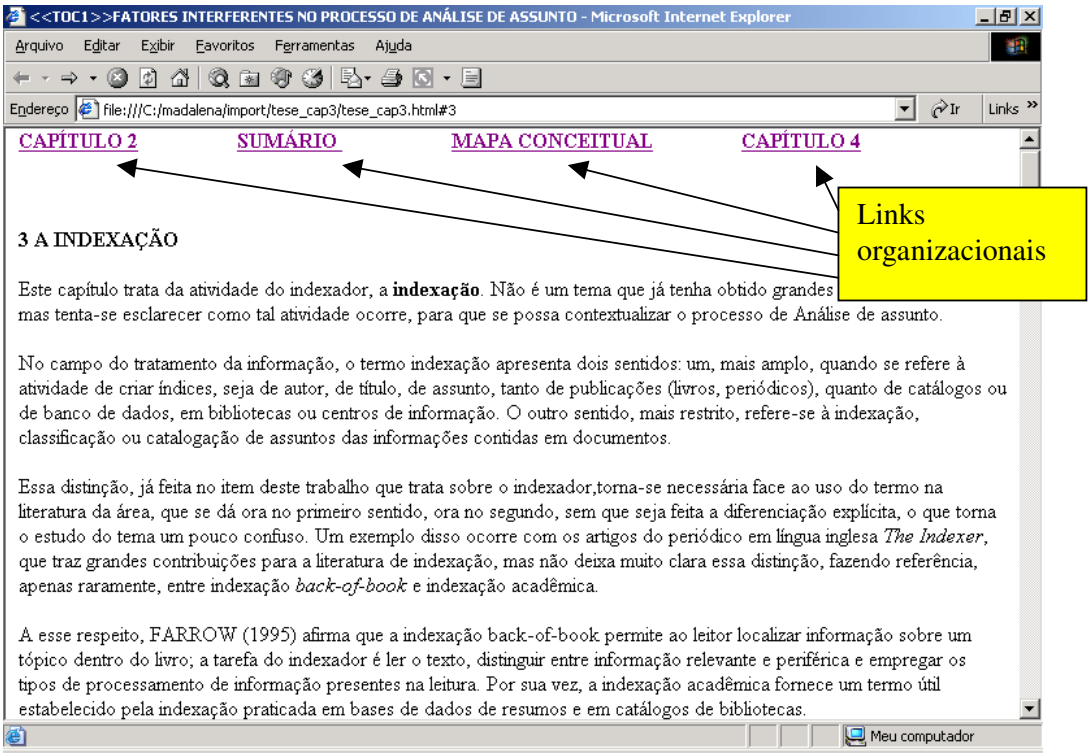



### 3.2.1.2.2 A navegação no MC

O navegador gráfico é um esquema de representação da estrutura hipertextual, que tem como objetivo fornecer ao usuário meios amigáveis de localizar informações específicas. A estrutura navegacional do **MC** demonstra ao usuário como seu conteúdo total está organizado semanticamente e como se conecta internamente, facilitando a movimentação de um nó para outro. Mostra que o documento pode ser representado graficamente em função dos nós de informação e dos *links* entre eles.

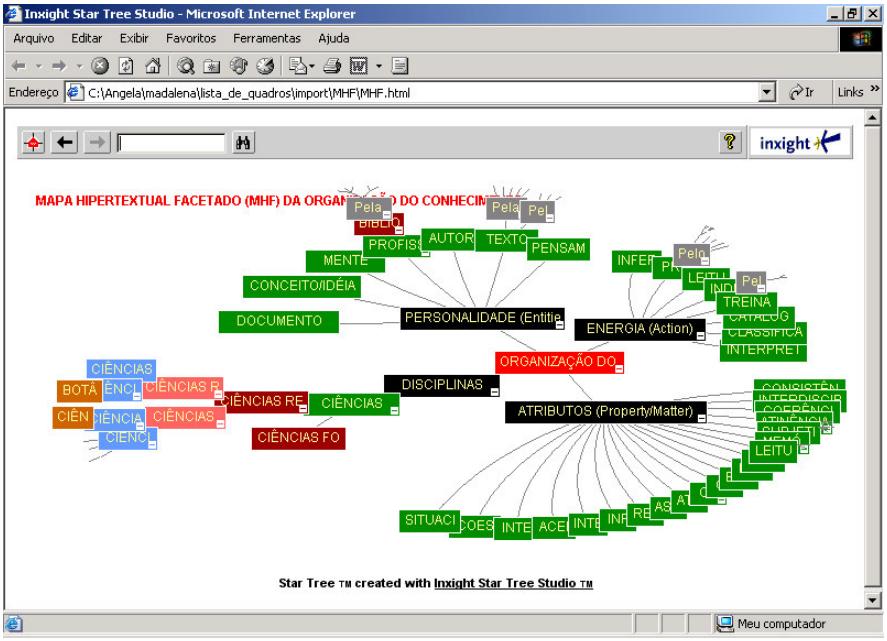
A navegação do **MC** é caracterizada pelo mecanismo de folheio e pelo mecanismo de pesquisa de conceitos. O acesso ao texto da tese realça a palavra-chave pesquisada em negrito (veja o termo ”indexação” na Fig.34). A partir daí, o usuário poderá navegar dentro do texto, com a opção de ir para capítulos seguintes e anteriores ou, ainda, acessar novamente o mapa.

FIGURA 34 – Navegação dentro do texto da tese através da palavra-chave “indexação” no **MC**




A visualização hiperbólica (*Fisheye*) na navegação do **MC** exibe os conceitos dos *links* em foco sempre próximos a outros, proporcionando maior orientação e seletividade entre conceitos. De fato, a primeira tela do navegador hiperbólico se assemelha muito à vista aérea de uma árvore, com um tronco único e central, ramificando para galhos cada vez menores, como mostra a FIG.35. Durante a operação deste mapa, os elementos situados nos nós de informação diminuem e aumentam de tamanho exponencialmente, ocasionando uma distorção à maneira do olho de peixe (*fisheye*), o que o permite abrigar estruturas enormes.

FIGURA 35 - Primeira tela do navegador hiperbólico do MC




### 3.2.2 Estruturação formal do MHTX

Além da estrutura conceitual, já descrita, o **MHTX** é também formado por uma estrutura formal que compreende a parte clássica do documento - o sumário, que inclui sua estrutura organizacional. Teoricamente, esta parte interrelaciona informações pré-textuais, elementos textuais e pós-textuais.

Sumário é uma listagem das principais divisões, seções e outras partes de um documento que reflete a organização do seu conteúdo. Segundo FRANÇA et al. (1998, p.74), o sumário deve conter apenas as partes da publicação que se sucedem (listas, prefácio, resumo, texto, referências bibliográficas, anexos e índice), exatamente como aparecem no texto. Deve indicar a numeração dos capítulos e suas divisões, o título de cada parte e a respectiva paginação. Adota-se aí a numeração progressiva, sendo que são numeradas somente as partes do texto que constituem elementos textuais, a partir da introdução e as partes pós-textuais, tais como referências bibliográficas, anexos, glossário, apêndice e índice. Aconselha-se usar letras maiúsculas para indicar títulos de partes ou capítulos (seções primárias) e apenas a inicial maiúscula para os títulos das divisões dos capítulos (seções secundárias, terciárias e assim por diante).

Como elemento do modelo proposto, o Sumário da Tese foi expandido tendo sido incluídos pontos de acesso de recuperação não listados no sumário original, considerados relevantes. A concepção do SE teve como objetivo dar ao leitor do hiperdocumento um maior detalhamento de informações contidas no documento, incluindo informações que não estavam expressas nas seções do sumário, criando-se *links* de acesso dessas informações e enriquecendo o mapa hipertextual proposto.

#### 3.2.2.1 O Sumário Expandido (SE)

A estrutura do sumário é apresentada conforme a Norma da ABNT sobre Informação e documentação, a NBR 6024/2003 (antiga NBR 6027/1989) – Sumário/Apresentação. A tese analisada possui 10 capítulos divididos em seções secundárias e terciárias, apresentando ainda Lista de Quadros, Resumo/*Abstract* e quatro anexos: Anexo 1: Entrevistas; Anexo 2: Protocolos Verbais; Anexo 3: Glossário; e Anexo 4: Cópias dos textos de sociologia e botânica. Para a construção do seu SE foram definidas algumas regras:

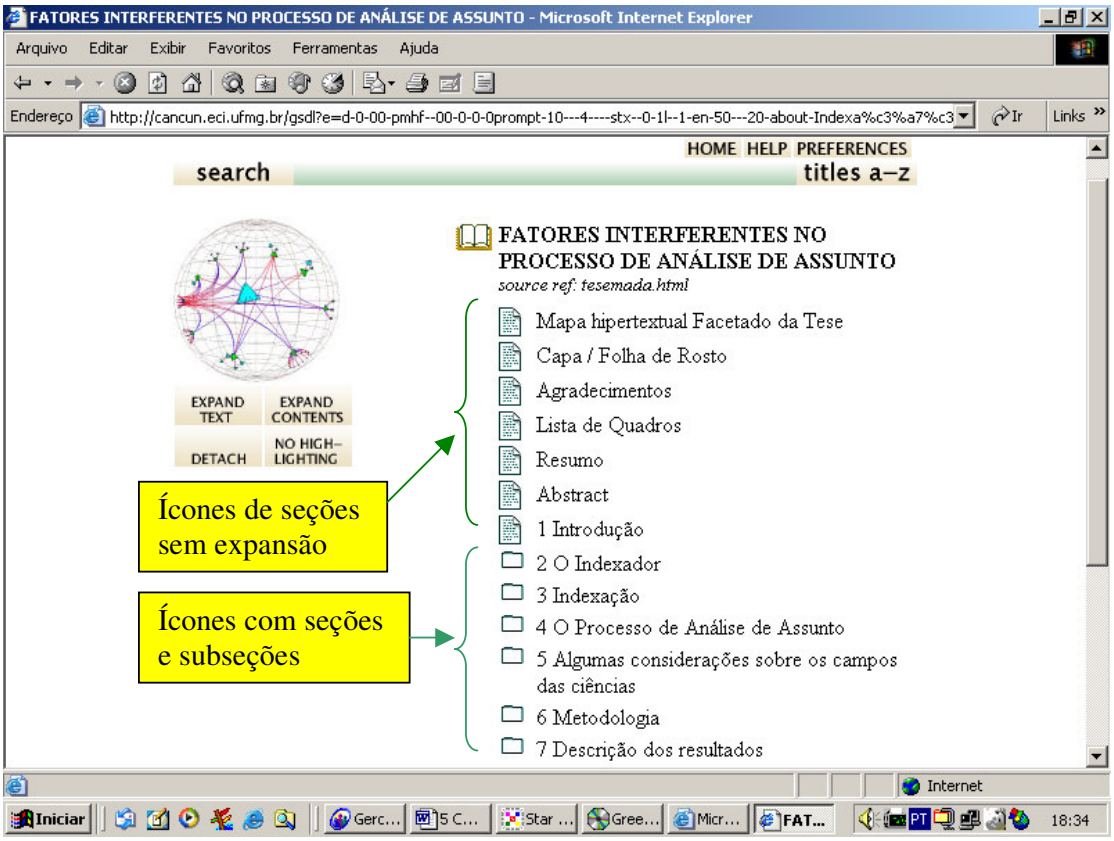
1. Manutenção da mesma forma dos títulos utilizados pelo autor nos tópicos do sumário;
2. Inserção, no SE de seções que não estão incluídas nas subdivisões do sumário (FIG.36 abaixo) e de quadros e figuras no contexto de sua utilização e localização no capítulo a que se referem.

Para criar a estrutura navegacional seqüencial do SE (apresentado integralmente no Anexo 3), da tese (FIG.36), foi também utilizado o *Greesntone digital library software gsdl 2.50*, descrito mais adiante.

A estrutura formal do **MHTX** correspondente ao SE foi criada partindo-se do sumário da tese, com todas as suas subdivisões, acrescidas de outros elementos informacionais julgados relevantes para o entendimento de contextos não inseridos no mapa conceitual, compensando assim, sua ausência na estrutura do sumário. Nesse momento, foram utilizados *links* do tipo organizacional, que permitem descrever a estrutura surperficial dos documentos, compreendendo todos os

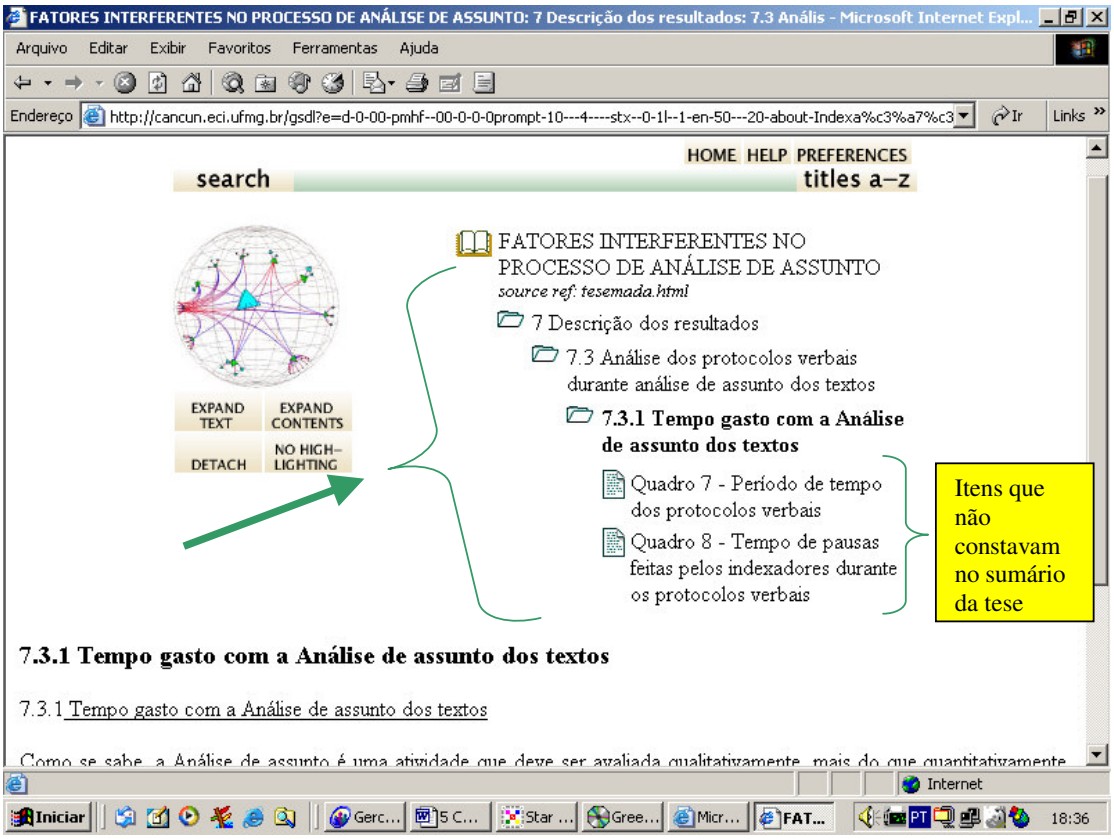
elementos da macro-estrutura sintática de apresentação do sumário. Para cada divisão principal e suas subdivisões, foram criadas *urls* dos *links*, sempre dentro de cada subdivisão primária. A cada *link* acrescentado que não estava presente nas seções primárias, secundárias ou terciárias do sumário, foi dado um número seqüencial, correspondendo àquela seção a que ele pertence, dentro da numeração correspondente, para facilitar a mnemonização da estruturação do sumário. Nesta estrutura hierárquica, a informação foi, então, introduzida do geral para o específico.

FIGURA 36 - Primeira tela do Sumário Expandido (SE) do MHTX




Enquanto que o MC tem 91 *links* conceituais (ativos), o SE tem 107 *links* organizacionais. Os *links* do SE, com suas URLs que levam às informações correspondentes, expandem-se dinamicamente em forma de árvore ao serem selecionados (FIG.37).

FIGURA 37 - Expansão do Sumário Expandido (SE)




## 3.3 A implementação tecnólogica: construindo o protótipo

O desenvolvimento do protótipo do **MHTX** foi o segundo passo no desenvolvimento desse estudo, considerando-se como primeiro passo a implantação conceitual. Protótipo é uma representação visual da implementação de um projeto que ajuda a analisar seu conjunto de necessidades funcionais (QUINN, 2002, p.1). Ele é importante porque permite ver a estrutura construída, analisar as necessidades dos usuários, antes de testar, generalizar e disseminar os resultados. Pelo seu caráter exploratório, protótipos geralmente não incluem todo o conteúdo do seu potencial de aplicação. Assim, o protótipo proposto foi construído, utilizando-se somente informações que eram necessárias para testar o **MHTX**, tendo como componentes o banco de dados referente à tese armazenada, o mapa conceitual e o **SE** da referida tese. A eficiência do protótipo foi testada ainda sem o delineamento tradicional de uma pesquisa experimental. Devido à limitação de tempo e escopo, testes experimentais sistemáticos com usuários

não envolvidos na condução da pesquisa só poderão ser realizados em etapas posteriores.

A implementação tecnológica, nessa fase compreendeu as seguintes etapas:

1- Identificação e análise de perfis de pacotes de *software* disponíveis para implementação do protótipo **MHTX**;

2- Escolha do *software* para organizar a base de dados Banco de Teses e Dissertações da Escola de Ciência da Informação da UFMG;

3- Estudo dos *softwares* e das linguagens de marcação *HTML* e *XML*;

4- Conversão do texto da tese objeto desta pesquisa para o formato das linguagens de marcação;

5- Escolha de *software* e criação do MC (*Star Tree Studio*);

6- Marcação dos *links* com suas respectivas URLs no mapa conceitual;

7- Marcação dos *links* no texto da tese em linguagem *HTML*;

8- Escolha de *software* e criação da "pedra fundamental" da base de dados da BTDECI (*Greesntone digital library software*);

9 - Escolha de *software* e criação do SE (*Greesntone digital library software*);

10- Marcação dos metadados no *XML*;

11- Organização dos aplicativos (o incipiente Banco de Teses, o SE e o MC) no protótipo **MHTX**;

12- Avaliações iniciais do protótipo.

### 3.3.1 Os *softwares* utilizados para criação do MHTX

#### 3.3.1.1 Escolha dos *softwares* para criação do banco de dados

Com o advento da Internet e dos chamados *softwares open source* (ou fonte aberta), várias alternativas de *software* para a criação e gestão de bancos de dados e bibliotecas digitais tornaram-se disponíveis, na sua grande maioria sem custos. O fato do *software* ser de fonte aberta, não significa que ele seja de uso gratuito, mas que o seu código fonte seja aberto. Os seguintes critérios foram adotados para a escolha do pacote de *software* que seria utilizado no desenvolvimento do protótipo:

1- Multi-plataforma tecnológica, em especial a plataforma *windows*;

2- Índices a partir de metadados e dos conteúdos dos documentos;
3- Flexibilidade de parametrização;
4- Facilidade de manipulação;
5- Processamento de documentos em múltiplos formatos, em especial o *HTML*;

Após amplo levantamento, foram identificadas três alternativas: *Phonesis*, *Nou-Rau* e *Greenstone*. O primeiro pacote de *software Phonesis*, que foi desenvolvido na *Universidad de Monterrey*, México, só pode ser executado em ambiente *Linux* ou *Unix Compatível* e trabalha apenas com metadados contemplados no formato *Dublin Core*. Afora essas restrições, ele atende a todos os outros requisitos, e parece ser o de mais fácil utilização.

O segundo pacote de *softwares* testados, o *Nou-Rau*, foi desenvolvido por um grupo da UNICAMP e segue, como os outros dois pacotes descritos acima, os princípios do *software open source*. Uma das restrições impostas por esse pacote é que sua plataforma tecnológica só pode ser executada em ambiente *Linux* ou *Unix*. Outra restrição diz respeito ao fato de que trabalha apenas com metadados contemplados pelo *Dublin Core*.

O terceiro pacote de *software*, o *Greenstone* foi desenvolvido na *University of Waikato*, na Nova Zelândia, no contexto do projeto de Bibliotecas Digitais daquele país, e foi o escolhido para implantação do **MHTX** por atender a todos os requisitos acima.

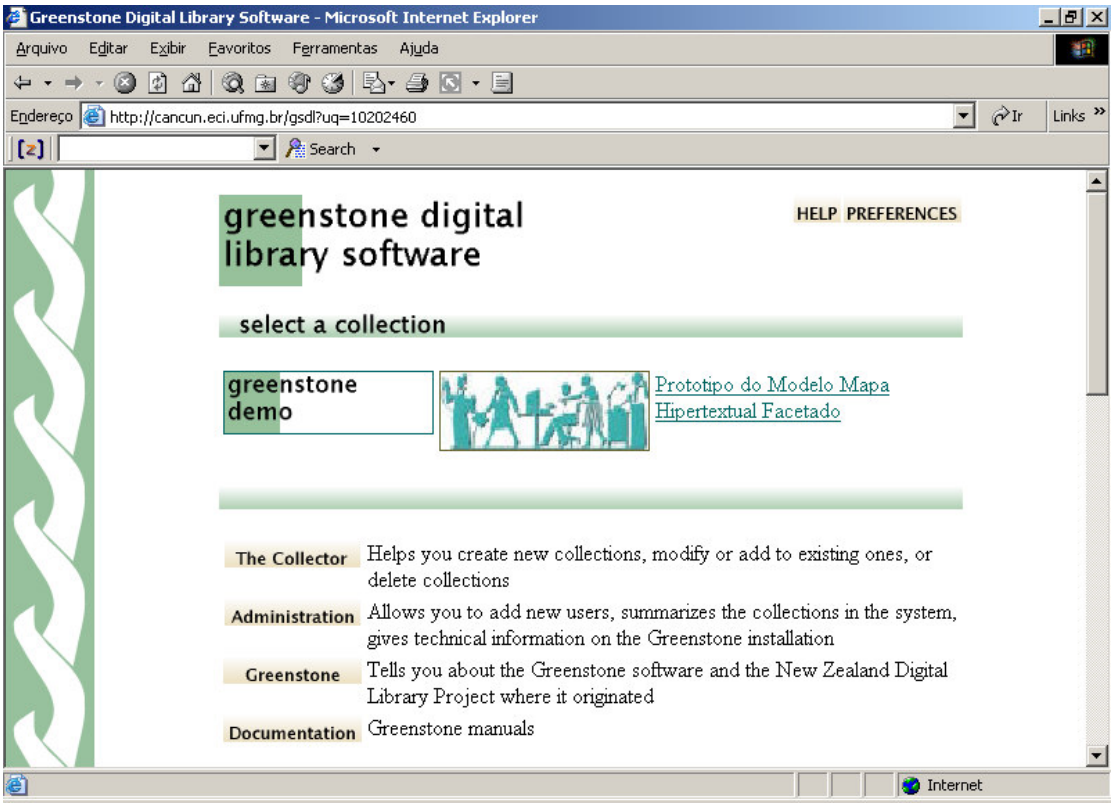
O *software Greesntone Digital library* (FIG.38) é um pacote de *software* que oferece muita flexibilidade de utilização, trata qualquer tipo de documento e, funciona em ambiente *Linux*, *Windows* e *Unix Compatível*. Além disso, está disponível em três versões, permitindo tanto o acesso remoto por um servidor quanto *offline*:

i) ***Greenstone* local**, pode ser executado a partir de um microcomputador *desktop* ou *laptop*, portanto, monousuário;

ii) ***Greenstone Web***, que deve ser instalado em um servidor conectado à Internet, pode disponibilizar as bases de dados na rede;

iii) ***Greenstone* código fonte**, o que significa que esse pacote segue a tendência de *software open source*, e por isso é distribuído com o seu código fonte. Além disso, esse *software* permite exportar bases de dados ou bibliotecas digitais criadas para serem disponibilizadas em um disco CD-ROM, no qual é incluído um executável do programa, de forma que o usuário poderá levar a sua cópia da biblioteca digital para qualquer outro computador, mesmo que ele não esteja conectado à Internet.

FIGURA 38 - Tela inicial do *Greesntone Digital Library*




O *Greenstone* possui várias ferramentas que demandam, muitas vezes, um conhecimento maior de programação. No entanto, ele oferece uma série de facilidades que independem de conhecimentos de programação, o que possibilitou a sua adoção para o desenvolvimento do protótipo que será mostrado a seguir. Por exemplo, o seu acesso combinando pesquisas em textos completos com folheio de índices baseados em diferentes tipos de metadados, é de fácil manuseio, o que torna essas pesquisas mais atraentes. Além disso, é um sistema de fácil manutenção, e que pode ser reconstruído de maneira totalmente automática. Ressalte-se que o *Greenstone* oferece facilidades de busca que agregam qualidade e eficácia ao modelo proposto.

Com vistas a estudar a viabilidade de implementação do protótipo que se propôs, a base de dados deve ser capaz de disponibilizar, no futuro, textos integrais em diferentes tipos de arquivo. A escolha do pacote de *software Greenstone* se mostrou acertada porque trabalha com diferentes formatos, como por exemplo, *.html*, *.pdf*, *.rtf*, *.ps*, *.txt* e *.doc*.

Além desses formatos, caso o usuário queira tratar um outro tipo de documento, ele pode desenvolver o *plugin*[26]. De uma forma geral, esse pacote de *software* permite a indexação dos metadados, assim como do próprio conteúdo dos documentos, por meio da extração de palavras do texto. Após o tratamento dos documentos, estes são convertidos para um formato compatível com o *XML*. Essa facilidade possibilitou a criação dos índices extraídos dos textos e dos metadados, assim como a inserção de uma interface de navegação hipertextual, organizada por meio de árvores hiperbólicas, contendo o mapa conceitual. Dessa maneira, o usuário tem a facilidade de realizar buscas como em um mecanismo de busca convencional, inclusive utilizando operadores booleanos, e explorar o documento por meio da navegação, utilizando os *links* inseridos e organizados em uma árvore hiperbólica. Essas características do *Greenstone* permitiram vislumbrar e comprovar a viabilidade de se implementar o modelo proposto.

Foram realizadas várias experimentações preliminares utilizando-se o *Greenstone*, até se chegar ao protótipo que está sendo apresentado nessa tese. Por uma questão de síntese, somente os passos principais são aqui apresentados. Essas experimentações foram possíveis graças à facilidade de parametrização do *Greenstone*, uma vez que ele não necessita ou não depende de programação. Manipulando um conjunto de parâmetros, o usuário descreve sua coleção, seu formato e seus metadados, indicando os índices a serem criados e a forma de apresentação da coleção em *html*. Esses parâmetros, informados por meio de um arquivo do tipo texto, chamado COLLECT.CFG, são ilustrados na Fig.39, à frente.

Outra característica importante que diferencia o *Greenstone* dos outros pacotes de *software* é a indexação por seções. O *Greenstone* indexa o documento como um

[26] *Plugin* é um conjunto de instruções escritas por meio de linguagem de programação. No caso do Greenstone, utiliza-se a linguagem PERL.

todo, assim como cada seção individualmente. A facilidade de indexação permite ao usuário recuperar partes do documento, que se considerou ideal para tratar documentos muito grandes como teses. Isso permitiu a apresentação de um documento organizado pelo seu SE, atendendo às necessidades do modelo proposto.

Na versão final do protótipo, o documento da tese objeto dessa pesquisa foi formatado em *HTML*, utilizando-se as seguintes *tags*:

- ✓ **<section> ... </section>**, para individualizar cada uma das partes lógicas de conteúdos da tese escolhida;
- ✓ **<Metadata name="...">...</Metadata>**, para descrição dos metadados de seções ou capítulos;
- ✓ **<description> ... </description>**, para delimitar os conteúdos de cada uma das seções.

Além do documento da tese formatado em *HTML*, foi criado também um arquivo contendo os metadados relativos à tese como um todo, denominado METADATA.*XML* (FIG.39)

FIGURA 39 – Arquivo contendo os metadados relativos à tese como um todo

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE DirectoryMetadata SYSTEM
"http://greenstone.org/dtd/DirectoryMetadata/1.0/DirectoryMetadata.dtd">
<DirectoryMetadata>
<FileSet>
   <FileName>tesemada</FileName>
   <Description>
     <Metadata name="Title">Fatores interferentes no processo de análise de
assunto</Metadata>
   </Description>
</FileSet>
</DirectoryMetadata>
```

### 3.3.1.2. Escolha do *software* para criação do MC e do SE

A representação visual pode auxiliar o gerenciamento de estruturas de informação grandes e complexas, em que o estilo da interface é um facilitador na manipulação de dados. Os *browsers* gráficos fornecem uma visão espacial da estrutura de nós e

*links*, mostrando ao usuário onde ele se encontra, que outras informações estão disponíveis e como elas podem ser acessadas.

Os seguintes critérios foram adotados para a escolha do pacote de *software* que seria utilizado no desenvolvimento do MC e do SE:

- Compatibilidade com diversos tipos de b*rowsers* (*Internet Explorer x*, *Netscape x*, *Opera Mozilla x*, em especial na plataforma *Windows)*;
- Geração de diferentes estilos de representação gráfica (dinâmicos, estáticos, expandidos, diagramas hierárquicos, hiperbólicos);
- Flexibilidade de parametrização;
- Facilidade de manipulação com controle interativo e programático;
- Processamento de documentos em múltiplos formatos, em especial o *HTML*, *DHTM*, *JavaScript*, *XML*;
- Interface amigável;
- Geração de uma estrutura navegacional eficiente.

A criação do **MC** demandou uma ferramenta que possibilitasse a visualização da estrutura dos nós e *links*, explicitando as relações entre eles. Para a estruturação do **SE**, buscou-se um *software* que possibilitasse a visualização no estilo em forma de árvore (*tree style*), reproduzindo a estrutura do sumário. Identificaram-se três alternativas: *SiteXpert*, *Star Tree Studio* e o *Greenstone*.

O primeiro programa, *SiteXpert*, foi desenvolvido pela *Xtreeme GmbH*, companhia fundada em 1997 em Delemont, Suíça. Possui um código compatível com muitos *browsers*, entre eles *Internet Explorer*, *Nestescape*, *Opera* e *Mozilla*. Possui diferentes estilos de navegação, tais como interface para portal, menu *DHTML*, mapa de *site*, índice do *site*, estilo em forma de árvore, estilo estático ou dinâmico, diagrama hierárquico, *Java applet tree* e suporte para *XML* (importação e exportação de dados). Outro ponto positivo é sua fácil utilização. Apesar de atender a todos os requisitos do protótipo, o *SiteXpert* tem a restrição de criar mapas somente em forma hierárquica, quer seja dinâmico ou estático.

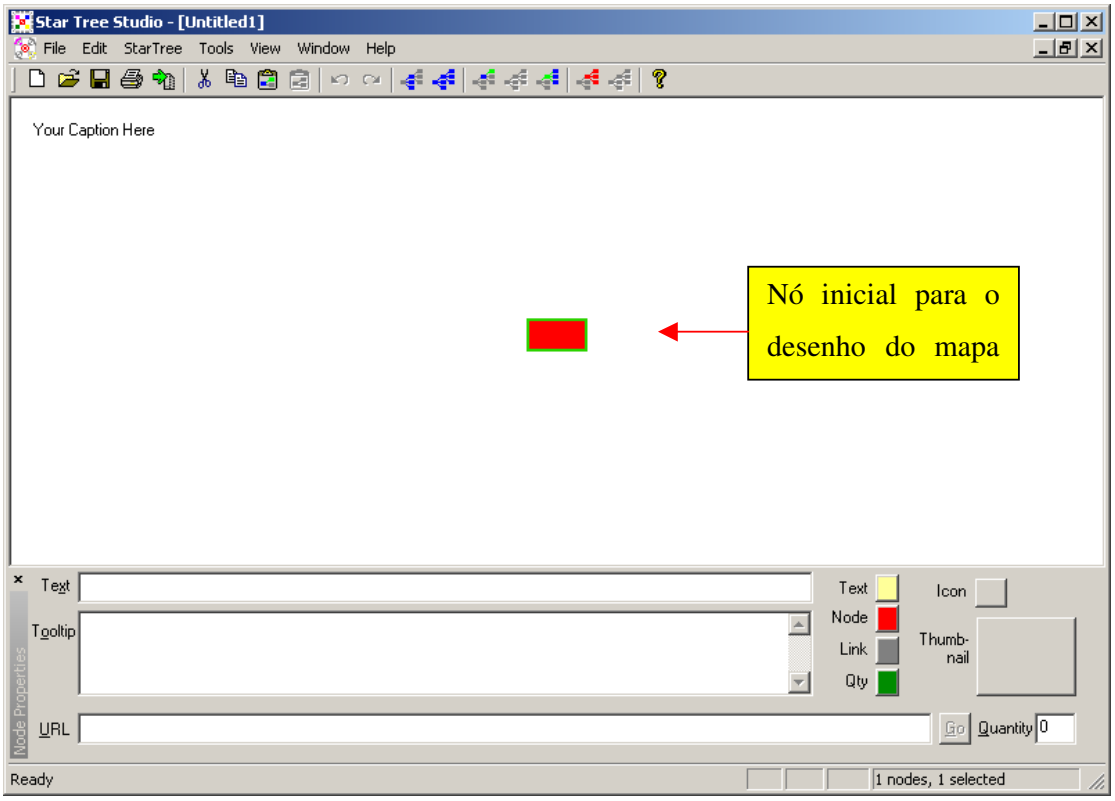
O segundo programa, *Star Tree Studio*, é produzido pela companhia *Inxight Software, Inc.* em *Sunnyvale,* Califórnia, USA. Esse *software* oferece um

mecanismo de pesquisa chamado *"spotlighting",* que melhora a navegação mesmo ao lidar com grandes quantidades de informações, auxiliando o rápido acesso à informação desejada. A desvantagem deste *software* é que ele só funciona em ambiente *Windows*. O terceiro *software*, *Greenstone*, foi descrito acima.

O ideal para implementação das estruturas do **MC** e do **SE** seria a existência de um único *software* que pudesse atender a todos os requisitos de ambos, possibilitando tanto a construção de uma estrutura gráfica hierárquica quanto a construção de uma estrutura que mostrasse as relações associativas entre os conceitos facetados da tese. Assim, o *Greenstone* parecia ser forte candidato para atender ao MC e ao SE. Foram realizadas várias simulações e chegou-se à conclusão de que nenhum dos três *softwares* atenderia simultaneamente a essas duas demandas. Assim, para a criação do **MC** foi escolhido o *Star Tree Studio* e para a criação do **SE** foi escolhido o *software Greenstone*. A desvantagem do *Star Tree Studio* não afeta a maioria dos usuários, devido à popularização do ambiente *Windows*.

Como o objetivo da construção do mapa conceitual é buscar uma solução para o problema da desorientação do usuário, o *software Star Tree Studio* (FIG.40) foi escolhido por ser específico para organização gráfica da informação e possuir vários aplicativos que interpretam e disponibilizam informações desestruturadas. Com uma interface intuitiva que guia o usuário no processo de criação de mapas tridimensionais, seu potencial cognitivo de espacialização melhora a navegação e o acesso à informação. Além disso, esse *software*, apesar de ser estruturado hierarquicamente, partindo sempre de um nó inicial, permite criar relações hierárquicas entre os nós através de características interativas. A inserção de termos hierárquicos pode ser feita a partir de comandos com alusões às relações genealógicas, como "*new parents"* ("novos pais"), *"add children"* ("adicionar filhos"), *"sibling"* ("irmão" ou "irmã"), *"grandchildren"* ("netos"), *"second generation"* ("segunda geração" abaixo de um nó "pai") etc. A inserção de termos com relações associativas pode ser feita através do comando. Essas características possibilitaram demonstrar as relações existentes nos renques e cadeias da análise facetada da tese através de uma estrutura com relações hierárquicas e associativas.

FIGURA 40 - Tela inicial do *Star Tree Studio* na criação do MC




Em relação à navegação hiperbólica, LAMPING et al (1995, p.1) explica que

> *"Inicialmente, o navegador hiperbólico mostra a hierarquia em forma de árvore com seu nó principal no centro, mas a interface pode ser suavemente transformada para trazer outros nós para o foco... Em qualquer posição, a quantidade de espaço disponível para os nós recuarem é uma função contínua de sua distância da hierarquia ao nó central em forma de árvore".*

Para a criação do SE, utlizou-se o *software Greenstone*, o mesmo utilizado na criação da base de dados. Sua característica de indexar por seções permitiu a criação da estrutura hierárquica do sumário da tese. A criação da estrutura do sumário dentro da própria base de dados na qual a tese está inserida, além de facilitar a navegação, resultou em uma interface virtual mais próxima da tese tradicional, com todos os seus elementos organizacionais.

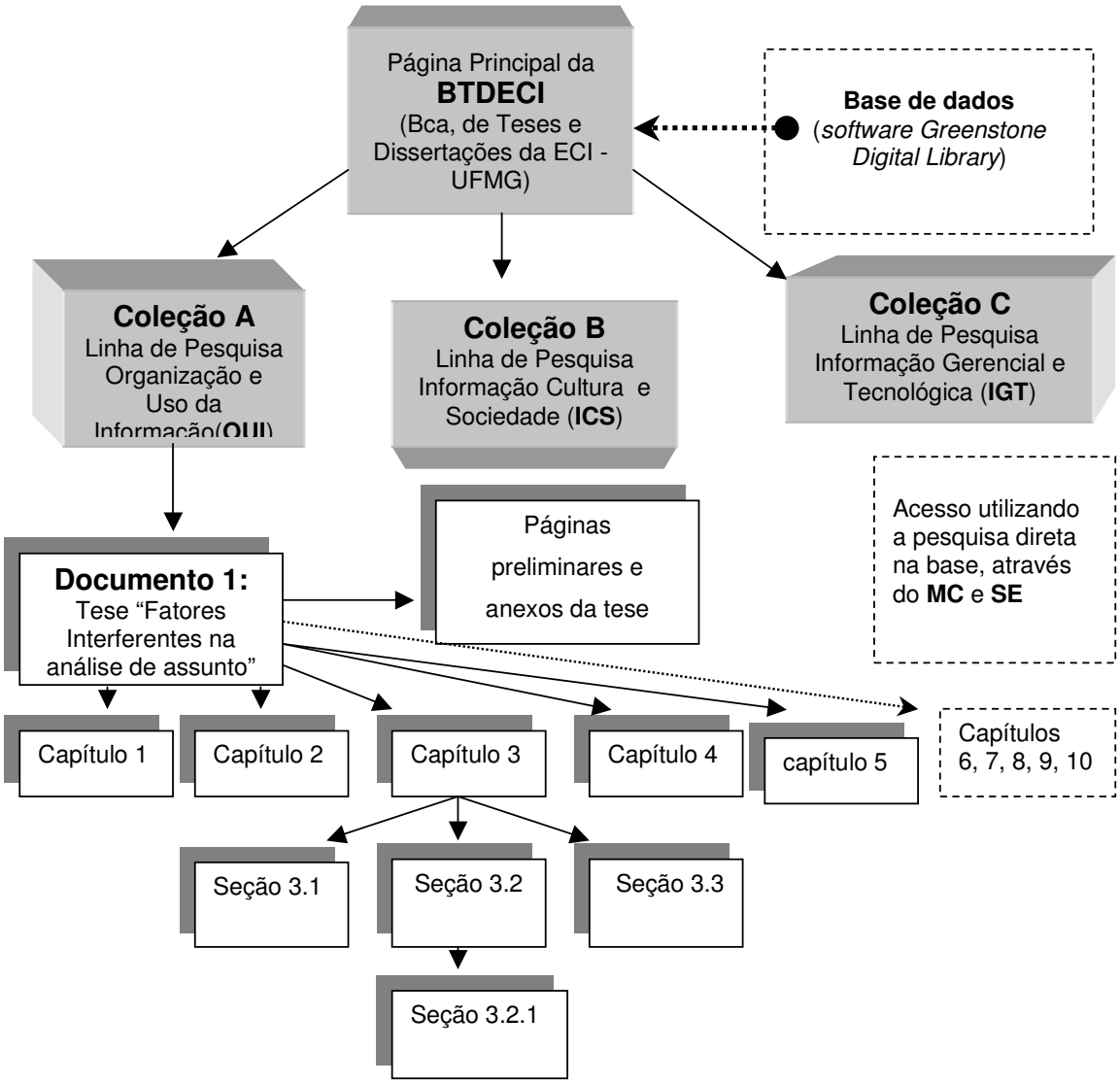
Esta implementação foi também testada com o *software SiteXpert*, que mostrou-se inicialmente adequado às demandas do protótipo, permitindo a experiência de

criação da estrutura do **SE** hierarquicamente em diferentes estilos. Porém, a utilização deste *software* não foi possível por causa da dificuldade de se criarem *links* externos ao *software Greenstone*, o que aumentaria o número de *links* externos para a navegação do usuário.

### 3.3.2 Criação da base de dados de coleções de teses e dissertações

Não se pretende, aqui, descrever a implantação ainda incipiente do projeto de criação da base de dados de textos completos de teses e dissertações da Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da UFMG. Não obstante, essa base foi criada preliminarmente com o intuito de abrigar a tese que é objeto do presente estudo e simular a situação que se pretende concretizada em futuro próximo. Com o pressuposto de que a base de dados, na qual vão inserir-se os documentos, já esteja construída, o foco dirige-se à organização dos documentos a serem inseridos e à criação de alternativas de pesquisa em documentos completos, com o apoio do MC e do SE propostos. A FIG.41 apresenta a estrutura futura da base de dados a ser criada para implementação completa do protótipo, composta de uma coleção de documentos virtuais resultantes de três linhas de pesquisa (teses e dissertações) e de uma coleção estruturada de informações auxiliares.

FIGURA 41 - Estrutura da base de dados de textos digitais completos de teses e dissertações da Escola de Ciência da Informação da UFMG




Na arquitetura proposta, as três coleções têm papéis diferentes. No primeiro nível, a coleção de documentos é implementada. No segundo nível, que é o nível das informações auxiliares, onde os conceitos se relacionam semanticamente, ocorre a interface entre o usuário e a coleção, através da pesquisa direta, utilizando o **MHTX** .

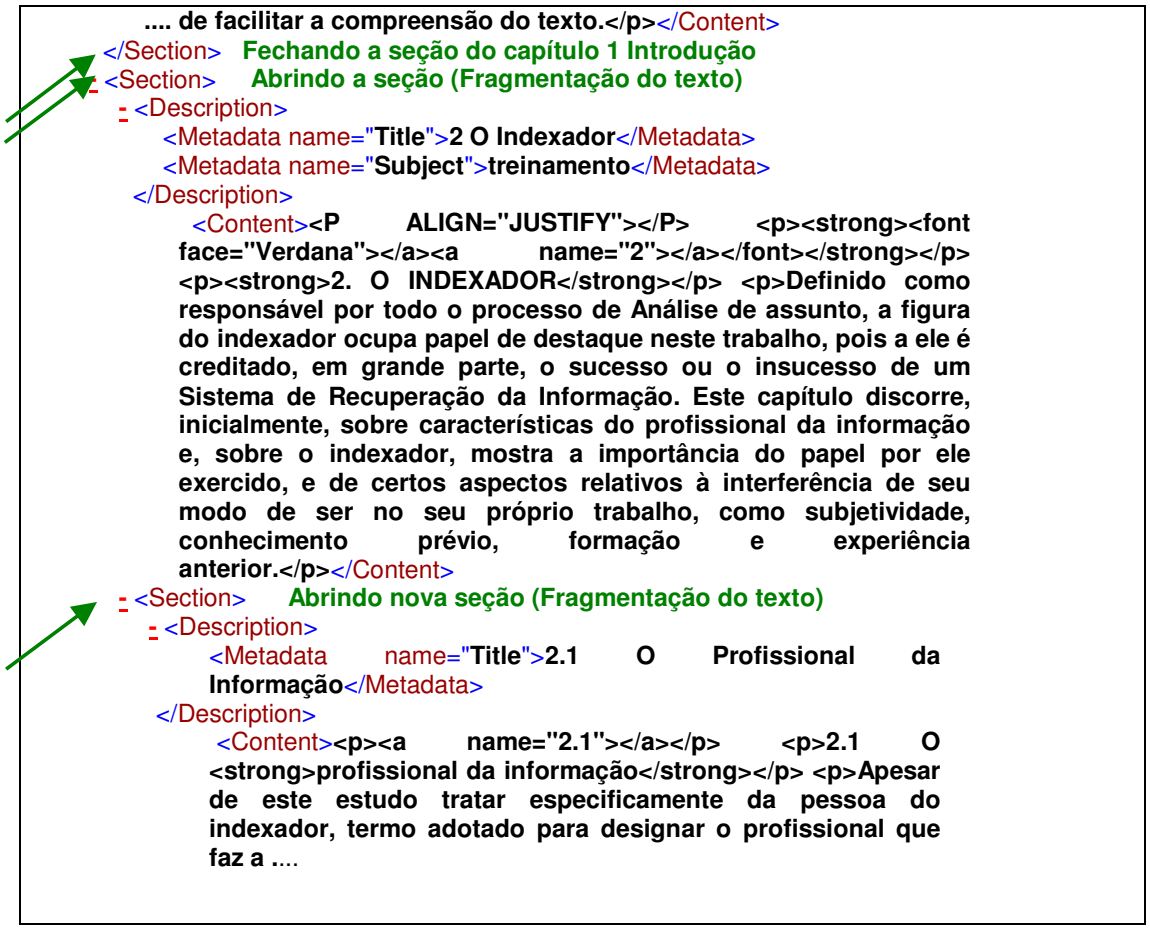
A tese foi indexada por seções utilizando a fragmentação flexível que a linguagem de marcação *XML* permite. Essa fragmentação foi definida e aplicada em pontos

apropriados para a modelagem da estrutura. O documento foi organizado hierarquicamente em seções lógicas, cujas unidades são os capítulos, seções e subseções. As FIG.42 e FIG.43 mostram como o espaço da futura coleção foi configurado no *software Greesntone*, onde se pode observar a criação de índices, a definição de seções que geram os índices para serem pesquisados e os *plugins* que permitem importar os documentos e transformá-los em padrão *XML* com os metadados incluídos. Os metadados foram arquivados com o documento, para utilização apenas interna à base de dados.

FIGURA 42 – Arquivo com os parâmetros de configuração da base de dados para o protótipo MHTX

```
public        true

indexes       section:text section:Title section:Subject document:text

plugin        GAPlug
plugin        HTMLPlug -description_tags -cover_image
plugin        ArcPlug
plugin        RecPlug -use_metadata_files

classify      AZList -metadata Title -sort Title -buttonname Title

# To build this collection using the List clasifier for the Howto
# listing (as shown in the Greenstone Developer's Guide) you should
# uncomment the following two lines and comment out the line above
# before rebuilding.
#classify      List -metadata Keyword -buttonname Howto
#format        CL4VList     "<br>[link][Keyword][/link]"

format SearchVList "<td
valign=top>[link][icon][/link]</td><td>{If}{[parent(All':
'):Title],[parent(All': '):Title]:}[link][Title][/link]</td>"
format VList "<td valign=top>[link][icon][/link]</td><td
valign=top>[highlight]{Or}{[Title],Untitled}[/highlight]<i><small>{If}
{[Date],<br>_textdate_[Date]}{If}{[NumPages],<br>_textnumpages_[NumPag
es]}{If}{[Source],<br>_textsource_[Source]}</small></i></td>"

format        DocumentText    "<h3>[Title]</h3>\\n\\n<p>[Text]"
format        DocumentImages  true
format        DocumentButtons "Expand Text|Expand
Contents|Detach|Highlight"

collectionmeta collectionname      "Prototipo do Modelo Mapa
Hipertextual Facetado"
collectionmeta collectionextra     "Trata-se de um protótipo do Modelo
Mapa Conceitual por Análise Facetada,\ndesenvolvido por Gercina Lima."
collectionmeta iconcollectionsmall
"_httpprefix_/collect/pmode2/images/tesis.jpg"
collectionmeta iconcollection
"_httpprefix_/collect/pmode2/images/tesis.jpg"
collectionmeta .section:Title      "titulos de secao"
collectionmeta .document:text      "todo o documento"
collectionmeta .section:Subject    "assuntos"
collectionmeta .section:text       "capitulos ou secoes"
```

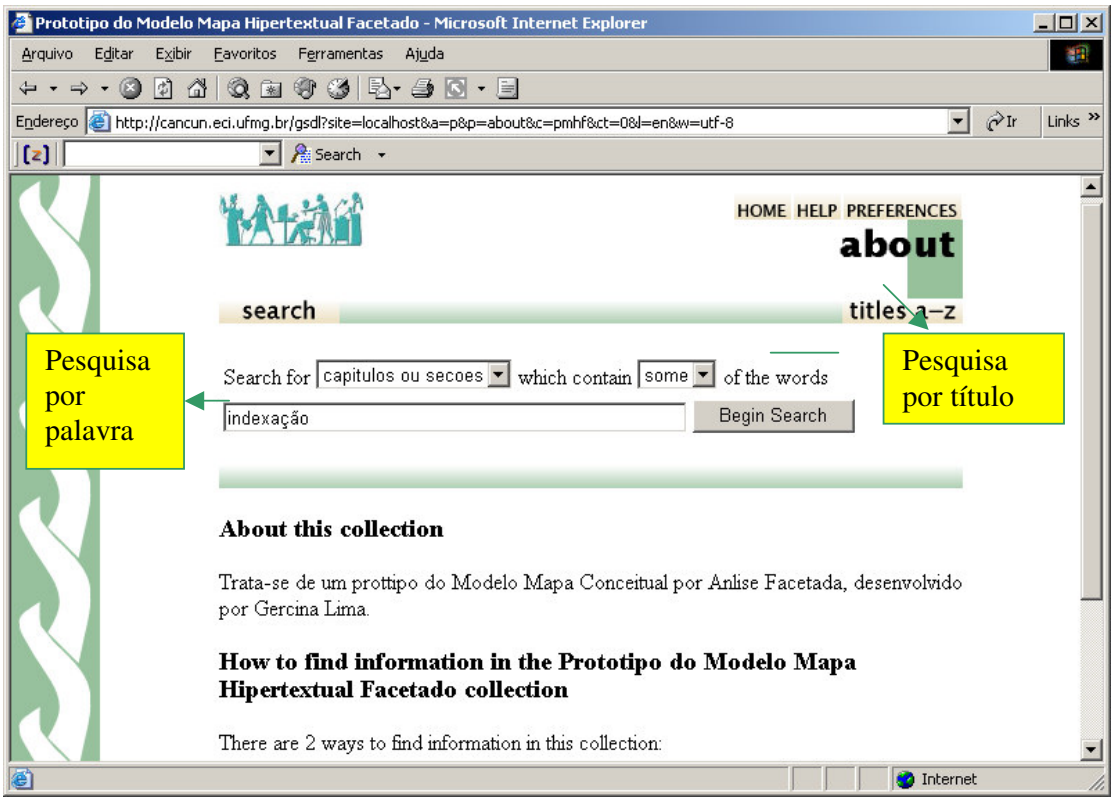
FIGURA 43 - Subdivisões de seções dentro do Capítulo 2 da tese em *XML*




### 3.3.3 Navegação na base de dados do MHTX

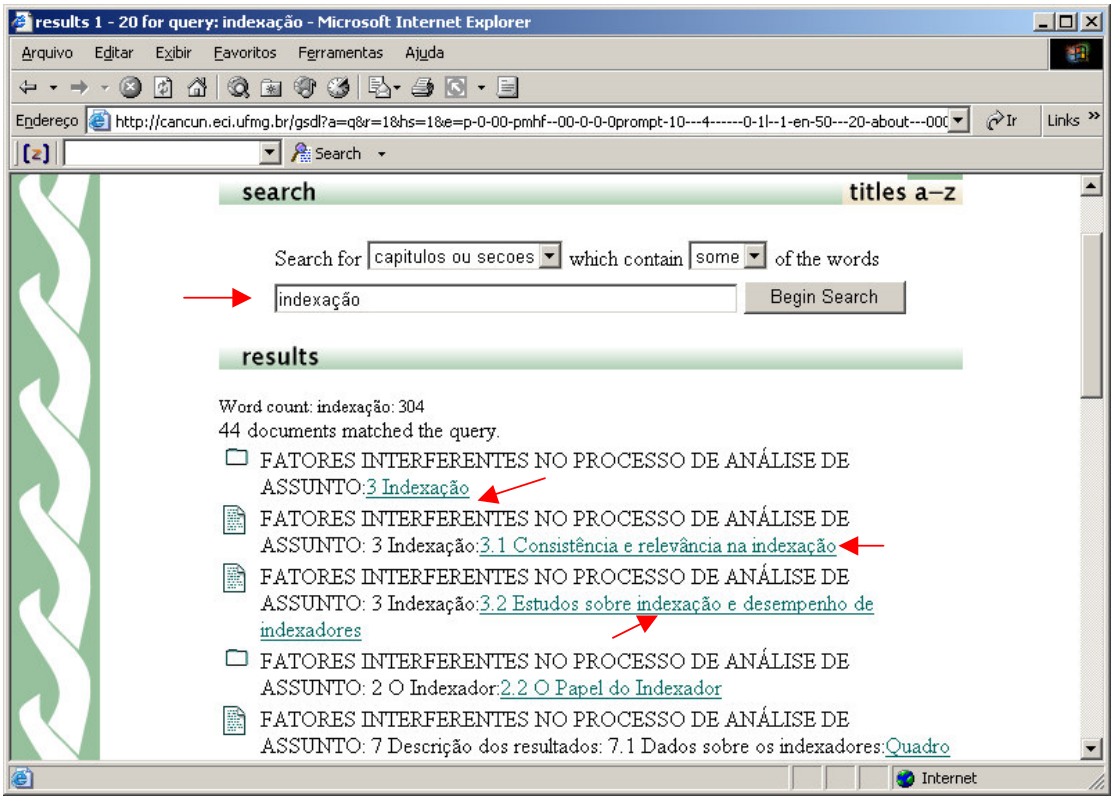
A navegação na base de dados do **MHTX** se inicia a partir das facilidades de busca oferecidas pelo próprio *Greenstone*. O protótipo hipertextual **MHTX** permanente disponibiliza dois tipos de navegação na base de dados da BTDECI: a pesquisa direta e o folheio. A pesquisa direta das teses e dissertações é feita por palavras contidas no texto colocadas no campo "*Search*", na primeira tela que aparece, quando se inicializa a base de dados (FIG.44). Essa tela também pode ser acessada por meio do ícone "*Search*" através de outras páginas. A outra modalidde de acesso às teses e dissertações pode ser feita através do seus títulos, acionando o ícone "*Titles A-Z*". Esse comando leva a uma lista em ordem alfabética com os títulos de todos documentos contidos na base. A partir daí, pode-se iniciar a navegação por folheio.

FIGURA 44 - Pesquisa direta em capítulos ou seções por palavras-chave ou por título



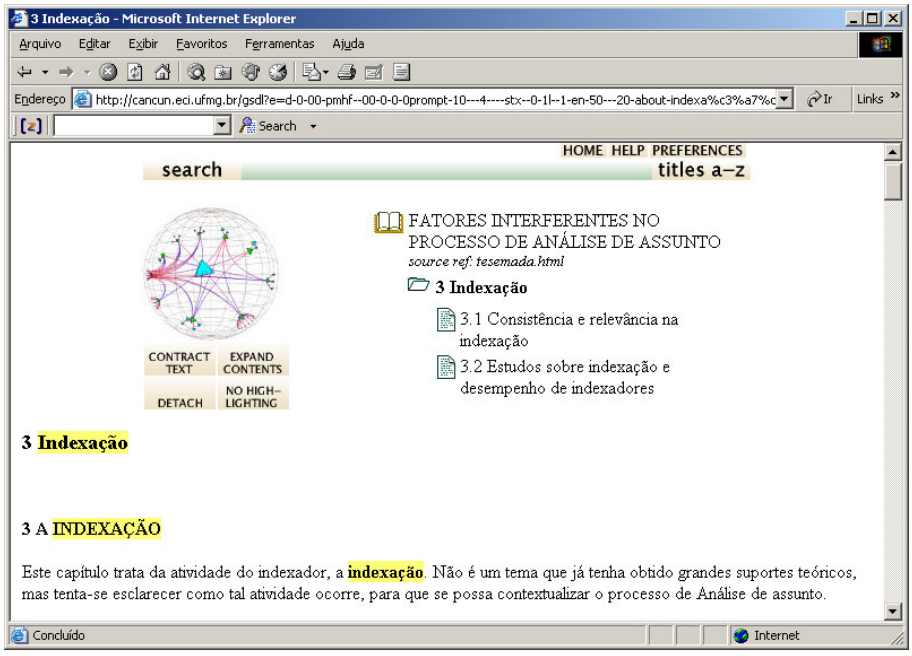

A busca direta é auxiliada pelo índice de assunto do banco de dados criado com o objetivo de possibilitar o acesso direto ao recurso desejado. Por exemplo, se a pesquisa é sobre a palavra-chave “indexação”, o programa *Greenstone*, através de seu mecanismo de coleta de informações, localiza dentro do texto integral todos os lugares onde a palavra-chave “indexação” aparece independentemente do contexto. Deve-se observar que, como a base de dados, até a presente etapa, dispõe apenas da tese estudada, todas as vezes que apareceu, o termo “indexação” este estava no mesmo contexto. No futuro, com a inserção de outras teses isto poderá diferir. O resultado lista os títulos e subseções dos capítulos da tese, com a ordem de relevância do assunto dentro do documento, onde a palavra é citada (FIG.45).

FIGURA 45 - Resultado de pesquisa direta com o termo “indexação”




A escolha e o acionamento de um dos *links* em destaque na figura acima permitem acessar o conteúdo textual e, ao mesmo tempo em que a palavra pesquisada aparece marcada na cor amarela, como mostra a FIG.46.

FIGURA 46 - Recuperação do conteúdo sobre o termo “indexação”



Somente após o acesso à primeira página, em que ocorre a pesquisa das teses e dissertações por palavra ou título (FIG.44 acima), é que o usuário terá a opção de navegar no **MHTX** através do **SE** (capítulos e subseções da tese) e através do **MC** (a estrutura facetada da tese). Nesse tipo de busca por conceito, diferentemente do que ocorre na pesquisa por palavra, o usuário é direcionado para um único lugar no texto, onde se encontra o conteúdo mais relevante sobre aquele determinado assunto. O fluxograma da FIG.47, à frente, mostra esquematicamente as três modalidades de pesquisa no **MHTX**: pesquisa direta no **SE**, folheio no **SE** e navegação no **MC**. Buscou-se aqui preservar a integridade do conteúdo do documento, evitando que as complexidades das informações fragmentadas, inerentes ao sistema de hipertexto, pudessem comprometer o entendimento temático da tese.

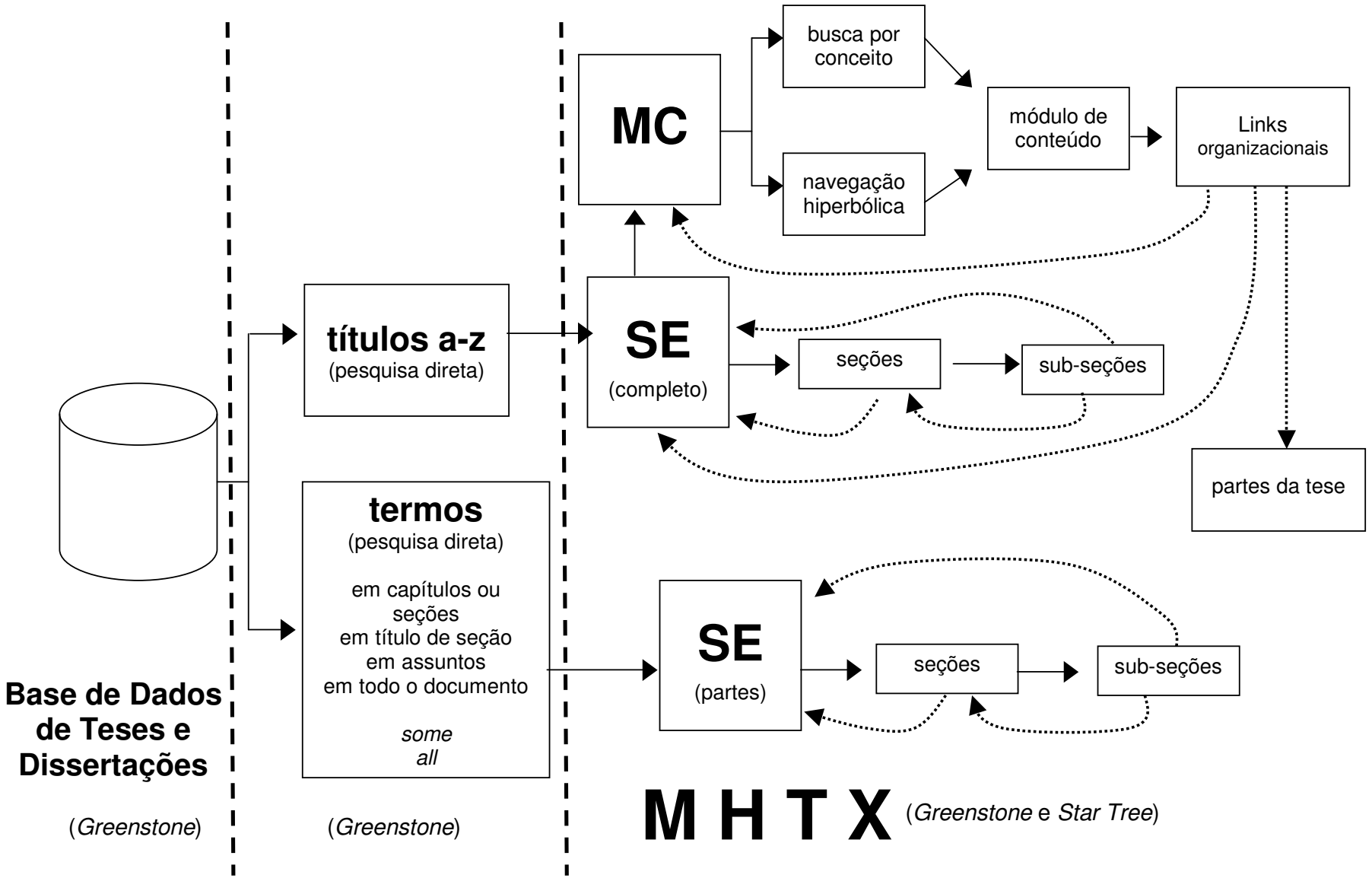
### 3.3.4 Da tese à hipertese

A tese *"Fatores interferentes no processo de análise de assunto"*, escrita sob a forma de documento *Word* (*.doc*), foi convertida em hiperdocumento através da linguagem de marcação *HTML* (*Hypertext Markup Language*), para navegação no **MHTX**. O arquivo *.doc* da tese foi também convertido em hiperdocumento através do *software Bonfire Studio*, que utiliza a linguagem de marcação *XML* para construção dos metadados na base de dados construída, através do *software Greenstone Digital Library*.

Para que os documentos eletrônicos, como o hiperdocumento, sejam disponibilizados via Internet e para que possam ser pesquisados e recuperados *a posteriori*, é necessário um formato padrão que permita sua descrição. Este formato padrão é formado por conjunto de descritores conhecidos como metadados. Se etmologicamente o termo "metadados" é "uma informação que descreve informação" ou "um dado sobre dado", ele incluiu, de um ponto de vista prático, informações descritivas, tais como autor, título, palavras-chave ou outras informações que são associadas ao documento.

O metadado é arquivado com o documento. Por exemplo:
*<Metadata name="Title">FATORES INTERFERENTES NO PROCESSO DE ANÁLISE DE ASSUNTO</Metadata>.* Podemos ver que a *tag <Metadata>*

FIGURA 47 – Fluxograma de busca direta e navegação (hierárquica e hiperbólica) na base hipertextual do MHTX



especifica o nome do tipo de metadado e agrega valor a ela. Nesse exemplo, o título do documento é um dos metadados associados ao documento.

As linguagens de marcação permitem explicitar características que são implícitas para o leitor. A escolha dessas características depende de quem cria o texto eletrônico e os objetivos para os quais servirá. Seu conteúdo pode se constituir de parágrafos, listas, abreviações, nomes, datas ou textos completos. A preocupação para que a conversão do texto linear em hipertexto não perdesse a noção do conteúdo no contexto geral foi um direcionador constante na estruturação do hiperdocumento. Houve a preocupação na estruturação da navegação dentro do contexto e o acesso de qualquer lugar da navegação ao texto completo, seja na seção, na subseção, no capítulo ou no todo da tese. A seguir, serão descritas as formas utilizadas na estruturação do texto da tese.

Como dito anteriormente, a linguagem *HTML* é uma linguagem de marcação voltada para a estruturação de documentos e criação de hipertextos que podem ser interpretados por navegadores. O *HTML* é uma linguagem simples, que não diferencia letras maiúsculas de minúsculas e é flexível quanto ao uso de aspas (simples, duplas, sem aspas). Cada parte da tese (primeiras páginas, Introdução, capítulos e anexos) foi salva na linguagem de marcação *HTML*, momento em que é criado um vínculo de navegação automaticamente. Cada conceito pertencente ao **MHTX** foi devidamente marcado dentro do texto conforme regras já mencionadas.

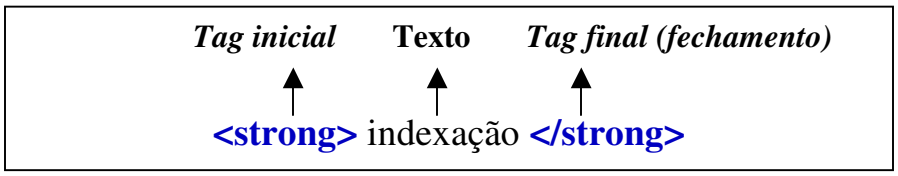
Em algumas circunstâncias, nas quais havia uma proximidade entre conceitos e para evitar subdivisões seqüenciais grandes, estabeleceu-se o mesmo *link* para diferentes conceitos na mesma página de interface da navegação. Por exemplo, os conceitos “macroestrutura“, “microestrutura“ e “hiperestrutura“ compartilham a mesma URL c:\madalena\import\tese_cap4\tese_cap4.html#4.1.1.1.1 (FIG.48).

FIGURA 48 - Marcação de três conceitos próximos com o mesmo *link* na linguagem *html*

**eles constituem os documentos, que são a matéria prima da análise documentária.  <a name="4.1.1.1.1"></a> Para facilitar seu trabalho, é necessário que o indexador conheça estruturas e tipos de textos. O sentido geral do texto é baseado na seguinte trilogia estrutural: microestrutura(estrutura superficial, que corresponde à realidade física do texto e seus símbolos de significação, as palavras), macroestrutura (concebida como um tópico representativo hierárquico e coerente da unidade textual, envolvendo mínima estrutura da representação textual sintática-semântica), e a superestrutura(estrutura retórica-esquemática, um tipo de esquema de produção convencional para o qual o texto é adaptado, podendo ser considerado como transição entre estruturas de superfície e de profundidade). Para CINTRA (1987), os constituintes básicos de um determinado tipo de texto é que definem a sua superestrutura.**

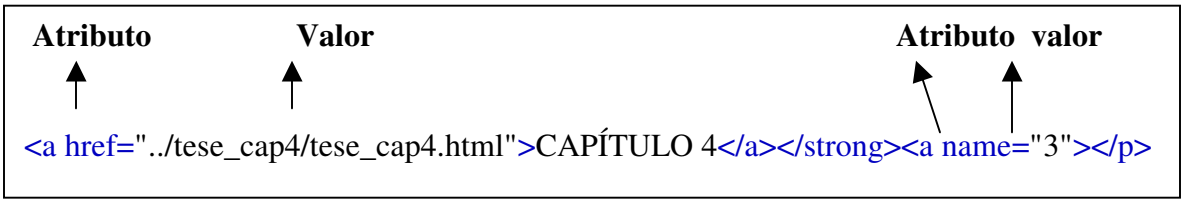
A estrutura do documento *HTML* é formada por elementos denominados *tags*, que possuem *atributos* e *valores*. Ao contrário da *SGML*, a linguagem *HTML* oferece um conjunto de elementos de marcação padronizados, não permitindo aos usuários a criação de seus próprios rótulos. Cada elemento é demarcado por dois *tags*: o *tag inicial* (ou *tag de abertura*) e o *tag final* (ou *tag de fechamento*) escritos entre os sinais de comparação *menor que* (<) e *maior que* (>), indicando como o navegador (*browser*) deve ler o texto, sendo que os *tags* de fechamento têm uma barra (/) colocada logo após o sinal menor que (FIG.49).

FIGURA 49 - Exemplos de *tag inicial* e *tag final*



O conteúdo do texto a ser demarcado por *tags* pode constituir-se de títulos, textos, parágrafos de textos, listas, *links* e âncoras, imagens inseridas, formulários, etc. Os *atributos* definem características ou propriedades de elementos a serem inseridos nas *tags iniciais*, por meio da sintaxe: **nome_do_atributo= "valor"**. Um elemento em *HTML* pode ter vários *atributos* ao mesmo tempo, separados por espaços em branco. Nesse caso, cada valor deve ser sempre delimitado por aspas duplas (FIG.50).

FIGURA 50 - Exemplos de atributo e valor em uma *tag inicial*




A FIG.51 mostra um excerto da tese *Fatores interferentes em análise de assunto* na linguagem de marcação *HTML*.

FIGURA 51 – Excerto da tese "Fatores interferentes em análise de assunto" como documento *HTML*

```
<html>

<head>
<meta HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=windows-1252">
<meta NAME="GENERATOR" CONTENT="Microsoft FrontPage 4.0">
<title>&lt;&lt;TOC1&gt;&gt;FATORES INTERFERENTES NO PROCESSO DE ANÁLISE
DE ASSUNTO</title>
</head>
<body>
<!--
<Section>
-->
<a name="3"><a/>
</a>
<p><strong> <a href="../tese_cap2/tese_cap2.html">CAPÍTULO 2</a>
            
```

```
&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;
<a href="file:///C:/Madalena/sitemap.html">SUMÁRIO </a>
&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;
&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;
<a href="file:///c:/madalena/análisefacetada.html">MAPA CONCEITUAL</a>
&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;
&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;
<a href="../tese_cap4/tese_cap4.html">CAPÍTULO 4</a></strong><a name="3"></p>

<p></a><strong>&amp;nbsp;</strong><a name="3"></p>

<p><strong>3 A INDEXAÇÃO</strong></p>

<p>Este capítulo trata da atividade do indexador, a <strong>indexação</strong>. Não é
um tema que já tenha obtido grandes suportes teóricos, mas tenta-se esclarecer como tal
atividade ocorre, para que se possa contextualizar o processo de Análise de assunto.</p>
```

A linguagem de marcação *XML* é o subconjunto da *SGML*, que permite a criação de marcações de idéias específicas a serem compartilhadas na rede. A *XML* tem as mesmas virtudes da *SGML* e da *HTML*, mas sem qualquer de suas óbvias limitações, permitindo uma linguagem de marcação personalizada, contendo *tags* descritivas sobre os dados nela contidos. Segundo CASTRO (2001, p.13), essa é a principal característica do potencial da *XML*. Para BOSAK (1997, p.2),

> *"A XML deriva da filosofia em que a informação pertence ao seus criadores e que os provedores de conteúdos são melhores servidos por formatos informacionais que não os amarram a uma linguagem específica de sinais, ferramentas de autoria ou ferramentas de busca, mas sim fornecem um campo padronizado, independente de fornecedores, sobre o qual diferentes ferramentas de autoria e de busca possam competir livremente.*

*Tags* com esse nível de identificação tornam os dados disponíveis para outras tarefas, o que permite uma escrita mais econômica do hipertexto, em que é possível extrair apenas as informações necessárias. Isto não ocorre com uma

página da Web escrita em *HTML*, em que muitas informações ficam perdidas. Na linguagem *XML*, as informações rotuladas podem ser reutilizadas sempre que necessário através de método estruturado e consistente para descrever e transferir a informação. Além disso, a *XML* possui a característica de separar o formato visual da informação propriamente dita.

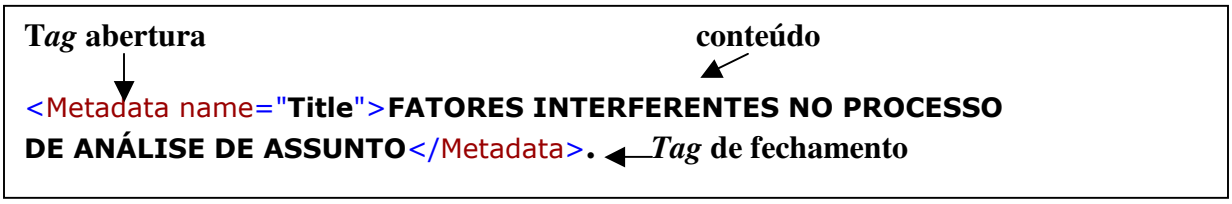
Entretanto, para facilitar a leitura e a interpretação dos dados escritos nessa linguagem pelos programas, a *XML* exige padronização quanto ao uso de maiúsculas e minúsculas, aspas, *tags* do fechamento e outras minúcias ignoradas pelos autores na linguagem *HTML*.

A *XML* utiliza os mesmos blocos de construção que a *HTML*: *elementos* e *tags* com *atributos* e *valores*. O elemento *XML* é a unidade mais básica de um documento, podendo conter qualquer tipo de informação, inclusive informação semântica, fator fundamental para a presente pesquisa, que propôs a criação um modelo de metadados semânticos, modelados através da análise facetada de uma tese.

O elemento *XML* deve possuir uma *tag de abertura* com seu nome escrito, exatamente como nas *tags HTML*, entre os sinais *menor que* (<) e *maior que* (>) com a barra (/) após o sinal menor que na *tag* de fechamento, podendo, às vezes, incluir também os *atributos*.

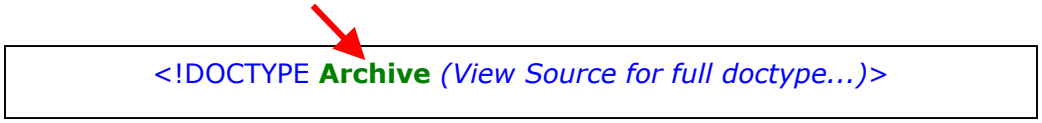
Os *atributos* que estão contidos na *tag de abertura* do elemento têm valores delimitados por aspas que descrevem a finalidade e o conteúdo (quando existir) de um determinado elemento (FIG.52). As informações contidas em um atributo geralmente são consideradas metadados, ou seja, dados de informações sobre os dados no documento *XML*, e não os dados com a informação final em si mesma.

FIGURA 52 - Exemplo da marcação de metadado no título da tese



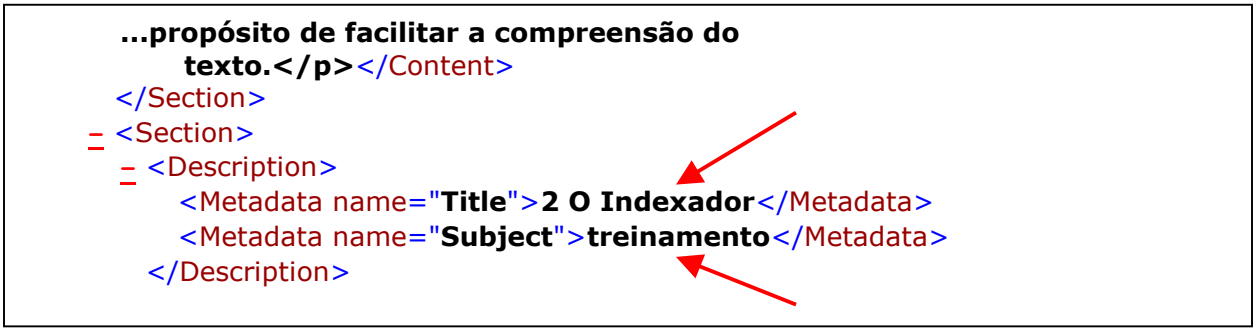

Para criar um documento em *XML*, deve-se indicar qual a versão *XML* está sendo utilizada, o que geralmente se dá na primeira linha do documento, chamada de *declaração XML* ou *documento raiz.* Após esse comando, passa-se ao *DTD* (*Document Type Definition* ou Definição de Tipo de Documento), que é uma gramática que confere estrutura ao documento *XML*. No protótipo desta pesquisa, foi realizado este procedimento para criar um DTD do tipo *archive* ("arquivo") com a raiz do documento, como mostra a FIG.53.

FIGURA 53 - Declaração de tipo de documento em *XML* no documento raiz




Como o *software Greesntone Digital Library* utiliza a *XML* como linguagem de marcação, gerou-se um arquivo convertendo o documento *.html* para documento *.xml.* Cada parte da tese (primeiras páginas, introdução, capítulos e anexos) foi salva na linguagem de marcação *XML*, permitindo depois a marcação dos metadados semânticos dentro de cada seção correspondente à sua localização no texto. Cada conceito pertencente à tese foi devidamente marcado dentro da seção correspondente para recuperação futura. Na FIG.54, pode-se ver o metadado criado para o título do *Capítulo 2 - O Indexador* e o metadado criado para o assunto "treinamento" que está inserido dentro deste capítulo.

FIGURA 54 – Criação de metadado para “título” e “assunto” da tese com valor semântico em *XML*




A FIG.55 mostra o excerto da tese *Fatores interferentes em análise de assunto* na linguagem de marcação *XML*.

FIGURA 55 – Exemplo de marcação *XML* da tese Fatores interferentes em análise de assunto

```
ALIGN="JUSTIFY"> </p>                                    <p
ALIGN="JUSTIFY"> </p>
<pALIGN="JUSTIFY"> </p><p
ALIGN="JUSTIFY"> </p>                                    <p
ALIGN="JUSTIFY"> </p>                                    <p
ALIGN="JUSTIFY"> </p></Content>
</Section>
```

# 4 CONSIDERAÇÕES FINAIS

A revisão de literatura do presente estudo, tanto em relação à pesquisa em estudos de caso quanto à pesquisa sobre tendências mais gerais observadas nos meios profissionais, mostrou que há dois problemas organizacionais básicos que ocorrem com o leitor hipertextual, problemas que são diretamente ligados à inerente fragmentação dessa forma de texto: a desorientação e o transbordamento cognitivo durante a navegação.

Esses dois problemas decorrem principalmente ou do fato de os *Web designers* não se preocuparem em explicitar o conteúdo semântico dos textos originais que o geraram ou pela não cooperação entre o autor temático e o autor tecnológico. A maioria desses profissionais não é da área de ciência da informação e recorrem principalmente aos mecanismos de recuperação automática. Esta deficiência foi a principal motivação para o presente estudo, que resultou na criação de uma estrutura conceitual,com sua correspondente implementação tecnológica, que permitiu uma estruturação e posterior navegação ao mesmo tempo flexível, coerente e objetiva. A revisão de literatura mostrou também que as pesquisas na área de ciência da informação ainda não privilegiam, como objeto de pesquisa, o momento de estruturação do hipertexto, mas sim as formas de leitura no momento da recuperação da informação, as quais, em muitos casos, se mostra ineficaz porque se resumem à utilização de algoritmos.

Como o presente estudo constituiu-se de uma pesquisa aplicada, a fundamentação teórica da tese teve dois momentos. O primeiro delineamento no qual foram estudados os fundamentos teóricos e referenciais que serviram de base para a idealização do modelo conceitual. No segundo momento, que é a parte prática do projeto e que envolveu a implementação tecnológica do modelo e a verificação de seu funcionamento, buscou-se na fundamentação explicar seu construto e sua aplicabilidade em bases teóricas.

Espera-se que o presente estudo contribua para a formação de uma teoria de estruturação hipertextual a partir do presente modelo proposto, fundamentado nas teorias e princípios, tal como expresso nesta tese. A literatura comprova que já há

massa crítica de pesquisadores interessados em estabelecer princípios, metodologias e procedimentos organizacionais para dar conta da crescente preferência e da dependência do homem pelo texto virtual. As metodologias e modelos descritos na fundamentação mostram que, apesar do objetivo comum de produzir hipertextos e *sites* da *Web* bem estruturados e organizados, estes ainda não satisfazem os paradigmas autorais de lidar com as interconexões dos fragmentos semânticos contidos nos textos. Mesmo as pesquisas mais recentes que tentam, através da ontologia de um domínio, representar um modelo conceitual sofisticado com os termos dos documentos e suas relações, não apresentaram ainda a saída para minimizar os problemas da fragmentação e da recuperação semântica nos seus contextos.

Entre as abordagens metodológicas de estruturação hipertextual apresentadas, a proposta de CAMPOS (2001) apresenta recomendações construtivas que visam resolver o problema da fragmentação das unidades de informação. Seu estudo destaca-se na literatura pela atenção às necessidades e abordagens teóricas da ciência da informação no desenvolvimento da modelagem semântica do sistema de hipertexto, mostrando preocupação com a preservação do conteúdo, os problemas da fragmentação textual e a organização de hiperdocumentos a partir da autoria dos mesmos. Entretanto, a autora da presente tese reconhece a necessidade de preencher o hiato entre suas recomendações de desenvolvimento de metodologias de autoria de hipertexto e a aplicação prática com o *design* e a implementação tecnológica desse tipo de documento.

À exceção da abordagem de Maria Luíza Almeida Campos, única proveniente da área de ciência da informação, todos os demais modelos aqui apresentados foram criados por pesquisadores da ciência da computação (especialmente da comunidade de engenharia de *softwares*), exemplificando bem dois aspectos do mesmo quadro: a incipiência de estudos correlatos na área de ciência da informação e a ainda difícil integração entre os profissionais dessas duas áreas. Além disso, nota-se ainda grande dificuldade desses modelos em ultrapassar as fronteiras da academia ou dos laboratórios onde são criados, não alcançando ainda o *status* de produto comercial ou de fácil acesso ao público em geral.

Para o desenvolvimento desta tese partiu-se dos pressupostos de que um sistema de hipertexto não apresenta um nível desejável de eficácia no processo de acesso às informações e sua recuperação por parte dos usuários se os hiperdocumentos que o compõem não se apresentarem estruturados conceitualmente. A Teoria de Análise Facetada (TAF), com raízes em Aristóleles, proposta por Ranganathan, adaptada pelo *CRG* (*Classification Research Group*) e simplificada por Louise Spiteri, apresenta-se como inquestionável potencial para compor a estrutura lógica de um hipertexto. Já os princípios de construção de mapas conceituais, proveniente da área da educação e psicologia, tornaram a implementação do modelo proposto mais cognitivo visualmente e de utilização mais amigável.

O protótipo **MHTX** foi composto de uma base de dados, um esquema de representação com nós, *links*, relações entre os *links* e interfaces, com mecanismos de busca e folheio. Sua elaboração se deu, primeiramente, com a implementação conceitual na modelagem do MC e na estruturação do SE. Na etapa seguinte, o modelo foi implementado tecnologicamente com os recursos técnicos dos *softwares Greenstone Digital Library*, para construção da base de dados e do SE e do *software Star Tree Studio* para a construção do MC.

Os nós representam as unidades semânticas do conhecimento da tese, equivalendo às seções ou subseções que tratam de determinado conceito. Assim, os nós podem conter outros nós, dependendo de sua especificidade. Os *links* de conteúdo fazem as conexões entre os nós e os *links* organizacionais estabelecem as relações entre os tópicos do sumário e suas subdivisões.

Estudos da psicologia cognitiva mostram que o fato de a memória humana compreender e reter melhor as informações organizadas espacialmente em diagramas e em mapas conceituais manipuláveis favorece o domínio mais rápido e fácil das informações, em relação aos meios áudio-visuais tradicionais.

A estrutura de conhecimento utilizada no protótipo desenvolvido nesta tese reflete fortemente o modelo cognitivo no qual se basearam o processo de organização e a escrita. A estrutura semântica de rede, na arquitetura do sistema, foi inspirada e

construída com base no modelo cognitivo. Isso foi alcançado através da indexação do texto e sua organização em seções modulares que representassem o conteúdo semântico de cada nível do texto, que foi progressivamente organizado até o produto final pela marcação dos *links* conceituais e dos *links* organizacionais da estrutura da tese facetada.

Na construção do Mapa Conceitual (MC), estas relações puderam ser evidenciadas visualmente, mostrando-se a precedência entre dois conceitos ou as relações associativas através de dois conceitos no mesmo nível de hierarquia. No estabelecimento de relações entre os termos, o controle de sinônimos e quase-sinônimos se deu pela escolha dos termos encontrados na tese estudada. Já em relação aos homógrafos, estes não ocorreram. Em etapa posterior, quando houver necessidade desses relacionamentos, como, por exemplo, a inclusão de assuntos compostos no MC, deverá ser estudada uma forma para que o termo possa ter mais de um conceito, os quais remeterão a mais de uma parte do documento.

A conversão da tese no formato *.doc* para documento no formato *.xml* possibilitou a marcação dos metadados dentro de cada seção e subseções da tese, permitindo mostrar o conteúdo de suas partes sem perder a visão do todo. Permitiu, ao mesmo tempo, mostrar as relações hierárquicas (do geral para o específico), gerando renques e cadeias.

Os mecanismos de navegação implementados no protótipo são três: (1) o folheio e (2) a pesquisa direta, os quais possibilitam ao usuário pesquisar por palavras-chave indexadas, ou (3) navegar por meio do mapa conceitual (interface hiperbólica) e navegação no sumário expandido por meio da interface em forma de árvore.

O potencial da TAF para compor a estrutura lógica do hipertexto serviu como ponto de partida para a modelagem da estrutura cognitiva da tese e o trabalho de análise de sua área de assunto. A abordagem estruturada da TAF permitiu a identificação de características individuais de vários conceitos, que puderam ser agrupados de maneira analítica, permitindo ao usuário ver um assunto sob diferentes pontos de vista. Essa característica satisfez a não linearidade do hipertexto, facilitando a criação de estrutura

fluída e interativa como a do **MHTX**. Observa-se que nem todos os métodos para a formação de assuntos propostos por *Ranganathan* foram utilizados. Dos cinco princípios propostos por ele (dissecação, laminação, desnudação, agregação e superposição) somente dois foram utilizados: o método de dissecação e o método de desnudação. Isso ocorreu devido ao fato de a amostra, uma tese apenas, ter domínio específico, cujos conceitos identificados estavam contidos em linguagem específica, não ocorrendo oportunidade de se aplicarem os outros métodos.

Para a formação das categorias, com base no Modelo Simplificado de Spiteri, foram aplicados somente sete princípios do plano das idéias na escolha das facetas e subfacetas: diferenciação, relevância, verificação, permanência, homogeneidade, exclusividade mútua e categorias fundamentais. Desses princípios, Spiteri adotou (1) o principio de diferenciação com base na proposta de Ranganathan; (2) o princípio das categorias fundamentais com base no *CRG*; e (3) os princípios de relevância, verificação, permanência, homogeneidade, exclusividade mútua com base em ambos Ranganathan e *CRG*. Como resultado, a tese selecionada apresentou apenas quatro facetas na sua estruturação conceitual: personalidade, atributo, energia e disciplina.

As recomendações para estruturação hipertextual de CAMPOS (2000) foram adotadas em seus três níveis, sendo que a implementação do **MHTX** foi um primeiro passo no cumprimento do Requisito 7 do terceiro nível (“Nível de estabelecimento de um veículo de comunicação e expressão sobre a temática do hiperdocumento”) de sua abordagem metodológica. Esse passo fez parte do percurso maior que é, primeiro, a realização de um teste experimental e aperfeiçoamento do **MHTX** e, segundo, a utilização desse protótipo na criação e organização da BTDECI – UFMG (Biblioteca de Teses e Dissertações do Programa de Pós-Graduação da Escola de Ciência da Informação da UFMG).

A utilização do mapa conceitual como técnica de representação gráfica, principalmente com o uso da interface tridimensional em formato hiperbólico “olho de peixe”, melhorou a navegação, mesmo no caso de se lidar com grandes quantidades de informações, auxiliando no rápido acesso à informação desejada através do *link* conceitual escolhido. O mapa conceitual, assim como a TAF, também possibilitou estruturação semântica da

tese objeto desta pesquisa, permitindo explicitar as relações de semelhanças e diferenças e facilitando o agrupamento. Assim, foi possível demonstrar as relações existentes nos renques e nas cadeias produzidas pela facetação da referida tese, através da estrutura de relações hieráquicas e associativas. O documento da tese foi formatado na linguagem HTML e cada *link* do mapa conceitual teve sua marcação inserida dentro do texto. A ordem de citação das facetas não foi aplicada porque todo conceito (*link*) tem o mesmo valor semântico na estrutura navegacional.

No processo de seleção de *softwares* para a implementação do **MHTX**, buscou-se a otimização das ferramentas tecnológicas. O software *Greenstone* é especifico para criação de bibliotecas digitais, que apesar de algumas limitações, possibilita a criação de uma base de dados com textos completos. O *software Star Tree Studio* utilizado na construção e na estruturação hierárquica do MC, também serviu para demonstração não só das relações hierárquicas como também das relações associativas, permitindo a inserção de mais de um termo no mesmo nível.

Alguns problemas encontrados durante a pesquisa são dignos de nota. O desenvolvimento do protótipo demonstrou a carência de suporte tecnológico para sua estruturação. As dificuldades de escolha e coordenação entre os dois *softwares* poderão diminuir em aplicações futuras, com o surgimento de pacotes de *softwares* que contemplem a implementação de todas as fases do **MHTX**. Deve-se ressaltar que além dos aplicativos criados nesta tese, o software *Greenstone* possui características que possibilitarão gerar aplicativos suficientes para montar uma biblioteca digital, como a BTDECI - UFMG. Por outro lado, apresenta limitações quanto à inserção de aplicativos em ferramentas construídas com outros *softwares*, necessitando da criação de *links* para remeter a esses aplicativos. Isso gerou dificuldades no acesso direto ao MC, que teve de ser abrigado dentro do SE, o que força o usuário a passar por uma página intermediária antes de acessá-lo.

Pode-se concluir que a técnica da análise facetada foi eficiente na modelagem conceitual da tese, proporcionando dinâmica metódica, desde a identificação dos termos relevantes até a formação das categorias. O MC proveu uma alternativa para o problema de desorientação do usuário, apesar de não incluir a técnica de rastreamento

da navegação. Por outro lado, criou-se uma navegação orientada com código de cores representando cada nível hierárquico do mapa, explicitando ao usuário como todo o conteúdo semântico é organizado e como se conecta internamente, facilitando a navegação hiperbólica.

Entre as novas perspectivas nessa área de pesquisa, destacam-se duas. A recente pesquisa desenvolvida no projeto *Facet Analysis*, sob a coordenação de Douglas Tudhope (Escola de Computação, *University of Glamorgan*, Inglaterra; *http://www.glam.ac.uk/soc/research/hypermedia/facet_proj/index.php*) e a *XFML* (*eXchangeable Facet Metadata Language*; *www.xfml.org*), que é um formato aberto da linguagem *XML* para publicação e compartilhamento de metadados hieráquicos facetados e processos de indexação.

Finalmente, as sugestões para continuação deste trabalho, em âmbito mais geral, dizem respeito a duas áreas especificas. Em relação à área de ciência da informação, destaca-se a necessidade de construção de linguagens que melhor atendam o tratamento semântico vislumbrado na *Web* semântica, e a utilização de assuntos compostos na modelagem conceitual facetada. Já em relação à área de ciência da computação, é imperativo o desenvolvimento de pacotes de *software* que possibilitem a organização da informação e a inserção de metadados sem perda semântica e que permitam a utilização de mapas conceituais para navegação em sistemas de hipermídia com interfaces de fácil compreensão. Talvez os esforços específicos nessas duas frentes se tornem importante passo para a almejada integração entre essas áreas do conhecimento: CI e CC.

# 5 REFERÊNCIAS


ABOUD, M. et al. Querying a hypertext information retrieval system by the use of classification. *Information Processing & Management*, v.29, n.3, p.387-396, May/June, 1993.

AIGRAIN, Philippe, LONGUEVILLE, Véronique. Evaluation of navigational links between images. *Information Processing & Management*, v.28, n.4. p. 517-528, 1992.

ALLEN, Bryce L. Cognitive research in information science: implications for design. *Annual Review of Information Science and technology*, v.26, p.3-37, 1991.

AMORETTI, Maria S.M; TAROUCO, Liane M. R. Mapas conceituais: modelagem colaborativa do conhecimento. *Revista Informática na Educação: Teoria & Prática*, v.3, n.1, p.67-71, set., 2000.

AMORETTI, Maria Suzana Marc. Protótipos e estereótipos: aprendizagem de conceitos. Mapas conceituais: experiência em educação à distância. *Revista Informática na Educação: Teoria & Prática*, PGIE, UFRGS, v.4, n.2, p. 49-55, dez., 2001.

ANDRIS, J.F. The relationship of indices of student navigational patterns in a hypermedia geology lab simulation to two measures of learning style. *Journal of Educational Multimedia and Hypermedia*, v.5, n.3/4, p.303-315, 1996.

*ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, NBR 6024: informação e documentação - numeração progressiva das seções um documento escrito - apresentação*. Rio de Janeiro: ABNT, 2003. 3 p.

*ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS.NBR 6027: informação e documentação, sumário, apresentação*. Rio de Janeiro: ABNT, 2003. 2 p.

AUSUBEL, D. P. *The psychology of meaningful verbal learning*. New York: Grune and Stratton, 1963.

AUSUBEL, D. P. *Educational psychology*: a cognitive view. New York: Holt, Rinehart and Winston, 1968.

AUSUBEL, D. P.; NOVAK, J.D.; HANESIAN, H. *Educational psychology*: a cognitive view. 2ed. New York: Holt, Rinehart and Winston, 1978.

AUSTIN, Derek. *Fields, categories and general systems theory.* London: BNB, 1968.

AUSTIN, Derek. Prospects for a new general classification. *Journal of Librarianship*, n.1, v.3, p.149-169, 1969.

AVISON, D. ; FITZGERALD, G. *Information systems development*. London: McGraw-Hill. 1995.

BALASUBRAMANIAN, V. *State of the art review on hypermedia issues and applications*. Newark, New Jersey: Rutgers University, Graduate School of Management, 1993.

BARBOSA, Alice Príncipe. Classificações facetadas. *Ci. Inf*. Rio de Janeiro, v.1, n.2, p.73-81, 1972.

BARON, Lisa et al. Labeled, typed links as cues when reading hypertext documents. *Journal of the American Society for Information Science*, v. 47, n.12, p.896-908, 1996.

BARTHES, R. *From work to text in textual strategies:* perspectives in post-structuralist criticism. Ithaca: Cornell University Press, 1979.

BEASLEY, R.E.; VILA, J.A. The identification of navigation patternship in a multimedia environment: A case study. *Journal of Educational Multimedia and Hypermedia,* v.1, n.2, p. 209-222. 1992.

BEASLEY, R.E.; WAUGH, M.L. Predominant initial and review patterns of navigation in a fully constrained hypermedia hierarchy: An empirical study. *Journal of Educational Multimedia and Hypermedia,* v.6, n.2, p.155-172, 1997.

BOECHLER, P.M.; DAWSON, M.R.W. *Connections between performance, path patterns and mental representation in hypertext navigation,* 2002. (Inédito)

BOGMAN, Christine L.; CHIGNELL, Mark H.; VALDEZ, Felix. Designing an information retrieval interface based on children's categorization of knowledge: a pilot study. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE 52, 1989. Washington, DC. *Proceedings…* Medford, NJ: Learned Information, 1989. p. 81-87.

BOLTER, J. D. *Writing space*: the computer, hypertext, and the history of writing. Hillsdale, N.J: Lawrence Erbaum, 1991.

BORDONARO, K. E LYNCH, P. P. Search engines and subject searching of the World Wide Web. *The Internet Homesteader*, v.2, n.8, p.14-16.1995.

BOSAK, Jon. *XML, Java, and the future of the Web*. 9p. 1997. Disponível em <http://www.ibiblio.org/pub/sun-info/standards/xml/why/xmlapps.htm >. Acesso em: 10 nov. 2003.

BROCOLLI , Kevin. *Indexes:* an old tool for a new medium. [online]. 8p. Disponível em <http://www.contentious.com.articles/1-8/>. Acesso em: 17 nov. 1998.

BROUGHTON, Vanda; LANE, Heather. Classification schemes revisited: applications to web indexing and searching. *Journal of Internet Cataloguing*, v.2, n.3/4, p.143-155, 2000.

BUSH, Vannevar. As we may think. *Atlantic Monthly*, v.176, n.1, p.101-108. 1945.

CAMPOS, Maria Luiza Almeida. *A organização de unidades do conhecimento em hiperdocumentos*: o modelo conceitual como um espaço comunicacional para realização da autoria. 2001. 190p. Tese (Doutorado em Ciência da Informação) - CNPq/IBICT-URFJ/ECO, Rio de janeiro, 2001.

CARLSON, P. A. New Metaphors for Electronic Text. *Impact Assessment Bulletin, v.9, n.*1/2, p.73-88, 1991.

CASANOVA, M. A. et al. The Nested Context Model for hyperdocuments. In: ACM CONFERENCE ON HYPERTEXT, 3, 1991, San Antonio. *Proceedings…* San Antonio: EUA, 1991, p.193-201.

CASTRO, Elizabeth. *XML para a Word Wide Web*. Trad. Hugo de Souza Melo. Rio de janeiro: Campus, 2001. 269p.

CERI, S.; FRATERNALI, P.; BONGIO, A. Web Modeling Language (WebML): a modeling language for designing Web sites. *Computers networks*, v.33, p.137-157, 2000.

CHAIBEN, Hamilton. *Hipermídia na educação*. Disponível em <http//:www.cce.ufpr.br/ ~hamilton/hed/hed.htm04/>. Acesso em: 15 nov. 2003.

CHEN, C.; RADA, R. Interacting with hypertext: A meta-analysis of experimental studies. *Human-Computer Interaction,* v.11, p.125-156, 1996.

CLASSIFICATION RESEARCH GROUP. The need for a faceted classification as the basis of all methods of information retrieval. In: CHAN, L. M. et al. (Ed.). *Theory of subject analysis*. Littleton, Col.: Libraries Unlimited, 1985, p.154-167.

*THE CONCEPTUAL open hypermedia project.* Highfield, UK: University of Southampton, Department of Electronics & Computer Science, 2002. Disponível em <http://cohse.semanticweb.org/> Acesso em: 10 jul. 2003.

CONKLIN, J. Hypertext: An introduction and survey. *Computer*, n.20, n.9, p.17-41, 1987.

COUGO, Paulo. *Modelagem conceitual e projeto de banco de dados*. Rio de Janeiro: Campus, 1997.

COURTOIS, M. P.; BAER, W. M.; STARK, M. Cool tools for searching the Web: A performance evaluation. *Online: The Magazine of Online Information Systems*, v.19, n.6, p.14-32, 1995.

CROUCH, D. B., CROUCH, C.J., ANDREAS, G. The use of cluster hierachies in hypertext information retrieval. In MEYROWITZ, N. (Ed.), HYPERTEXT'89, *Proceedings*... Pittsburgh, PA, Nov., 5-8. New York: ACM Press. 1989. p. 225-237.

DALAL, Nikunj; P.QUIBLE, Zane; WYATT, Katherine. Cognitive design of home pages: an experimental study of comprehension on the World Wide Web. *Information Processing and Management*, n.16, 2000, p.607-621.

DANIEL, R.; MILLS, J. *Classification of library of information science*. London: Library Association, 1975.

DAVID, Claire et al. Indexing as problem solving: a cognitive approach to consistency. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE ANNUAL MEETING, 58, 1995, Washington. *Proceedings...* Washington, DC, 1995, p.49-55.

DERRIDA, J. Signature event context. In: MARIN, S.; SUSSMAN, H. *Glyph*. Baltimore, Maryland: Johns Hopkins University, 1977. p. 172-197

*DEWEY decimal classification and relative index*. 22ed. Dublin, OH: Forest Press/OCLC, 2003. Disponível em <http://www.oclc.org/oclc/fp/>. Acesso em: 12 jan. 2002.

DE TROYER, O; LEUNE, C. WSDM: a user-centered design method for web sites. In: 7TH INTERNATIONAL WORD WEB CONFERENCE. *Proceedings*... 1997. Disponível em <http://www7/ scu.edu.au/programme/fullpapers/1853/com1853.htm>. Acesso em: 25 nov. 2003.

DIAS, Paulo. *A abordagem da comunicação multidimensional na concepção e desenvolvimento de interfaces hipermedia*. Disponível em <http:// niee.educom.ufrgs.br/ribie98/CONG1994/volumeII/II 3040.html> Acesso em: 05 dez. 2000.

DIAS, Cláudia Augusto. Hipertexto: evolução histórica e efeitos sociais. *Ci. Inf*. Brasília, v.28, n.3, p.267-275, set./dez. 1999.

DICIONÁRIO AURÉLIO eletrônico; século XXI. Rio de Janeiro: Nova Fronteira, Lexicon Informática, 1999. CD-rom, versão 3.0.

DUNCAN, Elizabeth B. Structuring knowledge bases for designers of learning materials. *Hypermedia*, v.1, n.1, p.20-33, 1989.

DUNCAN, Elizabeth B. A faceted approach to hypertext? In: McALEESE, Ray. *Hypertext:* theory into practice. Nowood, NJ: ABLEX, p.157- 163, 1989.

EDWARDS, Deborah M., HARDMAN, Lynda. Lost in hyperspace: Cognitive mapping and navigation in a hypertext environment. In: McALEESE, Ray. *Hypertext:* theory into practice. Norwood, NJ: ABLEX, p.105-125, 1989.

EKLUND, John. *Cognitive Models for structuring hypermedia and implications for learning from the world-wide web*. Disponível em <http://elmo.scu.edu.au/sponsored/ausweb95/papers/hypertext/eklund/ index.html> Acesso em: 15 set. 2002.

ELGELBART, Douglas. Englebart, D. Augmenting human intellect: a conceptual framework. SRI Report. 1962. Disponível em <http://www.ibiblio.org/pioneers/englebart.html>. Acesso em: 07 jan. 2004

ELLIS, David; VASCONCELOS, Ana. The relevance of facet analysis for World Web subject organization and searching. In: THOMAS, Alan R.; SHEARER, James R. *Internet searching and indexing*: the subject approach. New York: The Haworth Press, 2000. p.97-114.

ELLIS, David. *Progress and problems in information retrieval*. 2 ed. London: Library Association Publishing, 1996. 220p.

ELLIS, David; FORD, Nigel; FURNER, Jonathan. In search of the unknown user: indexing, hypertext and World Wide Web. *Journal of Documentation*, v.54, n.1, p.28-47, Jan. 1998.

ESCHER M.C. Antisymmetrical arrangements in the plane and regular three-dimensional bodies as source of inspiration to an artist. *Acta Cryst,* v.13, p.1083. 1960.

FIDERIO, J. A Grand Vision: Hypertext mimics the brain's ability to access information quickly and intuitively by reference. *Byte Magazine*, vl.13, n.10, pp.237-244, Oct. 1988.

FISHER, Scott. *Multimedia Authoring*. London: Academic Press, 1994.

FORD, N. e CHEN, S.Y. Individual differences, hipermedia navegation and learning: an empirical study. *Journal of Educational Multimedia and Hypermedia*, v.9, n.4, p.281-311, 2000.

FOSKETT, A. C. *The subject approach to information*. London: Library Association Publishing, 1996.

FOSKETT, D. J. *The London education classification*. London: Institute of Education, University of London, 1963.

FOSKETT, D. J. *Classification for a general index language*: a review of recent research by the Classification Research Group. London: Library Association, 1970.

FOSKETT, D. J. Classification Research Group, 1952-1962. In: *ENCYCLOPEDIA of library and information science.* Ed. Allen Kent and Harold Lancour. v.5. New York: Marcel Dekker, 1971.

FOSKETT, D. J. *Classification and indexing in the social sciences*. London: Butterworths, 1974.

FOSS, Carolyn L. Tools for reading and browsing hypertext. *Information Processing & Management*, v.25, n.4, p.407-418, 1989.

FOUCAULT, M. What is an author? In: RABINOW, P. (Ed.). *The Foucault Reader*. London: Penguin Books, 1979. p. 101-120.

FOWLER, R. H.; BRADELEY, A. W.; FOWLER, W.A.L. *Information navigator: an information system using associative networks for display and retrieval*. Disponível em <http://bahia.cs.panam.edu/info_vis/inf_nav_tr92.html> Acesso em: 20 dez. 2002.

FRANÇA, J. et al. *Manual para normalização de publicações técnico-científicas*. Belo Horizonte: Ed. UFMG, 1998.

FREI, H.P.; STIEGER, D. The use of semantic links in hypertext information retrieval. *Information Processing & Management*, v. 31, n. 1. p. 1-13, 1995.

FURNAS, George, W. Generalized fisheye views. In: ACM CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 1986 Boston, MA. *Proceedings...* Boston, MA: ACM, 1986. p.16.

GAINES, B. R.; SHAW, M. *Concept maps as hypermedia components*. Alberta, Canadá: Knowledge Science Institute, University of Calgary, 1995. Disponível em <http://citeseer.nj.nec .com/ gaines95concept.html> Acesso em: 13 nov. 2003.

GARDNER, H. *A nova ciência da mente*. São Paulo: Edusp, 1996.

GARZOTTO, F.; PAOLINI, P.; SCHWABE, D. HDM: a model-based approach to hypertext application design. *ACM Transactions of Information Systems*, v.11, n.1, p.1-26, Jan. 1993.

GLASSEL, Aimee. Was *Ranganathan a Yahoo!.* Mar. 1998. Disponível em <http://scout.cs.wisc.edu/addserv/toolkit/enduser/archive/1998/euc-9803.html>. Acesso em: 14 Mai. 2001.

GLENN, B. T; CHIGNELL, M. H. Hypermedia: Design for browsing. In: Hartson, H.R; Hix, D. (Eds.), *Advances in Human-Computer Interaction*. Nerwood, New Jersey: Ablex Publishing Corporation, v.3, p.143-183,1992.

GLUSHKO, Robert J. Transforming text into hypertext for a compact disc encyclopedia. In: CONFERENCE IN HUMAN FACTORS IN COMPUTERS SYSTEMS, 1989, Austin, TX. Apr. *Proceedings*… New York: ACM Press, 1989. p.293-298.

*GREENSTONE Digital Library Software, vers.2.50*. New Zealand, Waikato: New Zealand Digital Library/UNESCO. Disponível em <http://www.greenstone.org/cgi-bin/library>. Acesso em: 13 out. 2003.

GUIMARÃES, Célio. *Introdução às linguagens de marcação*: HTML, XHTML, SGML, XML. Disponível em <http://www.dcc.unicamp.br/~celio/inf533/docs/markup.html>. Acesso em: 24 out. 2003.

GUINCHAT, Claire. *Introdução geral às ciências e técnicas da informação e documentação*. 2 ed. Brasília: IBICT, 1994. 540p.

HALASZ, Frank; SCHWARTZ, Mayer. The Dexter hypertext reference model. *Communications of the ACM*, v.37, n.2, p.30-39, Feb. 1994. Disponível em <http://citeseer.nj.nec.com/halasz94dexter.html>. Acesso em: 24 nov. 2003.

HANDSCHUH, S.; STAAB, S.; MAEDCHE, A. *CREAM- Creating relational metadata with a component-based, ontology-driven annotation framework*, 2002. Disponível em <http://www.ai.mit.edu/people/jimmylin/papers/ Handschuh01.pdf>. Acesso em: 25 nov. 2003.

HANNEMANN, J.; THÜRING,M.; HAAKE, J.M. Hypermedia and cognition: designing for comprehension. *Communications of the ACM*, v.38, n.8, Aug. 1995. p. 57-66.

HANNEMANN, J.; THÜRING,M.; HAAKE, J.M. *Hyperdocument presentation*: facing the interface. GMD Report, n.784. Sept. 1993. Disponível em <http://www.citeseer.nj.nec.com/ hannemann93hyperdocument.html>. Acesso em: 19 nov. 2003.

HEINOMEN, Oskari. Optimal multi-paragraph text segmentation by dynamic programming. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 36., 1998, Québec, Canada. *Proceedings…* INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 17., 1998, Québec, Canada. *Proceedings…* Québec, Canada: L' Université de Montréal, 1998. Disponível em <http://citeseer.ist.psu.edu/heinonen98optimal.html> Acesso em: 13 out. 2003

HORNBAEK, K; FROKJAER, E. Reading of eletronic documents: the usability of linear, fisheye, and overview+detail interface. In: ACM CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS,2001, Seattle, WA. *Proceedings...* Seattle, WA, Mar./Apr. 2001. p.293-300.

HOFMAN, R. Van; OOSTENDORP, H. Cognitive effects of a structural overview in a hypertext. *Bristish Journal of Educational Technology*, v.30, n.2, p.129-140, 1999.

INGWERSEN, Peter. Cognitive perspectives of information retrieval interactions : elements of a cognitive IR theory. *Journal of Documentation*, v.52, n.1, p.3-50, Mar. 1996.

INGWERSEN, Peter. *Information retrieval interaction*. London: Taylor Graham, 1992.

INGWERSEN, Peter. Search procedure in the library – analysed from the cognitive point of view. *Journal of Documentation*, v. 38, n. 3, p. 165-191, Sep. 1982.

INGWERSEN, Peter; WORMELL, Irene. Ranganathan in the perspective of advance information retrieval. *Libri,* v.42, n.3, p.184-201, July/Sept. 1992.

ISAKOWITZ T.; STOHR, E. A.; BALASUBRAMANIAN P. A design methodology for structured hypermedia design. *Communications of the ACM*, v.38, n.8, p.34-44, Aug.1995.

INXIGHT Star Tree Software. California: Inxight Software, Inc. 2003. Disponível em <http://www.inxight.com>. Acesso em: 10 out. 2002

JACOB, Elin K.; SHAW, Debora. Sociocognitive perspectives on representation. *Annual Review of Information Science and technology*, v.33, p.131-185, 1998.

KEMP, B.; BUCKNER, K. Towards a hypertext design methodology for end-user. *The New Review of Hypermedia and Multimedia*, v. 5, p.9-27, 1999.

*KINDS OF CONCEPT MAPS.* Disponível em <http://www.classes.aces.uiuc.edu/ACES100/Mind/c-m2.html> 2002. p.1. Acesso em: 5 abr. 2003

KOBASHI, Nair Yumiko. *A Elaboração de informações documentárias: em busca de uma metodologia*. 195p. Tese (Doutorado em Ciência da Comunicação) - ECA/USP, São Paulo, 1994.

KOCH, I. G. V. *Desvendando os segredos do texto.* São Paulo: Cortez Editora, 2002. 168p.

KOEHLER, Wallace. *Concepts*. Disponível em <http.//www.ou.edu/cas/slis/courses/LIS5990A/slis5990/catalog/coordination/concepts.htm>. Acesso em: 06 jun. 2001.

KORTHAUER, R. D.; KOUBEK, R. J. An empirical evaluation of knowledge, cognitive style, and structure upon the performance of a hypertext task. *International Journal of Human-Computer Interaction,* v.6, n.4, p.373-390, 1994.

LAMPING, John; RAO, Ramana; PIROLLI, Peter. *A focus+context technique based on hyperbolic geometry for visualizing large hierarchies.* 1995 Disponível em <http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/jl_bdy.htm>. Acesso em: 6 nov. 2003.

LANCASTER, F.W. *Indexação e resumos*: teoria e prática. Brasília: Briquet de Lemos, 1993.

LANGE, D. An object-oriented design approach for developing hypermedia information systems. *Journal of Organizational Computing and Eletronic Commerce*, v.6, n.3, p.269-293, 1996.

LANZING, J. *What is concept mapping?* March, 1997. Disponível em <http://users.edte.utwente.nl/lazing/cm_home.htm>. Acesso em: 29 out. 2002.

LARA, Isabela. *Tese hipertextual sobre os hipertextos*. 50p. Dissertação (Mestrado em Comunicação). Faculdade de Comunicação, Universidade de Brasília, 2003. Disponível em <http://www.unb.br/fac/ncint/site/index.htm>. Acesso em: 23 set. 2003.

LAWLESS, K.A.; KULIKOWICH, J.M. Domain knowledge, interest and hypertext navigation: a study of individual differences. *Journal of Educational Multimedia and Hypermedia,* v.7, n.1, p.51-69, 1998.

LEE, H.; LEE, C.; YOO, C. A scenario-based object-oriented hypermedia design methodology. *Information & Management*, v.36, p.121-138, 1999.

LEVY, Pierre. *As tecnologias da inteligência*: o futuro do pensamento na era da informática. Rio de Janeiro: Ed.34, 1993. 203p.

*LIBRARY of congress subject heading.* Disponível em <http://geography.miningco.com/library/congress/bllc.htm>. Acesso em: 11 nov. 2003.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: uma revisão da literatura. *Ci. Inf.*, Brasília. v.31, n.1, p.41-51, jan./abr.2002.

LUCAS, Clarinda Rodrigues. A organização do conhecimento e tecnologias da informação: oráculos humanos e inteligência artificial. In: *ORGANIZAÇÃO do conhecimento e sistemas de classificação*. Brasília: IBICT, 1996. p.69-73.

MARCHIONINI, Gary. Designing hypertext; start with an index. In: FIDEL, Raya et al. *Challenges on indexing eletronic text and images*. Medford, NJ: ASIS, 1994. Cap. 4, p.77- 89.

MARQUES, Eugênia Vale. Introdução aos sistemas de hipertexto*. Revista da Escola de Biblioteconomia da UFMG*, v. 24, n.1, p.85-111, jan./jun. 1995.

MARTINS, James. *Hiperdocumentos e como criá-los*. Trad. Marcelo Bernstein. Rio de Janeiro: Campus, 1992. 222 p.

McDONALD, S.; STEVESON, R. Spatial versus conceptual maps as learning tools in hypertext. *Journal of Educational Hypermedia and Multimedia*, v.8, n.1, p.43-64, 1998.

McGARRY, Kevin. *O contexto dinâmico da informação*: uma análise introdutória. Brasília, DF: Briquet de Lemos/Livros, 1999.

McGRATH, Sean. *XML: aplicações práticas.* Ed. Campus, Disponível em <http://www.ebookcult.com.br/ebookzine/linguagem.htm> Acesso em: 23 out. 2003

MILLARD, David E. *Navigating Hyperspaces with Contextual Structured Linking*, 1999. Disponível em <http://citeseer.nj.nec.com/millard99navigating.html> Acesso em: 15 dez. 2003.

MILLS, J.; BROUGHTON, V. *Bliss bibliographic classification.* 2ed. London: Butterworths. 1977.

MOREIRA, Marco Antonio; BUCHWEITZ, Bernardo. *Mapas conceituais: instrumentos didáticos de avaliação e análise de currículo.* São Paulo: Editora Moraes, 1987.

MOREIRA, Marco A., *A Teoria de Educação de Novak e o Modelo de Ensino-Aprendizagem de Gowin.* Fascículos da IFUFRGS, Série Ensino-Aprendizagem, n.4, 1993.

NAVES, Madalena L. N. *Fatores interferentes no processo de análise de assunto*: estudo de caso de indexadores. 254p. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação da UFMG, Belo Horizonte, 2000.

NELSON, Ted. *Project Xanadu.* Disponível em < http://xanadu.com/ >. Acesso em: 13 set. 2003

NOVAK, Joseph D. *The theory underling concept maps and how to construct them*. 11p. Disponível em <http://cmap.coginst.uwf.edu/info>. Acesso em: 29 out. 2002

*ONTONPORTAL Project*. IAM Group, Department of Electronics & Computer Science, University of Southampton, Highfield, Southampton, UK. Disponível em <http://www.ontoportal.org.uk>. Acesso em: out. 2003.

OUTING, Steve. *Does your web site need an index?* 5p. Disponível em <www.mediainfo.com/ephone/news/newshtm/>. Acesso em: 30 out. 1998.

PERLMAN, G. Asychronous design/evaluation methods for hypertext development. IN: ACM CONFERENCE ON HYPERTEXT, Nov. 5-8, 1989. Pittsburgh, Pennsylvania. *Proceedings*... Pittsburgh, Pennsylvania: ACM, 1989. p.61-68.

PIEDADE, Maria Antonieta Requião. *Introdução à teoria da classificação*. 2.ed. rev. e aum. Rio de Janeiro: Interciência, 1983. 221 p.

POLLITT, A. Steven. *Interactive information retrieval based on faceted classification using views*. Disponível em <http./www.hud.ac.uk/school/cedar/dorking.htm>. Acesso em: em 07 out.1997

PRISS, Uta; JACOB, Elin. Utilizing faceted structures for information systems design. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. 62, Oct.31-Nov.4, 1999. Washington, D.C. *Proceedings*... Medford: Learned Information, 1999. p.203-212.

QUINN, Laura S. *Defining Feature Sets Through Prototyping*. November 11, 2002. Disponível em <http://www.boxesandarrows.com/archives/defining_feature_sets_through_prototyping.php>. Acesso em: 15 out. 2003.

RADA, Roy. Focus on links: a holistic view of hypertext. *International Classification*, n.18, v.1, p.13-18, 1991.

RAGHAVAN, K S. The General theory of classification as the basis for structuring of subject headings. In: REGIONAL CONFERENCE OF THE INTERNATIONAL FEDERATION FOR DOCUMENTATION COMMITTEE ON CLASSIFICATION RESEARCH (FID-CR), 2, 15 Nov. 1985. New Delhi. *Proceedings...* New Delhi: Delhi Library Associations, 1985.

RANGANATHAN, S. R. Faceted analysis. In: CHAN, L.M. et al. (Eds.) *Theory of subject analysis*: a sourcebook. Littleton, CO: Libraries Unlimited, 1985. p. 86-93.

RANGANATHAN, S. R. *Prolegomena to library classification*. 3ed. London: Asia Publishing House,1967.

REED, S. K. *Cognition*: Theory and applications. 3ed. Brooks/Cole: Pacific Groce, 1992.

ROSENFELD, Louis; MORVILLE, Peter. *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilley, 1998. 202p.

ROWLEY, J. E. *Knowledge organization*: an introduction to information retrieval. 2 ed. Hampshire, England: Grower, 1992. 510 p.

RYDER, Martin. *Instructional design models*. Disponível em <http://carbon.cudenver.edu/ ~mryder/itc_data/idmodels.html>. Acesso em: 11 dez. 2003

SANTOS, C. A. S. et al., *Um modelo formal para verificação da consistência de documentos hipermídia NCM.*1998. Disponível em <http://www.telemidia.puc-rio.br/english/researches/smh/projects/current/ hyperprop/publications.htm> Acesso em: 24 nov. 2003

SANTOS. Paula Xavier. *Engenharia da informação para sistemas de hipertexto*. 83p. Dissertação (Mestrado em Ciência da Informação) - CNPq/IBICT-URFJ/ECO, Rio de Janeiro: 1996.

SCHWABE, D.; ROSSI, G. The object-oriented hypermedia design model. *Communications of the ACM*, v.38, n.8, p.45-46, 1995.

SEARLE, John. *Minds, brains and science*. Cambridge, MA: Harvard University Press, 1984. 107 p.

SHERRATT, C. S.; SCHLABACH, M. L. The application of concept mapping in reference and information services. *RQ*, v.30, p.60-69, 1990.

SHNEIDERMAN, Ben. Reflections on authoring, editing, and managing hypertext. In: BARRET (Ed.), *The Society of Text*. Cambrige: MIT Press, 1989. p.115-131.

SHNEIDERMAN, Ben; KEARSLEY, Greg. *Hypertext hands-on*: an introduction to a new way of organizing and accessing information. Reading, Addison-Wesley, 1989. 165p.

SIQUEIRA, I. S. Próspero; PEREIRA, Antônio Eduardo da Costa. Perspectivas de aplicação da inteligência artificial à biblioteconomia e à ciência da informação. *R. Bras. Bibliotecon. e Documentação*, São Paulo, v.22, n.1/2, p.39-80, jan./jun.1989.

SMITH, Edward E.; MEDIN, Douglas L. *Categories and concepts*. Cambridge, Massachusetts: Harvard University Press, 1981. 203 p.

SPEZIALI, P. *Classifications of the sciences*. In: DICTIONARY of the history of ideas. New York: Scribners, 1973. p.462-467.

SPITERI, Louise. A Simplified Model for facet analysis: Ranganathan 101. *Canadian Journal of Information and Library Science*, v.23, p.1-30, April/July, 1998.

SONG, Inseok; BAYERL, Petra S. *Semantics of XML documents*. Internal Working Paper, University Gçessen, Germany, April, 2003. Disponível em <http://coli.lili.uni-bielefeld.de/Texttechnologie/ Forschergruppe/publications/inseok03a.pdf>. Acesso em : 21 set. 2003.

STAR, Susan Leigh. *Grounded classification*: grounded theory and faceted classification. 1996. Disponível em <http://alexia.lis.uiuc.edu/~star/gt.html> Acesso em: 13 ago. 2002.

TAROUCO, Liame. *Mapas conceituais*. Disponível em <http//penta2.ufrgs.br/edu/ambvirt/sld058.htm> Acesso em: 5 abr. 2003.

THÜRING, M.; HANNRMANN, J.; HAAKE, J.M. Hypermedia and cognition; designing for comprehension. *Communications of the ACM*, v.38, n.8, p.57-66, Aug. 1995.

UDC Consortium**.** *Classificação Decimal Universal*. Edição-padrão internacional em língua portuguesa. Brasília : IBICT, 1997. 1026p. Parte 1 Tabelas sistemáticas.

UDC Consortium. *Classificação Decimal Universal*. Edição-padrão internacional em língua portuguesa. Brasília : IBICT, 1999. 176p. Parte 2 Índice alfabético.

VERNON, K. G. B. *London classification of business studies*. London: Aslib, 1979

VICKERY, B. C. *Classificação e indexação nas ciências*. Rio de Janeiro: BNG/Brasilart, 1980. p.187-224: Aspectos históricos da classificação da ciência.

VICKERY, B. C. *Faceted classification: A guide to the construction and use of special schemes*. London: Aslib,1960

VICKERY, B. C. *Faceted classification schemes*. New Brunswick, NJ: Graduate School Library Science, Rutgers, the State University, 1966.

VICKERY, B. C. *Classification and indexing in science*. London: Butterworths, 1975.

WENGER, M.J.; PAYNE, D.G. Effects of a graphical browser on readers' efficiency in reading hypertext. *Technical Communications,* v. 41, p.224-233, 1994.

WILSON, T. D. The Work of the British Classification Research Group. In: WELLISCH, Hans; WILSON, Thomas D. (Eds.) *Subject retrieval in the seventies*. Westport, CN: Greenwood. 1972.

VILAN FILHO, Jayme Leiro. *Sistemas hipertexto para microcomputadores: uma aplicação em informação científica e tecnológica*. 105p. Dissertação (Mestrado em Biblioteconomia e Documentação) - Faculdade de Estudos Sociais Aplicados da UNB, Brasília, 1992.

VILAN FILHO, Jayme Leiro. Hipertexto: visão geral de uma nova tecnologia de informação. *Ciência da Informação*, v.23, n.3, p.295-308, set./dez. 1994.

WEINBERG, Bella Hass. *Complexity in indexing systems-abandonment and failure*: implications for organizing the Internet. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE, Oct., 19-24, 1996 Baltimore, MD. *Proceedings...,*.Baltimore, MD: ASIS, 1996. Disponível em <www.asis.org/annual-96/EletronicProceedings/weinberg.html>. Acesso em: 16 mar. 2003.

WOODHEAD, Nigel. *Hypertext and hypermedia*: theory and applications. Wilmslow, England: Sigma Press, 1990. 231 p.

WOUKEU, A. et al. *Rethinking web design models; requirements for addressing the content*. Disponível em <http://wick.ecs.soton.ac.uk/papers/TR-WWW2003-RethinkingWebDesignFinal.pdf> Acesso em: 25 nov. 2003

YUHUI, J. DECKER S., WIEDERHOLD, G. *OntoWebber:* Model-Driven Ontology-Based Web Site management. In: INTERNATIONAL SEMANTIC WEB WORKING SYMPOSIUM,1, July 29-Aug.1, 2001. Standford, CA. *Proceedings...* Standford, CA: Standford University, 2001. 16 p.

# ANEXO 1 – GLOSSÁRIO

**Análise de assunto:** Etapa de extração de conceitos que refletem a essência, o conteúdo de documentos. É considerada a operação base da indexação de assuntos.

**Análise facetada:** Análise de vários aspectos de um conceito para identificar suas características básicas pelas quais ele pode ser dividido em subclasses, é a primeira etapa para o desenvolvimento de um sistema de classificação facetado. (ODLIS: Online Dictionary of Library and Information Science, 2002, http://www.wcsu.edu/library/odlis.html)

**Âncora ou botão:** Jargão da HTML que denota uma marca de referência cruzada entre duas partes de uma página *Web*. Trata-se de um ponto sensível marcado por algum efeito visual (negrito, cores, tipos, etc) ou figura que indica a origem ou o destino de uma ligação, e que, quando acionado, ativa esta ligação.

**Arquitetura de metadados:** *". . .mecanismos que permitem a codificação e o transporte de uma grande variedade de metadados desenvolvidos de forma independente., maximizando a interoperabilidade do sistema através do uso de convenções comuns a respeito da semântica, sintaxe e estrutura do metadado"* (IANELLA, 1998).

**Atinência:** (*Abouteness)* Termo adotado por alguns especialistas da área como tradução do termo em inglês: no entanto, são sugeridas outras traduções. Trata-se da terceira fase do processo de Análise de assunto, a dizer "sobre o que se trata o documento" (NAVES, 2000, p.249).

**Atributos:** o termo "atributo" é utilizado no contexto do XML, para especificar propriedades ou características do elemento.

***Browser*: *"Programa aplicativo que permite pesquisar informações na World Wide Web, carregando documentos, imagens e arquivos correlatos. Também chamado de navegador Web".(***VENETIANER,p.226,1996)

**Cadeia: "***Seqüência formada por classes e seu universo de deslocamento (1, 2, 3 etc.) até um ponto desejado"* (RANGANATHAN, 1967, p.63).

**Característica:** "*A qualidade ou o atributo escolhido para servir de base à classificação ou à divisão."* (PIEDADE, p.17,1983)

**CD-ROM:** Do inglês *Compact Disk with Read Only Memory,* é um disco óptico para leitura, onde dados gravados são recuperados com o emprego de raio laser.

**Ciência cognitiva:** "um esforço contemporâneo, com fundamentação empírica, para responder questões epistemológicas de longa data - principalmente aquelas relativas à natureza do conhecimento, seu desenvolvimento e seu emprego." Gardner (1996, p.19).

**Classe:** “conjunto de coisas ou idéias que possuem um ou vários atributos, predicados ou qualidade em comum”. (PIEDADE, p.19,1983)

**Classificação:** *". . .processo mental de agrupamento de elementos portadores de características comuns e capazes de ser reconhecidos como uma entidade ou conceito, constitui uma das fases fundamentais do pensar humano"* (CAMPOS, 1973, p.15).

**Classificação facetada**:*". . . sistema que agrupa termos estruturados, na base da análise de um assunto, para identificação de suas facetas, isto é, dos diferentes aspectos contidos"* (BARBOSA. 1972, p.74-75).

**Classificar** – Do latim *"classis",* que designa grupos sociais em que se dividia o povo romano. "*Classificar é dividir em grupos ou classes, segundo as diferenças. É dispor os conceitos, segundo as suas semelhanças e diferenças, em certo número de grupos metodicamente distribuídos. É um processo mental habitual do homem, pois vivemos automaticamente classificando coisas e idéias, a fim de as compreender e conhecer".* (PIEDADE, 1983, p.16).

**Coerência** : Fator que se manifesta no nível macrotextual, e refere-se aos conceitos e às relações subjacentes ao texto de superfície.

***Coesão :*** Fator que se manifesta no nível microtextual, e refere-se à maneira como as palavras se ligam entre si, numa seqüência.

**Conceito** : É a representação dum objeto pelo pensamento, por meio de suas características; é a compilação de enunciados verdadeiros sobre um determinado objeto, fixada por um termo lingüístico.

**Conhecimento prévio** : É o conhecimento que o indivíduo vem acumulando ao longo de sua vida. Também chamado conhecimento anterior, conhecimento básico e conhecimento tácito.

**Consistência** : Na indexação, manifesta-se na similaridade dos termos de indexação determinados a um dado documento por diferentes indexadores. Alguns autores utilizam o termo coerência na indexação.

**Contexto** : Descrição do contexto representa o contexto no qual o documento é produzido e para o qual ele existe em determinado momento.

**Documento** : Todo suporte físico que contém informação escrita, seja livro, periódico, base de dados computadorizada, dentre outras.

**Especificidade** : É a extensão permitida por sistemas para se obter precisão na especificação do assunto de um documento.

**Elemento:** componente básico da linguagem XML que um “pedaço” de texto intercalado pelos sinais “< “e “>”.

**Esquemas** : *(Schemata)* São unidades que organizam seqüências de eventos.

**Estruturas** : *(Frames)* Nos domínios do conhecimento complexo, os conceitos são freqüentemente representados por uma série de estruturas associadas

**Exaustividade** : Relaciona-se com a possibilidade de se indexar o tema principal e os sub-temas de um documento, também chamada indexação em profundidade.

**Faceta: "***um termo genérico usado para denotar algum componente- pode ser um assunto básico ou isolado- de um assunto composto, tendo ainda, a função de formar renques, termos e números*." (RANGANATHAN,1967,p.88) ou
*". . . lista de termos mantendo entre si as mesmas amplas relações com a classe que lhes deu origem, ou como um conjunto de termos produzidos pela aplicação de um amplo princípio de divisão.* (PIEDADE,1983, p.)

**Frontpage:** veja *Homepage*.

**Hiperdocumento:** Documento eletrônico construído segundo a filosofia do hipertexto, formado pelo conjunto dados e relacionamentos em forma de texto, diagramas, imagens e som, e que permite que um usuário possa deslocar-se no seu interior realizando consultas personalizadas.

**Hipermídia:***"Hipermídia é o hipertexto enriquecido com outros meios, como imagem e som."* (MARQUES 1995,p.85)

**Hipertexto:** *"Janelas na tela* [que] *são associadas com objetos na base de dados e ligações* [que] *são estabelecidas entre estes objetos, tanto graficamente, na forma de marcas rotuladas, como na base de dados, na forma de ponteiros."* (CONKLIN, 1987,p17)

**Hipertextualidade:"***é uma forma de leitura-escritura que tem por base o texto tradicional e a cognição humana auxiliados sistemas de informações relacionais que aproximam ao máximo de nossos processos cognitivos e representacionais."* (SOUZA,p.1, s.d.)

**Homepage:** primeira página de um *Web sites*, cujo acesso é feito por meio de uma URL na Internet, geralmente contendo um sumário do *site.*

**HTML:** Do inglês *HyperText Markup Language*, é uma linguagem de marcação baseada na linguagem SGML (*Standard Generalized Markup Language*), criada especialmente para codificação dos documentos disponibilizados através dos servidores *Web* e interpretados por navegadores como *Nestscape Navigator, Internet Explorer* etc.

**Indexação** : No contexto do tratamento da informação, o termo possui dois sentidos: um, mais amplo, quando se refere à atividade de criar índices, seja de autor, título, assunto, tanto de publicações (livros, periódicos), quanto de catálogos ou banco de dados, em bibliotecas ou centros de informação. O outro sentido, mais restrito, se refere apenas à indexação ou catalogação de assuntos das informações contidas em documentos.

**Indexação acadêmica**: Formação de índices de assuntos baseados numa determinada coleção de documentos. (FARROW, 1995)

**Indexação de livros**: Trata-se da indexação *back-of-book*, cujo índice é feito para os assuntos contidos num único livro, devendo ser localizado no final da publicação.(FARROW, 1995)

**Indexador**: O profissional da informação que exerce a função de fazer a indexação do assunto contido em documentos, independente do tipo de instituição que atua. Também chamado de catalogador e classificador.

**Indexador experiente**: O indexador com mais de três anos exercendo a atividade de indexação, independente da área de atuação.

**Indexador pouco experiente** : O indexador com experiência de mais de um ano a três anos exercendo a atividade.

**Indexador novato**: O indexador recém-formado, até um ano de experiência.

**Índice:** *"Lista detalhada de assuntos, nomes de pessoas, nomes geográficos, acontecimentos, etc. (entradas) ordenados normalmente por ordem alfabética, com indicação de sua localização na publicação em que aparecem"* (AURÉLIO, 1986). Comumente confundido com o sumário, deve vir após as referências bibliográficas e ser elaborado com base na ABNT 6034 (FRANÇA et al, 1998, p.150)

**Inferência** : É a operação pela qual, utilizando seu conhecimento de mundo, o leitor de um texto estabelece uma relação não explícita entre dois elementos (normalmente frases ou textos) deste texto que ele busca compreender e interpretar.

**Interdisciplinaridade** : Significa a existência de interseção entre disciplinas ou áreas do conhecimento.

**Interface*:*** recurso utilizado pelos computadores para a visualização ilustrada de programas.

**Internet:** é um conjunto de redes de computadores interligadas geograficamente pelo mundo inteiro. Os usuários a ela conectados podem utilizar diversos serviços de comunicação de alcance mundial.

**Interpretação**: Na interpretação de um texto, o leitor dá sentido a esse e realiza a sua atualização, ou seja, captação de frases, compreensão dos fragmentos do texto e relacionamentos entre passagens.

**Intertextualidade**: Relação existente entre o texto e os outros textos derivados desse.

**Isolados**: como *termos que são vistos fora do contexto ou conceitos desorganizados antes de serem reunidos em facetas*.

**Leitura** : Prática de dar sentido perceptivo e intelectual a um texto.

**Leitura para fins documentários** : É o tipo específico de leitura que visa extrair informações de um documento, para inseri-lo num sistema de recuperação de informação. Também denominada leitura técnica.

**Lingüística textual** : Estudo das operações lingüísticas, cognitivas e argumentativas e argumentativas reguladoras e controladoras dos processos de produção, constituição, funcionamento e compreensão dos textos escritos ou orais.

**Linguagem de marcação:** um conjunto de convenções utilizadas para a codificação de textos. Uma linguagem de marcação deve especificar que marcas são permitidas, quais são exigidas, como se deve fazer distinção entre as marcas e o texto e qual o significado da marcação (ALMEIDA, 2002, p.5).

***Link*:** Também chamado de "ligação" ou "elo", é o elemento que estabelece ligações entre os nodos, independente de serem lineares ou não-lineares e hierárquicas ou não-hierárquicas. É representado por um botão ou marca, que conecta os nós entre si, ou partes deles.

**Macroestrutura** : Identificada como a estrutura profunda do texto, é considerada como a "forma lógica" ou estrutura subjacente de um texto. Consiste na representação semântica global que define a significação do texto como um todo (FÁVERO*, 1993)*..

**Mapa do *site*:** Forma visual de descrever a estrutura de um *site* na *Web*. Os mais comuns apresentam interface gráfica ou textual ou uma combinação de ambas.

**Mapa conceitual:** "*uma forma de diagrama especificamente direcionado para fornecer uma linguagem visual parecida com as características da linguagem natural do texto, no sentido de que eles possam estar sujeitos às limitações sintática e semântica, e sua capacidade de representação pode variar de uma forma muito informal a uma forma extremamente formal". (*GAINES e SHAW (2002, p.1)

**Memória episódica**: É o conhecimento privado do indivíduo.

**Memória semântica**: É o conhecimento formal também compartilhado pela educação.

**Memória LTM**: *(Long-Term Memory)* A LTM armazena informações de maneira permanente.

**Memória STM**: *(Short-Term Memory)* É vista como o espaço de trabalho onde a informação é processada, e é muito limitada e transitória.

**Microestrutura**: Identificada como a estrutura superficial do texto, consiste em um n-duplo ordenado de sentenças subseqüentes (FÁVERO, 1993).

**Metadado:** A clássica definição sobre metadados é "informação que descreve informação" ou "dado sobre dado". Metadado é uma informação descritiva tais como autor, titulo, palavra-chave ou outra informação que são associadas ao documento.

**Modelo:** *"Representação abstrata e simplificada de um sistema real, com a qual se pode explicar ou testar o seu comportamento, em um todo ou em partes"* (COUGO, 1997, p.7)

**Modelo conceitual:** Modelo que expressa informações semânticas representando associações entre conceitos dentro de uma área específica e é considerado como a primeira etapa da modelagem de dados.

**Navegação:** é conceituada por como direcionamento das mudanças de foco ou movimentação em relação ao conhecimento disponível na base por meio da utilização de características abstratas, estruturas ou outras ferramentas para orientação geral. (WOODHEAD 1990, p.4)

**Nó:** Também chamado de "nodo", é a unidade de informação de um hiperdocumento que pode conter um ou mais tipos de dados como: texto, figuras, fotos, sons, seqüências animadas, código de programas, e outros. É conectado por ligações numa variedade de estruturas possíveis. Geralmente, descreve um único conceito ou tópico.

**Objeto digital:** Estrutura de dados especificamente projetada para representar uma vasta categoria de objetos em formato digital que são acessados através da *Web* (ARMS, 1997).

**Objeto eletrônico:** Conjunto de objetos de hipertexto, que representam os nós e as ligações de um documento, consistindo de um conteúdo, sua organização e apresentação (THURING, 1991).

**Ontologia:** Modelo formal de conceitos e objetos como são apresentados no mundo real, juntamente com as relações entre eles.

**Palavras-chave:** Termos que expressam algum significado semântico do texto do qual são derivadas.

**Processamento analítico-sintético** : Neste tipo de processamento, os seres humanos explicam o significado de coisas relacionando a seus elementos às estruturas às quais elas pertencem ou às suas propriedades.

**Processamento *bottom-up :*** usa informações que estão no texto, também chamado ascendente, indutivo e perceptivo.

**Processamento *top-down*** : usa informações que não estão no texto, mas que fazem parte do conhecimento prévio do leitor. Também chamado descendente, dedutivo ou conceitual.

**Processos cognitivos**: São atividades mentais como pensamento, imaginação, lembrança e solução de problemas. Como outras atividades humanas, são desempenhadas diferentemente por indivíduos que têm diferentes níveis de

habilidade em raciocínio lógico, memória visual ou experiência em vocabulário, o que pode afetar o desempenho da recuperação da informação.

**Profissional da informação**: bibliotecário que atua como especialista na função de análise da informação, para o que será necessário, além dos conhecimentos na área de biblioteconomia, conhecimento específico do assunto, e que atuará como consultor, em relação à busca e à seleção da informação a ser recuperada.

**Pertinência:** Relação de necessidade entre um sujeito e uma fonte de informação em um determinado momento.

**Precisão:** Extensão com que os itens recuperados durante uma busca numa base de dados são considerados relevantes ou pertinentes. Numericamente, o coeficiente de precisão é calculado dividindo-se o número de itens relevantes (pertinentes) recuperados pelo número total de itens recuperados.

**Psicologia cognitiva** : É uma ciência que lida com os processos e estruturas mentais implicadas na aquisição, processamento e uso de conhecimento ou informação.

**RDF:** *Resource Description Framework*, constitui-se em uma arquitetura genérica de metadados que permite descrever recursos no contexto da *Web*, através de padrões de metadados. È uma tecnologia RDF é recomendada pelo *Word Wide Web Consortium* – 3WC.

**Recuperação da informação:** Conjunto de operações consecutivas, executadas para localizar uma informação necessária ou documentos que a contenham, com a recuperação subseqüente desses documentos.

**Redes semânticas:** são modelos de representação de conhecimento cuja a principal vantagem é a de poderem explicitar as associações entre conceitos.

**Relevância**: No contexto da recuperação da informação, é o julgamento feito pelo indivíduo ao se confrontar com o resultado de sua busca em um sistema de recuperação da informação; seria a avaliação da satisfação do usuário com relação à representação de documentos.

**Renques:** (fileiras, linhas) Do inglês "*array*", é um conjunto de classes derivadas de uma aplicação de uma única característica, ou no mesmo princípio de divisão. (RANGANATHAN, 1967, p.61).

**Semântica:** Estudo da relação de significação nos signos e da representação do sentido dos enunciados.

**Significado**: É aquilo que é representado por um texto; considera-se significado como a representação, na linguagem, do significante, correspondendo o primeiro ao conceito ou à noção, e o segundo à forma.

**Sintaxe** : Parte da gramática que estuda a disposição das palavras na frase e a das frases no discurso, bem como a relação lógica das frases entre si e a correta construção gramatical.

***Site*:** Do jargão da Internet, e também chamado de "sítio", é um espaço virtual na Internet com um endereço específico, administrado por um computador hospedeiro (*host*) com uma página de apresentação WWW.

**Sistema de classificação:** *"Um mapa completo de qualquer área do conhecimento, mostrando todos os seus conceitos e suas relações"* (LANGRIDGE, 1977).

**Storyboard:** Conceito emprestado da área de cinema, é uma estrutura abrangente das páginas de um sistema de hipertexto, com respectivos tópicos, vínculos primários e, às vezes, elementos do *design* gráfico.

**Subjetividade** : Diferentes indivíduos criam diferentes figuras ou idéias de uma mesma informação externa, por causa de suas inclinações pessoais, afetivas e sociais.

**Superestrutura** : Forma global do texto, pode ser descrita em termos de categorias e regras de formação.

***TAG:*** são comandos escritos entre os sinais (<) e maior (>), que indicam como o *browser* deve mostrar o texto.

**Terminologia:** Conjunto de termos utilizados em um campo determinado do conhecimento

.

**Termo** : É um símbolo com que se expressa um conceito, é portador de informação e possui um único significado.

**Texto**: O termo muitas vezes é confundido com documento, informação, dado, conhecimento ou até mesmo como literatura. Neste estudo, texto é considerado como aquele que permite a comunicação de idéias entre o sujeito que cria e dissemina informação e o sujeito que necessita e adquire informação.

**Textualidade** : Conjunto de propriedades que fazem com que um texto seja realmente um texto.

**Tesauro:** *" é uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos lingüísticos que o compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente".* (CURRÁS,1995, p.88)

**Tratamento da Informação**: Conjunto de *". . . todas as disciplinas, técnicas, métodos e processos relativos à descrição física e temática dos documentos num sistema de informação ou de recuperação de informação . . ."*

**Trilha:** uma seqüência de nodos interligados, que pode ser predeterminada pelo autor do hiperdocumento ou determinada pelo próprio usuário.

**URL:** *Uniform Resource Locator*, denominação genética de um endereço lógico na Internet. (VENETIANER,p.233,1996)

***Web*:** Sinônimo abreviado de *World Wide Web*, também representado por WWW e, menos ocasionalmente, por W3. (VENETIANER, p.232,1996)

***Web page*:** Página da *Web*.

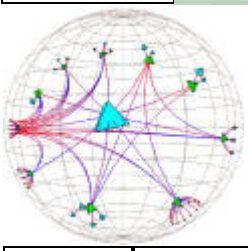
***Web* semântica:** ver Rede Semântica

***Web site*:** Conjunto de páginas da *Web linkadas* a outras páginas da *Web*.

***XFML*:** (eXchangeable *Facet Metadata Language*) é um formato aberto da linguagem *XML* para publicação e compartilhamento de metadados hieráquicos facetados e processos de indexação.

***XML*:** (*eXtensible Markup Language*) é um subconjunto da *SGML* (*Standard Generalized Markup Language*) que permite que uma marcação específica seja criada para especificar idéias e compartilhá-las na rede. Ela tem as virtudes da *SGML* e da *HTML* sem qualquer das limitações óbvias.

## Anexo 2 - Sumário Expandido (SE)



**FATORES INTERFERENTES NO PROCESSO DE ANÁLISE DE ASSUNTO - Sumario**

*source ref: tesemada.html*

- Mapa Conceitual
- Capa / Folha de Rosto
- Agradecimentos
- Lista de Quadros
- Resumo
- Abstract
- 1 Introdução
- 2 O Indexador
  - 2.1 O Profissional da Informação
  - 2.2 O Papel do Indexador
    - 2.2.1 Subjetividade
    - 2.2.2 Conhecimento Prévio
    - 2.2.3 Formação e experiência do indexador
- 3 Indexação
  - 3.1 Consistência e relevância na indexação
  - 3.2 Estudos sobre indexação e desempenho de indexadores
- 4 O Processo de Análise de Assunto
  - 4.1 Fases do processo de Análise de assunto
    - 4.1.1 A Leitura do texto de análise de assunto
    - 4.1.2 Extração de Conceitos
    - 4.1.3 Determinação da Atinência
  - 4.2 A interdisplinaridade em análise de assunto

- 4.2.1 Fatores Linguísticos
- 4.2.2 Fatores Lógicos e Cognitivos

5 Algumas considerações sobre os campos das ciências

- 5.1 Classificação das ciências
- 5.2 O papel da terminologia no meio científico

6 Metodologia

- 6.1 Universo de estudo
- 6.2 Coleta de dados
  - Entrevista semi-estruturada
  - Protocolo verbal

7 Descrição dos resultados

- 7.1 Dados sobre os indexadores
  - Quadro 1 - Formação acadêmica dos indexadores, por ano
  - Quadro 2 - Tempo de experiência dos indexadores
  - Quadro 3 - Síntese das atividades desenvolvidas pelos indexadores
- 7.2 Questões sobre análise de assunto
  - Quadro 4 - Síntese dos requisitos para o indexador fazer análise de assunto
  - Quadro 5 - Procedimentos a serem seguidos na extração do conteúdo de um documento
  - Quadro 6 - Fatores Interferentes no trabalho do indexador
- 7.3 Análise dos protocolos verbais durante análise de assunto dos textos
  - 7.3.1 Tempo gasto com a Análise de assunto dos textos
    - Quadro 7 - Período de tempo dos protocolos verbais
    - Quadro 8 - Tempo de pausas

durante os protocolos verbais

- 7.3.2 Discussão dos fatores identificados durante os protocolos verbais
  - Categoria 1: Fatores relativos ao texto
    - Título do artigo
    - Autoria do Artigo
    - Resumo do artigo
    - Leitura do texto
    - Referências bibliográficas
    - Terminologia
    - Ilustrações
    - Termos definidos como assuntos dos textos
    - Quadro 9 - Termos definidos pelos indexadores para o texto 1 - Sociologia
    - Quaro 10 - Termos definidos pelos indexadores para o texto 2 - Botânica
    - Facilidades e dificuldades para definir o assunto
    - Título do periódico
  - Categoria 2: Fatores relativos ao indexador
    - Conhecimento prévio
    - Formação, experiência e conhecimento na área do texto analisado
    - Orientação de especialistas na área de atuação
    - Interação com o usuário
  - Categoria 3: Fatores observados no

comportamento do indexador durante os protocolos verbais
Medo
Ansiedade
Atenção
Auto-crítica
Preguiça
Quadro 11 - Fatores da categoria 1 identificados na análise do texto 1
Quadro 12 - Fatores da categoria 1 identificados na análise do texto 2
Quadro 13 - Fatores da categoria 3 identificados na análise do texto 1
Quadro 14 - Fatores da categoria 2 identificados na análise do texto 2
Quadro 15 - Fatores da categoria 3 identificados na análise do texto 1
Quadro 16 - Fatores da categoria 3 identificados na análise do texto 2
8 Análise Comparativa dos resultados
Pefil dos indexadores
8.2 Análise comparativa dos fatores interferentes no processo de análise de assunto
Quadro 17 - Fatores interferentes na análise de assunto dos indexadores
8.3 Análise comparativa entre o processo de Análise de assunto de textos das disciplinas Sociologia e Botânica
Quaro 18 - Fatores da categoria 1, do texto 1, por grupos de indexadores

Quadro 19 - Fatores da categoria 1, do texto 2, por grupos de indexadores

Quadro 20 - Fatores da categoria 2, do texto 1, por grupos de indexadores

Quadro 21 - Fatores da categoria 2, do texto 2, por grupos de indexadores

Quadro 22 - Fatores da categoria 3, do texto 1, por grupos de indexadores

Quadro 23 - Fatores da categoria 3, do texto 2, por grupos de indexadores

9 Conclusões e recomendações

10 Referências bibliográficas

Anexos

Anexo 1: Roteiro da entrevista

Anexo 2: Protocolos verbais

Anexo 3: Glossário

Anexo 4: Cópias dos textos de Sociologia e Botânica

## Anexo 3 - Lista de *links* do MHTX: SE e MC com respectivas URLs

| LINKS DO SUMÁRIO | **LINKS DO MAPA CONCEITUAL** |
|---|---|
| PÁGINA PRINCIPAL<br>c:\Madalena\capa\capa.html#capa<br>c:\Madalena\capa\capa.html#folha de rosto<br>c:\Madalena\capa\capa.html#autor<br>c:\Madalena\capa\capa.html#título<br>c:\Madalena\capa\capa.html#orientador<br>c:\Madalena\capa\capa.html#banca avaliadora<br>c:\Madalena\capa\capa.html#linha de pesquisa<br>c:\Madalena\capa\capa.html#local e data<br>c:\Madalena\capa\capa.html#unidade<br>c:\Madalena\capa\capa.html#universidade<br>AGRADECIMENTOS<br>c:\Madalena\agradecimentos\agradecimento.html#agradecimentos<br>LISTA DE QUADROS<br>c:\Madalena\lista_de_quadros\lista_de_quadros.html#lista de quadros<br>RESUMO/ABSTRACT<br>c:\Madalena\resumo\resumo.html#resumo | |
| Cap. 1 - INTRODUÇÃO<br>C:\Madalena\tese_cap1\tese_cap.1html#1 | Ciência da Informação #1<br>Catalogação #1<br>Classificação #1 |
| Cap. 2 - O INDEXADOR | **O Indexador #2** |

| | |
|---|---|
| C:\Madalena\tese_cap2\tese_cap2html#2<br>C:\Madalena\tese_cap2\tese_cap2html#2.1<br>C:\Madalena\tese_cap2\tese_cap2html#2.2<br>C:\Madalena\tese_cap2\tese_cap2html#2.2.1<br>C:\Madalena\tese_cap2\tese_cap2html#2.2.2<br>C:\Madalena\tese_cap2\tese_cap2html#2.2.3 | **Profissional da Informação #2.1**<br>Bibliotecário #2.1.1<br>**O papel do indexador #2.2**<br>Indexador #2.2<br>**Subjetividade #2.2.1**<br>**Conhecimento prévio # 2.2.2**<br>**Formação e experiência do indexador #2.2.3**<br>Prática #2.2.3.1<br>Especialização #2.2.3.2<br>Treinamento #2.2.3.3 |
| Cap. 3 – A INDEXAÇÃO<br>C:\Madalena\tese_cap3\tese_cap3html#3<br>C:\Madalena\tese_cap3\tese_cap3html#3.1<br>C:\Madalena\tese_cap3\tese_cap3html#3.2 | Indexação #3<br>Documento #3.0<br>Tradução para um vocabulário controlado #3.0<br>Consistência e relevância na indexação #3.1<br>Consistência #3.1<br>Relevância #3.1.1<br>Estudos sobre indexação e desempenho de indexadores #3.2 |
| Cap. 4 – O PROCESSO DE ANÁLISE DE ASSUNTO<br>C:\Madalena\tese_cap4\tese_cap4html#4<br>C:\Madalena\tese_cap4\tese_cap4html#4.1<br>C:\Madalena\tese_cap4\tese_cap4html#4.1.1<br>C:\Madalena\tese_cap4\tese_cap4html#4.1.2<br>C:\Madalena\tese_cap4\tese_cap4html#4.1.3<br>C:\Madalena\tese_cap4\tese_cap4html#4.2<br>C:\Madalena\tese_cap4\tese_cap4html#4.2.1<br>C:\Madalena\tese_cap4\tese_cap4html#4.2.2 | O Processo de análise de assunto # 4<br>Análise de assunto #4.0<br>Fases do processo de análise de assunto #4.1<br>A leitura do texto pelo indexador #4.1.1<br>Texto # 4.1.1.1<br>Microestrutura #4.1.1.1.1<br>Macroestrutura #4.1.1.1.1<br>Superestrutura #4.1.1.1.1<br>Narrativo #4.1.1.1.2<br>Informativo #4.1.1.1.2<br>Autores #4.1.1.1.3<br>Coesão #4.1.1.1.4<br>Coerência #4.1.1.1.5<br>Intencionalidade #4.1.1.1.6<br>Aceitabilidade #4.1.1.1.6<br>Situacionalidade #4.1.1.1.6<br>Informatividade #4.1.1.1.6<br>Intertextualidade #4.1.1.1.7<br>Texto primário #4.1.1.1.8<br>Texto secundário #4.1.1.1.8<br>Hipertexto #4.1.1.1.8<br>Leitura #4.1.1.2<br>Cognitiva #4.1.1.2.1<br>Metacognitiva #4.1.1.2.1<br>Extração de conceitos #4.1.2<br>Conceito #4.1.2.1<br>Processos mentais #4.1.2.2<br>Análise #4.1.2.2<br>Síntese #4.1.2.2<br>Abstração #4.1.2.2<br>Generalização #4.1.2.2<br>Percepção #4.1.2.3<br>Emoção #4.1.2.3<br>Atitude #4.1.2.3<br>Assunto #4.1.2.4 |

| | |
|---|---|
| | Atinência #4.1.3<br>Intrínseca #4.1.3.1<br>Extrínseca #4.1.3.1<br>Significado #4.1.3.2<br>Semântica #4.1.3.3<br>A interdicisplinaridade em Análise de Assunto #4.2<br>Interdisciplinaridade #4.2<br>Fatores lingüísticos #4.2.1<br>Lingüística #4.2.1.1<br>Sintaxe #4.2.1.2<br>Fatores lógicos e cognitivos #4.2.2<br>Lógica #4.2.2.1<br>Raciocínio #4.2.2.2<br>Dedutivo #4.2.2.2<br>Indutivo #4.2.2.2<br>Ciência Cognitiva #4.2.3<br>Psicologia Cognitiva #4.2.3.1<br>Assimilação #4.2.3.2<br>Bottom-up #4.2.3.3<br>Top-down #4.2.3.4<br>Interativo #4.2.3.5<br>Compreensão #4.2.3.6<br>Interpretação #4.2.3.7<br>Inferência #4.2.3.8<br>Pensamento #4.2.3.9<br>Mente #4.2.3.10<br>Memória #4.2.3.11<br>Long-Term memory #4.2.3.12<br>Short-Term Memory #4.2.3.12 |
| Cap. 5 – ALGUMAS CONSIDERAÇÕES SOBRE OS CAMPOS DAS CIÊNCIAS<br>C:\Madalena\tese_cap5\tese_cap5html#5<br>C:\Madalena\tese_cap5\tese_cap5html#5.1<br>C:\Madalena\tese_cap5\tese_cap5html#5.2 | Algumas considerações sobre os campos das ciências #5<br>Ciência #5.0<br>Classificação das ciências #5.1<br>Ciências formais #5.1.1<br>Ciências reais #5.1.1<br>Ciências reais naturais #5.1.1<br>Ciências exatas #5.1.1<br>Ciências biológicas #5.1.1<br>Botânica #5.1.1.2<br>Ciências reais humanas #5.1.1<br>Ciências sociais #5.1.1<br>Sociologia #5.1.1.1<br>Ciência da Informação #1.0<br>Ciências humanas #5.1.1<br>Lingüística #4.2.1.1<br>Lógica #4.2.2.1<br>Ciências cognitivas #4.2.3<br>Psicologia cognitiva #4.2.3.1<br>O papel da terminologia no meio científico # 5.2<br>Terminologia #5.2.1 |
| Cap. 6 – METODOLOGIA<br>C:\Madalena\tese_cap6\tese_cap6html#6<br>C:\Madalena\tese_cap6\tese_cap6html#6.1<br>c:\Madalena\tese_cap6\tese_cap6.html#6.1.<br>c:\Madalena\tese_cap6\tese_cap6.html#6.1.<br>c:\Madalena\tese_cap6\tese_cap6.html#6.1. | |

| | |
|---|---|
| c:\Madalena\tese_cap6\tese_cap6.html#6.1.<br>C:\Madalena\tese_cap6\tese_cap6html#6.2<br>c:\Madalena\tese_cap6\tese_cap6.html#6.2.<br>c:\Madalena\tese_cap6\tese_cap6.html#6.2.1 | |

| | |
|---|---|
| Cap. 7 – DESCRIÇÃO E DISCUSSÃO DOS RESULTADOS<br>C:\Madalena\tese_cap7\tese_cap7html#7<br>C:\Madalena\tese_cap7\tese_cap7html#7.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.1.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.1.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.1.3<br>C:\Madalena\tese_cap7\tese_cap7html#7.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.2.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.2.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.2.3<br>C:\Madalena\tese_cap7\tese_cap7html#7.3<br>C:\Madalena\tese_cap7\tese_cap7html#7.3.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.1.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.1.2<br>C:\Madalena\tese_cap7\tese_cap7html#7.3.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.3<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.4<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.5<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.6<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.7<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.8<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.9<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.10<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.11<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.1.12<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.2.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.2.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.2.3<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.1<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.2<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.3<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.4<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.5<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.6<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.7<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.8<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.9<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.10<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.2.3.11<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.3<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.4<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.5<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.5<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.6<br>c:\Madalena\tese_cap7\tese_cap7.html#7.3.7 | |
| Cap. 8 – ANÁLISE COMPARATIVA DOS RESULTADOS | |

| | |
|---|---|
| C:\Madalena\tese_cap8\tese_cap8html#8<br>C:\Madalena\tese_cap8\tese_cap8html#8.1<br>C:\Madalena\tese_cap8\tese_cap8html#8.2<br>c:\Madalena\tese_cap8\tese_cap8.html#8.2.1<br>C:\Madalena\tese_cap8\tese_cap8html#8.3<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.1<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.2<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.3<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.4<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.5<br>c:\Madalena\tese_cap8\tese_cap8.html#8.3.6 | |
| Cap. 9 – CONCLUSÕES E RECOMENDAÇÕES<br>C:\Madalena\tese_cap9\tese_cap9html#9 | |
| Cap. 10 – REFERÊNCIAS BIBLIOGRÁFICAS<br>C:\Madalena\tese_cap10\tese_cap10html#10 | |
| ANEXOS<br>c:\Madalena\Anexos\Anexo.html#anexo | |
| ANEXO 1 -ROTEIRO<br>c:\Madalena\Anexos\Anexo1html#anexo1 | |
| ANEXO 2 – PROTOCOLO VERBAIS<br>c:\Madalena\Anexos\Anexo2html#anexo2 | |
| ANEXO 3 – GLOSSÁRIO<br>c:\Madalena\Anexos\Anexo3html#anexo3 | Indexador experiente #an3.1<br>Indexador pouco experiente #an3.1<br>Indexador novato #an3.1 |
| ANEXO 4 – CÓPIAS DOS TEXTOS<br>c:\Madalena\Anexos\Anexo4html#anexo4 | |

# Anexo 4 - Modelo Simplificado para Análise Facetada de Spiteri

## PRINCIPLES FOR THE IDEA PLANE

**1. Principles for choice of facets**

a) Principle of Differentiation
b) Principle of Relevance
c) Principle of Ascertainability
d) Principle of Permanence
e) Principle of Homogeneity
f) Principle of Mutual Exclusivity
g) Principle of Fundamental Categories

**2. Principles for citation order of facets and foci**

a) Principle of Relevant Succession
   I. Chronological Order
   II. Alphabetical Order
   III. Spatial/Geometric Order
   IV. Simple to Complex Order
   V. Complex to Simple Order
   VI. Canonical Order
   VII. Increasing Quantity
   VIII. Decreasing Quantity

b) Principles of consistent succession

## PRINCIPLES FOR THE VERBAL PLANE

1. Principle of Context
2. Principle of Currency

## PRINCIPLES FOR THE NOTATIONAL PLANE

1. Principle of Synonym
2. Principle of Homonym
3. Principle of Hospitality
4. Principle of Filing Order

# Anexo 5 - Cânones da Teoria da Análise Facetada de Ranganathan

## CANONS FOR THE IDEA PLANE

1 **Canons for Characteristics**
   a) Canon of Differentiation
   b) Canon of Relevance
   c) Canon of Ascertainability
   d) Canon of Permanence

2 **Canons for Succession of Characteristics**
   a) Canon of Concomitance
   b) Canon of Relevant Succession
   c) Canon of Consistent Succession

3 **Canons for Array**
   a) Canon of Exhaustiveness
   b) Canon of Exclusiveness
   c) Canon of Helpful Sequence

4. **Canons for Chain**
   a) Canon of Decreasing Extension
   b) Canon of Modulation
   d) Canon of Consistent Sequence

5. **Canons for Filiatory Sequence**
   a) Canon of Subordinate Classes
   b) Canon of Coordinate Classes

## CANONS FOR THE VERBAL PLANE

1. Canon of Context
2. Canon of Enumeration
3. Canon of Currency
4. Canon of Reticence

## CANONS FOR THE NOTATIONAL PLANE

1. Canon of Synonym
2. Canon of Homonym
3. Canon of Relativity
4. Canon of Uniformity
5. Canon of Hierarchy
6. Canon of Non-Hierarchy
7. Canon of Mixed Base
8. Canon of Pure Base
9. Canon of Faceted Notation
10. Canon of Non-Faceted Notation
11. Canon of Co-Extensiveness
12. Canon of Under-Extensiveness

## CANONS OF MNEMONICS

1. Canon of Alphabetical Mnemonics
2. Canon of Systematic Mnemonics
3. Canon of Seminal Mnemonics

**PRINCIPLES FOR HELPFUL SEQUENCE**

1. Principle of Later-in-Time
2. Principle of Later-in-Evolution
3. Principles of Spatial Contiguity
   a) Principle of Bottom Upwards
   b) Principle of Top Downwards
   c) Principle of Left to Right
   d) Principle of Clockwise Direction
   e) Principle of Counter-Clockwise Direction
   f) Principle of Periphery to Centre
   g) Principle of Centre to Periphery
4. Principles of Quantitative Measure
   a) Principle of Increasing Quantity
   b) Principle of Decreasing Quantity
5. Principle of Increasing Complexity
6. Principle of Canonical Sequence
7. Principle of Literary Warrant
8. Principle of Alphabetical Sequence

**POSTULATES**

1. Postulate of Five Fundamental Categories
2. Postulate of Basic Facet
3. Postulate of Isolate Facet
4. Postulates for Rounds of Manifestation
   a) Postulate of Rounds for Energy
   b) Postulate of Rounds for Personality and Matter
   c) Postulate of Rounds for Space and Time
5. Postulates for Levels of Manifestation
   a) Postulate of Level
6. Postulates for Facets
   a) Postulate of First Facet
   b) Postulate of Concreteness
   c) Postulate of Facet Sequence Within a Round
   d) Postulate of Facet Sequence Within the Last Round
   e) Postulate of a Level Cluster

**PRINCIPLES FOR FACET SEQUENCE**

1. Wall-Picture Sequence
2. Whole-Organ Principle
3. Cow-Calf Principle
4. Actand-Action-Actor-Tool Principle

# Anexo 6 - Princípios da Teoria da Análise Facetada do CRG

**PRINCIPLES FOR THE CHOICE OF FACETS**

1. Principle of Division
2. Principle of Homogeneity
3. Principle of Mutual Exclusivity
4. Principle of Relevance
5. Principle of Ascertainability
6. Principle of Permanence
7. Principle of Fundamental Categories

**PRINCIPLES FOR CITATION ORDER**

1. Schedule Order
2. Order in Array

**PRINCIPLES FOR NOTATION**

1. Filing Order
2. Hospitality
3. Ordinal/Expressive Notation