

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

Benildes Coura Moreira dos Santos Maculan

**TAXONOMIA FACETADA NAVEGACIONAL:  
construção a partir de uma matriz categorial  
para trabalhos acadêmicos**

**Belo Horizonte  
Junho, 2011**

Benildes Coura Moreira dos Santos Maculan

**TAXONOMIA FACETADA NAVEGACIONAL:  
construção a partir de uma matriz categorial  
para trabalhos acadêmicos**

Dissertação apresentada para exame da banca do Programa de Pós-Graduação em Ciência da Informação, da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Mestre em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação

Orientadora: Gercina Ângela Borém de Oliveira Lima

**Belo Horizonte  
Junho, 2011**

M175t  
2011 Maculan, Benildes Coura Moreira dos Santos  
Taxonomia facetada navegacional : construção a partir de uma matriz categorial para trabalhos acadêmicos [manuscrito]. / Benildes Coura Moreira dos Santos Maculan. – 2011.  
191f. : enc.

Orientadora: Gercina Ângela Borém de Oliveira Lima.  
Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.  
Bibliografia: f. 141

1. Organização da informação. 2. Recuperação da informação. 3. Bibliotecas digitais. 4. Ferramentas de busca. 5. Biblioteca digital de teses e dissertações. 6. Taxonomia facetada navegacional. 7. Navegação facetada. 8. Matriz categorial para trabalhos acadêmicos. 9. Teoria da análise do domínio. 10. Análise categorial temática. I. Lima, Gercina Ângela Borém de Oliveira. II. Universidade Federal de Minas Gerais, Escola de Ciência da Informação. III. Título.

CDU: 025.49:004.55



UFMG

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

**FOLHA DE APROVAÇÃO**

"TAXONOMIA FACETADA NAVEGACIONAL: CONSTRUÇÃO A PARTIR DE UMA  
MATRIZ CATEGORIAL PARA TRABALHOS ACADÊMICOS"

Benildes Coura Moreira dos Santos Maculan

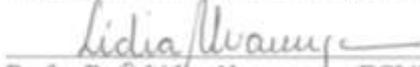
Dissertação submetida à Banca Examinadora, designada pelo Colegiado do  
Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de  
Minas Gerais, como parte dos requisitos à obtenção do título de **"Mestre em  
Ciência da Informação"**, Linha de Pesquisa: **"Organização e Uso da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

Dissertação aprovada em: 20 de junho de 2011.

Por:

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

  
\_\_\_\_\_  
Prof. Dr. Eduardo José Wense Dias - Prof. Aposentado - ECI/UFMG

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

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

  
\_\_\_\_\_  
Prof. Dra. Cristina Dotta Ortega - ECI/UFMG

Aprovada pelo Colegiado do PPGCI

Versão final aprovada por

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

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



UFMG

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

ATA DA DEFESA DE DISSERTAÇÃO DE **BENILDES COURA MOREIRA DOS SANTOS MACULAN**, matrícula: 2009653542

As 14:00 horas do dia 20 de junho de 2011, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação em 23/05/2011, para julgar, em exame final, o trabalho intitulado *Taxonomia facetada navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos*, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, Área de Concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação - OUI. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Gercina Ângela Borém de Oliveira Lima, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

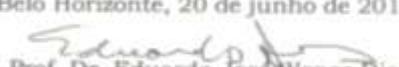
Profa. Dra. Gercina Ângela Borém de Oliveira Lima - Orientadora	APROVADA
Prof. Dr. Eduardo José Wense Dias	APROVADA
Profa. Dra. Lídia Alvarenga	APROVADA
Prof. Dr. Maurício Barcellos Almeida	APROVADA
Profa. Dra. Cristina Dotta Ortega	APROVADA

Pelas indicações, a candidata foi considerada APROVADA.

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

Belo Horizonte, 20 de junho de 2011

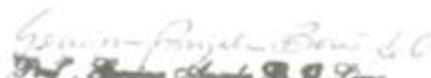
  
Profa. Dra. Gercina Ângela Borém de Oliveira Lima  
ECI/UFMG (orientadora)

  
Prof. Dr. Eduardo José Wense Dias  
Prof. Aposentado - ECI/UFMG

  
Profa. Dra. Lídia Alvarenga  
ECI/UFMG

  
Prof. Dr. Maurício Barcellos Almeida  
ECI/UFMG

  
Profa. Dra. Cristina Dotta Ortega  
ECI/UFMG

  
Prof. Gercina Ângela B. O. Lima  
Coordenadora do Programa de Pós-Graduação  
em Ciência da Informação - ECI/UFMG

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

Dedico este trabalho  
ao meu filho Sandro,  
que me encanta,  
que entende o caos de meus pensamentos e ideias,  
ordenando-os,  
que compartilha minhas dores e muitas alegrias:  
“te amo, filhinho”.

Dedico, ainda,  
a meu pai (*in memoriam*)  
em quem sempre me espelhei.

## **AGRADECIMENTO ESPECIAL**

Agradeço especialmente à professora Gercina A. B. de O. Lima, que inicialmente foi orientadora, depois amiga, e, hoje, é um gostoso misto disso tudo.

Sou sua orientada desde 2007, ainda na iniciação científica, e, desde então, venho aprendendo muito com ela, minha eterna Mestre.

Em um dado momento, tornamo-nos amigas...

Como amiga, sentiu minhas dores... Também riu de minhas bobagens!

E nos divertimos bastante...

E trabalhamos muito!

Então, meu agradecimento é para a amiga e para a orientadora...

Como amiga esteve sempre por perto, efetiva e afetivamente entendendo o que passei no mestrado, e sendo solidária em todos os momentos.

Já como orientadora, se manteve firme nas exigências, mas constantemente paciente, encorajadora e generosa.

E sempre disposta a tirar o melhor de mim.

Nossa convivência é composta de liberdade, empatia e comprometimento.

Obrigada Gessy, pois você é especial para mim e abençoada pelos Anjos!

Obrigada, por me ensinar!

## AGRADECIMENTOS

Não me dêem um espaço para escrever livremente... eu aproveito!

À *Deus Superior*, que me conforta, ilumina e a quem sou grata pelo muito que tenho, ainda que, às vezes, não entenda seus desígnios. Agradeço a *Ele* por me fazer otimista!

Aos meus pais, de quem me orgulho e dos quais aprendi importantes lições (e ainda aprendo!).

Ao meu filho Sandro, com quem tenho muitas afinidades e admiro a inteligência, lealdade e companheirismo. Se chegar a um impasse, sei que posso contar com ele e com sua clareza de ideias. Sua contribuição neste estudo foi inestimável, principalmente nos últimos momentos. Obrigada, filhinho!

À Patrícia, querida mediadora, e Serginho, meus filhos por afeição. E ao Xande, por toda a atenção e consideração, o que não tem preço.

À Tininha, mais que uma estimada e querida irmã, uma filha por afeto. E ao Wagner, pela amizade e carinho sem medida.

À Libéria, meu anjo da guarda, que cuida de mim, principalmente nos últimos seis anos.

À minha família, tão numerosa que seria impossível citar: irmãos, irmãs, cunhados e cunhadas, com os quais festejo e decomponho tristezas.

À minha irmã Isabel, especialmente, que também leu este trabalho.

Aos sobrinhos e sobrinhas, suas namoradas e namorados, e queridos agregados! Adoro a todos vocês, ao que compartilhamos e a nossas saídas juntos!

Aos amigos de infância, dos tempos memoráveis em *Monlé*.

À turma de amigos do Sandro, aos quais verdadeiramente estimo, valorizo e aprecio. Adoro a companhia e convívio!

À Noca, amiga que sempre cuida de mim.

Ao marco ECI em minha vida, na graduação em Biblioteconomia, em 2005, que me fez renascer. Encontrei amigos, desafios e um mundo acadêmico desconhecido, que me salvou! Na trajetória até aqui, muitas pessoas foram importantes: obrigada! Estão todos guardados em meu coração.

À profa. Marlene, primeira incentivadora, acertando-me com palavras animadoras.

À profa. Beatriz, primeira orientadora: fui sua monitora e ela me incentivou a querer mais.

À profa. Adriana Bogliolo que estimulou-me a escrever o primeiro artigo.

Aos professores que instigaram-me em sala de aula: Adriana, Alcenir, Antônio, Bax, Beatriz, Bernadete, Casal, Cláudia, Devanir, Eduardo, Gercina, Guiomar, Helena, Leandro, Lídia, Márcia Milton, Marlene, Marta Araújo, Marta Aun, Maurício, Mônica, Oswaldo, Paulo da Terra, Renato, Ricardo, Vander e Vera.

Aos professores de fora de sala de aula, dos quais me orgulho: Ana Maria, Cida Moura, Conceição, Cristina, Jorge Tadeu, Júlia, Marília, Marta Pinheiro, Lígia e Rose.

Aos professores Alcenir, Bernadete, Casal, Cristina, Guiomar, Mônica e Paulo da Terra, pelo carinho especial.

Aos funcionários e colaboradores da ECI, sobretudo a Claudinha, Lucimeire, Luiz Henrique, Roberto, Sônia, Walisson.

Aos eternos amigos da sala dos bolsistas, com os quais aprendi demais, especialmente Aline, Aninha e Lorena.

À Carol Cota, estimada amiga, que esteve comigo toda a graduação e que é sempre generosa comigo. Que leu, palpitou e normalizou este trabalho. Que me aceita, implica, às vezes, e a quem abuso com meus pedidos de última hora: e ela nunca falha! Vale ouro!

Aos colegas de graduação, pela cumplicidade desse tempo inesquecível, em especial ao *Diplofundão*, Cris, Ju, Michella e Raquel.

Aos colegas de pós-graduação, pelo conhecimento compartilhado.

À Graciane, querida amiga, tão determinada e disciplinada! Alma boa!

À Fernanda, valiosa amiga, de todas as horas, pelos momentos divididos e pelas alegrias multiplicadas. *Ta-ta-de-mais!*

À Érica Sarsur, pelas correções, pontuações e melhoras linguísticas neste texto.

Aos colegas do MHTX, entre bolsistas e orientandos, pelo enriquecimento advindo das calorosas discussões, pela ajuda dos bolsistas e, também, pela saudosa *MHTXroska*.

À ajuda inestimável de Patrícia Penido.

Ao Flávio Pontes, por sua grande contribuição nesta pesquisa, pelas *figurinhas* trocadas, por sua seriedade, dedicação e comprometimento.

Às professoras Cristina, Lídia e Madalena, pelas valiosas contribuições na etapa de qualificação.

Aos amigos da secretaria do PPGCI, Célio, Gisele e Nely, pela competência do trabalho e pelo constante apoio administrativo e no atendimento de excelência. Vocês são demais!

Aos amigos da Biblioteca Etelvina Lima, especialmente à Elaine e Eliane, pelo suporte, sobretudo no momento de levantamento de dados deste trabalho.

À CAPES, pela contribuição financeira durante o desenvolvimento deste estudo.

Aos professores, Eduardo Wense, Cristina Ortega, Lídia Alvarenga e Maurício Almeida, que me honraram ao aceitar meu convite para compor a Banca de Defesa desta Dissertação.

*The value of knowledge is largely tied to the way in which that knowledge is organized. If you can't find it, it's not likely to be of much use to you.*

*Marc Rapport*

*Unfolding Knowledge  
Knowledge Management E-zine*

## RESUMO

Esse trabalho investigou o uso da taxonomia facetada como instrumento de organização da informação, para dar consistência e padronização aos conteúdos informacionais de trabalhos acadêmicos. A finalidade foi facilitar o acesso às informações, apresentando ao usuário a opção de uma navegação pelos termos da estrutura facetada, apresentados em uma interface de busca. O ambiente de pesquisa foi a biblioteca digital de teses e dissertações (BDTD), da Escola de Ciência da Informação (ECI), da Universidade Federal de Minas Gerais (UFMG), tendo como insumo as teses e dissertações oriundas da linha de pesquisa Organização e Uso da Informação. O objeto empírico foram as dez categorias fundamentais temáticas e os termos que as alimentaram, procedentes da análise conceitual realizada nos resumos informativos dos trabalhos acadêmicos do *corpus*. O objetivo norteador deste estudo foi criar uma taxonomia facetada navegacional, denominada TAFNAVEGA, a partir de uma matriz categorial desenvolvida para trabalhos acadêmicos, visando demonstrar a sua adequação para a representação sistematizada e semântica do conteúdo de teses e dissertações. Pretendeu, ainda, contribuir para maior visibilidade do conteúdo das pesquisas científicas e facilitar a busca e a recuperação dessas informações. A metodologia envolveu procedimentos de natureza empírica, exploratória, com uma abordagem qualitativa em relação à análise da literatura especializada da área. Além disso, recorreu ao método de pesquisa aplicada, visto que ofereceu a criação de um mecanismo para ser empregado na navegação facetada. Esse percurso metodológico incluiu o uso da teoria da análise de domínio para a identificação do domínio, do usuário da BDTD e de suas necessidades informacionais, assim como do Modelo de Leitura Documentária, de Fujita e Rubi (2006), para o desenvolvimento da matriz categorial para trabalhos acadêmicos. Para a criação das categorias fundamentais temáticas foi utilizado o método da análise de conteúdo, com a técnica da análise categorial temática, juntamente com os princípios da teoria da análise facetada de Ranganathan. A revisão de literatura foi pautada em trabalhos sobre navegações facetadas através de taxonomias também facetadas, utilizadas como instrumento facilitador da exploração, busca e recuperação de informações em diferentes bibliotecas digitais. Ademais, a fundamentação se baseou em conhecimentos relacionados com a busca e a recuperação em sistemas de recuperação da informação. Abordou, ainda, o vocabulário controlado do tipo taxonomia, as bibliotecas digitais, focando a atenção em bibliotecas de teses e dissertações, e a navegação hipertextual, mostrando a navegação pela taxonomia facetada. Os resultados demonstraram que o mecanismo TAFNAVEGA facilitou a tarefa de exploração, busca e recuperação de conteúdos dos documentos, dando acesso a dados tais como teorias, métodos e instrumentos de coleta de dados, possibilitando combinar essas diferentes informações. Conclui-se que a possibilidade de refinamento da busca através da estrutura da taxonomia facetada navegacional auxilia o usuário a explorar os resultados das pesquisas concluídas, uma vez que dá maior visibilidade ao conteúdo disponível na biblioteca digital de teses e dissertações, sem sobrecarregá-lo de informações.

**Palavras-chave:** Taxonomia facetada navegacional. Navegação facetada. Busca e recuperação da informação. Biblioteca digital de teses e dissertações. Matriz categorial para trabalhos acadêmicos. Teoria da análise de domínio. Análise categorial temática.

## ABSTRACT

This dissertation studies the use of faceted taxonomy as an instrument of organization of information in order to give consistency and standardization to informational contents of university papers. Its aim was to facilitate the access to information, offering to users the possibility to navigate through the terms of the faceted structure, presented in a search interface. We had as input thesis and dissertations which followed the line of research *Organization and Use of Information* of the digital library of thesis and dissertations (BDTD) of the Information Science College (ECI) of the Federal University of Minas Gerais (UFMG). The empirical object were the ten fundamental thematic categories and the terms which fed them, coming from the conceptual analysis carried out in the information summaries of the university papers of the *corpus*. The objective of this study was also to make up a navigating faceted taxonomy, called TAFNAVEGA, from a category matrix developed for university works, aiming to demonstrate its adequacy to the systematized and semantic representation of the content of thesis and dissertations. In addition to that, this work intends to contribute for more visibility of the content of scientific researches, and to facilitate the search and retrieval of this information. The methodology encompasses procedures of exploratory and empirical nature, with a qualitative approach in relation to the analysis of the literature of that field. Besides, it made use of the method of applied research, inasmuch as it showed a mechanism to be used in faceted navigation. This methodological route included the use of the theory of domain analysis to identify the domain, the BDTD user, and his information needs, as well as the Model of Documentary Reading, by Fujita and Rubi (2006) for the development of the category matrix for university works. For the creation of the fundamental thematic categories were used the method of content analysis, the technique of the thematic category analysis, together with the principles of the theory of faceted analysis of Ranganathan. The literature review was based on papers about faceted navigation through faceted taxonomies, used as instrument to explore, search and retrieve information in different digital libraries. The grounding based itself on knowledge related to search and retrieval in information retrieval systems, approaching the vocabulary of the taxonomy type, the digital libraries, focusing on thesis and dissertation libraries, and the hypertextual navigation, showing this through faceted taxonomy. The results demonstrated that the mechanism TAFNAVEGA facilitated the task of exploration, search and retrieval of content of documents, giving access to data such as theories, methods and instruments of data gathering, enabling the combination of these different pieces of information. We conclude that the possibility of search improvement through the structure of the navigating faceted taxonomy helps the user to explore the results of finished researches, inasmuch as it provides more visibility to the content available in the digital library of thesis and dissertations, without overloading it with information.

**Keywords:** Navigating faceted taxonomy. Faceted navigation. Search and retrieval of information. Digital library of thesis and dissertations. Category matrix for university papers. Theory of domain analysis. Thematic category analysis.

## LISTA DE FIGURAS

FIGURA 1 – Grau de complexibilidade em diferentes vocabulários controlados .....	45
FIGURA 2 – Esquema da “Árvore de Porfírio” .....	49
FIGURA 3 – Ilustrações de Árvore Baniana .....	50
FIGURA 4 – Exemplo de taxonomia navegacional.....	57
FIGURA 5 – O sujeito na Análise de Domínio.....	77
FIGURA 6 – Processo de extração de conceitos dos resumos informativos do <i>corpus</i> .....	126
FIGURA 7 – Envolvimento do pesquisador com respondentes .....	130
FIGURA 8 – Tabela para elaboração de resumo informativo .....	131
FIGURA 9 – Tabela de categorias fundamentais .....	134
FIGURA 10 – Campos e relacionamentos entre facetas .....	135
FIGURA 11 – Relações entre documentos do <i>corpus</i> .....	136
FIGURA 12 – Relação entre categorias .....	136
FIGURA 13 – Relação direta entre categorias e documentos .....	137
FIGURA 14 – Relação expandida entre categorias e documentos.....	138
FIGURA 15 – Resultado das interrelações disponíveis para recuperação .....	138
FIGURA 16 – Busca usando a TAFNAVEGA (1) .....	139
FIGURA 17 – Busca usando a TAFNAVEGA (2) .....	140
FIGURA 18 – Busca usando a TAFNAVEGA (3) .....	141
FIGURA 19 – Envolvimento do pesquisador com respondentes .....	189

## LISTA DE GRÁFICOS

GRÁFICO 1 – Percentual de preenchimento das CAFTE (resumos do <i>corpus</i> ).....	127
GRÁFICO 2 – Problemas em resumos informativos do <i>corpus</i> da pesquisa.....	128
GRÁFICO 3 – Relação dos temas abordados no <i>corpus</i> .....	184
GRÁFICO 4 – Relação dos objetos empíricos utilizados no <i>corpus</i> .....	186
GRÁFICO 5 – Métodos utilizados nas pesquisas do <i>corpus</i> .....	187
GRÁFICO 6 – Tipos das pesquisas referentes ao <i>corpus</i> .....	188
GRÁFICO 7 – Tipos de instrumentos de coleta de dados .....	188

## LISTA DE QUADROS

QUADRO 1 – Comparação entre as categorias de Ranganathan e CRG .....	51
QUADRO 2 – Síntese das metodologias da análise categorial temática .....	86
QUADRO 3 – Modelo de leitura Documentária de Fujita e Rubi (2006) .....	96
QUADRO 4 – Sugestões de perguntas da Norma NBR-12.676 .....	97
QUADRO 5 – Relação de documentos do <i>corpus</i> (por ano)* .....	105
QUADRO 6 – Consolidação da estrutura textual de documentos acadêmicos.....	109
QUADRO 7 – Etapa da categorização (Análise de Conteúdo).....	111
QUADRO 8 – Validação das "categorias iniciais".....	112
QUADRO 9 – Primeira coluna do MCTCA .....	119
QUADRO 10 – Primeira e segunda coluna do MCTCA.....	121
QUADRO 11 – Matriz categorial para trabalhos acadêmicos (teses e dissertações) .....	123
QUADRO 12 – Matriz categorial para extração de conceitos .....	125
QUADRO 13 – Paridade entre as categorias de Ranganathan, CRG e CAFTE.....	142
QUADRO 14 – Síntese estatística dos temas .....	185

## LISTA DE SIGLAS

ABNT	–	Associação Brasileira de Normas Técnicas
AC	–	Análise de Conteúdo
ACT	–	Análise categorial temática (semântica)
AD	–	Análise documentária
AF	–	Análise Facetada
BD	–	Biblioteca digital
BDS	–	Biblioteca digital semântica
BDTD	–	Bibliotecas digitais de teses e dissertações
BU	–	Biblioteca Universitária
CAFTE	–	Conjunto de categorias fundamentais temáticas
CAPES	–	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	–	Construção de Categorias
CC	–	Ciência da Computação
CI	–	Ciência da Informação
CMS	–	<i>CROSI Mapping System</i>
CRG	–	<i>Classification Research Group</i>
DW	–	<i>Data Warehouse</i>
ECI	–	Escola de Ciência da Informação
GIC	–	Gestão da Informação e do Conhecimento
GUI	–	<i>Graphical User Interface</i>
HIBROWSE	–	<i>High resolution Interface for Browsing and Searching</i>
IBICT	–	Instituto Brasileiro em Informação em Ciência e Tecnologia
ICS	–	Informação, Cultura e Sociedade
IEC	–	<i>International Electrotechnical Commission</i>
IIB	–	Instituto Internacional de Bibliografia
ISAGOGE	–	<i>Introductio in Praedicamenta</i>
ISO	–	<i>International Organization for Standardization</i>
LD	–	Linguagem documentária
LSP	–	<i>Language for Special Purpose</i>
MARC	–	<i>Machine Readable Cataloging</i>
MCTCA	–	Matriz Categorial para trabalhos acadêmicos do tipo teses e dissertações
MEC	–	Ministério da Educação
MHTX	–	Protótipo Mapa Hipertextual
MIT	–	<i>Massachusetts Institute of Technology</i>
MLD	–	Modelo de Leitura Documentária
MSN	–	<i>Microsoft Network</i>
NASA	–	<i>National Aeronautics and Space Administration</i>
NLS	–	<i>On-Line System</i>

NT	–	<i>New Technology</i>
OLAP	–	<i>On-line Analytical Processing</i>
OPACs	–	<i>On-line Public Access Catalogs</i>
OUI	–	Organização e Uso da Informação
OWL	–	<i>Web Ontology Language</i>
PDF	–	<i>Portable Document Format</i>
PMEST	–	Personalidade, Matéria, Energia, Espaço e Tempo.
POSEDU	–	Portal Semântico Educacional
PPGCI	–	Programa de Pós-Graduação em Ciência da Informação
RAVE	–	<i>Relation Browser</i>
RB	–	<i>Relation Browser</i>
RDF	–	<i>Resource Description Framework</i>
SeCo	–	<i>Semantic Computing Research Group</i>
SIMILE	–	<i>Semantic Interoperability of Metadata and Information in Unlike Environments</i>
SMART	–	<i>System for the Mechanical Analysis and Retrieval of Text</i>
SPO	–	<i>Subject, Predicate, Object</i>
SRI	–	Sistema de recuperação da informação
SWED	–	<i>Semantic Web Environmental Directory</i>
TAD	–	Teoria da análise de domínio
TAF	–	Teoria da análise facetada
TAFNAVEGA	–	Taxonomia facetada navegacional
UC	–	Unidades de Contexto
UFMG	–	Universidade Federal de Minas Gerais
UNESCO	–	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UR	–	Unidades de Registro
USP	–	Universidade de São Paulo
VC	–	Vocabulário controlado
W3C	–	<i>World Wide Web Consortium</i>
XFML	–	<i>Exchangeable Faceted Metadata Language</i>
XML	–	<i>Extensible Markup Language</i>
XSLT	–	<i>Extensible Stylesheet Language for Transformation</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
1.1	Problema e justificativa .....	19
1.2	Objetivo geral.....	22
1.3	Objetivos específicos .....	22
1.4	Estrutura da dissertação .....	23
<b>2</b>	<b>METODOLOGIA.....</b>	<b>25</b>
2.1	Características da pesquisa .....	25
2.2	O caminho metodológico.....	25
<b>3</b>	<b>REVISÃO DA LITERATURA .....</b>	<b>30</b>
<b>4</b>	<b>FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA.....</b>	<b>43</b>
4.1	Vocabulário controlado: a taxonomia – da hierárquica à facetada.....	43
4.2	A taxonomia facetada navegacional.....	55
4.3	Recuperação da informação .....	58
4.4	Bibliotecas digitais.....	66
4.5	Navegação hipertextual.....	70
4.5.1	Navegação por meio da taxonomia facetada.....	72
4.6	Definindo o domínio: a teoria da análise de domínio .....	75
4.6.1	Identificando o usuário: a comunidade científica .....	80
4.7	Método de análise de conteúdo: a técnica categorial temática .....	82
4.7.1	Análise categorial temática e análise facetada: similaridades .....	89
4.8	Modelo de Leitura Documentária de Fujita e Rubi .....	94
4.9	Documentos acadêmicos: teses e dissertações .....	98
<b>5</b>	<b>DESCRIÇÃO DOS PROCEDIMENTOS E ANÁLISE DOS RESULTADOS .....</b>	<b>103</b>
5.1	Recorte temporal e levantamento do material empírico.....	103
5.1.1	Codificação do <i>corpus</i> da pesquisa.....	105

5.2	Identificação do domínio, do usuário e de suas necessidades .....	105
5.3	Criação da taxonomia facetada navegacional .....	107
5.3.1	Desenvolvimento das categorias fundamentais temáticas.....	108
5.3.2	O algoritmo para o processo de extração de conceitos .....	118
5.4	O experiência do MCTCA na análise do <i>corpus</i> da pesquisa .....	125
5.5	Apresentação da implementação tecnológica da TAFNAVEGA .....	133
5.6	Paridade entre as categorias fundamentais de Ranganathan, CRG e da TAFNAVEGA .....	142
<b>6</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>144</b>
	<b>REFERÊNCIAS .....</b>	<b>150</b>
	<b>APÊNDICES .....</b>	<b>172</b>
	APÊNDICE A – Lista do <i>corpus</i> da pesquisa .....	172
	APÊNDICE B – Qualificação dos autores das fontes especializadas .....	174
	APÊNDICE C – Conjunto de categorias iniciais .....	175
	APÊNDICE D – Conjunto de categorias intermediárias .....	176
	APÊNDICE E – Definições para o conjunto de categorias finais .....	177
	APÊNDICE F – Dados estatísticos do processo de extração de conceitos .....	180
	APÊNDICE G – Termos de indexação dos resumos.....	181
	APÊNDICE H – Metodologias de análise categorial temática .....	183
	APÊNDICE I – Outros resultados da análise do <i>corpus</i> da pesquisa .....	184
	<b>ANEXOS .....</b>	<b>190</b>
	ANEXO A – Modelo de busca facetada: protótipo <i>TerveSuomi.fi</i> .....	190
	ANEXO B – Modelo de busca facetada: <i>Relation Browser Demo</i> .....	191

## 1 INTRODUÇÃO

*Nesta introdução, são apresentados o tema, o problema e as justificativas deste trabalho, assim como os objetivos e a descrição da estrutura desta dissertação.*

O saber científico é dinâmico e primordial no desenvolvimento do conhecimento humano, num ciclo aberto de reconstrução. A evolução da ciência acontece com a possibilidade de replicação, quando um novo conhecimento emerge a partir de algo já descoberto. Para obter o conhecimento desvendado, o conjunto de saberes deve ser registrado, com informações expressas em uma linguagem de comunicação que possa ser acessada e recuperada.

A ciência passa sempre por revisões em suas teorias, seja porque não se adaptam mais à realidade empírica existente ou devido à descoberta de novos fenômenos. Sendo assim, as teorias científicas são falíveis e estão, continuamente, sendo substituídas ou recondicionadas, principalmente para se ajustarem à época histórica ou científica. Assim, o conhecimento humano está em constante fluxo, sempre se transformando, aperfeiçoando e ampliando.

A comunicação científica permite perceber as diferentes abordagens teóricas, epistemológicas e práticas que caracterizam uma área do conhecimento. Dessa maneira, na academia são recorrentes as discussões acerca das formas de disponibilização, comunicação e difusão desse conhecimento, visando uma maior visibilidade institucional, que possa legitimar e promover maior prestígio do conhecimento produzido nesse meio.

A Ciência da Informação (CI) é disciplina especialmente voltada para entender os fluxos de informação em diferentes ambientes, a fim de facilitar seu acesso e uso (BORKO, 1968), promovendo a transferência da informação. Acredita-se que o ser humano é primordialmente social, ligado a grupos de interesses, tradicionalmente transmitindo seu conhecimento a outro, e quem o recebe soma-o aos seus próprios saberes. Assim acontece a evolução humana, e também dessa forma se desenvolve a ciência e tecnologia.

Nesse contexto, o tema da recuperação da informação é importante, pois percebe-se que a sociedade tem a informação como insumo essencial, e as funcionalidades oferecidas pelas tecnologias de comunicação expandiram as formas de compartilhamento e busca dessas informações, movimento intensificado com a expansão da Internet.

O compartilhamento de informações é intermediado por sistemas de informação, que têm origem, pode-se declarar, ainda com a biblioteca de Alexandria, que reuniu fontes de informação e as armazenou e organizou, visando a sua preservação e recuperação. Desde então, esses sistemas se desenvolvem e procuram se adaptar às necessidades e demandas informacionais de uma comunidade de usuários. Nesse sentido, os sistemas de

informação estão cada vez mais focados na flexibilidade<sup>1</sup> e na integração, o que permite diferentes interações entre usuários e sistemas. Nessa trajetória, de Alexandria à contemporaneidade, os sistemas de informação mudaram seu foco da organização do documento para a disponibilidade de informação, o que mudou a forma de gerenciamento do recurso informacional.

Em vista disso, afirma-se que o desenvolvimento de sistemas automatizados de armazenamento, de busca e de recuperação de informação, acontecimento percebido a partir dos anos 1950, tem envolvido profissionais de diferentes áreas (Ciência da Computação, *Web Design*, Arquitetura da Informação, Ciência da Informação, entre outras) em pesquisas que abordam o assunto sob diferentes aspectos.

A título de resgate histórico, é interessante observar que, já no início do século XX, Paul Otlet, com sua ideia de mesa móvel, recomendava o desenvolvimento de ferramentas que possibilitassem o acesso às informações, de modo a permitir sua transferência. Otlet imaginou uma espécie de carteira escolar com rodas, com a qual os usuários poderiam buscar, ler e anotar informações, retiradas das bases de dados, armazenadas em fichas de papel. Assim, pode-se afirmar que Otlet criou o primeiro esboço de um sistema de recuperação da informação (SRI) semiautomatizado.

Um SRI ocupa-se de parte significativa do processo que envolve os fluxos de informação: seleção, representação, organização, armazenamento, busca e acesso aos recursos informacionais. Nesse contexto, são preocupação constante dos pesquisadores as questões que envolvem a criação de mecanismos para acesso remoto às informações disponibilizadas, principalmente em ambiente de bibliotecas digitais, onde a organização da informação deve garantir ao usuário facilidade de acesso às informações de sua necessidade.

É importante mencionar também que a organização da informação pode ser feita a partir de diferentes abordagens metodológicas, com vistas a imprimir consistência e padronização aos conteúdos informacionais, facilitando sua organização e posterior recuperação. Nesse sentido, os vocabulários controlados, tais como os tesouros e as taxonomias, podem ser utilizados para estruturar certo domínio<sup>2</sup>, aprimorando a recuperação de informações.

---

<sup>1</sup> Esse termo tem sido muitas vezes utilizado como atributo dos SRI. *Flexibilidade* pode ser entendida como a qualidade de algo que tem maleabilidade, que é fácil de manejar, e que está apto para variadas coisas ou aplicações (FERREIRA, 2009). Assim, um sistema de informação flexível é aquele capaz de absorver mudanças em suas estruturas de informação, seja permitindo inclusão de novos termos ou substituindo termos obsoletos.

<sup>2</sup> *Domínio* pode ser entendido como uma área de conhecimento (especialidade), um conjunto literário ou, ainda, a tarefa de um grupo de pessoas trabalhando juntas em um projeto ou em uma organização (HJORLAND; ALBRECHTSEN, 1995). Esse tema será tratado mais detalhadamente no Capítulo Quatro, subcapítulo 4.5 deste documento.

Com relação às taxonomias, seu uso tem sido ampliado em ambiente digital, para a estruturação de informações organizacionais, com o objetivo de organizar as informações, orientar o usuário (cliente) e facilitar a busca por produtos e serviços, propiciando sua recuperação.

Para uma melhor compreensão, é oportuna uma definição para o termo “taxonomia”, trata-se de um conjunto de termos estruturados de forma hierárquica, que são utilizados para representar o domínio no qual a taxonomia é aplicada. Essa forma de organização é a adotada tradicionalmente. Contudo, nos dias atuais, verificam-se mudanças no conceito de taxonomia, e na sua forma de organização: elas podem, também, ser apresentadas em forma de listas, árvores, poli-hierarquias, matrizes, facetas ou sistemas de mapas (LAMBE, 2007). Sobre essa última forma (taxonomia por facetas), além das propriedades de representar e organizar um domínio, ela pode, ainda, correlacionar diferentes linguagens usadas em um ambiente e criar mecanismos de acesso, com agregação de metadados que podem possibilitar a localização de conteúdos informacionais.

Sabe-se que, em uma biblioteca digital, a taxonomia facetada atua como um mapa conceitual, já que cada tópico da estrutura se interrelaciona com outro e pode ser explorado através da navegação. Assim, essa forma de apresentação contribui para a comunicação e o compartilhamento de conhecimento.

Ainda sobre as bibliotecas digitais, tem-se constatado um desenvolvimento no sentido de aprimorar as formas de armazenamento, disponibilização e acesso às informações contidas em seus bancos de dados. Historicamente, esses esforços perpassam Paul Otlet, com suas ideias revolucionárias, o Memex<sup>3</sup> (Vannevar Bush), o projeto Xanadu<sup>4</sup> (Theodore Nelson), culminando com o desenvolvimento da Internet, sendo esta constituída por um conjunto de tecnologias que permite o compartilhamento e a disponibilização de informações.

É importante destacar também que, nas instituições de ensino, as facilidades tecnológicas têm contribuído para a disponibilização de informações acadêmicas em bibliotecas digitais. Ademais, observa-se a crescente migração de periódicos científicos para o meio digital. Então, é verificado um crescimento quantitativo da implantação de bibliotecas digitais de teses e dissertações (BDTD) em instituições de ensino superior, visando à divulgação de resultados detalhados de pesquisas já concluídas. Como consequência dessas iniciativas, evitam-se esforços replicados em pesquisa similares e a perda de energia intelectual com o uso de metodologias e teorias, talvez, já obsoletas.

---

<sup>3</sup> Bush imaginou uma espécie de máquina que facilitasse o armazenamento de informações, para posterior recuperação e uso.

<sup>4</sup> O projeto Xanadu refere-se à idealização de uma ampla rede de informações, sendo considerado o primeiro passo para o desenvolvimento daquilo que, atualmente, é conhecido como hipertexto.

Não se pode perder de vista, que o desenvolvimento de BDTDs é observado, principalmente, após a Portaria 13/2006-CAPES<sup>5</sup>, que estabelece a obrigatoriedade da divulgação digital das teses e dissertações defendidas nos programas de pós-graduação *stricto sensu* no país. Contudo, percebe-se que o acesso a conteúdos específicos dessas pesquisas ainda é precário, pois as teses e dissertações são documentos textuais, em linguagem natural e com forma semiestruturada, dificilmente acessível à compreensão e à inferência da máquina. Dessa forma, para que os sistemas computacionais possam recuperar essas informações com agilidade e eficiência, o conteúdo informacional dos documentos precisa ser organizado de modo estruturado.

Nesse contexto, entende-se que a taxonomia facetada pode auxiliar na estruturação do conteúdo informacional das teses e dissertações, através da representação dos conceitos em um sistema facetado de termos. Com isso, busca-se facilitar o acesso às informações, oferecendo ao usuário a opção de navegação pelos termos, apresentados em uma interface amigável<sup>6</sup> de busca.

Face ao exposto, apresentam-se, a seguir, o problema e as justificativas referentes à proposta desta pesquisa, assim como os objetivos geral e específicos.

## 1.1 Problema e justificativa

A partir do início do século XXI, é possível notar um aumento nos projetos de implantação de bibliotecas digitais de teses e dissertações (BDTD), tanto nacionais<sup>7</sup> quanto estrangeiras (ROSETTO, 2008), especialmente, como já afirmado, após a obrigatoriedade instituída pela Portaria 13/2006-CAPES. Vale registrar que o objetivo dessa portaria é regular a forma para disponibilização digital da produção científica, sobretudo das teses e dissertações, a fim de oferecer acesso público a essa produção, bem como divulgar o conhecimento científico acumulado nas pesquisas (SOUZA, A., 2006, *on-line*). Entretanto, essa obrigatoriedade não tem se mostrado suficiente para “divulgar a ciência em geral e o saber produzido pelos programas de pós-graduação”. Certamente, pode-se perceber que, somente a partir do acesso aos resultados das pesquisas, a sociedade observou um avanço em seu conhecimento científico, teórico ou aplicado, criando novas tecnologias ou produtos, como salientado por Mueller (2000).

Nesse sentido, de fato, há ferramentas tecnológicas e sofisticadas funcionalidades disponíveis. Porém, as formas de busca e recuperação de informações em BDTDs ainda não superam, em grande medida, as de uma biblioteca tradicional, uma vez

---

<sup>5</sup> CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

<sup>6</sup> Interface amigável é aquela considerada fácil e intuitiva de usar, sem poluição visual que desorienta o usuário.

<sup>7</sup> Como exemplo, pode-se citar o projeto da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT); a Biblioteca Nacional Digital Brasil; e a Rede Memória Virtual Brasileira.

que, geralmente, permitem o acesso apenas a elementos descritivos (autor e data, por exemplo), ao resumo, às palavras-chave e ao texto completo dos documentos. Nesse último caso, inclusive, o texto, em geral, é apresentado no formato *Portable Document Format* (PDF), que é aberto em aplicativo independente da biblioteca digital, impedindo que o sistema possa varrer seu conteúdo durante as buscas. Nesse caso, estando os outros elementos (descritivos, resumo e palavras-chave) incompletos, haverá perda de informações.

É importante destacar também, que há o problema de palavras-chave e títulos que não são representativos do conteúdo dos documentos, além do entrave provocado por resumos incompletos, observados na literatura por diferentes estudos (BORKO; CHATMAN, 1963; VIEIRA, 1988; NOGUEIRA, 1997; GIL LEIVA; FUJITA; RUBI, 2008; ELIEL, 2008; SOUZA, 2009).

Ademais, mostra-se pertinente apontar que, nas buscas por assunto, é comum as bibliotecas digitais oferecerem a “busca simples” e a “busca avançada”. Nessas duas formas, o sistema faz uma varredura nos elementos descritivos (título, autor e data), no resumo e nas palavras-chave. Entretanto, em ambas, há problemas para o usuário: na “busca simples”, geralmente, o usuário submete, no mínimo, uma palavra ao sistema, sendo que o resultado de sua busca pode apresentar problemas de ambiguidade, envolvendo questões de falta de precisão<sup>8</sup> e de revocação<sup>9</sup>. Com relação à busca avançada, há a possibilidade de buscas combinadas, visando a refinamento dos resultados, normalmente, através da busca *booleana*<sup>10</sup> (*or*, *and* e *not*). Porém, o maior problema nesta forma de busca, refere-se ao fato de que o usuário, recorrentemente, está inapto para utilizá-la, seja porque não consegue acompanhar as mudanças tecnológicas (barreira tecnológica), seja por comodismo (barreira psicológica) (FIGUEIREDO, 1987), ou mesmo pelo motivo de considerar que as buscas *booleanas* são de difícil compreensão (BAEZA-YATES; RIBEIRO-NETO, 1999).

Tendo isso em vista, acredita-se que é possível oferecer ao usuário, principalmente em uma biblioteca digital especializada, informações já estruturadas (organizadas), com resultados já refinados, sob a perspectiva das necessidades de informação desse usuário, que seja uma estrutura fácil de usar e interpretar, como forma de minimizar os problemas apontados na busca simples (ambiguidade) e na busca *booleana* (inaptidão do usuário em utilizá-la). Para tanto, a característica textual semiestruturada das teses e dissertações exige a criação de mecanismos que orientem a estruturação do recurso

<sup>8</sup> Precisão é o índice da relação entre o número de documentos recuperados considerados relevantes em relação ao total de documentos recuperados.

<sup>9</sup> Revocação é o índice da relação entre o número de documentos relevantes sobre um tema recuperados pelo sistema em relação ao total de documentos existentes no banco de dados sobre esse mesmo tema.

<sup>10</sup> A lógica *booleana* é derivada da teoria dos conjuntos e possui uso universal para ampliação ou refinamento de resultados.

informacional, melhorando a eficiência na localização de informações específicas. Então, devem-se reconhecer suas partes ou capítulos, além de sua sequência de ideias, a fim de organizá-los e classificá-los ordenadamente.

Nesse cenário, entende-se que a taxonomia facetada possui qualidade estruturante, que permite organizar as informações dispersas no texto. Além disso, visto que provém de um vocabulário controlado, ela possibilita a padronização da terminologia utilizada de modo que exclui as ambiguidades e, por essa razão, melhora a eficiência da recuperação dos conteúdos informacionais. A ambiguidade pode ser uma qualidade em textos poéticos, porém, em textos científicos e ambientes especializados ela deve ser sempre evitada, pois compromete o sentido e o significado expressos pelo autor.

Na mesma direção das ideias abordadas acima, Figueiredo (2006, *on-line*) afirma que “as taxonomias podem também assumir representações gráficas<sup>11</sup>, facilitando ainda mais a exploração e a compreensão do universo de conteúdo em questão”. Assim sendo, essa estruturação, concebida a partir da taxonomia facetada navegacional, permite que “as informações [...] [sejam] organizadas respeitando-se os temas, os assuntos e a hierarquia estipulados pela ferramenta, apresentando-se ao usuário aquilo que há de mais relevante naquele contexto” (NONATO *et al.*, 2008, p.132), orientando, dessa maneira, sua busca.

Outro aspecto relevante indica que o uso da taxonomia facetada navegacional permite operar e refinar a busca de forma progressiva, de modo que o usuário saiba, exatamente, quais informações estão disponíveis no banco de dados, sem que haja a possibilidade de um “resultado de busca vazio<sup>12</sup>”. Aliás, essa particularidade é confirmada por Prazeres *et al.* (2006, p.2), quando afirmam que

Uma das vantagens da navegação baseada em facetas é a possibilidade de se combinar um ou mais valores. Por exemplo, é possível obter todas as páginas escritas em português (faceta língua), escritas entre 2001 e 2002 (faceta data), do tipo texto (faceta tipo do conteúdo). Um exemplo deste tipo de navegação pode ser visto no projeto Simile-Longwell<sup>13</sup>.

Em face disso, com essa interface de navegação, o usuário tem os recursos de uma busca avançada, sem a necessidade de indicar as combinações lógicas<sup>14</sup>, pois isso já se apresenta, semanticamente, implícito na estrutura facetada. O usuário pode, então, escolher níveis de especificidade e restringir a busca gradativamente, usando os termos dispostos na taxonomia facetada, como se fosse um *menu* do SRI.

<sup>11</sup> É um tipo de interface que permite a interação do usuário com os recursos digitais a partir de elementos visuais. Essas representações são conhecidas como *Graphical User Interface (GUI)*.

<sup>12</sup> Entende-se aqui, que as categorias “vazias” não são apresentadas ao usuário, o que impede a obtenção de um resultado “vazio”, pois só aparece para escolha aquilo que, efetivamente, existe.

<sup>13</sup> Disponível em: <[http://simile.mit.edu/wiki/Faceted\\_Browser](http://simile.mit.edu/wiki/Faceted_Browser)>.

<sup>14</sup> Como a utilizada pela busca *booleana*, considerada difícil pelo usuário.

À luz dessas considerações, entende-se que a importância deste estudo na CI está na relevância do tema da organização da informação em ambientes digitais, sobretudo no que concerne à busca e ao aperfeiçoamento da recuperação da informação, o que, no caso específico desta pesquisa, pode impactar positivamente no processo de comunicação científica.

Por essa razão, a proposta desta pesquisa é criar uma taxonomia facetada para navegação, como opção de busca em bibliotecas digitais de teses e dissertações, priorizando as necessidades dos pesquisadores, com vistas a promover maior visibilidade ao conhecimento científico acumulado.

Para tanto, este estudo tem como expectativa responder às questões seguintes: Como evidenciar as informações disponibilizadas nas BDTDs, referentes aos resultados detalhados das pesquisas? De que modo é possível facilitar o acesso do usuário que busca esse tipo de informações acadêmicas? Como refinar essa busca, de maneira que as informações possam ser buscadas em formato estruturado e o usuário não seja sobrecarregado com informações?

Levando-se em consideração a proposta e a expectativa, a seguir são apresentados o objetivo geral e os objetivos específicos delimitados nesta pesquisa.

## **1.2 Objetivo geral**

Criar uma taxonomia facetada navegacional a partir de uma matriz categorial, desenvolvida para trabalhos acadêmicos, visando a uma representação sistematizada e semântica do conteúdo de teses e dissertações capaz de facilitar a busca e a recuperação das informações.

## **1.3 Objetivos específicos**

- 1) Demonstrar a adequação do uso de categorias conceituais como potencial mecanismo para estruturar o conteúdo textual dos trabalhos acadêmicos;
- 2) Contribuir para maior visibilidade do conteúdo das pesquisas científicas, evidenciando, principalmente, os resultados detalhados das pesquisas acadêmicas;
- 3) Aprimorar a busca por conteúdos disponíveis em bibliotecas digitais de teses e dissertações, através da navegação pela taxonomia facetada.

## 1.4 Estrutura da dissertação

Esta pesquisa possui seis capítulos, nos quais são abordados os assuntos considerados pertinentes à apresentação, análise e entendimento das questões discutidas, sendo organizado da seguinte maneira:

Capítulo 1: Introdução: com destaque para a comunicação científica, visando melhoria na qualidade dos sistemas de recuperação de informação, com a possibilidade do uso de taxonomias facetadas como mecanismo de navegação e acesso às informações. Foram apresentados também os problemas, as justificativas e a proposta (indicada no objetivo geral e nos objetivos específicos), incluindo a estrutura desta dissertação.

Capítulo 2: Metodologia da pesquisa: primeiramente foram apresentadas as características da pesquisa. Em seguida, no subcapítulo “caminho metodológico”, apresenta-se o universo e o recorte de investigação, com o *corpus* e o objeto empírico manuseado. Indicam-se, ainda, os métodos e técnicas aplicados, assim como a fundamentação teórico-metodológica que dá respaldo aos procedimentos. Essas informações são fornecidas de forma sintética, uma vez que as descrições detalhadas se encontram no Capítulo Cinco, de “Descrição dos procedimentos e análise de resultados” e no Capítulo Quatro, da “Fundamentação teórico-metodológica”.

Capítulo 3: Revisão da literatura: neste capítulo busca-se atualizar informações sobre o tema de busca e navegação através de estruturas facetadas, com foco no uso de taxonomias. Para isso, descrevem-se trabalhos que respondam a questões sobre quais procedimentos metodológicos e os fundamentos teóricos que vêm sendo empregados para a criação desse instrumento, de forma que possa ser um facilitador para a exploração, busca e recuperação de informações em bibliotecas digitais, principalmente nos últimos dez anos. Além disso, descrevem-se os possíveis avanços alcançados e as disciplinas que estão engajadas neste tipo de pesquisa e aplicação, assim como a Ciência da Informação se insere nesse contexto. Ao final, citam-se os três principais livros que nortearam esta pesquisa, descrevendo o conteúdo dos mesmos.

Capítulo 4: Fundamentação teórico-metodológica: este capítulo inicia abordando, brevemente, os fundamentos teóricos sobre os vocabulários controlados como instrumento de representação e organização da informação. Verticaliza sobre o tema da taxonomia, apresentando sua forma mais tradicional, a hierárquica, depois a forma facetada, até chegar ao produto final que é a taxonomia facetada navegacional. Nesse percurso são também indicados, além dos fundamentos teóricos, os princípios metodológicos de criação, uma vez que os dois estão imbricados. Expõe os fundamentos teóricos acerca da busca em sistemas de recuperação da informação (SRI) de bibliotecas digitais (BD), especialmente nas BDs de teses e dissertações. Depois, indica os princípios teóricos da navegação hipertextual e a

facetada para, em seguida, discorrer sobre a Teoria da Análise de Domínio (TAD) e a metodologia indicada para a identificação da comunidade científica e suas necessidades de informação. Logo após, descreve-se os fundamentos do método da Análise de Conteúdo, especificamente aprofundando sobre a aplicação da técnica da análise categorial temática e fornecendo um confronto entre os procedimentos metodológicos dessa técnica e da Teoria da Análise Facetada. Dando prosseguimento, apresenta-se o método utilizado para o desenvolvimento do Modelo de Leitura Documentária de Fujita e Rubi (2006). Finalmente, caracterizam-se documentos do tipo teses e dissertações, uma vez que eles são o material empírico manuseado nesta investigação.

Capítulo 5: Descrição dos procedimentos e análise dos resultados: esse capítulo descreve detalhadamente todos os procedimentos dos métodos e técnicas adotados nas diferentes etapas de desenvolvimento da proposta da taxonomia facetada navegacional, denominada de TAFNAVEGA, assim como a análise dos resultados. A proposta é apresentada desde a identificação do domínio, do usuário e de suas necessidades de informação, seguindo com a apresentação do conjunto de categorias fundamentais, depois com a formulação de um algoritmo para sistematizar o processo de extração de conceitos dos documentos e, finalizando, faz uma demonstração do aplicativo. O objetivo do mecanismo é possibilitar ao usuário uma forma mais fácil de acessar, além de elementos tais como o título, autor, ano e resumo do documento, informações estruturadas sobre temas, métodos, teorias e os resultados das pesquisas já concluídas. Capítulo 6: Considerações finais: neste último capítulo volta-se o olhar para os pressupostos e as questões que motivaram a pesquisa, buscando responder, analiticamente, aos objetivos determinados como meta de investigação, sempre indicando de que forma os resultados podem contribuir para o avanço do conhecimento e das discussões dentro da área da Ciência da Informação. Finalmente, apontam-se, a partir de observações durante a pesquisa, possibilidades para trabalhos futuros.

## 2 METODOLOGIA

O método científico envolve um conjunto de procedimentos sistemáticos e racionais, utilizados no intuito de alcançar o objetivo proposto, traçando um caminho mais seguro e econômico, capaz de permitir detectar erros e facilitar as decisões sobre a postura do investigador (MARCONI; LAKATOS, 2010, p.65). Nessa perspectiva, a pesquisa, para esses autores, “é um procedimento formal, com método de pensamento reflexivo, que requer um tratamento científico e se constituiu no caminho para conhecer a realidade ou para descobrir verdades parciais” (*ibidem*, p.139).

Nessa direção, o percurso metodológico, com a aplicação dos procedimentos necessários para a conclusão deste estudo, é informado a partir de agora.

### 2.1 Características da pesquisa

Esta pesquisa envolveu procedimentos de natureza empírica, uma vez que houve a manipulação de documentos para análise, cujos resultados dos dados foram usados na concretização do objetivo deste estudo. Segundo Demo (2000, p.21), a pesquisa empírica trata da “face empírica e factual da realidade; produz e analisa dados, procedendo sempre pela via do controle empírico e factual”.

Utilizou-se também da investigação do tipo exploratória, pois abrangeu levantamento bibliográfico para análise de exemplos de experiências e práticas, em relação ao problema pesquisado (GIL, 1994), ampliando o conhecimento do proponente sobre o tema estudado. Nessa exploração na literatura da área, houve a busca por conceitos sobre os temas discutidos, com o objetivo de obter informações para realizar inferências. Realizou-se, ainda, análise conceitual nos documentos do *corpus* e houve a identificação e descrição das necessidades informacionais do usuário final, caracterizando-se também como uma pesquisa de cunho qualitativo.

Além disso, recorreu-se ao método de pesquisa aplicada, visto que ofereceu a criação de uma taxonomia facetada navegacional, com o propósito de permitir a navegação por sua estrutura, com metadados embutidos.

### 2.2 O caminho metodológico

Nesta seção, serão brevemente apresentados os métodos e técnicas utilizados, com a fundamentação teórico-metodológica escolhida para a análise, quando for o caso, mas somente de forma sintética. Houve a escolha pela exposição dos métodos e das técnicas empregados no desenvolvimento da TAFNAVEGA ao longo do trabalho, na medida em que muitas vezes os fundamentos teóricos estão imbricados por fundamentos e

princípios metodológicos. Assim, muitas dessas informações estão apresentadas no Capítulo Quatro, da “Fundamentação teórico-metodológica”. Além disso, optou-se pela descrição detalhada dos procedimentos metodológicos, inclusive das ferramentas utilizadas, no Capítulo Cinco, de *Descrição dos procedimentos e análise de resultados*. Foi uma opção consciente de estruturação textual da dissertação, uma vez que parte da metodologia aplicada foi também composta por interpretação e análise de resultados, inviabilizando separar as duas seções. Acredita-se que a forma escolhida trará mais benefícios à compreensão do processo pelo leitor. Dessa forma, neste capítulo estão indicadas unicamente as bases da metodologia empregada nesta pesquisa.

Como primeiro passo, realizou-se “pesquisa bibliográfica”<sup>15</sup> (MARCONI; LAKATOS, 2010), sistematizada, no “Portal Capes”, que abrangeu, mas não se restringiu, a buscas por publicações nas diferentes bases de dados (tal como *Library and Information Science Abstracts* – LISA, por exemplo), em especial nas *Proceedings of the American Society for Information Science and Technology, Information Technology & Libraries, Computer Science Human Interface and the Management of Information, Computer Science Flexible Query Answering Systems, Computer Science Research and Advanced Technology for Digital Libraries, Proceedings of Annual International ACM e ACM SIGIR Workshop on Faceted Search, Citeseer, Annual Review of Information Science and Technology* – ARIST (artigos de revisão), entre outras. Em relação às publicações nacionais, foram utilizadas as Bibliotecas Digitais de Teses e Dissertações de diferentes instituições<sup>16</sup> e as revistas nacionais: *Informação & Sociedade*, *Revista da Ciência da Informação* (Instituto Brasileiro em Informação em Ciência e Tecnologia - IBICT), *Revista Encontros Bibli* e *DataGramaZero* – *Revista de Ciência da Informação*.

Além disso, utilizou-se o acervo da Biblioteca Professora Etelvina Lima, da Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG), e buscas na Internet (Google Acadêmico e Biblioteca Pública de Nova Iorque e, também, uso de acervo pessoal. Deve-se registrar, ainda, que essa atividade priorizou a análise da literatura sobre os fundamentos teórico-metodológicos e históricos dos temas abordados nesta pesquisa, tais como bibliotecas digitais, hipertextos, vocabulários controlados, categorização, análise de conteúdo e mecanismos de navegação, de modo que os mesmos pudessem ser articulados com os estudos sobre a criação de classificações facetadas navegacionais e com os estudos que abarcam as atividades de profissionais de indexação.

Ademais, para a revisão da literatura sobre o tema específico desta pesquisa – taxonomia facetada navegacional – foram utilizados os seguintes termos para busca: em

<sup>15</sup> Pesquisa bibliográfica é uma técnica que tem como finalidade explorar na literatura (livros, revistas, teses, entre outros) material sobre determinado assunto de interesse da pesquisa (MARCONI; LAKATOS, 2010). Sempre que o termo “pesquisa bibliográfica” for utilizado neste documento, o sentido será este.

<sup>16</sup> Tais como as do IBICT, UFMG, Universidade de São Paulo e Domínio Público.

português: “navegação”, “navegação facetada”, “busca”, “busca facetada”, “taxonomia facetada” e “taxonomia facetada navegacional (e de navegação)”; em inglês: “browsing”, “navigation”, “faceted browsing”, “faceted navigation”, “search”, “faceted search”, “faceted taxonomy”, “faceted navigation taxonomy”, “faceted browsing taxonomy”.

Ainda cabe esclarecer, que durante todo o percurso da pesquisa, houve pesquisa bibliográfica sobre fundamentos que respaldassem ou refutassem os caminhos escolhidos para solucionar os problemas encontrados, visando nortear e validar as decisões assumidas.

Salienta-se que o universo de investigação foram as teses e dissertações defendidas entre os anos 1998 e 2009, no Programa de Pós-Graduação em Ciência da Informação (PPGCI), da Escola de Ciência da Informação (ECI), da Universidade Federal de Minas Gerais (UFMG), até 31 de dezembro de 2009. De um total de 290 trabalhos analisados, 62 são teses, e 228, dissertações. É importante explicar que a escolha desse ambiente se deve a questões de ordem operacional, sobretudo aquelas relacionadas ao acesso ao material disponibilizado no banco de dados da biblioteca digital, assim como à proximidade com a instituição em questão.

Ressalta-se que a determinação do *corpus* foi feita por amostragem, sendo este um processo pelo qual se obtém informação sobre um todo (universo), examinando-se apenas uma parte do mesmo (amostra ou *corpus*). Nessa circunstância, há a amostragem probabilística (ou estatística) e a não-probabilística (ou subjetiva). A primeira fundamenta-se na distribuição amostral e na seleção aleatória. Já a outra não possui base estatística, sendo a amostra de tipo intencional (por julgamento), selecionada por critérios pessoais decorrentes da experiência profissional e do conhecimento da área de domínio (MARCONI; LAKATOS, 1990). Sendo assim, acredita-se que a aplicação da amostra do tipo intencional, neste estudo, é aceitável e, até mesmo, desejável, pois a linha de pesquisa Organização e Uso da Informação (OUI) é de familiaridade da proponente, sendo, ainda, a área de especialidade e de experiência profissional da sua orientadora, o que, certamente, confere condição de confiabilidade ao trabalho.

Além disso, a amostra por julgamento é indicada, exatamente, para investigações de cunho exploratório, principalmente se o universo é homogêneo (MALHOTRA, 2001, p.316). Elucida-se que a proposta apresentada aqui atende a este pressuposto, já que todos os documentos são trabalhos acadêmicos, do tipo teses e dissertações, produzidos numa única área de conhecimento, a Ciência da Informação (CI).

Finalmente, avalia-se, também como forma de argumentação, que, em alguns casos, a amostragem por julgamento pode ser mais fidedigna e representativa do que uma amostra probabilística (AAKER; KUMAR; DAY, 1995; HANSEN; HURWITZ; MADOW, 1996).

Tendo em vista essas colocações, o *corpus* foi constituído pelas teses e dissertações defendidas no PPGCI/ECI/UFMG, provenientes da linha de pesquisa OUI, no período compreendido entre os anos 1998 e 2009, disponibilizadas no banco de dados da biblioteca digital de teses e dissertações da instituição (BDTD/ECI/UFMG), totalizando 41 documentos. Desse total, 12 são teses e 29 são dissertações.

Quanto ao objeto empírico utilizado para a criação da taxonomia facetada navegacional, este foi constituído pelas categorias fundamentais temáticas e pelos termos procedentes da análise conceitual nos documentos acadêmicos, denominado como “processo de extração de conceitos”. Essa análise foi realizada nos resumos informativos disponíveis em cada documento do *corpus* (41 documentos) da pesquisa.

A determinação das categorias fundamentais temáticas (CAFTE) para o desenvolvimento da taxonomia facetada navegacional, denominada de TAFNAVEGA, estão apresentadas nos subcapítulos 5.3 e 5.3.1. Esses procedimentos se basearam na Teoria da Análise Facetada (TAF) e foram validadas utilizando o método da Análise de Conteúdo, com a técnica da análise categorial temática (ACT), descrito no subcapítulo 4.7. Já para a análise conceitual nos resumos informativos, sistematizando o “processo de extração de conceitos” foi utilizado o algoritmo da matriz categorial para trabalhos acadêmicos de documentos do tipo teses e dissertações, cuja fórmula está expressa no subcapítulo 5.3.2. A matriz contendo o algoritmo é denominada Matriz Categorial para trabalhos acadêmicos (MCTCA), do tipo teses e dissertações, e foi desenvolvida nesta pesquisa, com base no Modelo de Leitura Documentária (MLD) de Fujita e Rubi (2006), criado para a indexação de artigos da área das ciências biológicas. Essa metodologia está indicada no capítulo da fundamentação teórico-metodológica, no subcapítulo 4.8.

Para a atividade de tradução dos termos resultantes do “processo de extração de conceitos”, foi utilizado como instrumento a Taxonomia da Ciência da Informação, desenvolvida por Oddone e Gomes (2004). O uso desse instrumento teve a exclusiva finalidade de fazer o controle do vocabulário na extração<sup>17</sup> de conceitos, visando a compatibilização da linguagem natural empregada nos diferentes textos, principalmente para resolver problemas relacionados com sinonímias e ambiguidades.

Ademais, houve a necessidade da inserção de novos termos na taxonomia de Oddone e Gomes (2003). Essa prática é apoiada pela norma NISO AZ-39 (2005) quando afirma que, em instrumentos criados *ad hoc*<sup>18</sup>, novos termos encontrados no conteúdo dos documentos são potenciais candidatos à inclusão no vocabulário controlado e, se admitidos, são inseridos nas classes correspondentes. Isso mantém as características de uma

---

<sup>17</sup> O termo “extração” neste trabalho tem o sentido de retirar conceitos instituídos no texto, a partir de uma análise conceitual no documento, e não deve ser confundido com o processo de indexação por *extração*, que geralmente é realizada de forma automática.

<sup>18</sup> Para uma finalidade específica.

taxonomia facetada de ser flexível, dinâmica e, constantemente, atualizada. A política utilizada para a inserção de novos termos foi a consulta à literatura especializada sobre metodologia, especialmente nas fontes selecionadas para respaldar este estudo, descritas no Capítulo 5, no subcapítulo 5.3.

Após a análise conceitual nos resumos informativos, os termos obtidos no processo de extração de conceitos alimentaram o conjunto de categorias fundamentais temáticas da TAFNAVEGA.

Resta esclarecer, ainda, que para a identificação do usuário da BDTD, assim como de suas necessidades informacionais, foi utilizado o método da teoria da análise de domínio (TAD), principalmente a partir dos estudos de Hjørland e da literatura especializada sobre a comunidade científica.

Dentro desse percurso, a partir do próximo momento, propõe-se uma breve revisão de literatura, seguida pela fundamentação teórico-metodológica, os resultados e discussões, até chegar às considerações finais.

### 3 REVISÃO DA LITERATURA

Em um trabalho científico, a revisão de literatura busca obter informações mais atualizadas sobre o tema central da pesquisa, ainda que tenha sentido também histórico. Ela é parte integrante da revisão bibliográfica que tem como objetivo orientar o pesquisador durante todo o trajeto da investigação, desde o pré-projeto. A revisão bibliográfica é, assim, mais ampla e guia o pesquisador, principalmente evitando que repita trabalhos já concluídos ou em fase adiantada de investigação.

Segundo Luna (2009), uma revisão de literatura pode atender a diferentes objetivos: determinação do “estado da arte”, revisão teórica, revisão histórica ou revisão empírica. No caso desta pesquisa, optou-se pela revisão empírica<sup>19</sup> na qual, segundo Luna (2009), se espera demonstrar como o problema vem sendo pesquisado do ponto de vista metodológico, fornecendo

dados das pesquisas já analisadas, para responder a questões tais como: quais os procedimentos normalmente empregados no estudo desse problema? Que fatores vêm afetando os resultados? que propostas têm sido feitas para explicá-los ou controlá-los? que procedimentos vêm sendo empregados para analisar os resultados? há relatos de manutenção e generalização dos resultados obtidos? do que elas dependem? (LUNA, 2009, p.91).

Seguindo tais orientações, essas indagações possibilitaram a observação de como o objeto de estudo está factualmente sendo aplicado. Para proporcionar um recorte de análise, neste capítulo também estão determinadas as variáveis que serão estudadas, esclarecendo o significado de cada uma delas dentro da pesquisa .

Os trabalhos pautados na utilização de navegações através de taxonomias facetadas navegacionais, como instrumento facilitador para a exploração, busca e recuperação de informações em bibliotecas digitais ainda são incipientes no campo da Ciência da Informação (CI) brasileira. A revisão de literatura demonstrou que esse tema é investigado na área da Ciência da Computação (CC), e está presente também nas áreas de Engenharia de *Software*, Arquitetura da Informação e Informática, com predominância para trabalhos estrangeiros.

No âmbito da CI, muitos estudos enfatizam a elaboração de taxonomias facetadas para uma área do conhecimento, a partir dos princípios do método da análise facetada (ARAÚJO, 2005; NOVO, 2007; COSTA; RAMOS, 2008). Além disso, a literatura da área também indica aplicações práticas do modelo da lógica analítico-sintética de organização do conhecimento como ferramenta potencial para a descrição e recuperação de conteúdos em SRIs, como observado na revisão de literatura de Café e Bratfish (2007).

---

<sup>19</sup> A palavra “empírica” neste contexto tem o sentido de “observação da realidade” para identificar como o tema estudado vem sendo aplicado factualmente, respaldando as argumentações em relação à escolha da aplicação prática da pesquisa.

Contudo, a presente revisão não incluirá trabalhos dessa natureza, uma vez que o aspecto predominante na criação da taxonomia facetada abordada nesta pesquisa é o do acesso a informações contidas em sistemas de recuperação de informações (SRI) especializados, em ambientes digitais, nos quais é possível a navegação através da estrutura facetada da taxonomia, facilitando a busca e a recuperação de informações.

É preciso esclarecer, ainda, que também não serão apresentados trabalhos que versam sobre interfaces a partir de *clustering* (agrupamento), sintagmas nominais ou de iniciativas do comércio eletrônico (*e-commerce*), pois estes últimos nem sempre podem ser considerados como estruturas de navegação facetada (ver Capítulo 4, item 4.4.1).

Devido à abundância do material encontrado, principalmente em trabalhos estrangeiros, para a inclusão dos estudos nesta revisão, foram considerados os seguintes parâmetros: (1) estudos que utilizam a técnica da análise facetada na criação de taxonomias para navegação; (2) trabalhos que tratam especificamente de desenvolvimento de taxonomias facetadas para navegação, independente de indicar, explicitamente, a técnica utilizada. Em qualquer dessas situações, serão considerados trabalhos nacionais e estrangeiros, dando prioridade aos trabalhos desenvolvidos nos últimos dez anos, do mais antigo aos mais novos. Entretanto, houve a preocupação em apresentar trabalhos que podem ser considerados pioneiros na implantação da estratégia de navegação facetada, como forma de mostrar a evolução do tema de navegação e busca facetada, oferecendo uma localização temporal histórica do tema. Além disso, apresenta-se o *Relation Browser*, que é um projeto que está fora do recorte temporal estipulado para esta revisão, pois foi iniciado entre dezembro de 1998 e janeiro de 1999, mas que merece ser citado por ter sido criado dentro da área da CI e, também, porque está em evolução desde então.

Ressalta-se que não se pode afirmar, categoricamente, que não há qualquer trabalho anterior aos que serão mencionados, mas são projetos que merecem ser citados como pioneiros sobre o tema da navegação facetada: o trabalho de Ahlberg e Shneiderman (1994), da Universidade de Maryland, e o de Pollitt *et al.* (1996), da Universidade de Huddersfield.

No primeiro trabalho, da área da Ciência da Computação e Informática, foi criado o protótipo *FilmFinder* (<<http://www.film-finder.com/>>), de Ahlberg e Shneiderman (1994), que permite explorar uma base de dados de filmes por facetas. Os autores não utilizaram o termo “faceta”, entretanto, a modelagem informacional era em facetas, mas com uma “busca por parâmetros” (*parametric search*). Nesse tipo de busca, o usuário tem de pré-definir as variáveis, as quais devem ser escolhidas em uma única vez. Dessa maneira, o usuário pode obter uma resposta vazia (resultado de busca vazio<sup>20</sup>) à questão formulada. Procurando

---

<sup>20</sup> Essa característica significa que as categorias “vazias” não são apresentadas ao usuário, o que impossibilita um resultado também “vazio”, pois só aparece, para escolha, aquilo que, efetivamente, existe.

solucionar esse problema, Doan, Plaisant e Shneiderman (1995) desenvolveram um mecanismo que permitiu substituir a “busca por parâmetros” por uma navegação facetada, que foi implementado no portal da *National Aeronautics and Space Administration* (NASA), eliminando a possibilidade de um resultado de busca vazio. A NASA, ainda hoje, dispõe dessa navegação facetada em seu portal, que vem sendo sempre aperfeiçoado, desde esse primeiro mecanismo criado em 1995.

O segundo trabalho, também na área da Ciência da Computação, desenvolvido por Pollitt *et al.* (1996), foi um projeto de navegação facetada, denominado “busca baseada em visualização<sup>21</sup>”, chamado pelos autores de HIBROWSE (*High resolution Interface for BROWsing and SEarching*). O mecanismo foi criado para dois diferentes bancos de dados: para o EMBASE, que é a mais importante base de dados bibliográfica da área médica, e para o EPOQUE, uma biblioteca digital do Parlamento Europeu. O sistema foi desenvolvido tendo por base de conhecimento um tesouro, que pode-se denominar de “tesouro facetado”, uma vez que foi criado a partir das recomendações de Ranganathan e do *Classification Research Group* (CRG). O tesouro facetado foi empregado para indexar o recurso informacional e também como mecanismo de navegação. A deficiência desse projeto foi a grande extensão da estrutura do mecanismo de busca, que mostrava os termos mais gerais, os preferidos, os sinônimos, e os termos com relações associativas. Como exemplo, pode-se citar uma busca no EMBASE em que, se o usuário buscasse por uma doença, ele poderia indicar “alz” e o sistema responderia com uma lista para “alzheimer”. Nesse caso, o usuário teria como resultado 1297 itens (POLLITT *et al.*, 1996). A limitação apresentada pelo sistema é que ele não oferecia a possibilidade de iniciar a busca pela faceta “doenças”. Dessa forma, o resultado de busca apresentava itens sobre remédios e tratamentos alternativos, por exemplo. Para chegar à resposta que o usuário procurava, nesse caso do exemplo, era preciso fazer quatro operações de refinamento (POLLITT *et al.*, 1996). Apesar de tudo isso, sem dúvida, esse estudo representou um avanço para a época e facilitou bastante a recuperação de informações nessas bases de dados. Além disso, indicou o potencial das classificações facetadas para a estruturação do conteúdo de documentos e para melhorar a busca e a recuperação de informações em base de dados.

Um trabalho desenvolvido na Escola de Biblioteconomia e Ciência da Informação, da Universidade do Norte da Carolina, em *Chapel Hill*, com a participação do *Interaction Design Lab*, é o projeto denominado *Relation Browser*<sup>22</sup> (RB) ou RAVE (CAPRA; MARCHIONINI, 2008). Esse projeto foi iniciado em 1998-1999 e está em constante desenvolvimento e, em 2008, estava na versão RB07. O objetivo dessa ferramenta é

---

<sup>21</sup> *View-based searching systems.*

<sup>22</sup> Disponível em: <<http://idl89.ils.unc.edu/rb07>>.

oferecer ao usuário do *Fedstats*<sup>23</sup> o mapeamento dos tipos de informações disponíveis no acervo de portais de diversas agências federais americanas, oferecendo um mecanismo de busca e navegação. Dessa maneira, há a possibilidade de o usuário visualizar as interrelações entre os diferentes itens do acervo, oferecendo a exploração de todo o conteúdo informacional do banco de dados. Essa exploração é obtida através de uma interface dinâmica, na qual as informações estão estruturadas em facetas. A interface permite que o usuário faça sua busca diretamente nas facetas ou pela busca por palavras-chave. Esse projeto serve de laboratório para o desenvolvimento de novas ferramentas e para estudos de usabilidade, em relação ao sistema e aos usuários. O RB iniciou-se com a estruturação das informações em categorias e facetas, mas somente sob uma ou duas dimensões da informação disponibilizada. Com isso, o usuário podia buscar apenas em uma faceta por vez. Depois, o sistema passou a ser mais flexível na indicação dos relacionamentos entre categorias, com atribuição de maior quantidade de dimensões às informações disponibilizadas, mas ainda com navegação em apenas uma faceta por vez. Em sua última versão, o RB07, foi adicionado um mecanismo de filtro dinâmico, que suporta a exibição de múltiplas facetas, nas quais o usuário pode navegar usando a tecla “Tab” do teclado do computador. O usuário recebe o resultado de sua busca no formato de tabelas, tanto o resultado direto quanto a lista de resultados similares (relações associativas). Um dispositivo interessante adicionado ao novo modelo foi a possibilidade de o usuário saber o histórico de sua busca (facetadas e termos escolhidos), por meio dos “*breadcrumb*”. Além disso, também houve o acréscimo da funcionalidade de busca no texto completo, que pode ser combinada com a busca e navegação facetada. Dessa forma, o usuário recebe o resultado da questão formulada e vai, progressivamente, refinando sua busca, até chegar à informação que procura. As dificuldades encontradas no projeto RB, tais como a demora no processamento da informação ou capacidade de armazenagem de informação, estão todas relacionadas a soluções tecnológicas.

É preciso ressaltar que o projeto mais conhecido no âmbito de navegação e busca facetada é o *Flexible information Access using Metadata in Novel Combinations* (<<http://sourceforge.net/projects/flamenco>>) (FLAMENCO<sup>24</sup>), desenvolvido na Universidade da Califórnia, em Berkeley (HEARST, 2000), com uma equipe interdisciplinar, mas, basicamente, com membros da área da Ciência da Computação. Durante o experimento, segundo English (2002), a equipe desenvolveu duas interfaces: uma baseada em metadados hierárquicos, denominada *SingleTree*, e outra em metadados facetados, denominada *MatrixView*. O teste nas duas interfaces demonstrou a maior flexibilidade e

<sup>23</sup> Portal governamental que provê informações estatísticas de mais de 100 agências americanas (<<http://www.fedstats.gov/>>).

<sup>24</sup> Acesso a um “demo” do projeto: Disponível em: <<http://flamenco.berkeley.edu/demos.html>>.

também a preferência dos usuários pela *MatrixView*. A importância deste projeto, no âmbito desta pesquisa, é que o conteúdo do recurso informacional foi estruturado em diferentes categorias e em multi-hierarquias, isto é, sob cada categoria, há uma taxonomia hierárquica. Além disso, os constantes testes realizados com usuários demonstram que eles se satisfazem com a possibilidade de compor uma busca de forma fluida, com o refinamento progressivo. O FLAMENCO está há mais de dez anos sendo aprimorado, atualizado e mantido, desenvolvendo mecanismos de busca e navegação facetada, além de servir como laboratório para estudos de usabilidade com usuários finais. É importante, ainda, destacar que os resultados apresentados aos usuários não são ranqueados (*ranking*). A razão disso é que os metadados facetados (data, autor, título e qualificadores) permitem ao usuário escolher sob qual dimensão quer recuperar um documento, não havendo a necessidade desse tipo de interferência do sistema. O usuário pode, também, combinar a busca simples com a navegação facetada. Outro dado importante desse projeto, para o estudo realizado nesta pesquisa, é que as facetas e os metadados são organizados e determinados manualmente, por um engenheiro de sistemas, não sendo um sistema automático. Isso faz com que se acredite que a proposta desta pesquisa é exequível, pois os idealizadores do FLAMENCO afirmam que o modelo desenhado por eles, nessa etapa de desenvolvimento, pode suportar uma coleção de até 100.000 artigos (ENGLISH, 2002).

Iniciado em 2000 e com término em abril de 2003 tem-se o Projeto FACET<sup>25</sup>, coordenado por Douglas Tudhope, da Universidade de *Glamorgan*, que contou com o apoio do *National Museum of Science and Industry* (NMSI), de J. Paul Getty Trust (vocabulário), do *Museum Documentation Association* (MDA) e do *Canadian Heritage Information Network*. Neste projeto houve investigação sobre o uso de tesouros facetados na recuperação da informação, através de técnicas e interfaces de busca. O usuário não necessita formular sua busca por termos idênticos aos utilizados na indexação, uma vez que o sistema apresenta os resultados a partir de medidas de aproximação semântica, que são obtidos com base nos relacionamentos existentes no tesouro. Dessa forma, o usuário é automaticamente guiado na exploração dos termos do tesouro, podendo refinar ou expandir sua busca. Atualmente, os esforços deste projeto estão voltados para a digitalização de coleções para a *Web*, abrindo o horizonte para uma escala mais ampla de usuários. Além disso, estuda a possibilidade de integrar o projeto FACET à buscas em enciclopédias e às coleções multimídia.

Hyvönen *et al.* (2004), ainda na área da Ciência da Computação e Engenharia de *Software*, desenvolveram um portal semântico, o *MuseumFinland*, para publicar as coleções culturais, tais como artefatos têxteis, peças de mobília e ferramentas, na *Web*.

---

<sup>25</sup> Disponível em: <<http://www.comp.glam.ac.uk/~facet/>>.

Para ajudar no projeto, foi criado o *software OntoViews*, que é composto por três componentes: 1) o *Ontogator*, um motor de busca multifacetado, baseado em ontologias; 2) *Ontodella*, um servidor lógico, que armazena dados *Resource Description Framework* (RDF) apresentados no portal; 3) *OntoViews-C*, um sistema estruturado que transforma um RDF em dados navegáveis na páginas *Web*. Na versão piloto foram utilizadas as coleções do Museu Nacional<sup>26</sup>, do Museu da cidade de Espoo<sup>27</sup> e do Museu da cidade de Lahti<sup>28</sup>. Esses são museus de três cidades, e usam três distintos sistemas e bases de dados relacionais, além de utilizarem diferentes sistemas de gerenciamento de coleções (*Musketti*, *Escoll* e *Antikvaria*, respectivamente). Os metadados dos objetos das coleções (data, texto explicativo, etc.) foram atribuídos a partir dos 15 elementos do padrão *Dublin Core*<sup>29</sup>. Sempre que possível, foram atribuídas aos conteúdos palavras-chave, selecionadas de um vocabulário controlado ou tesauro, mantendo a padronização. Isso é importante para uma recuperação de informações mais acurada e para dar consistência à coleção. Apesar de não citarem Ranganathan em seus estudos sobre o método da análise facetada, o princípio de funcionamento do sistema é o mesmo. O conteúdo das coleções é sistematizado em categorias conceituais, e sob cada uma delas há uma hierarquia taxonômica, cujo conjunto é denominado “facetas” ou visões (*views*). Contudo, houve problemas relacionados com a transformação dos conteúdos das coleções em dados RDF, principalmente pautados em interpretações semânticas. Algumas expressões são complexas e originadas de diferentes campos de dados da coleção, podendo se referir a diferentes expressões sintáticas, dependendo do catálogo ou da anotação convencionalizada em uso. Esse problema pode ser minimizado pelo uso de padrões terminológicos, mas, segundo os autores, na prática, isso é impossível, pois há grande variação de descrições. Resta acrescentar, que, no *MuseumFinland*, a interface desenvolvida para o usuário final é similar a outro repositório qualquer na *Web*, pois possui um único *browser* de busca e navegação. Os autores conseguiram tornar as coleções heterogêneas e geograficamente distantes em um único sistema interoperável, através de um processo de *merging*. Cada museu pode acessar o sistema do portal e publicar seus conteúdos de forma fácil e independente.

Em âmbito governamental, de Sidoroff e Hyvönen (2005), na área da Ciência da Computação e Engenharia de *Software*, apresentaram um estudo de caso, no qual implementaram uma busca semântica (ou busca multifacetada) em portais governamentais da Finlândia, o *Suomi.fi*. No portal original, o conteúdo era apresentado em facetas, mas apenas indexado sob uma única dimensão, o que os autores denominaram “abordagem

<sup>26</sup> Disponível em: <<http://www.nba.fi>>.

<sup>27</sup> Disponível em: <<http://www.espoo.fi/museo/>>.

<sup>28</sup> Disponível em: <<http://www.lahti.fi/museot/>>.

<sup>29</sup> O padrão de metadados Dublin Core descreve recursos de informação multi-domínios, provendo um conjunto de convenções simples e padronizado para descrever objetos digitais, de modo a facilitar a sua localização.

*Yahoo!*<sup>30</sup>. Para os pesquisadores, a simples classificação taxonômica (hierárquica) dá uma visão unidimensional ao documento e não atende aos diferentes usos de que os cidadãos necessitam. O objetivo com o projeto foi o de agregar valor às informações disponibilizadas aos cidadãos, de forma que a recuperação do conteúdo pudesse ser otimizada. Nesse sentido, o projeto se baseou nos princípios da *Web* semântica, utilizando uma ontologia como base de conhecimento, e usando padrões *Resource Description Framework* (RDF) e *Web Ontology Language* (OWL). Dessa forma, o recurso informacional foi classificado sob diferentes dimensões ou categorias, e cada categoria é ligada a um conjunto de objetos (do recurso informacional). Com isso, ao formular uma busca, o usuário pode selecionar qualquer faceta dentro de uma categoria e o resultado apresentado será simplesmente a interseção<sup>30</sup> daquela faceta com as demais categorias disponíveis. Essa interseção é atribuída *a priori*, isto é, na entrada de dados já se especificou os relacionamentos entre os objetos, a partir dos objetivos e dos critérios pré-determinados. No caso desse trabalho, esclarece-se que os resultados apresentados ao usuário são *links* que direcionam para a informação procurada, e não para documentos. Além da navegação facetada, o portal também oferece a busca por palavras-chave, permitindo a combinação das duas. Todo o recurso informacional da *Suomi.fi* foi descrito usando um conjunto de ontologias, projetado dentro de múltiplas taxonomias na interface do usuário. Nessas taxonomias é que estão atribuídos os metadados que classificam o conteúdo do portal sob as distintas dimensões, dando suporte às diferentes escolhas do cidadão (usuário). A interface foi criada a partir do *OntoViews* que, segundo Sidoroff e Hyvönen (2005), é similar ao sistema utilizado no projeto FLAMENCO, pois, no lado esquerdo, o *menu* oferece as facetas disponíveis para a navegação e, a parte central, é reservada à apresentação dos itens de informação que combinam com a busca formulada. Finalmente, esse estudo concluiu que as facetas ajudam o usuário a recuperar informações de diferentes maneiras: oferece ampla visão de todo o conteúdo disponível no repositório; guia o usuário na formulação de buscas, fazendo-o utilizar os conceitos corretos; não admite problemas de homônimos; e nunca apresenta um resultado de busca vazio.

Suominen, Viljanen e Hyvönen (2007) criaram um protótipo, o *TerveSuomi.fi*<sup>31</sup> (ANEXO B), no qual há uma interface de busca (multifacetada), com uma equipe interdisciplinar na área de Ciência da Computação, Inteligência Artificial e Engenharia de *Software*. Nela, o usuário tem uma visão geral do conteúdo do portal e pode combinar um ou mais valores (facetas) na navegação. O avanço desse projeto, um esforço do grupo *Semantic Computing Research Group* (SeCo), criou vários portais, dentre os quais se

<sup>30</sup> É o mesmo princípio da lógica booleana *and*, pois apresenta os itens que, simultaneamente, pertencem a duas ou mais categorias

<sup>31</sup> Disponível em: <<http://demo.seco.tkk.fi/tervesuomi/home>>.

destaca o portal *HealthFinland*, que é um sistema semântico destinado a atender à população em geral. Ele é coordenado pelo *National Public Health Institute*, da Finlândia, e está no ar desde agosto de 2008. Em linhas gerais, o portal apresenta ao usuário final uma interface de busca facetada, oferecendo, atualmente, 21 categorias de conteúdo. Suominen (2009) descreve o portal *HealthFinland* como um sistema que utiliza tecnologia semântica para distribuir conteúdos e criar interoperabilidade, atribuindo valores (metadados) aos conteúdos automaticamente, como forma de minimizar o custo de cuidar da manutenção dos *links* e da atribuição de metadados referentes aos diferentes conteúdos. No portal *HealthFinland*, o sistema foi programado para ser capaz de alocar um novo termo em seu contexto correto no portal e fazer o *link* com o seu conteúdo específico, automaticamente. Ele utiliza um *software* colaborativo de criação de conteúdos sobre saúde, com o objetivo de evitar trabalhos replicados e redundantes dentro da própria instituição e entre diferentes instituições da rede. A interoperabilidade é obtida com a atribuição de metadados semânticos aos conteúdos, através de ontologias compartilhadas entre as instituições participantes. O diferencial desse projeto é que o vocabulário usado na interface do usuário final é diferente e independente do utilizado para indexar o conteúdo informacional (este último é indexado através de terminologia especializada). A ideia do protótipo foi implementada a partir de três elementos: 1) um vocabulário do domínio é usado para atribuir diferentes dimensões ao conteúdo informacional (assunto, gênero), que são publicadas no repositório de ontologias (*FinnONTO*); 2) um sistema de produção de conteúdo, que possui especificações e ferramentas que permitem fazer anotações, rastrear e verificar os conteúdos; 3) um portal semântico, no qual estão disponibilizados os conteúdos para uso humano, através da busca e navegação, e outro para as máquinas (ex: outras aplicações da *Web*). Deve-se acrescentar que outros estudos (STOICA; HEARST, 2004; STOICA; HEARST; RICHARDSON, 2007), também desenvolveram algoritmos para a criação de metadados facetados de forma automatizada como solução para a *Web*.

No campo da área de Sistemas e Ciência da Computação, no Brasil, Lachtim (2008), desenvolveu um mecanismo para portais semânticos denominado “Portal Semântico Educacional” (POSEDU). A finalidade desse projeto foi criar uma arquitetura capaz de gerar instâncias para a organização, integração e gestão dinâmica de informações advindas de diferentes instituições acadêmicas publicadas na *Web*, de forma a disponibilizá-las no portal. Para a conclusão do protótipo, foi utilizado como ferramenta de busca o *Yahoo* e o *Swoogle*, e, para o cálculo de similaridades, o *CROSI Mapping System* (CMS). Ademais, para a instanciação das informações, usou-se o gerador de portais *Semantic Web Environmental Directory* (SWED), e, para a criação da base de conhecimento, o editor de ontologias *Protegè*. O CMS foi estendido para atender a requisitos de eficiência que possibilitaram fazer o cotejamento entre os conceitos das ontologias recuperadas e os da ontologia base

do portal. Isso facilitou a transferência das informações dos diferentes locais na *Web* para o portal POSEDU. A partir dessas ferramentas, as informações foram estruturadas em facetadas, e, por meio de regras formais estipuladas na ontologia, o sistema forneceu ao usuário um mecanismo de busca e navegação facetada. A autora afirma em seu estudo que essa estratégia de busca e navegação facetada vem sendo empregada pela área de recuperação da informação, cujos princípios estão baseados na ideia de busca multifacetada proposta por Ranganathan. Segundo Lachtim (2008), o princípio de estruturas facetadas para representar informações é adequado para a *Web* semântica, pois oferece uma visão geral de todo o conteúdo do portal e guia o usuário em sua formulação de busca entre os termos e conceitos disponibilizados. A particularidade encontrada no POSEDU refere-se ao fato de que, para o sistema conseguir fazer a interseção entre as diferentes categorias, o usuário deve selecionar (marcar) aquelas que são de seu interesse, situação que não ocorre em outros exemplos apresentados (HEARST, 2000; HYVÖNEN *et al.*, 2004; SIDOROFF; HYVÖNEN, 2005).

Ainda no Brasil, na área da Informática, existe um protótipo para navegação facetada, desenvolvido por Araújo (2009), denominado *Explorator*. Trata-se de um mecanismo que permite a navegação e busca, construído com base em linguagem RDF. O autor criou um modelo de linguagem de especificação, denominado “vocabulário FACETO”, além de um algoritmo de relevância para a geração automática de facetadas, com o objetivo de possibilitar a criação de mecanismos de navegação facetada sobre dados *Resource Description Framework* (RDF<sup>32</sup>). A linguagem de especificação para extração das facetadas é independente do modelo de navegação facetada, que aceita visualizar operações em um conjunto de triplas do tipo diferença, interseção e união<sup>33</sup>. O arquétipo permite, ainda, operações *Subject, Predicate, Object* (SPO<sup>34</sup>) de consultas em uma base RDF, o que fornece exemplos de soluções para problemas, auxiliando o usuário a aprender sobre o domínio de dados RDF. Assim, é possível definir as facetadas a partir da combinação de valores e propriedades do modelo RDF. Uma limitação do *Explorator* é que a escolha automática das facetadas para a taxonomia não leva em consideração as relações definidas na ontologia dos dados RDF. E a principal contribuição do *Explorator* para o usuário é sua capacidade de permitir que ele manipule os recursos e triplas diretamente na interface de navegação facetada. Em ambientes complexos como os da *Web*, a determinação automática de facetadas parece ser a única solução provável, o que não se aplica ao ambiente restrito da BDTD utilizado nesta pesquisa. Entretanto, é interessante verificar o uso da

---

<sup>32</sup> RDF é uma linguagem para representar informação na Internet.

<sup>33</sup> A interseção e união se baseiam na lógica booleana *and* e *or*.

<sup>34</sup> SPO consiste em três elementos RDF: assunto, predicado e objeto. É uma tripla simples de uma *query language* (linguagem de computador para criar bancos de dados relacionais e manipular dados).

navegação a partir de taxonomias facetadas como solução para a busca de informação semi-estruturada.

Outro trabalho brasileiro que aborda o tema é o de Heuseler (2010), no Núcleo de Computação Eletrônica, no Instituto Militar, que desenvolveu um protótipo que foi aplicado no domínio de clínicas médicas. O autor criou diferentes mecanismos genéricos (baseados na leitura de arquivos *Extensible Markup Language* - XML), de domínio público, que podem ser utilizados em distintos domínios de informação, usando taxonomias facetadas como mecanismo exploratório. Na criação do protótipo, Heuseler procurou adaptar seu mecanismo, que trata informações não-estruturadas, às dimensões de um *Data Warehouse* (DW), que já são utilizadas em operações analíticas sobre dados de natureza estruturada. O autor criou uma interface própria com o DW, mas, segundo ele, adicionando-se uma ferramenta adaptada da *On-line Analytical Processing* (OLAP), as lacunas referentes a funcionalidades analíticas mapeadas em seu estudo podem ser suprimidas. O protótipo foi testado, e, como resultado de sua arquitetura, Heuseler (2010) conseguiu que o usuário de sistemas analíticos pudesse ter uma visão geral e conjunta de informações estruturadas e não-estruturadas, através da exploração pela navegação nos termos da taxonomia facetada. Essa composição entre os dois tipos de informação (estruturada e não estruturada) foi possível com a utilização de arquivos de configuração em XML e mecanismos de leitura do modelo dimensional, que podem se adaptar a outros tipos de domínio e, assim, para o domínio abordado nesta pesquisa. As informações geradas pelo processo de extração dos termos dos documentos são armazenadas em um repositório de metadados, desenvolvido especificamente para a proposta do protótipo, sendo esses os termos que alimentam a taxonomia facetada e nos quais o usuário pode explorar o recurso informacional da base. Um problema que o autor não conseguiu resolver foi em relação aos sinônimos, pois o mecanismo apresenta como resultado somente respostas idênticas ao termo buscado. Entretanto, o uso de um vocabulário controlado para o processo de extração de termos dos documentos e a implementação de uma base de conhecimento por trás do banco de dados pode solucionar esse problema.

No cenário de novas tecnologias, é significativo constatar que, quando são desenvolvidos mecanismos para ajudar o usuário em sua busca por informações, é preciso, fundamentalmente, pensar em estudos de usuário, como o realizado por Fagan (2010), que fez uma revisão de literatura que incluiu estudos sobre a avaliação de interface com navegação facetada. Em um dos trabalhos, por exemplo, Pratt, Hearst e Fagan (1999) realizaram um estudo de usuários com 50 pessoas, familiares e pacientes com câncer de mama, buscando por informações médicas sobre a doença. Eles determinaram três perguntas sobre câncer de mama para serem buscadas na *Web*, a partir da utilização de

três ferramentas para visualização de resultados de busca: por *clustering* (SONIA<sup>35</sup>) de resultados, por *rank*<sup>36</sup> de resultados e por navegação facetada (*DynaCat*<sup>37</sup>). O estudo concluiu que há larga preferência pela interface com navegação facetada, o que favorece a ampliação desse campo de estudo na CI. Segundo o autor, todos os 50 participantes afirmaram que a navegação facetada oferece uma visão geral do todo, dando sentido semântico ao resultado e, ainda, que é útil, pois possui rótulos objetivos e claros. A partir da análise dos dados coletados no estudo de usabilidade dessa interface de busca, Fagan (2010) demonstrou que a satisfação dos usuários no uso desse tipo de busca tem relação com o seguinte: facetas são úteis para estruturar a navegação; a categorização em facetas facilita e torna a recuperação de informações mais eficiente; a navegação facetada elimina resultados vazios; os usuários encontram a informação mais rapidamente quando usando sistemas facetados; e os usuários preferem que os resultados de suas buscas sejam organizados em hierarquias multidimensionais. O autor afirma, também, que sua revisão indicou que muitas bibliotecas estão pensando em uma nova forma de apresentar seus catálogos, em interfaces facetadas, para modernizar o visual dos catálogos e para melhorar a recuperação de informações (tais como as de artigos em suas bases de dados), o que pode ser mais detalhadamente conhecido em Anderson e Hofmann (2006), Calhoun (2006), Breeding (2007) e Olson (2007). Nas novas interfaces dos catálogos, denominadas “nova-geração de catálogos de bibliotecas”, as facetas<sup>38</sup> geralmente incluem campos para autor, assuntos e formatos, além de outros campos criados a partir do padrão *Machine Readable Cataloging* (MARC). Percebe-se assim, que há bastante espaço para a implementação de navegação facetada com vistas a aprimorar a busca e a recuperação de informações.

Vale ainda destacar dois *software* para aplicações na *Web* semântica, que dispõem de navegação facetada, permitindo a personalização da interface pelo usuário (definição do *layout* da página): o *Simile*<sup>39</sup>-*Longwell* e o *FacetMap*. O *Longwell* foi desenvolvido pelo *Massachusetts Institute of Technology* (MIT) e é um aplicativo para navegação facetada e manipulação de dados RDF. O *Longwell* utiliza formato proprietário para armazenamento dos dados, com arquivos de entrada RDF, XML, N3 ou NT<sup>40</sup>. Ele foi desenvolvido em linguagem de programação JAVA, e permite vários níveis de personalização da interface com o usuário. Possui um modelo para definir as facetas automaticamente, suportando a meta-linguagem *Fresnel*, que aceita a seleção e formatação

<sup>35</sup> Ferramenta que cria um conjunto de grupos-padrão de termos, em hierarquias, que são refinados por um método interativo.

<sup>36</sup> Foi utilizado um algoritmo padrão recomendado por Salton e Buckley (1988) para vocabulários técnicos.

<sup>37</sup> Ferramenta para automatizar a categorização de documentos através de critérios pré-determinados.

<sup>38</sup> O termo “facetadas” no âmbito dos trabalhos com os catálogos de biblioteca tem o sentido de “atributos” do documento, que incluem elementos referentes à descrição física do documento (autor, título, etc.) e também elementos referentes ao assunto do documento, aludindo aos campos existentes no MARC e no padrão Dublin Core.

<sup>39</sup> *Semantic Interoperability of Metadata and Information in unLike Environments*.

<sup>40</sup> *New Technology*.

de um conjunto de triplas RDF. Já o segundo, o *FacetMap*, foi desenvolvido por Travis Wilson<sup>41</sup>, acerca de dez anos, e é um programa para navegação em facetas, escrito em linguagem *Exchangeable Faceted Metadata Language* (XFML), com especificações próprias para a geração de facetas e armazenamento de dados, que permite visualizar como as informações estão conectadas. A personalização é obtida através de estilos *Extensible Stylesheet Language for Transformation* (XSLT), e não suporta RDF. Em 2006, uma equipe da *Microsoft* desenvolveu um *software* com o mesmo nome *FacetMap*, e a questão do uso do nome, patentado para Travis Wilson, está em litígio.

Por fim, também com certo destaque, podem-se citar três livros que abordam o tema em questão e que foram utilizados nesta pesquisa:

1) *Organising Knowledge: taxonomies, knowledge and organisational effectiveness*, de Patrick Lambe (2007), um estudioso no tema taxonomia. O livro é dedicado, quase exclusivamente, às taxonomias, indicando conceitos e distintas formas de apresentação. O autor apresenta a taxonomia como solução em ambientes organizacionais, provendo estruturação e gerenciamento de informações, oferecendo diferentes estudos de casos de aplicações de taxonomias em ambiente institucional. Proporciona um guia completo para a implementação de projetos de taxonomias em organizações, relatando a negociação com o alto escalão da empresa, os planos de abordagem e motivação dos auxiliares, a coleta da terminologia de especialidade até a implantação, propriamente dita, e sua integração ao sistema geral da organização. Ao final, discute outros temas adjacentes, tais como ontologias e folksonomias.

2) *Faceted Search*, de Tunkelang (2009), apresenta a busca facetada sob o aspecto histórico, teórico e prático, de forma bem didática e sucinta, dando um panorama geral de como o tema vem sendo abordado e utilizado, indicando uma extensa bibliografia sobre cada ponto discutido. O autor discorre sobre os princípios de Ranganathan, incluindo classificações, facetas e ontologias, além do sistema de recuperação da informação, com conceitos de relevância, precisão e revocação. Distingue os conceitos de busca por parâmetros, navegação facetada e busca facetada, mostrando alguns exemplos do uso de buscas facetadas em distintas iniciativas. Ao final do livro, o autor apresenta práticas técnicas para a implementação de sistemas facetados, com discussões acerca de temas como escala, eficiência, metadados, vocabulário controlado, padrões de modelagem, múltiplas entidades, valores e ordenação de facetas.

3) *Dynamic Taxonomies and Faceted Search: theory, practice and experience*, de Sacco e Tzitzikas (2009), é um referencial para os estudos sobre os temas de taxonomias dinâmicas e as buscas facetadas sob o paradigma de taxonomias

---

<sup>41</sup> Informação disponível em: <<http://facetmap.com/pub/>>.

multidimensionais, geralmente organizadas através de facetas. Os autores abordam a modelagem da informação, a interação do usuário, o desenho da taxonomia, a implementação do mecanismo facetado e o seu desempenho. Inclui, ainda, comparações com outras técnicas de recuperação de informação, tais como resultados por *clustering*, taxonomias estáticas e descrições lógicas, principalmente em relação à *Web* semântica. Oferecem, também, instruções técnicas e tecnológicas para a criação de modelos, inclusive exemplos de solução para manter a integração de diferentes fontes de informação. Além disso, proporcionam um manual para a construção automática de taxonomias dinâmicas, a partir de textos, e analisam as áreas nas quais a aplicação dessa solução vem sendo utilizadas, apresentando os desafios que deverão ser enfrentados.

Tendo em vista o que foi exposto, no capítulo seguinte discute-se a fundamentação teórico-metodológica, o que, certamente, constitui o alicerce para as análises pretendidas nesta pesquisa.

## 4 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA

Neste capítulo estão apresentadas a fundamentação teórica e a metodológica utilizadas para o desenvolvimento desta pesquisa. Toda a fundamentação se baseou em conhecimentos relacionados com a busca em sistemas de recuperação da informação (SRI), iniciando com uma breve discussão sobre instrumentos de organização e representação da informação, aqui denominados de “vocabulários controlados”. Aprofunda-se sobre o tema do vocabulário controlado do tipo taxonomia, que é o objeto de estudo desta pesquisa. Nesse sentido, apresenta a sua forma mais tradicional, que é a hierárquica, passando à forma facetada e, finalmente, à facetada navegacional. É preciso esclarecer que, ao apresentar os fundamentos teóricos da taxonomia facetada, indicam-se também os alicerces metodológicos de sua criação, uma vez que teoria e metodologia estão intrinsecamente ligadas. Em seguida, é abordado o assunto da recuperação da informação em SRIs, sendo apresentado um breve relato sobre as opções de buscas, passando para as bibliotecas digitais e focando o olhar sobre as bibliotecas de teses e dissertações, que é a ambientação desta pesquisa. Depois, descreve-se a navegação hipertextual, mostrando a navegação pela taxonomia facetada. Serão discutidos, ainda, os fundamentos da teoria da análise de domínio (TAD), indicando sua metodologia e finalizando com a identificação da comunidade científica. Além disso, houve a exposição do método da análise de conteúdo, do método utilizado para o desenvolvimento do Modelo de Leitura Documentária de Fujita e Rubi (2006). E, por fim, foram caracterizados os documentos referentes a trabalhos acadêmicos do tipo teses e dissertações, que são de interesse desta pesquisa, demonstrando a sua estrutura textual.

### 4.1 Vocabulário controlado: a taxonomia – da hierárquica à facetada

Na literatura da área, encontram-se vários sinônimos para o termo “vocabulário controlado”, que para alguns autores é “linguagem documentária (LD)” (LIMA *et al.*, 2006; LIMA; BOCCATO, 2009). Uma LD, por sua vez, é igualmente designada de “linguagem controlada” (LOPES, 2002). Segundo Wanderley (1973, p.180), Melton (1964) denomina LD de “linguagens de indexação”, B. Vickery (1970) de “linguagens descritoras” e Grolier (1962) de “codificações documentárias”. Além disso, Soergel (1997) considera “linguagem de informação”, “linguagem de recuperação” e “linguagem descritora” como termos sinônimos de LD.

Um vocabulário controlado (VC) é uma linguagem artificialmente desenvolvida, composto por termos que podem ser organizados em uma estrutura relacional ou alfabeticamente. Currás (1995) declara que um VC “é uma lista de termos elaborada para

identificar o assunto ou os assuntos de um documento com especificidade bastante para permitir sua recuperação rápida e eficaz”. Com essa afirmação, entende-se que um VC, de fato, permite controlar a terminologia de um domínio, instituindo regras para seu uso, seu compartilhamento e sua expansão. Para Smit e Kobashi (2003, p.14), esse controle possibilita organizar e recuperar informações de forma consistente, atribuindo confiança ao sistema de recuperação de informação (SRI).

Nesse contexto, observa-se que a norma ANSI Z39.19 (2005), *Guidelines for the construction, format, and management of monolingual thesauri*, desenvolvida pela *National Information Standards Organization* (NISO), que sugere regras para a criação de padrões de conteúdos, apresentação, testes, manutenção e gerenciamento de VC, o descreve como uma lista controlada de termos explicitamente listados, os quais devem ter uma definição clara e inequívoca, dentro do propósito para o qual foram criados. Assim, o seu objetivo primordial é o de organizar a informação. Os regulamentos da Norma Z39.19 são especialmente desenvolvidas para a criação de VCs para uso em pós-coordenação em sistemas de recuperação da informação, mas podem também ser utilizadas para a criação de VCs para compor sistemas de recuperação de informação (SRI) com pré-coordenação, tais como para sistemas de navegação na *web*, mecanismos de busca ou taxonomias navegáveis.

Outro ponto importante a esclarecer, é que os VCs devem ser constantemente atualizados no intuito de adequar os termos utilizados e de acolher novos termos. Essa prática, segundo a norma ANSI/NISO Z39.19 (2005), compreende um método empírico, que tem abordagem indutiva, que são características de instrumentos criados *ad hoc*<sup>42</sup>. Nessa abordagem, novos termos encontrados no conteúdo dos documentos são potenciais candidatos à inclusão no vocabulário controlado e, se admitidos, são inseridos nas classes correspondentes. Essa é uma operação constante nesse tipo de vocabulário controlado (ANSI/NISO Z39.19, 2005, p.92).

A fim de avançar neste debate, é preciso lembrar que os VCs possuem cinco funções, quais sejam: 1) Tradução: na etapa da indexação, a linguagem natural dos autores é convertida para a linguagem utilizada no sistema, fazendo a compatibilização das duas linguagens, o que favorece a recuperação de informações; 2) Consistência: promove a padronização entre o formato do termo e sua atribuição; 3) Indicação dos relacionamentos: indica o relacionamento semântico entre os termos; 4) Campo e busca: fornece hierarquias consistentes e claras para o sistema de navegação, ajudando o usuário a recuperar informações de sua necessidade; 5) Recuperação: usado como mecanismo de busca para localizar conteúdos informacionais (ANSI/NISO Z39.19, 2005, p.11). Em síntese, pode-se

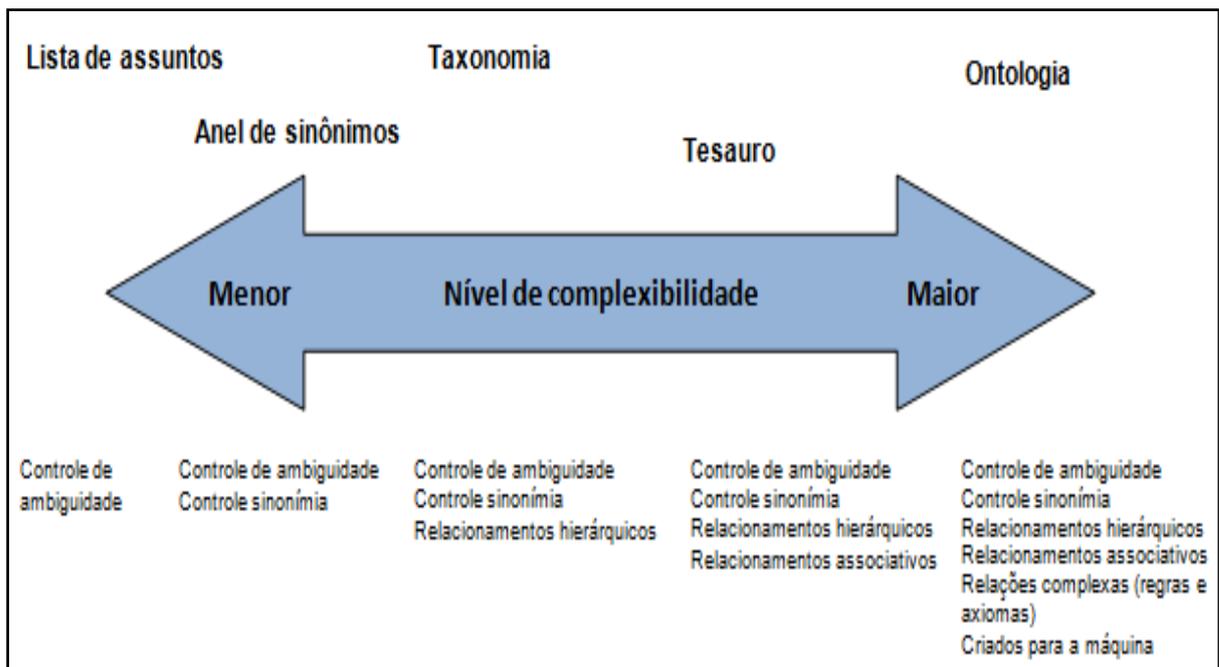
---

<sup>42</sup> Para uma finalidade específica.

entender que a principal função do VC é padronizar a entrada de dados em um sistema de recuperação da informação (SRI), de forma a trazer maior eficácia e precisão à saída do sistema, isto é, maior satisfação do usuário final.

Vale ressaltar que, segundo a norma ANSI/NISO Z39.19 (2005), há diferentes tipos de VC, que são os seguintes: lista de assuntos, anel de sinônimos<sup>43</sup>, tesouro e taxonomia, cujo grau de complexidade de controle e de representação pode ser observado na FIG. 1.

FIGURA 1 – Grau de complexibilidade em diferentes vocabulários controlados



Fonte: Traduzido de ANSI/NISO Z39.19 (2005, p.17).

Observa-se que o VC de tipo taxonomia está, em nível de complexidade, entre uma simples lista de assuntos e uma ontologia. Este último é um instrumento de uso exclusivo em ambiente digital, criado para ser “legível” pela máquina, capaz de proporcionar inferências a partir das relações atribuídas entre seus diferentes elementos. Como já antecipado, na presente pesquisa, toma-se como base o VC de tipo taxonomia que, segundo definição na norma citada, “é uma coleção de termos de um vocabulário controlado, organizados em uma estrutura hierárquica [na qual] cada termo [...] está em uma ou mais relações tipo pai/filho (geral/específico) em relação a outro termo nessa taxonomia” (p.9) ou “é um vocabulário controlado que consiste em termos preferidos, conectados em uma hierarquia ou poli-hierarquia” (p.18).

<sup>43</sup> Anel de sinônimos (*synonym ring*) é um grupo de termos considerados sinônimos, que têm o propósito de recuperação.

Quanto à estrutura deste tipo de VC, a norma acrescenta que “uma taxonomia é um vocabulário controlado constituído por termos preferidos, relacionados entre si, em hierarquia ou poli-hierarquia” (ANSI/NISO Z39.19, 2005, p.18). Acerca de seu uso, as taxonomias, geralmente, “são criadas e utilizadas em atividades de indexação e para navegação na *Web*, porque sua estrutura, geralmente simples estruturas hierárquicas, são eficientes em orientar o usuário, levando-o aos termos mais específicos em certo domínio” (p.136). Em linhas gerais, pode-se afirmar que uma taxonomia diz respeito a um conjunto de termos estruturados, recorrentemente, de forma hierárquica, representando o domínio no qual é aplicada, através da estruturação do mesmo.

Através de um olhar etimológico, a palavra *taxonomia* vem do verbo grego *tassein*, que significa “arranjos”, e *nomos*, que se refere à “lei”. Observa-se, ainda, que o termo foi empregado pela primeira vez em 1778, por Augustin Pyrame de Candolle, um biólogo suíço, especializado em botânica. Nesta ocasião, Candolle classificou e ordenou todas as espécies vegetais conhecidas, descrevendo mais de 80.000 espécies (BARQUÍN; MOREIRO GONZÁLEZ; PINTO, 2006). Entretanto, segundo esses mesmos autores, a mais conhecida taxonomia surgiu também no século XVIII, na área da biologia, com Linnaeus (Taxonomia de Lineu), para classificar organismos vivos. Seguindo tal propósito, Linnaeus criou um sistema hierárquico com uma nomenclatura binominal dos seres vivos, identificando, de forma unívoca, as espécies da seguinte maneira: filos, classes, ordens, famílias, gêneros e espécies. Por meio dessa classificação, Linnaeus resolveu os problemas de comunicação produzidos pela variedade de nomes locais.

Contudo, Tunkelang (2009, p.3) considera Aristóteles como o taxonomista pioneiro, uma vez que esse filósofo grego tem significativo papel na organização do conhecimento em hierarquias, pois criou um sistema de classificação que dividiu os organismos vivos em dois grupos: “plantas” e “animais”; sendo que os “animais” foram divididos em “com sangue” e “sem sangue”, os “com sangue”, em “mamíferos” e “ovíparos” e, assim, sucessivamente.

Entende-se que uma das primeiras definições para taxonomia pode ser a de Vickery, que a identifica como um instrumento para “organizar todas as entidades de um universo em uma simples hierarquia: uma grande árvore na qual cada entidade pertence apenas a uma única classe, cada espécie a um único gênero, e assim por diante” (VICKERY, 1975, p.10). Desde essa primeira definição o conceito de taxonomia vem sofrendo transformações. Na contemporaneidade, o termo é aplicado em sentido mais abrangente, de modo que a taxonomia é considerada como um instrumento de recuperação de informação, com possibilidade de agregar dados (metadados) na estrutura:

a taxonomia, em um sentido amplo, é a criação da estrutura (ordem) e dos rótulos (nomes) que ajudam a localizar a informação relevante. Em um sentido mais específico, é o ordenamento e rotulação de metadados, que

permite organizar sistematicamente a informação primária (MARTINEZ *et al.*, 2004, p.106).

Com base nisso, é possível a constatação de que a taxonomia tem apresentado destaque no meio digital, exercendo importante papel na representação, com a concepção de rótulos para os conceitos, e na organização das informações, a partir da criação de uma estrutura que dá ordenação aos rótulos. Essa característica auxilia a recuperação das informações porque apresenta ao usuário todo o conteúdo de forma lógica.

Não há mais como aceitar o termo “taxonomia” rigidamente relacionado a estruturas puramente hierárquicas. Nesse sentido, Campos e Gomes (2008, *on-line*) afirmam que

as taxonomias atualmente são estruturas classificatórias que têm por finalidade servir de instrumento para a organização e recuperação de informação em empresas e instituições. Estão sendo vistas como meios de acesso atuando como mapas conceituais dos tópicos explorados em um serviço de recuperação (CAMPOS; GOMES, 2008, *on-line*).

Dessa maneira, apesar de a literatura indicar, com mais frequência, a taxonomia hierárquica, é possível encontrar outros tipos de taxonomias sugeridos na literatura, principalmente no meio corporativo. Patrick Lambe, por exemplo, é um estudioso no tema taxonomia e, em seu livro *Organising Knowledge: taxonomies, knowledge and organisational effectiveness*, de 2007, apresenta alguns tipos de estruturas, além da hierárquica, sendo elas: listas, árvores, hierarquias, poli-hierarquias, matrizes, sistemas de mapas e facetas.

A partir dessa tipologia, neste trabalho será utilizada a taxonomia facetada que, segundo Lambe (2007), consiste na classificação expressa do conteúdo de um documento sob diferentes dimensões, ou seja, em facetas. Nesse caso, cada faceta na taxonomia é uma dimensão diferente ou um campo de registro de metadado distinto, que pertence a uma categoria específica, pois elas têm a característica de serem mutuamente exclusivas. Dessa forma, o conteúdo de um documento não está restrito a uma única dimensão, oferecendo diferentes opções de busca ao usuário, o que facilita a recuperação de informações. Em síntese, uma taxonomia facetada é composta por um conjunto de taxonomias e cada uma dessas taxonomias descreve o domínio (documentos) sob um aspecto (dimensão) diferente (TZITZIKAS *et al.*, 2004).

Segundo Terra (2005), a taxonomia facetada é particularmente útil em: 1) áreas emergentes de conhecimento ou quando os relacionamentos entre os termos ainda não podem ser completamente definidos; 2) áreas interdisciplinares; 3) ambientes digitais (TERRA, 2005). Em ambientes digitais, a estrutura da taxonomia facetada pode ser usada para navegação, dando ao usuário uma visão geral do conteúdo do banco de dados, sendo que cada faceta ou subfaceta pode ser escolhida para busca.

A possibilidade de ser navegável torna a estrutura facetada da taxonomia uma importante ferramenta para estruturar as informações semi-estruturadas de documentos do

tipo teses e dissertações. Isso pode ser explicado pelo fato de que, segundo Lambe (2007), Ranganathan construiu um esquema de classificação facetado que é capaz de expressar, com precisão, o conteúdo dos documentos, por meio de blocos de facetas, que estruturam suas partes e conceitos. Essa classificação é um esquema analítico-sintético, porque envolve dois processos distintos: a análise do assunto em facetas e a síntese dos elementos que constituem o mesmo, decompondo os conceitos mais complexos em simples. A síntese, em seu turno, é a recombinação das partes para a classificação, de acordo com as características do documento a ser descrito e representado (LAMBE, 2007). Além disso, essa abordagem permite ainda a inclusão de novos conceitos, sem que seja necessário mudar toda a estrutura do sistema classificatório.

Vale ressaltar que a literatura aponta que há, basicamente, dois importantes estudos sobre o tema: 1) os estudos de Ranganathan, cuja teoria possui 46 cânones, 13 postulados e 22 princípios, e está apresentada, basicamente, em cinco obras: *Five Laws of Library Science*, 1931, *Colon Classification*, 1933, *Prolegomena to Library Classification*, 1937, *Philosophy of Book Classification*, 1951, e *Elements of Library Classification*, 1962; 2) os estudos do *Classification Research Group* (CRG), criado na década de 50, no Reino Unido, incluindo catorze integrantes: 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 (LIMA, 2004a).

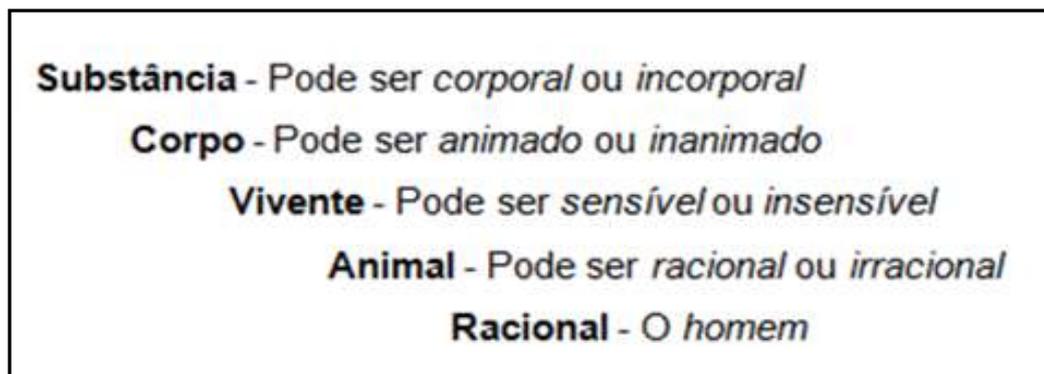
No contexto desses estudos, sobre a análise facetada, há conceitos que devem ser entendidos, tais como: 1) *isolado*: cada componente (termo/indivíduo) simples de um assunto, obtido a partir da divisão de uma faceta, mas ainda apartado da estrutura; 2) *foco*: é um *isolado* já acomodado na estrutura facetada, porém, sem preocupação sobre a relação que tem com outros componentes (termo/indivíduo) da estrutura; 3) *subfaceta* (ou *arrays*): grupos de termos coordenados, obtidos com a divisão de um assunto por meio de um mesmo princípio, sendo, mutuamente, exclusivos; 4) *faceta*: “qualquer componente – assunto básico ou isolado – de um assunto composto. [...] conjunto das subclasses obtidas pela divisão de uma classe, de acordo com determinada característica. Dentro da faceta, o membro individual, ou subclasse, denomina-se foco” (CUNHA; CAVALCANTI, 2008); faceta pode, também, ser entendida como um ponto de vista, categoria ou atributo usado para agrupar conceitos em uma área de assunto (domínio); 5) *categoria*: classe mais geral do sistema, as categorias fundamentais, usada para representar ideias fundamentais; 6) *divisão*: o processo pelo qual uma faceta se decompõe em diferentes focos, tal como a decomposição de um gênero em espécie; 7) *Renques*: divisão feita a partir de apenas uma característica, em uma série horizontal de conceitos; e 8) *Cadeias*: divisões sucessivas de um mesmo assunto, em uma série vertical de conceitos.

Acerca da teoria da análise facetada (TAF), Ranganathan propõe que se analise um dado domínio de conhecimento sob um princípio classificatório que não mais se pauta em uma metodologia dicotômico-binária, chamada de “Árvore de Porfírio”, mas, sim, numa policotomia<sup>44</sup> ilimitada, a “Árvore Baniana”.

Nessa direção, cabe acrescentar que Porfírio foi um filósofo neoplatônico, e, em sua obra *Introductio in Praedicamenta* (ISAGOGE), realizou comentários sobre a obra *Categorias*, de Aristóteles, quando desfaz o conceito de “categorias”, reduzindo-o, apenas, a gênero/espécie. Assim, Porfírio sistematizou a noção de ideias gerais até alcançar as mais específicas (gênero, espécie, diferença, próprio e acidente), o que originou a “Árvore de Porfírio”. De certo, essa árvore ilustra a classificação lógica da substância e pode ser considerada a precursora das classificações taxonômicas.

De uma forma simplista, o esquema pode ser apresentado conforme FIG. 2.

FIGURA 2 – Esquema da “Árvore de Porfírio”



Fonte: Elaborada pela autora, baseado no esquema de Porfírio.

Ranganathan, por sua vez, usou a ideia da “Árvore Baniana” (*banyan tree*), um tipo de figueira indiana (FIG. 3), que se espalha por uma grande área, enviando galhos para o solo. Os galhos criam raízes e formam vários troncos. Essas raízes são aéreas e crescem do tronco principal, tornando-se novos troncos adicionais. É importante destacar que, em uma árvore baniana, não há apenas relacionamentos hierárquicos, pois essa árvore sugere a ideia de que os assuntos (espécies de categorias) podem ser relacionados uns aos outros de diferentes, complexas e imprevistas formas, gerando novas categorias (RANGANATHAN, 1967, p.368), em um sentido multidimensional.

A fim de entender melhor o modelo baniano, verifica-se que ele configura uma representação simbólica para um conhecimento, como um organismo vivo, o qual está em constante crescimento e, portanto, sofre ramificações, transformações e desenvolvimento através do tempo.

<sup>44</sup> Sujeito a um ilimitado número de divisões das áreas do conhecimento.

FIGURA 3 – Ilustrações de Árvore Baniana



Fonte: Banco de imagens Google. Disponível em: <[www.thefreedictionary.com/banians](http://www.thefreedictionary.com/banians)>.

Dessa forma, os métodos de divisão para a organização do conhecimento são feitos a partir de categorias pensadas em sua abrangência conceitual, a qual não está relacionada somente a um núcleo, mas a diversos, dependendo da forma como as unidades de conhecimento se relacionam. Assim, o conhecimento caracteriza-se como multidirecional e multidimensional, pois pode se desenvolver em diferentes direções, com fertilizações cruzadas, multidisciplinares ou mesmo interdisciplinares. O sentido de multidirecional indica que o conhecimento pode ser orientado a diferentes propósitos (lazer, atividade profissional, credos), sejam eles convergentes ou divergentes. Já o sentido de multidimensional, significa que o conhecimento possui diferentes dimensões em relação espaço-tempo, que permite que um objeto possa ser observado sob mais de um aspecto, resultando em uma relação entre dimensões (por exemplo: teorias ou processos) e fatos, possibilitando combinações.

Ao desenvolver seu sistema facetado, Ranganathan adotou o sinal de dois pontos (:) para inserir uma faceta em outra, permitindo, assim, a inter-relação entre conceitos, o que desfez a abordagem, até então adotada, de sistemas hierárquicos. Essa opção se explica pelo fato de que o autor acreditava que o conhecimento não era estático e, sim, dinâmico, podendo gerar novos assuntos, por meio de: a) laminação; b) reunião livre de termos; c) desnudação; d) dissecação; e) interpolação. Para o arranjo de elementos nas facetas, forneceu, ainda, 18 princípios chamados de “princípios para sequência útil” (DAHLBERG, 1979). Essa mesma autora afirma que Ranganathan “matematizou” a classificação, utilizando essa ciência exata de forma qualitativa. A partir da aplicação de sua metodologia, Ranganathan criou a *Colon Classification*.

Com vistas ao trabalho de análise para a classificação, Ranganathan introduziu três níveis distintos: 1) plano da ideia: nível das ideias e conceitos; 2) plano verbal: nível da expressão verbal dos conceitos; 3) plano notacional: nível da fixação dos conceitos em formas tais como sinais, letras, etc. Para acomodar todo o conhecimento de um domínio, Ranganathan concebeu um conjunto de categorias fundamentais, que é denominado PMEST: Personalidade, Matéria, Energia, Espaço e Tempo.

Entretanto, o CRG divergiu da teoria de Ranganathan nesse ponto, afirmando que as categorias fundamentais deveriam ser identificadas e derivadas a partir do domínio a ser classificado e não *a priori*. Para esse grupo de pesquisa, o PMEST é muito restritivo, o que prejudica a flexibilidade do sistema classificatório e dificulta seu uso. Em contrapartida, as categorias sugeridas pelo CRG são as seguintes: Tipo de Produto Final, Partes, Materiais, Propriedade, Processo, Operação, Agente, Espaço, Tempo e Forma de Apresentação (LIMA, 2004b). Não obstante, observa-se que as categorias indicadas pelo CRG parecem se acomodar dentro das categorias indicadas por Ranganathan, sendo apenas mais segmentadas e subdivididas. O QUADRO 1 apresenta a visão pessoal desta pesquisa.

QUADRO 1 – Comparação entre as categorias de Ranganathan e CRG

RANGANATHAN	CRG
Personalidade	Tipos de Produto Final
Matéria	Partes Materiais Propriedades
Energia	Processos Operações Agentes
Espaço	Espaço
Tempo	Tempo

Fonte: Elaborado pela autora.

Para analisar o QUADRO 1 partiu-se da consideração de que as categorias fundamentais, tanto de Ranganathan quanto do CRG, estão a segmentar o conteúdo de um documento, considerando-o como um domínio em particular. Pondera-se, nesse sentido, que todo domínio possui, na visão de Ranganathan, as categorias: “Personalidade”, que é o assunto ou objeto; “Matéria”, que é a decomposição desse assunto ou objeto em diferentes atributos ou características (materiais, propriedades, qualidade); “Energia”, que são as ações, atividades ou processos em relação ao assunto ou objeto investigado; “Espaço”, que é a ambientação do assunto ou objeto, tal como seu lugar geográfico; e “Tempo”, que geralmente é a data ou período histórico-contextual do objeto estudado, ou mesmo o ano de

produção de tal conhecimento. As categorias “Espaço” e “Tempo” são iguais e por isso não causam polêmica, porém, as outras merecem melhor esclarecimento. Traçando um paralelo com as demais categorias indicadas por Ranganathan e o CRG, conforme QUADRO 1, consideram-se as categorias: “Personalidade” como “Tipos de Produto Final”, pois é o tema sobre o qual o autor irá investigar; “Matéria” como “Partes, Materiais, Propriedade”, que é a segmentação do tema pesquisado; e “Energia” como “Processos, Operações, Agentes”, uma vez que indica as formas de execução e as variáveis analisadas na ação. Desse modo, nota-se que mesmo ampliando o número de categorias, as mesmas podem ser consideradas especificidades das categorias representadas pelo PMEST, idealizado por Ranganathan.

Contudo, concorda-se que haverá casos nos quais é mais adequado que as categorias sejam subdivididas e que essa decisão deve advir do objetivo ao qual a classificação irá atender. Desmembrar um domínio categorias mais segmentadas ou em mais especificidades pode facilitar a percepção do usuário em relação a esse domínio. Porém, vale ressaltar que Ranganathan não determinou que as categorias fossem designadas exatamente pelos termos utilizados no PMEST, mas que todo domínio pode ser decomposto por esses cinco conceitos (categorias fundamentais). Um aspecto que merece ser considerado refere-se ao fato de que o CRG não é um grupo de pesquisa com propósito único, isto é, de método de tratamento da informação, porém, com pesquisas e iniciativas realizadas separadamente, uma vez que os seus estudos encontram-se espalhados em diferentes fontes individuais. Com isso, nota-se dispersão de seus resultados, dificultando uma avaliação mais acurada acerca dos mesmos. A dificuldade em reuni-los em um só documento fez com que a pesquisadora Spiteri (1998) desenvolvesse um “modelo simplificado para a análise facetada”, que tinha como objetivo principal o ensino do método, oferecendo aos alunos uma forma mais fácil de aprender o processo classificatório facetado. Segundo Lima (2004a), 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 subdivisões e (b) Princípio da Sucessão Consistente. Spiteri utiliza, ainda, quatro princípios do plano verbal e dois princípios do plano notacional [...] (LIMA, 2004a, p.86).

Analisando o *Prolegomena* de Ranganathan (1967), percebe-se que a síntese desse estudo, apresentada por Spiteri, não reduz o princípio e o sentido da análise facetada, como se pode observar pela explicação do seguinte raciocínio:

### **Plano das ideias:**

#### 1) Princípios do plano das ideias:

a) Diferenciação (Ranganathan): distinção precisa do critério utilizado para a divisão, de forma que não haja forma de agrupar elementos diferentes;

b) Relevância (Ranganathan e CRG): adequado ao objetivo proposto;

c) Verificação (Ranganathan e CRG): uso de facetas definitivas e que possam ter seu uso validado na literatura ou pelos especialistas do domínio;

d) Permanência (Ranganathan e CRG): a escolha das facetas deve obedecer ao critério de ter um sentido estável dentro do domínio;

e) Homogeneidade (CRG): ser coeso, com facetas divididas apenas sob um critério (Ranganathan estabelece a “exclusão mútua”, que parece reafirmar o caráter de homogeneidade);

f) Mútua exclusão (Ranganathan e CRG): os focos de uma faceta, que representam um dado conceito, não podem pertencer a mais de uma categoria, ou seja, não podem se repetir;

g) Categorias fundamentais (CRG): as categorias emergem em relação ao domínio e ao objetivo pretendido, e não por imposição de princípios rígidos de determinação;

#### 2) Princípios para a escolha da ordem das facetas e classes:

a) Sucessão relevante (CRG): a sequência dos assuntos deve ser útil e se acomodar ao objetivo proposto, podendo ser: do complexo ao simples, do simples ao complexo, espacial ou geométrica, canônica, cronológica, em quantidade crescente ou decrescente ou alfabética.

b) Sucessão consistente (Ranganathan): para classes semelhantes, preferencialmente, o uso de uma sequência de assuntos.

### **Plano verbal**

1) Contexto (Ranganathan): o significado de cada termo na classificação deve ser determinado em razão da classe à qual pertence, pois é preciso que todos eles denotem um só sentido, em completude.

2) Aceitabilidade ou terminologia usual (Ranganathan): cada termo usado para a faceta ou para um foco deve ser aceito no contexto de uso. Caso o termo escolhido se torne obsoleto, o mesmo deve ser trocado por outro mais adequado.

## Plano notacional

1) Sinônimo (Ranganathan): se houver mais de um termo representando um mesmo conceito (sinonímia), ainda assim haverá, apenas, uma única notação para indicar esse conceito, sejam quantos forem os sinônimos;

2) Homônimo (Ranganathan): nenhum número poderá representar mais de um conceito, ainda que sejam homônimos (mesma grafia ou pronúncia e significados diferentes);

3) Hospitalidade (CRG): ser flexível, sempre permitindo a inclusão de novos termos, pois o sistema notacional deve acomodar mudanças, atualizações e adições de novos termos;

4) Arquivamento (CRG): respeitar a ordem estabelecida pela tabela utilizada pelo sistema.

Sabe-se que esse modelo simplificado de Spiteri facilitou bastante os estudos sobre a análise facetada, tendo servido de base para a elaboração de “sistemas de classificação facetada e de tesouros para recuperação da informação” (LIMA, 2004a, p.87). Nessa circunstância, de acordo com Morville e Rosenfeld (2006, p. 221-222) “a diferença entre um sistema de classificação tradicional e uma classificação facetada é que esta última é baseada no princípio de múltiplas taxonomias”, que é a característica da taxonomia facetada desenvolvida nesta pesquisa.

Aliás, vale ressaltar que a característica de múltiplas taxonomias está presente nas ontologias, que vêm sendo aplicadas na CI desde a década de 90, como instrumentos de representação do conhecimento de um domínio, visando à recuperação de informações. Não restam dúvidas de que a estruturação ontológica de uma área de conhecimento é complexa e gera sistemas robustos e mais precisos, implicando formalização lógica-matemática de conceitos, o que possibilita um amplo relacionamento entre eles. Isso é possível porque uma ontologia é constituída por classes (conceitos), sob as quais há uma taxonomia, que é sua espinha dorsal, e pelas relações existentes entre as classes e os elementos da taxonomia (GRUBER, 1996). Acredita-se, que essa forma de estruturar um domínio, dividindo-o em classes que possuem uma hierarquia taxonômica, abriu precedentes para a evolução da taxonomia tradicional (hierárquica) para uma taxonomia facetada, uma vez que, em relação à sua estrutura, uma ontologia possui uma taxonomia hierárquica sob cada uma de suas categorias mais gerais.

Contudo, as ontologias exigem especificações formais e são capazes de realizar inferências sobre um domínio, mas elas ainda não apresentam uma interface muito amigável ao ser humano, sobretudo para o usuário final de uma biblioteca digital. Acredita-

se, entretanto, que, nesse nicho, as taxonomias facetadas são mais convenientes, uma vez que são menos complexas e mais intuitivas. Dessa maneira, é fácil entender a qual classe determinado elemento pertence, principalmente em uma BDTD, com informações especializadas para a comunidade científica.

Corroborando essa ideia, Simon, Valdes-Perez e Sleeman (1997) afirmam que nem sempre desenvolver um programa com algoritmos complexos é o mais acertado, pois “os sistemas complexos podem às vezes ser necessários, entretanto, a simplicidade, quando atingível, deve sempre ser preferida” (*ibidem*, p.181). Os autores alegam que a *Web* e grandes bancos de dados podem exigir complexidade, mas que simples soluções podem ser suficientes para que uma informação científica especializada esteja facilmente acessível para ser recuperada.

Dessa forma, é interessante apontar que a taxonomia facetada navegacional, desenvolvida nesta pesquisa, tem a característica de ser uma ideia simples e exequível, capaz de facilitar o acesso a conteúdos específicos dos resultados de pesquisas disponibilizados na BDTD, aprimorando, com isso, a recuperação de informação. Para melhor entendimento do tema, apresenta-se, a seguir, sua conceituação.

## 4.2 A taxonomia facetada navegacional

Neste ponto da fundamentação, apresentam-se os conceitos que caracterizam o objeto (tema) deste estudo, uma “taxonomia facetada navegacional”. Como já conceituado anteriormente, tradicionalmente, uma taxonomia é apresentada na forma de uma hierarquia, representando um domínio. Uma taxonomia facetada, por sua vez, apresenta um domínio segmentado em facetas, sendo que, em cada faceta, há uma hierarquia, e as possíveis relações entre as facetas podem indicar a multidimensionalidade de um termo nesse mesmo domínio.

Já o termo “navegação”, em ambiente digital, refere-se à técnica interativa, utilizada para acessar um conjunto de informações. O verbete “navegação” do Dicionário de Biblioteconomia e Arquivologia explica essa noção apresentando a definição:

pesquisar (ou navegar) em um programa, procurar comandos, percorrer um documento e buscar informações [...] [que é o] processo dos usuários interagindo com um sítio visando satisfazer suas necessidades de informação [...], [navegando] em sítios por meio de busca e folheio de objetos com conteúdo (CUNHA; CAVALCANTI, 2008, p.257).

Dessa maneira, percebe-se que o sentido de “navegação” indica a possibilidade de o usuário explorar o conteúdo de um banco de dados através da navegação. Nesse sentido, Bonilla (2005, p.139) afirma que “a navegação está baseada nas indexações e associações de idéias e conceitos, organizados sob a forma de *links*, os quais agem como portas virtuais que abrem caminhos para outras informações”. Na literatura nota-se que a

classificação de documentos de um domínio, visando à sua estruturação, é empregada como mecanismo para facilitar a recuperação de informações. Essa abordagem vem sendo denominada de “paradigma de navegação facetada<sup>45</sup>” (BROUGHON, 2006; TZITZIKAS; ANALYTI, 2007; SACCO; TZITZIKAS, 2009).

Tendo isso em vista, é correto pensar que uma navegação facetada é o procedimento que o usuário realiza, percorrendo, interativamente, um conjunto de informações que estão ordenadas em facetadas, de tal forma que possibilitam combinações multidimensionais, visando encontrar um conteúdo. Isso é corroborado por Tunkelang (2009, p.23), quando afirma que “a navegação facetada permite ao usuário elaborar uma busca progressivamente, e lhe permite perceber o efeito que a escolha que faz por uma faceta disponível opera nas outras facetadas”. Além disso, esse tipo de navegação, no qual o usuário pode ir refinando sua busca a partir das opções que lhe são apresentadas, elimina a possibilidade de “resultados vazios”, pois somente são apresentadas a ele as facetadas nas quais há conteúdos de informação.

Alguns autores fazem distinção entre “navegação facetada” e “busca facetada” (TUNKELAND, 2009), apontando que a “busca facetada” somente acontece quando é possível a combinação da busca direta (por palavras-chave) com a navegação facetada. Entretanto, a maior parte dos autores consultados usa ambos os termos como sinônimos (BROUGHON, 2006; OLSON, 2007; BREEDING, 2007; SUOMINEN; VILJANEN; HYVÖNEN, 2007; TVAROŽEK *et al.*, 2008; TVAROŽEK; BIELIKOVÁ, 2007), para indicar a característica e ter informações estruturadas através de taxonomias facetadas, nas quais é possível buscar informações e explorar o conteúdo do banco de dados pela navegação por elementos dispostos em facetadas.

Sendo um tema relativamente novo, ainda há divergência quanto ao melhor termo para indicar esse novo paradigma. Em vista disso, neste estudo consideram-se os termos “busca facetada” e “navegação facetada” como sinônimos, uma vez que são os termos mais comumente aceitos pela maior parte dos autores e porque considera-se que carregam um sentido mais adequado à CI, pois fazem alusão aos princípios da teoria da análise facetada.

Partindo do exposto, é preciso, ainda, apresentar um conceito para “taxonomia navegacional” que, segundo Conway e Sligar (2002, *on-line*), é criada para permitir que os usuários encontrem informações por meio da navegação. Para os autores, na criação da taxonomia navegacional há a decomposição do conteúdo de um documento ou domínio em categorias e facetadas, com base em conceitos, a partir de uma lógica de organização pré-estabelecida. Assim, tem-se o estabelecimento dos termos através de similaridades,

---

<sup>45</sup> *Faceted browsing paradigm.*

semelhanças, diferenças, afinidades e dependências, buscando associações. Os pesquisadores citam como exemplo de taxonomias navegacionais a página de busca da *Microsoft Network*<sup>46</sup> (MSN) e a *HomeAdvisor*<sup>47</sup>.

Conway e Sligar (2002) afirmam, ainda, que uma taxonomia navegacional é, geralmente, criada para propósitos específicos de uso. Dessa forma, o reuso de suas categorias por outros sistemas está condicionado àqueles que possuírem idênticos propósitos e tipos de conteúdo, como em qualquer outro tipo de estrutura de modelagem informacional. Ademais, a estrutura da taxonomia navegacional expõe e permite ao usuário perceber o conteúdo do domínio, conforme demonstra a FIG. 4.

FIGURA 4 – Exemplo de taxonomia navegacional

The screenshot shows a web interface with a navigation taxonomy. At the top, there is a 'Taxonomy' header with a 'Help' link. Below it are navigation tabs: 'Home', 'Search', 'Taxonomy', 'Document Library', and 'Subscriptions'. A search bar is present with a dropdown menu set to 'This site' and a 'Go' button. Below the search bar, the current category path is shown as ':SQL'. A breadcrumb trail indicates the path: 'Categories : Technologies : Databases : SQL'. There is a link to 'Up One Category' and a 'Subscribe to this category' option. Below this, there are links for '. Deployment', '. Design', and '. Management'. A 'Documents' section follows, with a 'Hide Descriptions' link and 'Items: 2'. The documents are sorted by 'Author', 'Title', and 'Date'. The first document is 'SQL 2000 Architectural Overview' by Samantha Smith, modified on 12/19/2001. The second document is 'SQL 2000 Designed for Availability' by Jon Morris, also modified on 12/19/2001. On the left side of the screenshot, there are two labels with arrows pointing to the category path and the document list: 'Navigational taxonomy' and 'Navigation to SQL content'.

Fonte: Conway e Sligar (2002, *on-line*).

Observa-se, a partir da FIG. 4, que uma taxonomia navegacional consiste em um conjunto de elementos passíveis de serem navegados em um SRI, de forma progressiva, restringindo a busca a um subconjunto de elementos que foi determinado com base nas relações existentes entre as facetas.

Partindo de tudo que foi exposto neste subcapítulo, pode-se chegar a uma conceituação para “taxonomia facetada navegacional”: é uma estrutura de organização

<sup>46</sup> Disponível em: <<http://search.msn.com/>>.

<sup>47</sup> Disponível em: <<http://homeadvisor.msn.com/>>.

composta de um sistema de categorias, sendo que, dentro de cada uma (categoria) é criada uma hierarquia de facetas e subfacetadas, obedecendo a critérios pré-estabelecidos para a concepção das mesmas (categorias), o que permite atribuir diferentes dimensões (multidimensional) a um objeto (documento<sup>48</sup>). É, ainda, um sistema organizado sob a forma de *links*, através dos quais o usuário tem a possibilidade de navegar, de forma interativa, filtrando e restringindo sua busca em conformidade com sua escolha de facetadas.

Por conseguinte, a busca realizada a partir de uma taxonomia facetada navegacional permite aos usuários explorar uma grande quantidade de informações, de forma flexível, através da navegação, de modo que a sobrecarga informacional possa ser minimizada. Sendo assim, o papel fundamental de uma taxonomia facetada navegacional é guiar o usuário na busca, explicitando as informações referentes ao conteúdo que realmente está disponível no banco de dados.

Dessa maneira, o usuário pode iniciar a formulação de sua questão a partir de qualquer uma das dimensões, o que facilita a busca e auxilia a recuperação de informações, tema este que merece uma atenção maior através do próximo tópico.

### 4.3 Recuperação da informação

Discorrer sobre o aumento da produção e da quantidade de informação disponibilizada nos dias atuais é tema recorrente na sociedade científica desde a década de 40 e, desde então, diferentes problemas recaíram no processo de recuperação de informações.

No âmbito da CI, o termo “recuperação da informação” tem usos distintos, pois pode ser empregado: na operação de seleção de documentos em um acervo, a partir da demanda de um usuário; no fornecimento de um tópico específico de informação, a partir da demanda de um usuário; para designar a elaboração de bibliografia ou notas de síntese para atender ao pedido de um usuário; para indicar um campo mais amplo, que inclui o tratamento dado à informação; e também como uma área de pesquisa (FERNEDA, 2003). Diante dessa diversidade de usos, esclarece-se que esta dissertação destaca a área de pesquisa em recuperação da informação (RI), focalizando o processo de busca da informação.

A área de pesquisa de RI tem Calvin Mooers como precursor, pois esse pesquisador cunhou o termo “*information retrieval*” ainda na década de 50, indicando que, aliado a questões referentes ao tratamento na entrada de dados, o processo de RI envolve

---

<sup>48</sup> Nesta dissertação, “documento” tem um sentido mais amplo de ser qualquer unidade informacional, impressa ou não, passível de catalogação ou indexação, seja qual for a sua natureza ou suporte (ver item 4.8). Entretanto, neste estudo trata-se de manusear documento especificamente do tipo teses e dissertações e é a esse tipo de objeto (documento) que nos referimos nesse exemplo.

“as especificações de busca, além de os sistemas, técnicas e equipamentos que são aplicados para realizar o processo (SARACEVIC, 1999, p.1057<sup>49</sup>).

Outro importante autor nessa área é G. Salton, principalmente em estudos sobre indexação automática para sistemas de recuperação de informações (SRI) (SAYÃO, 2008), na década de 60, com seu sistema *System for the Mechanical Analysis and Retrieval of Text* (SMART). É importante destacar, que a concepção de recuperação da informação nessa época estava centrada na construção de sistemas (abordagem centrada no sistema), fato que parece ser explicado porque, nesse período, outros pesquisadores, além de Calvin, envolvidos nas questões relativas à recuperação da informação advinham da área da computação: Mortimer Taube, James Perry, Allen Kent, Hans Peter Luhn, entre outros (SARACEVIC, 1999).

Desde então, à abordagem centrada no sistema foram sendo agregadas outras abordagens, com o desenvolvimento de estudos sobre os SRIs visando melhores soluções para permitir o acesso do usuário ao conteúdo de seus bancos de dados, centrando seus esforços também nos modelos mentais do usuário (abordagem alternativa) na busca por informações. Nesse processo de busca da informação, o usuário está interessado em identificar um conjunto de documentos em um SRI, que atendam à sua necessidade de informação. Esse usuário está, portanto, “interessado em recuperar ‘informação’ sobre determinado assunto e não em recuperar dados<sup>50</sup> que satisfazem sua expressão de busca, e nem tampouco documentos, embora seja nestes que a informação está registrada” (FERNEDA, 2003, p.15). Entende-se, assim, que a recuperação de um documento que contém determinada expressão de busca utilizada na busca do usuário não, necessariamente, significa que esse item contenha informação considerada relevante. Pode acontecer de este item apenas apresentar dados, isto é, conteúdos sobre “as palavras” da expressão de busca, talvez com a mesma construção sintática, mas sem valor semântico para que possa atender à necessidade do usuário. Portanto, quanto mais específica e contextualizada for a resposta oferecida pelo SRI em relação à necessidade informacional do usuário, maior poderá ser a satisfação dele.

Um SRI, na concepção de Lancaster (1993), é um sistema de informação que possui a propriedade de informar ao usuário quais documentos se referem ao tema buscado. Ademais, Kuramoto (1995) reuniu diferentes pontos de vista sobre o conceito de SRI:

<sup>49</sup> Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques or machines that are employed to carry out the operation.

<sup>50</sup> Dado é “necessariamente uma entidade matemática e, desta forma, puramente *sintática*. Isto significa que os dados podem ser totalmente descritos através de representações formais, estruturais. [...] *Informação* é uma abstração informal (isto é, não pode ser formalizada através de uma teoria lógica ou matemática), que representa algo significativo para alguém através de textos, imagens, sons ou animação.” (SETZER, 1999, *online*).

Harter define um sistema de recuperação de informação (SRI) como um **dispositivo que se interpõe** entre os usuários e a coleção de informação. Strzalkowski complementa esta definição, afirmando que tais sistemas têm como função típica a de selecionar documentos de uma base de dados, em **resposta a uma questão do usuário**, apresentando-os em ordem de pertinência, enquanto Salton e McGill conceituam tecnicamente um SRI como um sistema que trata da representação, do armazenamento, da organização e do **acesso aos itens** de informação (KURAMOTO, 1995, p.1, grifos nossos).

Partindo dessas conceituações, observa-se a importância do tema sob dois aspectos: a questão da necessidade do usuário e a questão do acesso à informação. É preciso entender, que os SRIs lidam com informação textual<sup>51</sup> e incorporam os problemas existentes no uso da linguagem natural (sinonímia, polissemia). Esses problemas podem ser minimizados na medida em que o SRI estabelece medidas de controle terminológico utilizando, por exemplo, vocabulários controlados para a padronização de seu vocabulário (FERNEDA, 2003).

Nesse cenário, cabe mencionar que existem diversos modelos apropriados para a representação da informação para a recuperação de conteúdos informacionais. Tais modelos auxiliam na compreensão de um dado domínio e possibilitam a elaboração de vocabulários controlados. Segundo Currás (1995, p.37), qualquer vocabulário controlado é “uma linguagem controlada, normalizada, usada com fins classificatórios, onde a linguagem natural é transladada para uma linguagem terminológica”. Essas linguagens objetivam a “recuperação de informações e a organização dos conteúdos informacionais de documentos” (CAMPOS, M., 2004, p.24). Essa atividade, entretanto, está ligada, geralmente, à entrada de dados no sistema (SRI), com o processamento das informações.

Entretanto, o processo completo de um SRI inclui basicamente uma entrada (coleta de informações), processamento (tratamento das informações), saída (formas de apresentar o resultado do processamento) e respostas (acesso às informações). Assim, é esperado que a preocupação com a recuperação de informações em relação à sua forma de saída também esteja presente nos estudos na área de CI. A forma de saída abordada nesta dissertação se refere à possibilidade de navegação entre os documentos, a partir da representação e organização de seus conteúdos, formando sistemas de hipertexto<sup>52</sup>. Considera-se que bibliotecas digitais (BD) de assuntos especializados, como é o caso das BDs de teses e dissertações, devem contar com interfaces que aliem a recuperação de informações com a possibilidade de navegação por *links* de conteúdos dos documentos.

Ainda pelos idos de 1895, Otlet, no Instituto Internacional de Bibliografia (IIB), no campo da Documentação, desenvolveu metodologias para a organização do conteúdo dos

<sup>51</sup> Ainda que hoje haja sistemas de recuperação áudio-visual (ou de quaisquer outros tipos), neste estudo serão tratadas apenas questões relativas a informações textuais.

<sup>52</sup> Possibilidade de associar diferentes partes de um mesmo documento ou de distintos documentos, através de links navegáveis para acesso ao conteúdo.

documentos, a fim de obter melhor acesso a esses conteúdos, aprimorando a recuperação de informações, tendo sido, talvez, o pioneiro nessa área em particular (ALVES, 2007).

Em seguida, pode-se destacar Bush que, em 1945, já tinha especial preocupação com as questões de acesso e recuperação de informações. Dessa maneira, apesar de se tratar de uma época marcada pelo fim da Segunda Guerra Mundial, focou seus estudos, sobretudo, em questões acadêmicas, pois

diagnosticou a dificuldade cada vez maior do desenvolvimento científico em função de ineficientes sistemas de produção, organização, acesso, recuperação e disseminação da informação devido ao crescimento do conhecimento humano. Alertou para a perda de conteúdos importantíssimos em meio a uma imensidade de outros textos e consequente duplicação de informações, e sim na maneira de processá-la, na demanda de pesquisas e experiências que pediam processamentos que fizessem circular eficazmente um volume infinito de informação. Para gerenciar este problema, havia a necessidade de novas tecnologias (novas metodologias, instrumentos e máquinas) (ALVES, 2007, p.32).

Como se nota, desde aquela época os estudos têm sido direcionados para as questões de acesso a informações estruturadas, tentando impedir a duplicação de pesquisas e a perda de parte do conhecimento já descoberto. Com isso, Bush almejava o acesso e a transferência da informação e do conhecimento, a partir de seu “sistema hipertextual Memex”.

Mais recentemente, com a ideia da *Web* semântica, desenvolvida por Tim Berners-Lee, no âmbito da *World Wide Web Consortium*<sup>53</sup> (W3C), os trabalhos em recuperação da informação na CI têm adotado essa linha condutora, buscando dar significado ao conteúdo disponibilizado, de forma a tornar o acesso mais facilitado e a recuperação de informações mais eficiente. Em face disso, Souza e Alvarenga (2004, p.140) concluem que

o estudo das possibilidades que se abrem e a compreensão de que todo o embasamento filosófico, metodológico e conceitual da *Web* Semântica parte do núcleo duro da ciência da informação nos impelem a demarcar e arrebatar os legítimos territórios do saber e a buscar ativamente uma atuação no desenho destes novos panoramas informacionais. Precisamos hoje trilhar um caminho rumo a uma nova e necessária valorização da área de ciência da informação, que oferece teoria, metodologias e competências que compõem a quintessência daquilo que se espera dos trabalhadores e pesquisadores de uma sociedade baseada em informação e conhecimento. E a importância da *Web* e das demais redes digitais de troca de informações no panorama mundial são amostras de como a atividade de organização da informação é necessária para a evolução dos indivíduos, organizações e da sociedade em geral.

Sendo assim, nota-se constante preocupação com melhores práticas que aprimorem a recuperação de informações, desde para Bush, em 1945, com a preocupação

---

<sup>53</sup> Consórcio internacional com cerca de 300 membros, que agrega empresas, órgãos governamentais e organizações independentes. Busca desenvolver padrões para a criação e a interpretação de conteúdos para a *Web*.

voltada para o reuso de conhecimentos já desvendados, através do uso de tecnologias, até para Souza e Alvarenga, em 2004, preocupados em encontrar soluções tecnológicas que estejam respaldadas em teorias e metodologias de organização da informação, propiciando o acesso e a troca de informações essenciais para o desenvolvimento da ciência e da sociedade como um todo.

Portanto, chega-se ao contexto da proposta apresentada nesta dissertação, de criar um mecanismo que se baseia em teorias e metodologias da área da CI, que organiza e estrutura o conhecimento contido nas teses e dissertações, através de uma taxonomia facetada navegacional. Dessa forma, pretende-se minimizar a possibilidade de duplicação de pesquisas e evitar a perda de conhecimentos valiosos para a área, usando, para isso, soluções tecnológicas já conhecidas e disponíveis, a fim de facilitar a busca e a recuperação de informações.

Neste contexto, é significativo, ainda, indicar que a “busca de informação<sup>54</sup>” retrata um “processo que consiste em situar a informação numa tabela ou arquivo por referência a um campo especial de cada registro, denominado chave. O objetivo da busca é descobrir um registro com determinada chave” (CUNHA; CAVALCANTI, 2008, p.59). Em contrapartida, a “recuperação de informação<sup>55</sup>” é a “restituição dos dados constantes do sistema, para obtenção de informações específicas ou genéricas. A restituição, ou recuperação, abrange o processo total de identificação, busca, encontro e extração da informação armazenada” (CUNHA; CAVALCANTI, 2008, p.307). Esse processo é parte de um processo mais amplo, que se denomina “sistema de recuperação da informação (SRI)”, que, para Lancaster e Warner (1993), refere-se à interface entre o acervo da biblioteca, tradicional ou não, e os usuários, e cujas funções incluem a aquisição, armazenamento, organização, distribuição e disseminação de documentos. Ademais, Baeza-Yates e Ribeiro-Neto (1999), indicam o SRI como um sistema que engloba funções de representação, armazenamento, organização e acesso às informações. Já Cunha e Cavalcanti (2008), conceituam este sistema como aquele que

armazena e recupera informação, a qual pode estar em textos completos, documentos substitutos (tais como resumos) ou referências bibliográficas. ‘Engloba os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou meios eletrônicos empregados para o desempenho da operação’ (ROWLEY, 1994, p.47 citado por CUNHA; CAVALCANTI, 2008, p.345).

Nesse sentido, vê-se que descrever o recurso informacional, visando às especificidades da busca, constitui uma tarefa esperada de um SRI, ainda que, para isso, seja preciso empregar esforços humanos de indexação e não apenas meios automáticos e inovações tecnológicas.

---

<sup>54</sup> Termo em inglês: *information searching*.

<sup>55</sup> Termo em inglês: *information retrieval*.

Com as inovações tecnológicas, os serviços de transferência de informação se expandiram e vêm sendo redimensionados e arquitetados para se adequarem às necessidades dos usuários. Nesses termos, as bibliotecas digitais podem ser consideradas como inovação tecnológica no gerenciamento de conteúdos digitais, justamente, e se preocupam com a questão do acesso e com a comunicação da informação. Como reconhecimento de sua importância, projetos de implantação de bibliotecas digitais têm crescido e já é possível verificar grandes variações nas opções de buscas disponibilizadas.

A fim de um aprofundamento neste contexto, destaca-se que houve significativa evolução nas opções de busca em bibliotecas digitais em geral, estratégias que já estavam presentes em alguns softwares de bibliotecas (como o CDS-ISIS, por exemplo, desenvolvido pela UNESCO). Em meio digital, pode-se dizer que essas alternativas se iniciaram com os catálogos eletrônicos *on-line*, conhecidos na Internet como OPACs (*On-line Public Access Catalogs*), nos quais se faz consultas padrões por palavra-chave, nome do autor ou título da obra. Atualmente, a biblioteca digital pode oferecer diversificadas opções que ultrapassam essas consultas padrão, dentre as quais, destacam-se:

**Boolean Logic** (busca *booleana*): busca por combinações lógicas de múltiplos termos, utilizando os operadores típicos *and*, *or* e *not*, em que, usando o operador *and*, o resultado deverá conter, pelo menos, uma ocorrência de cada palavra ou frase especificada; usando o operador *or*, o resultado apresentará pelo menos uma ocorrência de qualquer uma das palavras ou frases especificadas; e, usando o operador *not*, o resultado não deverá apresentar qualquer ocorrência do termo especificado para este operador.

**Fuzzy Expansion** (expansão *Fuzzi*): expande a busca por similaridade da escrita ou da pronúncia, para minimizar erros de grafia. Os operadores típicos são: *spelled like* (similaridade na escrita) e *sounds like* (similaridade na pronúncia).

**Wildcard**: expande a busca de uma palavra, partindo de uma origem comum, sem considerar se ela pertence ou não à mesma origem linguística. Há alguns sistemas que já embute esta propriedade como padrão, expandindo a busca do usuário, sem que ele especifique isso. Os operadores típicos são: (\*) encontra qualquer conjunto de 0 ou mais caracteres, até um espaço ou pontuação (há bibliotecas que utilizam o caractere (%)); (?) representa o número de caracteres que se deseja, após a palavra origem.

**Exactly like / stem expansion** (busca exata / busca estendida): expande a busca para variações derivadas da palavra originária. A variação inclui nome, pronome, adjetivo. Porém, todas as palavras derivadas têm a mesma origem linguística.

**Case Sensitivity**: serve para ignorar ou não o uso da palavra escrita em maiúscula, minúscula ou de ambas as formas. Geralmente é uma configuração padrão nos SRIs, e não está disponível para a escolha do usuário. Os operadores típicos são: *upper/lower case* no qual a palavra deve aparecer em maiúscula ou minúscula; e *ignore*

*case differences*, sendo aquele em que o sistema ignora o aparecimento de maiúscula ou minúscula na frase.

**Accent Sensitivity:** consultas insensíveis ou não ao uso de distinção de sinais diacríticos, que são sinais gráficos que permitem distinguir modulação das vogais, tais como a vírgula, acento agudo, ponto, etc.

**Stopwords:** há um conjunto de palavras que podem ser ignoradas nas buscas. Geralmente, também aparecem embutidas por padrão.

**Proximidade:** localiza documentos em que os termos especificados ocorram, no mínimo, uma vez. Quanto mais próximos estiverem os termos dentro de um mesmo documento, maior será sua relevância.

Nas BDTDs, de modo recorrente, empregam-se as seguintes formas de busca e recuperação da informação: buscas por elementos como título, autor e data, além da busca por assunto, que pode ser na forma simples ou avançada. Nas formas de busca por assunto, seja simples ou avançada, há uma varredura nas palavras contidas nesses elementos (título, autor e data), no resumo e nas palavras-chave. Com relação à busca simples, em geral, o usuário submete um termo ou mais ao sistema, podendo ocorrer, na recuperação de informações, o aumento da revocação e a falta de precisão do sistema. A título de definição,

a revocação, ou “recall” ou mesmo “abrangência”, é a razão do número de documentos atinentes recuperados sobre o total de documentos atinentes disponíveis na base de dados. A revocação mede o sucesso do SRI em recuperar documentos pertinentes.

[a precisão é a] razão do número de documentos atinentes recuperados sobre o total de documentos recuperados. A precisão mede o sucesso do SRI em não recuperar documentos que não sejam relevantes de acordo com a necessidade de informação (SOUZA, R., 2006, p.163).

Já na busca avançada, há vários campos, com diferentes possibilidades de busca, que podem permitir combinação. O tipo mais comum é a busca *booleana* (já citada anteriormente), que pesquisa combinações lógicas de múltiplos termos. É fato que são muitos os usuários que não se sentem à vontade nesse tipo de busca, geralmente, por falta de conhecimento sobre como a combinação dos operadores booleanos (*and*, *or* e *not*) se estabelece. Inclusive, isso é confirmado pelo estudo de Baeza-Yates e Ribeiro-Neto (1999), no qual os autores estabelecem as principais características dos usuários de bibliotecas digitais, dentre as quais, destacam-se: 1) preferem buscas pautadas em respostas, em detrimento a perguntas, pois têm dificuldades na formulação destas; 2) usam poucos termos ao formular suas consultas; 3) dispersam-se com facilidade ao serem sobrecarregados com informações; 4) não têm paciência em aguardar respostas de suas buscas; 5) acham as buscas *booleanas* difíceis de compreender.

Em vista disso, aponta-se que três das cinco características acima deram respaldo à proposta desenvolvida nesta dissertação e representam os pressupostos, quais sejam: 1) buscas pautadas em resposta; 2) sobrecarga de informações; 3) buscas *booleanas* são difíceis de compreender. Pretende-se que a taxonomia facetada navegacional ordene o conteúdo dos documentos, de forma que as perguntas estejam formuladas implicitamente, através de combinações do tipo das buscas *booleanas*, gerando respostas pré-estabelecidas, evitando sobrecarga de informações.

Não se pode esquecer que Kwasnick (1999, p.40-42) enumera algumas vantagens das classificações (taxonomias) facetadas: 1) não exigem que o classificador conheça todas as entidades ou as relações entre elas; 2) são flexíveis; 3) são hospitaleiras, uma vez que permitem a inserção de novas entidades; 4) são expressivas; 5) podem ser *ad hoc*, isto é, elaboradas para um objetivo específico; e 6) permitem diferentes perspectivas e abordagens dos objetos representados.

Portanto, uma taxonomia facetada pode ser construída sob múltiplos critérios, facilitando a recuperação da informação, pois cada documento pode ser etiquetado (atribuições) sob diferentes categorias. Em uma biblioteca digital, um mesmo item pode ser segmentado por categorias, por exemplo, pela cor, pela idade, pelo gênero, pelo tamanho, entre outros. Assim, cada faceta, na taxonomia, pode ser incorporada separadamente, na forma de metadados, e cada uma dessas facetadas oferece condições de ser buscada. Lambe (2007) afirma que “taxonomias facetadas evitam a ambiguidade, assegurando que cada faceta seja única e obviamente diferente de qualquer outra” (*ibidem*, p.37, tradução nossa). Essa característica facilita a comunicação entre os pares, fornece visibilidade ao conhecimento acumulado e permite a transferência de informação, preocupações ainda à época de Bush.

Já naquela época Bush abordou preocupações atuais da comunidade científica e da humanidade como um todo: Qual a maneira eficiente de armazenar e recuperar o conhecimento humano? Como fazer uso apropriado do crescente volume de informações? Como impedir que uma informação relevante passe despercebida em meio a uma abundante geração de conhecimento? (ALVES, 2007, p.33).

Nesse sentido, a taxonomia facetada navegacional proposta nesta dissertação pretende atuar como um mapeamento do conhecimento acumulado e registrado, facilitando a comunicação entre os membros da comunidade científica, fortalecendo os vínculos entre os diferentes atores e a área em geral, através do ambiente de uma biblioteca digital. Refletindo sobre essas questões, a seguir passa-se a uma breve apresentação do ambiente digital de uma biblioteca.

#### 4.4 Bibliotecas digitais

As bibliotecas digitais têm sua origem na década de 90, juntamente com a expansão comercial do uso da Internet. Com a facilidade no armazenamento e publicações digitais, houve a necessidade de que as bibliotecas, antes tradicionalmente em ambientes físicos, estabelecessem uma nova estrutura de serviços de informação. Dessa forma, a partir de então, o ponto central de preocupação dessas instituições deixou de ser somente o acervo físico e seu armazenamento, que cedeu espaço para as questões também relacionadas ao acesso às informações (ZAFALON, 2008).

Notou-se uma evolução nos serviços oferecidos pelas bibliotecas digitais, que passaram pelo acesso a apontadores, a obras digitalizadas, a catálogos e a informações institucionais (SILVA; MÁRDERO; CLAUDIO, 1997). Contudo, pode-se observar que o momento é de transição e adaptação, pois os serviços de informação oferecidos pelas bibliotecas digitais ainda não estão aplicando todo o conhecimento já acumulado sobre a organização para documentos em acervos físicos. Em ambiente digital há diferentes perspectivas para o gerenciamento do recurso informacional disponibilizado que podem ser melhor exploradas, utilizando produtos de representações documentárias já amplamente conhecidas, de forma que se antecipe às necessidades de diferentes comunidades de usuários.

Destaca-se que, até há pouco tempo, era tema de debate entre os pesquisadores o conceito de biblioteca digital, e, ainda hoje, percebe-se alguma polêmica entre os autores. Porém, na contemporaneidade, parece haver mais pontos semelhantes do que divergentes entre eles, apesar de denominações diversificadas tais como biblioteca virtual, biblioteca digital, biblioteca eletrônica e biblioteca sem paredes, termos que muitas vezes são empregados como sinônimos. Em um artigo publicado em 2002, Ohira e Prado levantaram a evolução dos conceitos de biblioteca digital e biblioteca virtual contidos em artigos entre os anos de 1995 e 2000. Nesse estudo, foram analisados 33 artigos de periódicos nacionais especializados em Biblioteconomia e Ciência da Informação. Como o intuito da presente pesquisa não é adentrar na polêmica de uso dos diferentes termos para biblioteca digital, serão fornecidos alguns conceitos elencados por Ohira e Prado (2002, *on-line*), com o objetivo de situar o leitor no tema de biblioteca digital:

Biblioteca digital difere das demais, porque a informação que ela contém existe apenas na forma digital, podendo residir em meios diferentes de armazenagem, como as memórias eletrônicas (discos magnéticos e óticos). Desta forma, a biblioteca digital não contém livros na forma convencional, e a informação pode ser acessada, em locais específicos e remotamente, por meio de computadores (MARCHIORI, 1997, *on-line*).

a biblioteca digital é também conhecida como biblioteca eletrônica (termo preferido pelos britânicos), biblioteca virtual (quando utiliza os recursos da

realidade virtual), biblioteca sem paredes e biblioteca conectada a uma rede (CUNHA, 1999, p.258).

[a biblioteca digital] não contempla materiais convencionais impressos como livros, já que estes seriam convertidos/digitalizados para o formato digital (MACEDO; MODESTO, 1999, p.64).

a biblioteca digital seria aquela que teria, além de seu catálogo, os textos dos documentos de seu acervo armazenados de forma digital, permitindo sua leitura na tela do monitor ou sua importação (*download*) para o disco rígido do computador [...] (LEMONS, 1998, p.358 citado por PEREIRA; RUTINA, 1999, p.14).

a biblioteca digital tem como característica uma coleção de documentos eminentemente digitais, independentemente se forem criados na forma digital ou digitalizados a partir de documentos impressos, e permite, por meio do uso de redes de computadores, compartilhar a informação instantânea e facilmente (MOREIRA, 1998, p.31 citado por MACHADO *et al.*, 1999, p.218).

bibliotecas digitais são simplesmente um conjunto de mecanismos eletrônicos que facilitam a localização da demanda informacional, interligando recursos e usuários (CUNHA, 2000, p.78).

Como se pode observar, a despeito de alguma polêmica sobre o tema, o ponto recorrente nas características descritas sobre “biblioteca digital” é que ela possui recursos informacionais na forma digital para acesso remoto. Portanto, nesta dissertação, uma biblioteca digital é considerada assim.

A título de maiores esclarecimentos, sabe-se que toda biblioteca, seja tradicional<sup>56</sup> ou digital, tem como funções primordiais: selecionar, armazenar, disponibilizar, preservar e dar acesso a documentos em qualquer tipo de mídia e suporte. Sendo assim, há, nos acervos de uma biblioteca, ou no banco de dados de uma biblioteca digital (BD), um patrimônio cultural e intelectual que devem ser representados, organizados e difundidos de forma que, efetivamente, possam ser aplicados pela sociedade.

Para atender a esse propósito, os estudos sobre o tema de bibliotecas digitais são feitos sob diferentes aspectos, que podem ser assim sintetizados:

a) arquiteturas, sistemas, ferramentas, e tecnologias b) conteúdos e coleções digitais; c) metadados; d) interoperabilidade; e) normas e padrões; f) organização do conhecimento; g) usuários e usabilidade; h) aspecto legal, organizacional, econômico e social (SAYÃO, 2007, p.18).

Em vista dessa amplitude, atualmente, muitos são os autores que se interessam pelo tema de bibliotecas digitais no Brasil. Contudo, em qualquer que seja o aspecto investigado, as pesquisas sempre têm em comum a característica de abordar a questão do acesso, com preocupação focada em uma recuperação de informação mais relevante e eficaz para o usuário. Nesse sentido, a partir do início do século XXI, nota-se um aumento nos projetos de implantação de bibliotecas digitais de teses e dissertações (BDTD), tanto nacionais quanto estrangeiras, conforme afirma Rosetto (2008):

<sup>56</sup> Considera-se biblioteca *tradicional* aquela na qual a maior parte de seus itens do acervo é composto por documentos impressos, ainda que muitas delas, hoje em dia, estejam cada vez mais possibilitando acesso digital ao acervo.

Nesses primeiros anos do século XXI, muitos projetos para a construção de bibliotecas digitais foram implementados, criando variados tipos de estruturas e acumulando experiências. As melhores práticas vêm se transformando em metodologias que podem subsidiar o desenvolvimento de implementações similares, como alguns exemplos de projetos de caráter internacional, regional e nacional (ROSETTO, 2008, p.109)

No artigo citado, a autora discorre sobre as diferentes iniciativas, de implementação de bibliotecas digitais, dentre as quais foram salientadas algumas iniciativas brasileiras: o Portal Prossiga, o projeto da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), do Instituto Brasileiro em Informação em Ciência e Tecnologia (IBICT), a Biblioteca Digital Portal da CAPES e a Biblioteca de Teses e Dissertações da Universidade de São Paulo. São iniciativas que visam dar acesso aos textos das pesquisas e aprimorar a recuperação das informações, uma vez que o meio digital facilita indicar outros pontos de acesso além daqueles já utilizados em bibliotecas tradicionais (ex. título, autor, cabeçalho de assunto).

Nessa perspectiva, alguns estudos têm se preocupado com a criação de modelos de bibliotecas digitais que visam o aprimoramento da recuperação da informação contida nos documentos, a partir de elementos semânticos, como é o caso do protótipo apresentado na tese defendida por Lima (2004a), que propõe uma busca a partir da navegação em contexto. Nessa pesquisa, a autora propõe uma forma de recuperação de informação em contexto, com a construção de um modelo hipertextual, para a organização de documentos acadêmicos do tipo teses e dissertações e, ainda, a sua implantação tecnológica. Com relação ao arquétipo, trata-se de um instrumento gerado a partir de uma modelagem conceitual, capaz de auxiliar o usuário em suas buscas, tendo em vista que o conhecimento está estruturado conceitualmente. É preciso ressaltar que essa estrutura é feita através da categorização dos conceitos, a qual possibilita organizar, semanticamente, o conteúdo dos documentos, contribuindo para que o usuário perceba, através da navegação nos mapas conceituais<sup>57</sup>, as relações existentes entre os conceitos (LIMA, 2004a). Essa navegação é feita, atualmente, de modo individual, através dos mapas conceituais dos documentos.

Diante das ressalvas descritas, percebe-se que o referido protótipo não se trata, apenas, de uma biblioteca digital. Pode-se entender que ele está próximo daquilo que se espera de uma biblioteca digital semântica. É importante, porém, esclarecer que há diferenças conceituais entre biblioteca digital (BD) e biblioteca digital semântica (BDS). Segundo Kruk, Decker e Zieborak (2005), a BDS possui as seguintes características: a) permite compartilhar diversificados padrões de metadados, com comentários/anotações de usuários; b) estabelece interoperabilidade com quaisquer tipos de sistemas de informações,

---

<sup>57</sup> Os mapas conceituais assemelham-se a diagramas, nos quais são indicadas relações entre conceitos, em uma estrutura que perpassa os conceitos mais abrangentes até abarcar os conceitos mais específicos.

incluindo os que possuem metadados de diferentes origens; c) propicia o desenvolvimento de interfaces mais amigáveis, baseando as buscas e a navegação em elementos semânticos.

Além disso, também existe o conceito de biblioteca híbrida, o qual, de modo inovador, parece englobar os conceitos de biblioteca tradicional, biblioteca digital e biblioteca digital semântica, como se observa abaixo:

no desenvolvimento e aplicação de uma biblioteca híbrida, [...] [é preciso estabelecer] metodologias para que seja desenvolvido um conjunto de padrões para identificar, armazenar e disponibilizar e [sic] gerenciar as informações em diversas mídias que compõe [sic] seu acervo num ambiente digital, possibilitando a busca e a recuperação deste tipo de material, resultando na interoperabilidade entre interfaces abertas e protocolos de comunicação de dados, onde os usuários poderão realizar suas consultas de forma unificada em uma única interface e possibilitando a disponibilização dos textos para *download* (FIGUEIREDO *et al.*, 2004, *online*).

Nota-se, então, que a biblioteca híbrida procura integrar todos os serviços das bibliotecas tradicionais e das digitais, ampliando seu alcance, ao oferecer seus serviços e acervo no formato digital e, organizados semanticamente. Por essa visão, verifica-se que o conceito de biblioteca híbrida ultrapassa, em larga escala, o conceito de automação de bibliotecas tradicionais, na medida em que o primeiro permite acesso ao catálogo do acervo e a alguns serviços ou ao seu acervo digitalizado. De forma ainda mais reluzente, na biblioteca híbrida, há o interesse em manter a interoperabilidade entre os diferentes sistemas, o que expande, consideravelmente, as possibilidades de compartilhamento de informações.

Em face do que foi exposto acima, nesta dissertação entende-se que a construção de bibliotecas digitais de teses e dissertações (BDTD) deve levar em consideração a busca documental a partir de elementos semânticos, constituindo, para tanto, seu eixo conceitual. Nesse contexto, a “*web semântica*”<sup>58</sup> abre novas perspectivas para a criação de bibliotecas digitais semânticas, nas quais se pode melhorar o sistema de informação, principalmente nas atividades de busca e navegação. Assim sendo, o desenvolvimento e uso de instrumentos semânticos poderão resultar em uma recuperação de informações mais relevante ao usuário.

Como reflexo dos estudos realizados até o momento, esta pesquisa aponta para a ideia de que uma taxonomia facetada para navegação, usada como interface de busca, pode ser mecanismo facilitador para a recuperação de informações relevantes aos usuários. Em vista disso, o princípio de navegação utilizado é correspondente à navegação em hipertexto, o que será desenvolvido a seguir.

---

<sup>58</sup> A “*web semântica*” pretende fornecer meios para que se possa estruturar o conhecimento, dando significado (semântica) aos conteúdos disponibilizados em ambiente digital, de forma que esse conhecimento seja processado (entendido) pela máquina.

## 4.5 Navegação hipertextual

O hipertexto (ou hiperfídia) é um sistema para visualização de informações, construído pelos usuários em meios digitais. Pode conter alguns itens, tais como textos, áudio, vídeos ou imagens, e por ligações (*links*), de forma a facilitar a busca de informação. Quanto ao princípio de ligação, é possível que seja unidirecional ou multidirecional, numa rede interligada de informações. O nó de origem, por sua vez, é denominado “referência” e o de destino, “âncora”.

Para Conklin (1988, p.424), “o conceito de hipertexto é muito simples”, pois são janelas que, através de *links*, representados graficamente por seus rótulos, remetem a conteúdos que estão associadas em um banco de dados. Assim, segundo o autor, essa nova tecnologia proporcionou diversificadas possibilidades, sobretudo em relação à comunicação e à navegação através de *links*.

Nesse aspecto, os sistemas de hipertextos permitem aos usuários que naveguem por nodos de informação, obtendo acesso a informações contidas em bancos de dados, a partir de mecanismos de navegação (*browsers*). Aliás, possuir a característica de navegação é fundamental para que um sistema seja considerado, realmente, hipertextual.

A partir de um resgate histórico, compreende-se que a ideia de hipertexto teve origem na proposta de Vannevar Bush, datada de 1945. Nessa ocasião, o autor idealizou uma biblioteca que tinha por finalidade armazenar o conhecimento acadêmico registrado, designada por ele por Memex. Dessa maneira, o foco principal do projeto de Bush era a organização da grande quantidade de informação dispersa nas diferentes bibliotecas, especialmente, aquela relativa aos interesses de pesquisa dos estudantes, de modo que facilitasse o acesso e a recuperação de informações.

Entretanto, somente em 1965 o termo “hipertexto” foi cunhado por Theodor Nelson, com o Projeto Xanadu (NELSON, 1999), projetado para ser um *software* capaz de criar, hospedar e gerir um banco de dados hipertextuais. Porém, considera-se, como o primeiro sistema operacional, detentor de funcionalidades hipertextuais, o sistema de hipertexto *oN-Line System* (NLS), que foi desenvolvido para facilitar a criação de bibliotecas digitais, armazenar e recuperar documentos eletrônicos e apresentado ao público em 1968.

Desde então, existem duas vertentes sobre o conceito de hipertexto: uma de associação de ideias e outra relacionada à conexão de ideias. Para Bush, o hipertexto é uma associação de ideias, quase casual, tal como acontece na mente humana, cuja possibilidade de leitura não linear se assemelha com o tipo de rede associativa que um indivíduo faz (FURTADO, 2010). Por outro lado, o conceito de conexão se apresenta como um tipo de associação controlada, pois está estreitamente ligada à necessidade de comunicação de ideias, conforme pontua o teórico citado a seguir:

Um dos critérios necessários da conexão reside em ser inteligível para os outros, do mesmo modo que a singularidade do sujeito tem de ser limitada na sua tendência para a livre associação, já que a sua parte substancial se exerce ao nível do conjunto de concepções comuns à maioria das pessoas (WHORF, 1956 citado por FURTADO, 2010, p.37).

Segundo essa premissa, por maior que seja a liberdade do usuário em traçar seu caminho hipertextual, ele está limitado às possíveis associações existentes no sistema, sendo ela controlada pelo número e tipos de conexões imputadas pelo *designer*. Dessa forma, o usuário está restrito à intenção de comunicação do autor da mensagem, ainda que tenha certa autonomia para escolher um entre os admissíveis caminhos.

Se o que realmente define hipertexto é a existência de *links* organizando a informação, independentemente do processo pelo qual esses *links* foram criados, [...] então a “associação” é apenas um tipo de “conexão” e, na verdade, o menos desejável quando o objetivo é a comunicação (FURTADO, 2010, p.37).

Nesse sentido, não há como imaginar que a tecnologia, sozinha, seja capaz de transformar um texto em hipermídia. Então, haverá, sempre, um propósito de comunicação por trás de todo o esforço em fornecer nodos e *links*. Mesmo porque, todo sistema hipertextual oferece a possibilidade de navegação por esses nodos, através dos *links*. Assim, o hipertexto, como sistema de navegação, consolida o bônus de possuir uma estrutura não sequencial, permitindo distintos caminhos e em diferentes ordens, dentro do que foi projetado para o sistema.

Nessa linha de pensamento, o hiperdocumento, segundo Marchionini (1994), deve ter seus nós indexados (metadados) para possibilitar acesso ao conteúdo, sugerindo oito regras para a criação de hipertexto: 1) identificação das facetas mais relevantes; 2) lista exhaustiva de termos e frases; 3) mapeamento e revisão de termos para facetas; 4) criação de termos preferidos; 5) introdução de nós de informação, com remissivas; 6) revisão do conjunto de nós, de acordo com certos critérios (gramaticais, estilo); 7) manutenção e atualização do sistema de hipertexto; 8) edição e teste do hiperdocumento final.

Partindo disso, percebe-se a característica de uma navegação por facetas (já pré-estabelecidos), na qual um usuário tem capacidade de filtrar a informação desejada, dentre um conjunto de itens possíveis, através da seleção gradual. Esta seleção progressiva é possível a partir das facetas com metadados (valores) atribuídos, pois o sistema somente apresenta os itens, efetivamente, existentes no hiperdocumento e, conseqüentemente, no banco de dados.

Se os *links* proporcionam a interligação dos nodos (metadados), formando um conjunto coeso com a ideia concebida, esse sistema hipertextual terá característica semântica, na qual

os nós são os conceitos ou categorias relativos a um determinado conhecimento e os arcos explicitam as relações entre os nós. Nesta

representação as classes herdam atributos e comportamentos das superclasses que as contém (PINHO; KUX; ALMEIDA, 2007, p. 638)

Dessa forma, a apresentação do conteúdo de um banco de dados na forma de um sistema hipertextual permite que as informações sejam organizadas de acordo com o objetivo do arranjo, facilitando o acesso e a recuperação das mesmas. Além disso, o conjunto de facetas possibilita a visualização do todo, evidenciando esses mesmos conteúdos e oferecendo a navegação pelos *links*.

De modo a fixar, cada vez mais, a análise no objetivo central proposto, tem-se a oportunidade, agora, de refletir, especificamente, sobre a navegação realizada por circunstância da taxonomia facetada.

#### 4.5.1 Navegação por meio da taxonomia facetada

A navegação por facetas já vem sendo usada em alguns projetos de bibliotecas digitais, como demonstrado na revisão de literatura. Nesse sentido, o sistema FLAMENCO destaca-se como um dos primeiros e o mais citado exemplo de navegação desse tipo. Em função dessa importância, cabe salientar que a busca e navegação facetada disponibilizada neste *site* permitem organizar grandes coleções de dados, ajudando o usuário a não se sentir deslocado do sistema de busca. Para tanto, o FLAMENCO compõe-se de um conjunto de textos e, na entrada de dados desses textos, usa um modelo de representação em facetas que foi pré-determinado. Dessa forma, os arquivos dos textos são processados à medida que o modelo e os dados são inseridos em tabelas localizadas num banco relacional, o que permite definir termos e criar hierarquias entre eles. Não se pode deixar escapar, também, que, neste *site*, a busca facetada pode ser combinada com a busca livre, sendo esta uma ação que facilita, ainda mais, a busca.

Outro exemplo de navegação por facetas é o sistema da biblioteca temática disponível no *site* da *National Aeronautics and Space Administration*<sup>59</sup> (NASA). Nela, é possível fazer uma busca facetada, em razão de o conteúdo estar dividido em categorias e subcategorias. Com esse mecanismo, pode-se refinar a busca, que se inicia por um escopo mais abrangente até alcançar informações específicas.

Com o intuito de ampliar as ideias em trânsito, deve-se esclarecer que um sistema de recuperação da informação (SRI), em ambiente digital, somente pode ser considerado facetado quando for constituído por um número mínimo de operações, além de um conjunto de elementos nos quais o usuário possa navegar de forma arbitrária<sup>60</sup> para fazer sua busca, tendo possibilidade de, aos poucos, restringi-la, até um subconjunto mínimo de elementos, o qual é definido a partir de relações (interseções) previamente

<sup>59</sup> Disponível em: <<http://www.nasa.gov/home/index.html>>.

<sup>60</sup> Aqui, de forma *arbitrária* significa conforme o desejo do usuário.

determinadas entre as facetas (SUOMINEN; VILJANEN; HYVÖNEN, 2007). Dessa forma, a navegação facetada “permite que o usuário elabore uma pergunta progressivamente, e que vá observando o efeito que sua escolha por uma faceta disponível causa nas outras facetas” (TUNKELANG, 2009, p.23).

É importante detectar, ainda, que, no contexto de um sistema navegacional relacional, uma faceta constitui um conjunto de termos (valores), que representa uma propriedade (característica) desse conjunto, e seus valores, por sua vez, representam os possíveis valores que essa propriedade pode possuir. Por exemplo, uma faceta de “alunos” pode ser dividida em características de “homens” e “mulheres”, que, em seu turno, podem incluir valores como: “homens – Marco / Moisés / Paulo” e “mulheres – Paula / Ana / Carla”. Ademais, cada um desses valores pode manter relações com outras facetas, por exemplo, com a faceta “professores”.

Nesse sentido, o uso da taxonomia facetada navegacional permite nomear e organizar entidades em grupos que compartilham características similares, através de determinados procedimentos, tais como correlacionar diferentes linguagens usadas num ambiente (compatibilizar), criar mecanismos de acesso (*links*), normalizar informações disponibilizadas (vocabulário controlado) ou criar palavras-chave (termos) e conceitos que categorizam os conteúdos em ramos.

Cabe ressaltar, que essa forma de estruturação de informações, através de taxonomias facetadas, vem sendo comparada com as interfaces de busca disponibilizadas em *sites*, principalmente em *sites* comerciais (*e-commerce*). Contudo, sob o ponto de vista teórico da área de Ciência da Informação (CI), nem sempre o que há nesses portais pode ser denominado navegação facetada. Na CI, uma taxonomia facetada é uma cadeia de relações, na qual cada item de uma faceta, sob determinada categoria, está conectada a outros itens das facetas de diferentes categorias, com interligações sob diferentes tipos de relações. Dessa forma, perde-se a noção de ordem hierárquica, porque não há mais como indicar que uma faceta é mais importante ou mais complexa hierarquicamente, surgindo, assim, o sentido de cadeia de relações multidimensional e multidirecional.

Navegando em *sites* comerciais, o que se percebe é que muitas vezes as informações (ou produtos) estão estruturadas sob um único aspecto (atributo), ou sob até no máximo três aspectos, engessando a visão do conteúdo do portal, de modo a impedir a visão multidimensional do objeto em questão. É preciso esclarecer, que permitir as visões multidimensional e multidirecional de um objeto é necessário para que uma navegação possa ser considerada facetada, dentro dos princípios das teorias na CI.

Partindo dessa declaração, explica-se que no desenvolvimento desta dissertação cada documento foi indexado sob diferentes aspectos, permitindo uma visão multidimensional do seu conteúdo, através da estrutura da taxonomia facetada. Essa

estruturação foi realizada obedecendo à estrutura textual desse tipo de documento, possibilitando a navegação (*browsing*<sup>61</sup>) pelos itens sob cada categoria da estrutura. Assim, o usuário pode “navegar” pela taxonomia facetada, escolher um aspecto sob o qual pretende efetuar sua busca e localizar a informação desejada, sem que, em tempo algum, encontre um resultado vazio.

Tunkelang (2009, p.24) também participa dessa discussão e salienta que “a abordagem da navegação facetada é indicada para textos semi-estruturados<sup>62</sup> [caso das teses e dissertações], nos quais se supõe que seja possível representar seus assuntos em um sistema de classificação facetado” (tradução livre). Nesse caso, o documento pode ser segmentado por suas partes estruturais, pelo atributo de cada parte, formando um conjunto de categorias. Sob cada categoria, insere-se uma taxonomia hierárquica (as facetas e subfacetadas), que interliga os assuntos semelhantes naquele atributo específico, sendo que cada um deles é um metadado. Corroborando e acrescentando dados, Tunkelang (2009, p.24) afirma, ainda, que “o metadado de um documento consiste nos atributos da estrutura do próprio tipo de documento que, logicamente, segue a estrutura de produção do próprio texto” (tradução livre).

Um mecanismo de navegação facetada com essas características pode indicar, ao usuário final, as ocorrências do tipo de informação específica para cada categoria, facilitando as buscas do usuário final. Nessa diretriz, Tunkelang (2009) aponta algumas vantagens da navegação facetada em relação às buscas tradicionais: oferece orientação ao usuário, a formulação de questões de busca pode ser feita progressivamente, permite a exploração e a recuperação do recurso informacional através dos elementos em sua estrutura de forma fácil.

Pelo exposto, vê-se que as vantagens da navegação facetada estão, basicamente, ligadas à exploração do conjunto de informações disponíveis em cada categoria, que reflete o que está, efetivamente, disponível no banco de dados. Dessa maneira, faz-se necessário conhecer a natureza (domínio) desse tipo de informação disponibilizada e a quem ela se destina. Nesse momento, mostra-se oportuno apresentar teoria da análise de domínio, na qual se apóia essa averiguação.

---

<sup>61</sup> No jargão da biblioteca, o termo *browsing* consiste na exploração dos itens do acervo (livros, periódicos, folhetos), que estão dispostos nas prateleiras, examinando-os quase ao acaso, usando a intuição e os sentidos.

<sup>62</sup> Entende-se como texto “semi-estruturado” aquele que apresenta elementos estruturados (autor, data, título da obra, títulos das seções) e elementos sem estrutura (resumo, conteúdo).

## 4.6 Definindo o domínio: a teoria da análise de domínio

Sabe-se que o autor Hjørland foi o pioneiro na formulação explícita da teoria da análise de domínio<sup>63</sup> (TAD). Porém, Jesse H. Shera, no início da década de 70, já pensava a informação como originada e dependente do contexto social e cultural.

De modo a alinhar o eixo de discussão, salienta-se que, para Hjørland e Albrechtsen (1995), a TAD configura-se a forma mais adequada para identificar um domínio. Nessa perspectiva, o conhecimento é um produto histórico, cultural e socialmente desenvolvido, portanto, focado nos domínios de conhecimento (HJORLAND, 2002a).

É necessário esclarecer que o termo “domínio” é polissêmico, sendo, então, empregado em diversas disciplinas e em distintas situações, dentre as quais se podem citar:

- na disciplina Biologia: é aplicado para identificar a categoria mais alta da taxonomia de classificação dos seres vivos;
- na área do Direito: refere-se ao poder, como por exemplo, o domínio público, que significa estar sob o poder/direito de todos;
- no campo da Matemática: é usado no estudo de funções, sendo o conjunto de todos os possíveis valores de um conjunto “A” que a função leva até um elemento do conjunto “B”.
- em Computação, o termo “domínio” representa um grupo de computadores na *web*, os quais compartilham um mesmo endereço de comunicação;
- no campo da Engenharia de *Software*: é empregado para abrigar um conjunto de sistemas que possuem função semelhante, dentro de outro sistema, ou é usado como denominação a um conjunto de características que se referem a um problema, cuja solução será fornecida por uma determinada aplicação.

No *The American Heritage Dictionary of the English Language*, o termo “domínio” é definido como um espaço (*sphere*) de atividade, de negócio ou de função (ex.: o domínio da História). Além disso, possui sinônimos como *area* (área), *arena* (área de atividade), *field* (campo) e *sphere* (um espaço social particular; um extrato social; um campo específico de conhecimento; uma comunidade discursiva).

Entretanto, no campo disciplinar da CI, o entendimento e uso do termo “domínio” são explicitados por Hjørland e Albrechtsen (1995), quando afirmam que o domínio pode ser uma área de conhecimento (especialidade), um conjunto literário ou, ainda, a tarefa de um grupo de pessoas trabalhando em conjunto em um projeto ou em uma organização. Nessa mesma linha de pensamento, Lloréns (2004, p.846) acrescenta que um domínio pode ser

<sup>63</sup> A análise de domínio não deve ser confundida com a análise de assunto. Esta se refere a uma das etapas da indexação, no tratamento temático da informação documental, que visa identificar termos que melhor representem os assuntos tratados no documento, para posterior recuperação no sistema de informação. Já a análise de domínio, caracteriza-se como uma abordagem que pretende entender as funções implícitas e explícitas da informação, para a comunicação, dentro de uma comunidade discursiva.

uma área de conhecimento, uma atividade (projeto, manutenção, etc.), um interesse específico qualquer ou uma aplicação com fronteiras definidas. Além desses, Thellefsen e Thellefsen (2004, p.179) determinam que os domínios de conhecimento são “uma demarcação de um determinado conhecimento, seja ele fixado num contexto profissional ou não”.

É possível perceber a pluralidade de significações e conceitos atribuídos ao termo “domínio”. No entanto, na presente dissertação, o emprego do termo tem em vista o sentido adotado pela CI, o qual, em síntese, entende que o domínio pode compor desde uma disciplina, no todo ou em parte, ou apenas representar uma específica atividade profissional ou de lazer, sendo preciso, portanto, estabelecer e delimitar seu âmbito de abrangência para efeito de identificação.

De modo a verificar essas considerações, aponta-se que, segundo Hjørland (2002a), cada domínio do conhecimento possui uma estrutura conceitual própria, o que o difere de outro domínio. Para tanto, conforme esse autor, a abordagem da TAD alia teoria e prática, uma vez que o que busca é um método funcional, visando à compreensão das funções explícitas e implícitas da informação e as razões subjacentes aos fluxos de comunicação. Além disso, é um método filosófico que se contrapõe ao paradigma cognitivo, por este ter como objetivo encontrar elementos externos à percepção subjetiva do usuário. Para Capurro (2003, p.9), uma consequência prática do paradigma sócio-cognitivo é o de desistir da busca por uma linguagem ou algoritmo universal para a modelagem informacional nos sistemas de recuperação da informação, e pensar em soluções práticas para uma comunidade de usuários.

Nesse cenário, deve-se perceber que, dentro da organização e representação do conhecimento, Hjørland (2002b) sugeriu modos específicos de estudar os domínios: produção de classificações especiais, pesquisa especializada em indexação e recuperação, estudos da história da transferência de informações, estudos terminológicos, linguagens para objetivos especiais (LSP), estudos do discurso, entre outros. Apresentou, ainda, quatro métodos gerais para adquirir o conhecimento sobre o domínio: 1) métodos racionalistas (Ranganathan e a análise facetada); 2) métodos empíricos (bibliometria, co-citação); 3) métodos históricos; e 4) métodos pragmáticos (concentrando-se em objetivos, metas e utilidade).

Com propósito argumentativo, destaca-se a seguinte asseveração de Hjørland:

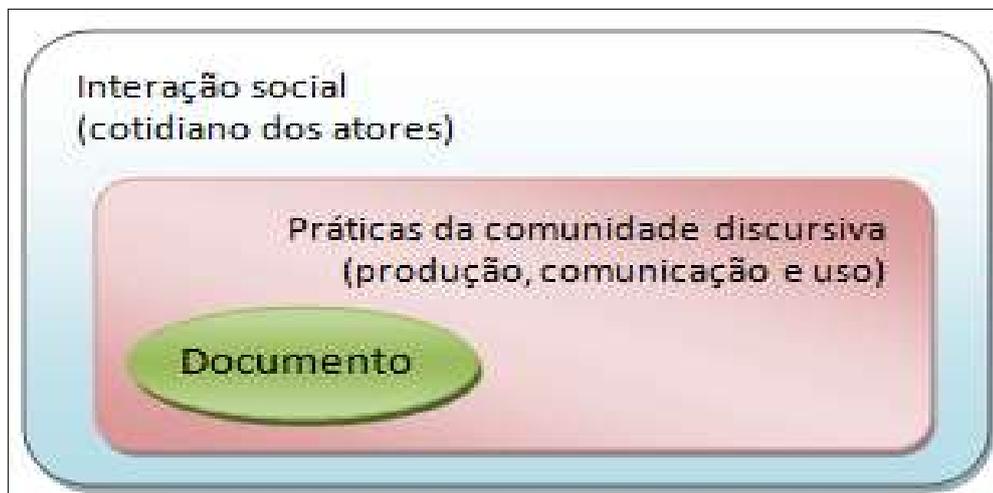
Um ponto central em minha abordagem é a alegação de que as ferramentas, conceitos, significados, estruturas de informação, necessidade de informação e critérios de relevância são moldados em comunidades de discurso, por exemplo, em disciplinas científicas, que são partes da divisão do trabalho feita pela sociedade. Uma comunidade de discurso sendo uma comunidade na qual ocorre um processo de comunicação ordenado e delimitado (HJORLAND, 2002a, p. 258, tradução nossa).

A partir dessa citação, é possível o entendimento de que a abordagem da TAD leva em consideração a comunidade de discurso na qual a informação é produzida, uma vez que toda prática social está também inserida em práticas informacionais. Comunidades discursivas, para Hjørland e Albrechtsen (1995, p.400), são distintos grupos sociais sincronizados em pensamento, linguagem e conhecimento, constituintes da sociedade moderna. Em vista disso, o ponto central da TAD consiste em estudar suas atividades e produtos, pois os “instrumentos, conceitos, significados, estruturas de informação, necessidades informacionais e critérios relevantes estão refletidos nas comunidades discursivas” (HJORLAND, 2002b, p.258).

Nesse contexto, nota-se que a TAD configura-se como uma proposta nova para a CI, pois sua essência de análise está no seio da comunidade de discurso, na qual o fluxo da informação busca um sentido prático de produção, de comunicação e de uso. A propósito, esse modo de perceber a realidade se aproxima do pragmatismo, que acolhe a ideia de que a verdade reside no sentido de utilidade, na concepção prática das coisas, ou seja, qualquer ato, objeto ou proposição deve ser útil ou gerar um efeito prático.

Por conseguinte, vem à tona a ideia de que, em qualquer nível, as interações humanas compreendem a geração (produção), recepção (uso) e transferência de informação (comunicação). Esse cenário de discussão pode ser retratado pela FIG. 5.

FIGURA 5 – O sujeito na Análise de Domínio



Fonte: Elaborada pela autora.

Nesse percurso, é cabível o registro de que, no momento da geração, as condições de produção da informação, sejam elas em relação aos temas escolhidos, aos objetos utilizados, aos valores ou aos seus significados, estão intensamente envolvidas na dimensão social do espaço/lugar de produção e de uso. Portanto, nas práticas

informativas, é preciso ter em vista o domínio de discurso e as interações sociais nas quais essas práticas acontecem.

A partir disso, pode-se entender que os conceitos passam a ser considerados como construídos socialmente, apesar de cada usuário vivenciar situações e circunstâncias próprias. Nesse sentido, o foco são os paradigmas (suposições e ideias sobre o que é significativo e como o conhecimento deve ser produzido) e não os modelos mentais. Essa ideia é corroborada através do pensamento de Hjørland (2002b), que afirma que os indivíduos não processam a informação segundo um sistema de regras individualistas e pessoais, mas, ao contrário, essas regras são fundadas na/pela sociedade e na cultura. Assim, o autor afirma que os conceitos são formados dentro das comunidades de discurso, isto é, dentro de cada domínio.

Para identificar as necessidades específicas de cada comunidade discursiva, nem sempre é preciso recorrer a estudos com os usuários, pois, segundo Hjørland (2002b) os usuários são incapazes de exprimir suas necessidades e critérios de relevância de forma real. Para o mesmo autor, relevante não é somente aquilo que é apontado pelos usuários, pois os critérios de relevância deles podem se modificar ao receberem uma nova informação. É por essa razão que Hjørland (2002b) acredita que, ao oferecer um serviço de informação e prescrever critérios de indexação e recuperação da informação, o profissional de informação deve criá-los a partir dos paradigmas, das normas metodológicas e epistemológicas. No entanto, afirma que a epistemologia permite a fundamentação teórica geral sobre a organização do conhecimento, mas não pode substituir o conhecimento do assunto, do domínio. Então, sugere que a área de CI deve estar aberta a diferentes visões de organização e representação da informação, de modo a auxiliar os usuários, oferecendo-lhes essa diversidade de visões nos sistemas de informação.

Ademais, os princípios da TAD entendem o indivíduo como um ser coletivo, inserido em relações intersubjetivas (entre diferentes sujeitos), e não um ser individual. Essa premissa sugere que as relações do ser humano com o mundo (realidade) estão sempre mediadas por informações produzidas, armazenadas, transferidas e compartilhadas no meio social, as quais ordenam e fornecem sentido a essa relação. É, basicamente, um sentido de alimentação e retroalimentação, em uma aceitação cíclica de comunicação.

Deste modo, encontrar a comunidade discursiva pressupõe responder a questões relativas a “quem produziu”, “em qual contexto social produziu”, “que finalidade tem a produção” e “quem usa o produto”. Entende-se que para criar um serviço de informação, em SRIs de bibliotecas digitais de teses e dissertações, é preciso responder a esses pontos, identificando o usuário e suas necessidades de informação. Considera-se que um “serviço de informação” tem como “função básica [...] fornecer informações relativas aos

respectivos acervos, bem como torná-los acessíveis aos usuários” (CUNHA; CAVALCANTI, 2008, p.333-334).

É preciso explicar que não há, na CI, a explicitação metodológica (um passo a passo) para ser utilizada na aplicação da TAD. Usualmente, quando se aplica essa metodologia, emprega-se a metodologia de análise de domínio referente à área de engenharia de *software*, como pode ser visto nos trabalhos de Guizzardi (2005) e de Dias (2010). Em vista disso, esta metodologia (da análise de domínio da engenharia de *software*) foi adaptada para a área da CI e é composta por quatro etapas, apresentadas a seguir.

#### **a) Identificação do domínio**

Nesta etapa deve-se caracterizar o domínio a ser analisado, a partir das seguintes indagações: qual é o domínio? o domínio é conhecido? que tipo de informações esse domínio produz?

#### **b) Análise de dados**

Nesta etapa, identificam-se os pontos centrais que caracterizam o domínio, estabelecendo relações e funções entre as informações, a partir dos seguintes questionamentos: quem produz as informações identificadas na etapa anterior? com qual propósito a informação é produzida? a quem interessa o uso das informações produzidas? A análise para responder a essas questões deve se basear nos princípios da TAD que afirma que o fluxo de informação é determinado no seio da comunidade de discurso, a partir do estudo de quem a produz, comunica e usa.

#### **c) Modelagem do domínio**

Nesta etapa, selecionam-se os tipos de informações consideradas importantes para o serviço de informação a ser criado, que estejam alinhados ao objetivo proposto, a partir dos dados coletados na análise anterior.

#### **d) Coleta e seleção de dados e validação**

Nesta etapa, há a identificação, na literatura especializada, tais como literatura técnica, histórico do domínio, questionário com especialistas do domínio ou outro procedimento capaz de respaldar a fase anterior, da escolha das informações mais importantes. A pergunta que norteia esta etapa é a seguinte: Quem é este “não-usuário total” a quem o conteúdo dos trabalhos acadêmicos, do tipo teses e dissertações, pode

interessar?. É preciso esclarecer, que o termo “não-usuário total” tem um sentido específico que é o daquele usuário que não usa um serviço porque acredita que tem outros meios de conseguir a informação ou porque não está ciente de que há aquele tipo de serviço disponível para ele.

No escopo desse levantamento, e no contexto desta dissertação, revela-se que a criação da taxonomia facetada navegacional, como mecanismo de busca na BDTD, da ECI-UFGM, tem como comunidade discursiva a “comunidade científica”, que é apresentada a seguir.

#### 4.6.1 Identificando o usuário: a comunidade científica

A comunidade científica também pode ser denominada comunidade acadêmica. Segundo Kuhn (2006), uma comunidade científica possui, entre os seus integrantes, interesses científicos semelhantes, uma bibliografia similar, vocabulário próprio (ainda que divergente em alguns pontos), compartilha modelos e abordagens metodológicos, partindo, geralmente, do resultado de pesquisas já realizadas. Para esse autor, ainda, a comunidade é fortalecida na medida em que o conhecimento produzido pode ser transferido e utilizado em novas abordagens de investigação dentro dela, intensificando o uso de suas próprias teorias, métodos e técnicas.

Nesse campo de análise, há dois termos com conceitos distintos – “comunicação” e “difusão” científica – que devem ser ressaltados. Ainda que, às vezes, sejam usados como sinônimos, há diferença entre eles: a “comunicação científica” é “[...] processo específico de produção, consumo e transferência de informação no campo científico” (CUNHA; CAVALCANTI, 2008, p.97), enquanto a “difusão” é “a disseminação da informação” (CUNHA; CAVALCANTI, 2008, p.125), sendo que essa disseminação pode ser entendida como “documentos distribuídos a pessoas ou entidades, a partir de um ponto central de armazenamento” (*ibidem*, p.130). Esse armazenamento, por sua vez, pode ser o acervo de uma biblioteca tradicional ou o banco de dados de uma biblioteca digital.

Nessa linha de raciocínio, tem-se que, com a implantação de bibliotecas digitais, de repositórios institucionais e de iniciativas de *open archives*<sup>64</sup>, a disseminação da informação científica vem sendo ampliada em níveis nacional e internacional, permitindo divulgar as publicações e resultados de pesquisas à sociedade em geral.

Resgatando a ideia de comunicação científica, Targino (2000, p.54) anuncia que é “a comunicação científica que favorece ao produto (produção científica) e aos produtores (pesquisadores) a necessária visibilidade e possível credibilidade no meio social em que

<sup>64</sup> *Open Archives* (arquivos abertos) são, geralmente, criados a partir de *software open source*, usando padrões de interoperabilidade, que possibilitam a disseminação da informação científica com acesso livre ao conteúdo que foi disponibilizado.

produto e produtores se inserem”. Em outros termos, entende-se que essa comunicação, ao indicar a credibilidade ou a refutação de um conhecimento desenvolvido em determinada comunidade discursiva, promove a consolidação dessa mesma área ou atividade e, por consequência, concede-lhe visibilidade.

Com finalidade didática, expõe-se o que pode ser encontrado no Dicionário de Biblioteconomia e Arquivologia, em que Cunha e Cavalcanti (2008) conceituam a comunidade científica como

**grupo social** formado por indivíduos que têm como profissão a pesquisa científica e tecnológica [...] as comunidades científicas são, sobretudo, **redes de organizações e relações sociais** formais e informais que desempenham várias funções. Uma das funções dominantes é a de **comunicação**. O papel da comunicação consiste em assegurar o **intercâmbio de informações sobre os** trabalhos em andamento, colocando os cientistas em contato entre si (*ibidem*, p.98-99, grifos nossos).

Percebe-se, assim, uma ideia semelhante à discutida no subcapítulo anterior, referente à teoria da análise de domínio, pois indica a comunidade científica como um grupo social, no qual os cientistas estão ligados por redes organizacionais e sociais, o que demonstra um caráter interativo entre a instituição e a sociedade. Ademais, a citação acima estabelece que a função primordial dos pesquisadores, em sentido *strictu sensu*, é produzir e usar informações dentro do grupo (seus pares), ou seja, comunicando e intercambiando informações concernentes a trabalhos em andamento ou já concluídos. Nesse sentido, reafirma-se o pensamento de que o produto das pesquisas tem especial importância e relevância dentro da própria comunidade acadêmica (domínio).

É necessário acrescentar que a comunidade científica é constituída e representada por cientistas, sendo que um cientista configura

pessoa que possui treinamento, habilidade e desejo de procurar novos conhecimentos, novos princípios e novos materiais em algum campo da ciência (DICTIONARY OF SCIENTIFIC AND TECHNICAL TERMS, 1984<sup>65</sup>). E entendo também como aqueles que fazem pesquisa para tentar entender e ampliar o conhecimento científico já conhecido. São aqueles que fazem trabalho criativo na ciência e tratam de alguma maneira de ampliar o conhecimento ou de melhorar as teorias científicas existentes (STEVENSON; BYERLY, 1995) (ALVARADO; OLIVEIRA, 2008, p.14).

Por meio disso, verifica-se a natureza investigativa do trabalho dos cientistas, que são movidos muito mais pelas formulações de questões, perguntas e interlocuções, do que pelas respostas obtidas. Para atender a esse propósito, em geral, os cientistas, “independentemente de seu campo de conhecimento, têm necessidades de diferentes espécies de informação, a cada estágio da pesquisa, no processo de gerar, disseminar e usar a informação” (FIGUEIRA NETTO, 1994, p.29). Nessa busca por informações, esse pesquisador muitas vezes se depara com o problema de filtrar, das diferentes bases de dados disponíveis, as informações que lhe são convenientes.

<sup>65</sup> DICTIONARY OF SCIENTIFIC AND TECHNICAL TERMS. 6<sup>th</sup> ed. New York: McGraw-Hill, 1984.

Ainda tendo em vista a citação de Alvarado e Oliveira (2008), o pesquisador necessita de três informações básicas, em distintos momentos de sua pesquisa: sobre temas já pesquisados anteriormente; 2) sobre métodos e técnicas utilizados em pesquisas; 3) resultados alcançados de pesquisas já concluídas. Analisando essas necessidades básicas, pode-se concluir que elas têm relação estreita com a atividade profissional (ocupação) desses cientistas. Segundo Case (2002), pesquisas sobre o comportamento informacional “por ocupação” indicam que as buscas por informação para o exercício da profissão (pesquisador) são investigações orientadas às tarefas e as buscas por soluções para estruturar (organizar) essas informações conforme a demanda do profissional são investigações orientadas aos sistemas.

Na perspectiva da busca por informação para o pesquisador, a informação, insumo para toda atividade humana, pessoal, técnica e para a comunidade científica, é essencial. Não podendo ser de outra maneira, dentro da sua comunidade discursiva, os cientistas geram, comunicam e usam informação. Sendo assim, no momento de produção das teses e dissertações, é preciso saber acerca de “quem” e “por que” produziu determinado conteúdo, pois, segundo Franco (2008),

1. Toda mensagem [...] contém, potencialmente, uma grande quantidade de informações sobre o autor: suas filiações teóricas, concepções de mundo, interesses de classe, [...] representações sociais, motivações, expectativas, etc. 2. O produtor/autor é antes de tudo um selecionador e essa seleção não é arbitrária. Da multiplicidade de manifestações [...], seleciona o que considera mais importante para “dar o seu recado” e as interpreta de acordo com seu quadro de referência. Obviamente, essa seleção é preconcebida. Sendo o produtor, ele próprio, um produto social, está condicionado pelos interesses de sua época, ou da classe a que pertence. E, principalmente, ele é formado no espírito de uma teoria da qual passa a ser o expositor. 3. A ‘teoria’ da qual o autor é o expositor orienta sua concepção da realidade. Tal concepção (consciente ou ideologizada) é filtrada mediante seu discurso e resulta em implicações extremamente importantes (FRANCO, 2008, p.25-26).

Então, os resultados das pesquisas expõem as “filiações teóricas” do autor, assim como seus interesses dentro da comunidade discursiva (interesses de classe). Por essa razão, é preciso descobrir o conteúdo manifesto (explícito) do texto produzido, de forma que ele possa ser acessado, quando, então, poderá ter seu conteúdo latente (implícito) desvendado pelo uso. Entretanto, para a apreensão do conteúdo manifesto das teses e dissertações, deve-se recorrer ao método de Análise de Conteúdo (AC), usando a técnica da análise categorial temática, a qual será discutida a partir deste momento.

#### **4.7 Método de análise de conteúdo: a técnica categorial temática**

A Análise de Conteúdo (AC) é um método bastante utilizado na investigação empírica realizada nas ciências humanas e sociais. Consiste na análise textual, com ilações

qualitativas e dados quantitativos, para investigação histórica ou para estudos bibliométricos, por exemplo.

Esse método se destacou, a partir do século XX, como forma de validação da análise realizada em textos, utilizando a sistematização para tornar o diagnóstico interpretativo mais objetivo (ROCHA; DEUSDARÁ, 2005).

Avançando nesta reflexão, verifica-se que, segundo Campos C. (2004), Bernard Berelson, um cientista comportamental americano, na década de 40, entendia a AC como baseada no modelo cartesiano de pesquisa, conceituando-a como uma descrição do conteúdo manifesto dos textos, de maneira objetiva, sistemática e quantitativa. Atualmente, essa visão vem sofrendo críticas, pela rigidez do método, o que levou Bardin a acrescentar um sentido qualitativo na análise de conteúdo:

um conjunto de técnicas de análise das comunicações visando obter por procedimentos sistemáticos e objectivos [sic] de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (BARDIN, 2009, p.44).

Assim, a AC se mostra como um conjunto sistemático de técnicas que visam a analisar o conteúdo de documentos, com o objetivo de identificar o uso de determinadas palavras ou conceitos e as relações existentes entre eles, através de análise de dados qualitativos, quantitativos ou um misto de ambos. Com relação às condições de produção/recepção, são denominadas por Bardin “variáveis inferidas”, pois podem advir da inferência de variadas formas, tais como “variáveis psicológicas do indivíduo emissor, variáveis sociológicas e culturais, variáveis relativas à situação de comunicação ou do contexto de produção da mensagem” (BARDIN, 2009, p.42).

O método da AC abriga diferentes técnicas de análise: análise de avaliação (de atitudes), análise da anunciação (processo de comunicação), análise proposicional do discurso (significado dos enunciados), análise léxica (vocabulário/ideias), análise da expressão (indicadores formais), análise das relações (intratextual) e análise categorial (temática/semântica).

Em virtude disso, observa-se que a AC pode ser adequada para distintas aplicações, pois

a análise de conteúdo é um instrumento de pesquisa científica com múltiplas aplicações. Os procedimentos utilizados podem variar em função dos objetivos da pesquisa, entretanto, sejam quais forem suas finalidades, é preciso que ela se submeta, para que tenha valor científico, a algumas regras precisas que a diferenciem de análises meramente intuitivas (OLIVEIRA, 2008, p.569-570).

Oliveira (2008) ressalta, assim, o caráter versátil desse método e, também, o rigor de sua metodologia, o que confere confiabilidade aos resultados obtidos por sua utilização. Dessa maneira, na presente dissertação, a aplicação do método foi efetuada a

partir da abordagem da técnica de análise categorial temática (semântica) (ACT), que “funciona por operações de desmembramento do texto em unidades, em categorias segundo reagrupamentos analógicos [agrupar objetos diferentes por sua semelhança, por sua similaridade]” (BARDIN, 2009, p.199), por ser considerada a mais adequada para analisar o conteúdo manifesto (explícito) dos textos. Em linhas gerais, o que se pretende é determinar as categorias temáticas que são capazes de revelar o conteúdo de um texto por completo, e não as ideias contidas nos documentos. Nesse sentido, entende-se que uma categoria, representada por um termo, carrega um sentido semântico que a torna única em dado contexto.

Atendendo a essa expectativa, na análise categorial temática, o discurso (textos em geral, como respostas a questionários, documentos, cartas) é fragmentado em categorias, por ocasião de critérios, objetivos e recorte de pesquisa. Assim sendo, a análise categorial temática (ACT) consiste em desmembrar e categorizar o texto, a partir de agrupamentos por temas, para análises temáticas ou para outra propriedade qualquer que se queira estudar.

Antes de passar às etapas da aplicação desta técnica, é necessário apresentar alguns conceitos relevantes no contexto da AC, segundo alguns autores que abordam o assunto (MORAES, 1999; OLIVEIRA, 2008; BARDIN, 2009).

**Objetividade:** as categorias devem ser definidas de tal forma que seja possível a outro pesquisador chegar, a partir dos mesmos critérios, ao mesmo desmembramento, sendo que o conceito em questão pode ser verificado e replicado.

**Sistematicidade:** a análise deve ater-se ao objetivo proposto e ao conteúdo do texto, a fim de minimizar a influência dos pressupostos pessoais do pesquisador, evitando-se a ocorrência de escolha arbitrária de categorias, ainda que sejam determinadas *a priori*.

**Conteúdo Manifesto:** a análise deve limitar-se ao que está explicitado no texto, efetivamente expresso, e não ao conteúdo latente (implícito) do documento. Mesmo que, no tipo de análise aplicada, se levem em consideração os conteúdos latentes, estes devem estar baseados no conteúdo real do texto.

**Unidades de Registro (UR):** é a menor parte do conteúdo, sendo a que será selecionada de acordo com as categorias levantadas. Consiste na unidade de fragmento de texto que será classificado, podendo ser uma palavra, um termo, um parágrafo do texto ou parte de uma frase. Essas unidades, por sua vez, devem ter sentido completo em si mesmas. A UR também pode ser denominada unidade de análise ou unidade de significado.

**Unidades de Contexto (UC):** contextualizam as unidades de registro, dando significado a elas e limitando sua interpretação. Cada unidade de contexto pode ter diversas unidades de registro.

Construção de Categorias (CC): atividade de classificação de acordo com critérios pré-estabelecidos que busquem atender ao objetivo pretendido. Esses critérios podem ser semânticos (temáticos), léxicos (as palavras e seus sentidos), expressivos (problemas de linguagem) ou, ainda, construídos a partir de verbos, adjetivos, substantivos, entre outros. Entretanto, o conjunto de categorias deve obedecer somente a um critério. Isso significa impor uma organização intencional ao conteúdo do texto, determinando um discurso diferente do original.

Ainda na categorização, segundo Bardin (2009), há a classificação dos elementos do texto em agrupamentos segundo suas semelhanças e diferenças, em um processo estruturalista de classes reunidas sob um termo genérico. Por meio desse arranjo, a autora salienta que as categorias têm a função de condensar o texto completo, constituindo uma representação do mesmo. Além disso, através das relações entre as diferentes categorias é possível que se criem inferências e interpretações. A propósito disso, Bardin afirma que a AC se “assenta implicitamente na crença de que a categorização (passagem de dados brutos a dados organizados) não introduz desvios (por excesso ou por recusa) no material, mas que dá a conhecer índices invisíveis, ao nível dos dados em bruto” (*ibidem*, p.147). Contudo, para atenderem a esse postulado, as categorias devem ter como características a adequação (válidas e pertinentes), exaustividade, homogeneidade, exclusividade e objetividade, as quais serão explicadas a seguir.

Adequação (válidas e pertinentes): todas as categorias devem ter um significado útil para o objetivo pretendido, adequando-se ao problema e à fundamentação teórica da pesquisa realizada. Do mesmo modo, todo o conteúdo do texto precisa estar expresso nas categorias, as quais devem permanecer na fronteira de dois extremos: criar novas categorias até que seja o suficiente para representar todo o texto e, ao mesmo tempo, manter um número mínimo possível de categorias. Se criadas *a priori*, as categorias devem estar baseadas em uma fundamentação teórica, com critérios de classificação pré-estabelecidos.

Exaustividade (ou inclusividade): as categorias devem ser capazes de abrigar todas as unidades de registro (UR) selecionadas no texto, permitindo a inclusão de todo dado significativo, de acordo com o objetivo pretendido.

Homogeneidade: é necessário que o conjunto de categorias obedeça a apenas um critério de categorização, em uma dimensão única de fragmentação do texto, ou seja, satisfaça a somente uma variável.

Exclusividade (ou exclusão mútua): cada unidade de registro (UR) somente poderá pertencer a uma categoria, a partir de regras precisas e claras. Dessa forma, não há dúvidas quanto ao pertencimento de cada UR a uma determinada categoria.

Objetividade (ou consistência e fidedignidade): diferentes pesquisadores têm de chegar a resultados semelhantes, com base nos mesmos critérios de análise estabelecidos, os quais, efetivamente, devem ser explicitados. Ainda que outro pesquisador nomeie as categorias com termos diferentes, a fragmentação do conteúdo (do todo) será o mesmo.

É preciso destacar, ainda, que na construção das categorias os itens elencados anteriormente devem ser continuamente revistos durante a análise, visando ao ajustamento contínuo das categorias.

Feitas essas necessárias explicações, indica-se, no QUADRO 2, uma síntese das metodologias para a análise categorial temática apresentadas pelos autores Moraes (1999), Oliveira (2008) e Bardin (2009).

QUADRO 2 – Síntese das metodologias da análise categorial temática

MORAES (1999)	OLIVEIRA (2008)	BARDIN (2009)
<b>Etapa 1:</b> Preparação das informações		
<b>Etapa 2:</b> Unitarização ou transformação do conteúdo em unidades	<b>Etapa 1:</b> Pré-análise	<b>Etapa 1:</b> Pré-análise
<b>Etapa 3:</b> Categorização ou classificação das unidades em categorias	<b>Etapa 2:</b> Exploração do material ou codificação	<b>Etapa 2:</b> Codificação <b>Etapa 3:</b> Categorização
<b>Etapa 4:</b> Descrição <b>Etapa 5:</b> Interpretação	<b>Etapa 3:</b> Tratamento dos resultados, inferência e interpretação	<b>Etapa 4:</b> Inferência

Fonte: Baseado em Moraes (1999), Oliveira (2008) e Bardin (2009).

Observa-se pelo QUADRO 2 que as três metodologias são similares, diferenciando-se apenas na apresentação de suas subdivisões e, em geral, abarcam as seguintes etapas: 1) pré-análise; 2) exploração do material; 3) tratamento dos resultados; 4) inferência; e 5) interpretação. Essas etapas serão, abaixo, vistas de modo mais detalhado.

1) Pré-análise: consiste, basicamente, em três elementos: escolher os documentos para a análise, cuja escolha pode ser *a priori*, determinar o objetivo pretendido e formular as hipóteses ou pressupostos. Esses procedimentos não são lineares, mas são interdependentes;

2) Exploração do material: analisar e conhecer os documentos (o *corpus*) escolhidos para a análise, decidindo se eles, realmente, enquadram-se no objetivo pretendido e, ainda, se cobrem o domínio investigado;

3) Tratamento dos resultados: inclui a codificação (ou unitarização) dos documentos, a identificação das unidades de registro (UR), definição das unidades de contexto e a categorização. De maneira conceitual têm-se os seguintes procedimentos:

- codificação: estabelecer códigos que identifiquem cada texto (documento);
- identificação das unidades de registro (UR): retirar do texto suas URs, estabelecendo códigos para cada uma delas. O tipo de UR utilizado será determinado pelo objetivo pretendido, com a condição de que tenha um significado em si mesmo, pois será empregado fora do contexto do documento, conservando, naturalmente, o sentido original;
- definição das unidades de contexto (UC): definir (conceituar) cada uma das unidades de contexto, a fim de fornecer uma referência de significado às URs, limitando sua interpretação. Essa é a unidade de compreensão para codificar as UR, pois cada UC terá diferentes URs;
- categorização: agrupar os dados, segundo suas similaridades, seguindo um único critério de divisão, previamente determinado. Nessa etapa o conjunto de categorias temáticas é adequado continuamente, pois primeiramente há o conjunto de “categorias iniciais”, depois o conjunto de “categorias intermediárias” e, finalmente, o conjunto de “categorias finais”. Além disso, deve-se obedecer às características já discriminadas e explicadas anteriormente (adequação, exaustividade, homogeneidade, exclusividade e objetividade). Explica-se, também, que a categorização é um esforço contínuo de formulação e reformulação, em um processo cíclico. Nesse sentido, as categorias devem ser, sempre, analisadas e revisadas, o que ultrapassa a etapa três, de tratamento dos resultados. Geralmente, somente se chega ao conjunto das “categorias finais” ao término da etapa seguinte, de inferência (ou descrição).

4) Inferência (ou descrição): descrever os resultados obtidos através de gráficos ou tabelas, seja para análises quantitativas (frequências e percentuais) ou qualitativas (definições para as categorias, de preferência com citações diretas). Sobre a análise qualitativa, sabe-se que ela funciona bem em *corpus* reduzidos, quando se lida com categorias mais específicas, nas quais o contexto (de produção e de uso) tem importância capital. A inferência pode ocorrer em relação à mensagem (significado e código) ou no que se refere ao emissor/receptor. Em relação ao emissor (produtor da mensagem), que pode ser um indivíduo ou um grupo de indivíduos, a função expressiva é de comunicação, pois a mensagem representa o emissor. Já quando se trata do receptor, que também pode ser um indivíduo ou um grupo de indivíduos, restritos ou não, a função é de instrumentação (dar condição de ação) ou de adaptação (de aceitação pelo indivíduo ou grupo).

5) Interpretação: ultrapassa a descrição e analisa se os resultados alcançados são passíveis de generalização. Está presente com mais vigor na pesquisa qualitativa do que na quantitativa, uma vez que na primeira a ACT recomenda que se analise o conteúdo

manifesto (explícito; de produção) e o latente (implícito; de uso). Finalmente, esta etapa pode ser realizada, basicamente, por duas vias: 1) fundamentação teórica claramente explicitada *a priori*, com a análise comparativa entre o resultado e a literatura; 2) nos dados e nas categorias da análise: a construção da teoria, a partir da análise dos documentos, o que também já constitui a interpretação.

Ainda é preciso esclarecer, que as categorias podem ser estabelecidas *a priori*, desde que obedecem aos critérios já aventados anteriormente

Entretanto, seja com categorias definidas *a priori*, seja com uma categorização a partir dos dados, o estabelecimento de categorias necessita obedecer a um conjunto de critérios [...] ser *válidas, exaustivas e homogêneas* [...] ser *mutuamente exclusiva* [...] ser *consistente* (grifos do autor) (MORAES, 1999, p.15).

Dessa forma, desde que atendam aos critérios prescritos, o conjunto de categorias utilizado pode ser determinado *a priori*. Corroborando com essa afirmativa, Rocha e Deusdará (2005, p.321), afirmam que a ACT pode ser a “reprodução e disseminação de uma realidade *a priori*” e que, como concepção de ciência, é um “instrumento neutro de verificação de uma determinada realidade”. Portanto, a prática de conceber o conjunto de categorias *a priori* é possível e aceitável.

Ressalta-se o fato de que o método AC e a análise documentária (AD) ou documental (BARDIN, 2009, p.47) são, em algumas situações, tratados de modo indistinto e, portanto, como sinônimas. Contudo, mesmo havendo semelhanças entre essas duas técnicas, sobretudo nos processos intelectuais empregados em ambas, tais como os recortes de informação, analogias ou na representação condensada das informações (BARDIN, 2009, p.47), as duas técnicas apresentam diferenças que devem ser consideradas.

Nesse sentido, aponta-se como primeira diferença básica o fato de que, enquanto a AC se ocupa do conteúdo da mensagem, a fim de demonstrar os indicadores que deixam perceber outra realidade que não a da mensagem em si, a AD trabalha documentos e sua representação, visando à recuperação de informações pelo usuário (BARDIN, 2009, p.47).

Com esse mesmo olhar investigativo, Bardin (2009, p.47) sugere que o propósito da análise documentária é obter “o máximo de informação (aspecto quantitativo), com o máximo de pertinência (aspecto qualitativo)”. Em seu turno, Tálamo *et al.* (1992, *on-line*), afirma que a AD “tem como objetivo propor princípios para a elaboração de informações documentárias, definidas como múltiplas representações de textos para fins de recuperação face às diferentes demandas de informação”. De acordo com essa definição, entende-se que a análise documentária é realizada para a indexação de documentos com objetivos

específicos que se voltam para a prática de atender às demandas informacionais de um usuário, facilitando acesso e recuperação de informação.

No processo de indexação de documentos, destaca-se o importante papel desempenhado pelos vocabulários controlados, no sentido de controlar os significados dentro de um domínio e, assim, padronizar a entrada de dados no SRI, melhorando a qualidade da recuperação da informação.

Voltando a atenção ao tema desta pesquisa, que é a criação de uma taxonomia facetada, sente-se a necessidade de destacar as similaridades existentes entre a técnica da análise categorial temática (ACT), do método da análise de conteúdo (AC), com o método da análise facetada, da teoria da análise facetada desenvolvida por Ranganathan. Essa necessidade advém do fato de que a ACT foi utilizada para a atribuição e validação do conjunto de categorias fundamentais criado para compor a taxonomia. Passa-se, assim, a apresentar essas similaridades.

#### **4.7.1 Análise categorial temática e análise facetada: similaridades**

De antemão, é válido registrar que toda análise textual pressupõe a decomposição ou fragmentação de um texto em suas partes constitutivas, buscando seus elementos essenciais, assim como a identificação das relações existentes entre os mesmos. Com isso, pretende-se interpretar e apreender o texto como um todo, determinar seus temas, seu enredo, suas argumentações, explicações, descrições e conclusões, que, juntos, compõem a base de sustentação do texto.

Em face disso, ao estudar as etapas da análise categorial temática (ACT), percebeu-se similaridades com a análise facetada, princípio que norteia o desenvolvimento de taxonomias facetadas. Então, esse foi o principal motivo para a escolha do método da Análise de Conteúdo, utilizando a técnica da ACT, como metodologia de validação do conjunto de categorias fundamentais temáticas, determinado a partir de levantamento de normas e da literatura especializada da área.

Dessa maneira, já se verifica o fato de que tanto a Análise Facetada (AF) quanto a Análise de Conteúdo (AC), com sua técnica de análise categorial temática (ACT) possibilitam a análise de um mesmo conjunto documental sob diferentes aspectos e distintos objetivos. Com o objetivo de estender esta constatação, apresenta-se, a seguir, uma breve comparação, com apontamentos mais precisos referentes aos elementos similares entre essas duas formas de análise.

Primeiramente, é preciso ressaltar que, na atividade de AF, Ranganathan (1967) indicou três níveis distintos: o plano das ideias, para a análise conceitual; o plano verbal, concernente à expressão verbal (termos) dos conceitos; e o plano notacional, determinando

códigos para a localização dos termos e sinais que pudessem sugerir as relações entre os conceitos tratados nos documentos. Da mesma maneira, na ACT, podem ser ressaltados esses três planos, ainda que não sejam discriminados dessa forma. Para tal entendimento, basta refletir que o plano das ideias, na ACT, é representado pelas regras estabelecidas como critérios para a escolha das categorias, das unidades de registro e unidades de contexto, que devem, sempre, estar atreladas ao objetivo e à finalidade pretendidos. No plano verbal, também na ACT, tem-se a escolha do termo mais adequado, pertinente e válido, para representar um conceito. E, no plano notacional, mesmo que não se enquadre no sentido de armazenamento, a ACT determina critérios para a codificação do material a ser analisado, facilitando sua identificação e localização.

Para atender a esses três planos, necessita-se de três etapas para a AF e para a ACT, que são indicadas na sequência, fazendo-se considerações sobre suas similaridades.

### **Primeira etapa**

Análise facetada: identificação do documento para análise.

Análise categorial temática: pré-análise com a escolha dos documentos, que sejam representativos em relação ao objetivo proposto. Essa escolha pode ser *a priori*.

Considerações: A análise facetada sempre será realizada para um domínio já definido *a priori*, pois se pressupõe que a análise seja efetuada em documentos de um domínio específico e para um objetivo pré-definido. Do mesmo modo, a análise categorial temática também almeja a análise de um domínio específico, com um objetivo pré-definido e formulação de hipóteses (ou pressupostos), mas difere quanto à seleção dos documentos que poderá acontecer *a priori* ou em uma seleção por amostragem. Isso, no entanto, é feito simultaneamente com a determinação dos objetivos. Como afirma Bardin (2009, p.121), “a escolha dos documentos depende dos objetivos, ou, inversamente, o objetivo só é possível em função dos documentos disponíveis”. Assim, essa particularidade não as torna diferentes sob esse aspecto.

### **Segunda etapa**

Análise facetada: leitura do documento.

Análise categorial temática: leitura flutuante, para conhecer o texto.

Considerações: Na análise facetada, durante a leitura do documento, há o processo constituído pela identificação do assunto e sua constituição, que seguem orientações do tipo: dissecação, para segmentar o domínio em termos de mesmo nível; laminação, para combinar termos isolados, formando termos compostos; desnudação, para segmentar, ainda mais, o domínio e, assim, obter maior especificidade nos assuntos;

reunião livre de termos, para permitir a formação de assuntos complexos (isolado + composto); interpolação, que é a possibilidade de reunião de ideias isoladas (multidimensionalidade). Nota-se que algumas dessas orientações somente serão evidenciadas durante o uso da classificação facetada, tais como a laminação, reunião livre dos termos e a interpolação. Assim, tem-se, nesta fase, a dissecação e a desnudação, que são atividades de leitura do documento, que buscam a segmentação do domínio de forma que atenda ao objetivo proposto. Com relação à análise categorial temática, a leitura flutuante consiste em analisar e conhecer o documento, identificando e desmembrando o texto em “unidades de registro – UR”, quando é decidido se estas serão representadas por frases, expressões, palavras ou por categorias de palavras (substantivo, verbos, adjetivos), descobrindo os núcleos de sentido. Isso remete à dissecação e desnudação, da análise facetada, pois decide-se o grau de especificidade da representação, a partir do objetivo pretendido. Além disso, nesta análise, as outras orientações (laminação, reunião e interpolação) também podem ser evidenciadas ao final da atividade classificatória, pois a representação é referente a um texto completo, no qual as partes se combinam, juntam-se e interpõem.

### **Terceira etapa**

Análise facetada: seleção de termos relevantes e categorias (facetas e subfacetas), ordenação das facetas, subfacetas e focos, e organização final.

Análise categorial temática: tratamento dos resultados, inferência e interpretação.

Considerações: Nas duas análises, este é o momento da atividade de categorização, na qual o método da análise facetada é desenvolvido até que se obtenha a estruturação final.

Na análise categorial temática (ACT), o “tratamento dos resultados” há a identificação das unidades de registro (UR), que são os conceitos relevantes dentro do domínio, que é representado pelas unidades de contexto (UC), e as UCs são as categorias mais gerais que contextualizam e dão significado às URs. É elaborada uma definição para as categorias de forma que o sentido das URs fique restrito à demarcação de cada UC, uma vez que uma UC (categoria) pode possuir diferentes URs (facetas, subfacetas e focos). Há, ainda, a categorização, realizada em um processo cíclico de análise, validação e reformulação. Essa é a atividade de agrupar as URs conforme suas semelhanças, a partir de critério único de divisão, obedecendo às características de adequação, exaustividade, homogeneidade, exclusividade e objetividade, em um contínuo que ultrapassa esta etapa três.

Em Ranganathan e CRG, a “seleção de termos relevantes e categorias” é a atividade de aplicar o método da análise facetada que fornece os princípios para a categorização de dado domínio, com o agrupamento de conceitos semelhantes (facetas e subfacetas) em classes mais gerais (categorias). Nesse processo de agrupar, ou de categorizar, são realizadas análises do domínio de maneira que se possa identificar os conceitos mais gerais (categorias) e os mais específicos (facetas e subfacetas). Para Ranganathan, todo o conhecimento humano, representado pelos assuntos (ideias, conceitos) tratados nos registros de documentos, em qualquer domínio a ser investigado, pode ser acomodado, e estruturado em cinco categorias fundamentais, o PMEST: Personalidade, Matéria, Energia, Espaço e Tempo. As facetas e subfacetas, que alimentam as categorias fundamentais, são os assuntos (ou conceitos) inscritos nos documentos, que estão representados por seus termos.

Em relação à ordem dentro de cada categoria (das facetas, subfacetas, cadeias, focos e renques), que é recomendada nos princípios do Ranganathan e do CRG, os procedimentos da técnica da análise categorial temática (ACT) a desconsidera. Entretanto, essa diferença não apresenta gravidade no contexto desta dissertação, pois em ambiente digital a ordem de apresentação não se mostra de grande importância. Houve a escolha por apresentar as facetas, subfacetas e focos na ordem alfabética.

Em seus estudos, Ranganathan (1967) ressalta que há liberdade para a criação de um maior número de categorias, desde que o uso delas seja empiricamente avaliado com a categorização de documentos utilizando tais categorias e que o resultado seja satisfatório (critério de utilidade). Para atender a essa recomendação, a técnica da ACT foi empregada para essa validação do conjunto de categorias estabelecidas *a priori*, assim como houve uma experiência utilizando esse conjunto na indexação dos documentos do *corpus*. Adicionalmente, a ACT indica que seja realizada a “inferência e interpretação” na qual se deve re-validar o conjunto de categorias a partir de um respaldo teórico ou sustentar uma nova teoria ou modelo desenvolvido.

De modo a ampliar o campo de discussão aqui proposto, resgatam-se os cânones do Plano das Ideias e do Plano Verbal de Ranganathan e do CRG, referentes às características e ao número de facetas, em relação às condições exigidas para as categorias na análise categorial temática (ACT), os quais podem ser considerados conforme o que se segue:

### **Plano das ideias**

a) Diferenciação (Ranganathan): equivale à Sistemática e à Categorização da ACT;

- b) Relevância (Ranganathan e CRG): equivale à Objetividade da ACT;
- c) Verificação (Ranganathan e CRG): equivale à Adequação (válidas e pertinentes) da ACT;
- d) Permanência (Ranganathan e CRG): equivale à Adequação (válidas e pertinentes) e à Objetividade (ou consistência e fidedignidade) da ACT;
- e) Homogeneidade (CRG): equivale à Homogeneidade da ACT;
- f) Mútua exclusão (Ranganathan e CRG): equivale à Exclusividade da ACT.

### **Plano verbal**

- a) Contexto (Ranganathan): equivale à Unidade de Contexto e à Adequação (válidas e pertinentes) da ACT;
- b) Enumeração (Ranganathan): equivale à Codificação da ACT;
- c) Aceitabilidade ou terminologia usual (Ranganathan): equivale à Adequação (válidas e pertinentes) e à Objetividade (ou consistência e fidedignidade) da ACT;
- d) Restrição (Ranganathan): equivale à Objetividade (ou consistência e fidedignidade) da ACT.

Outro aspecto a ser considerado diz respeito à Exaustividade. Neste contexto, entende-se que “as classes, as cadeias de classes e a lista de isolados nas cadeias devem ser totalmente exaustivos em relação a seus respectivos universos em comum” (RANGANATHAN, 1967, p.158). Isso remete à Exaustividade (ou inclusividade) estabelecida pela ACT que determina que as categorias devem ser capazes de abrigar todas as unidades de registro (UR) selecionadas no texto, permitindo a inclusão de todo dado significativo, de acordo com o objetivo pretendido. Além disso, a Exaustividade, segundo Ranganathan, no que tange ao significado de “ser exaustivo em relação aos universos em comum”, alude ao sentido de “Conteúdo Manifesto” da ACT, quando esta última afirma que a análise deve limitar-se ao conteúdo explicitado no texto.

Finalmente, é preciso ressaltar que Ranganathan desenvolveu sua teoria em meio aos problemas de classificação de registros do conhecimento, elaborando melhores práticas para a indexação do conteúdo temático dos documentos, com a finalidade de destacar as ideias explicitadas no texto. Nesse sentido, há estudos que criam modelos para orientar os indexadores na análise documentária, tema que será apresentado no subcapítulo a seguir.

## 4.8 Modelo de Leitura Documentária de Fujita e Rubi

A análise documentária<sup>66</sup> é atividade essencial de um bibliotecário-indexador, que consiste em estabelecer os assuntos que representam o conteúdo de um determinado documento, através de uma leitura técnica. Nessa leitura técnica o indexador examina o texto sistematicamente, através dos elementos que o compõem (tais como o sumário, a folha de rosto, a orelha, o resumo ou a apresentação), com o objetivo de entender como o todo foi organizado e quais as ideias principais do texto. Nesse processo, a atenção do bibliotecário tem em vista a recuperação das informações pelos usuários da biblioteca, seja ela tradicional ou digital. Daí, a importância da disciplina de estratégias de análise documentária na formação profissional dos bibliotecários.

No âmbito da CI, Tálamo *et al.* (1992, *on-line*) afirmam que a análise documentária “tem como objetivo propor princípios para a elaboração de informações documentárias, definidas, por sua vez, como múltiplas representações de textos para fins de recuperação face às diferentes demandas de informação”. Em face dessa asserção, entende-se que a CI busca por métodos e técnicas para representar o conteúdo dos documentos visando aprimorar a recuperação de informações.

É preciso destacar que em toda análise documentária estão embutidos o objetivo pretendido e as necessidades informacionais dos usuários. Nesse sentido, Albrechtsen (1993, p.220) indica que há três formas para a análise de um documento: forma simplista, forma orientada ao conteúdo ou forma orientada à demanda. Na forma simplista, os conceitos são atribuídos estatisticamente, podendo ser uma tarefa completamente automatizada. Já na análise orientada ao conteúdo, atenta para a interpretação temática implícita e explícita contida no texto. E, finalmente, na análise orientada à demanda, é regida pelo objetivo e necessidade *ad hoc*, isto é, para um fim específico, que visa à transferência de informação e conhecimento.

Vale lembrar neste momento, que a análise de assunto (ou conceitual) é uma das etapas da indexação, constituída basicamente por duas fases: a análise conceitual e a tradução. Na análise de assunto (ou conceitual) há a identificação e seleção dos conceitos que representam os assuntos tratados no documento. Já a tradução consiste em transmutar a linguagem natural utilizada para apresentar nos conceitos do texto para uma representação de uma linguagem controlada, preferencialmente, de um vocabulário controlado empregado pelo sistema. Esse controle terminológico permite melhorar a recuperação das informações, através dos termos indexadores que padronizam,

---

<sup>66</sup> Há certa polêmica sobre o uso da expressão “análise documentária” (AD) que, na corrente francesa, é considerada como um campo mais amplo no qual o processo de indexação está inserido. A corrente espanhola, por sua vez, considera a AD como composta de forma (descritivo) e conteúdo (temático). Já na corrente inglesa considera a AD e a indexação como processos idênticos (SILVA; FUJITA, 2004).

principalmente, problemas relacionados com a sinonímia e ambiguidades. Segundo Dodebei (2002, p.79), “a unidade informacional é o conceito, embora é claro, ela deva estar expressa por termos, simbolicamente representados por palavras”. Nesse sentido, o termo é uma representação conceitual, controlada, do conteúdo de um texto.

A atividade de indexação é uma prática sujeita à subjetividade, pois envolve o discernimento e conhecimento do profissional indexador, o que pode gerar certa inconsistência. Na literatura há relatos desse tipo problema, tais como a inconsistência intraindexador e inconsistência interindexador. A inconsistência intraindexador, acontece quando um mesmo profissional indexa um documento de diferentes formas, em distintas ocasiões temporais, e a inconsistência interindexador acontece quando distintos profissionais indexam um mesmo documento de maneiras diferentes (BORKO, 1977).

Como pode ser percebido nos últimos parágrafos, há distintas maneiras para a análise documentária, sendo imprescindível que o indexador esteja preparado para enfrentar as dificuldades e subjetividades envolvidas na atividade de leitura e interpretação de textos. Na literatura da área da CI, encontram-se propostas de metodologias que têm por finalidade orientar a leitura técnica dos indexadores. Como exemplo, há a metodologia de Kobashi (1994), que sugere formas de identificação do tema do documento, a partir da leitura do texto. Além dessa, há a proposta metodológica de Tálamo (1994), que consiste em determinar a estrutura temática a partir de “categorias temáticas”, que ofereçam respostas às perguntas: Quem (ser), O quê (tema), Como (modo), Onde (lugar), e Quando (tempo). Segundo a autora, a estrutura temática de um texto “consiste na distinção do objetivo principal do texto, daquilo que é acessório ou secundário” (TÁLAMO, 1994, p.24).

Outra importante metodologia foi apresentada por Fujita e Rubi (2006), que desenvolveram um Modelo de Leitura Documentária (MLD) que teve como ponto de partida o aperfeiçoamento da formação de indexadores. As autoras tiveram como objetivo criar um instrumento capaz de orientar e sistematizar a leitura técnica de textos pelos indexadores, profissionais ou em formação. Assim, o MLD é uma metodologia para a análise de assunto nos documentos, de forma a facilitar a identificação e seleção de conceitos, sobretudo em textos científicos. É preciso salientar, que esse modelo foi elaborado para a análise de artigos científicos, especificamente, da área biológica, e

consiste, fundamentalmente, da combinação das sistemáticas de identificação de conceitos análise conceitual (primeira coluna) e abordagem sistemática da Norma 12.676 (segunda coluna) com a localização dos conceitos em parte da estrutura textual (terceira coluna) (FUJITA; RUBI, 2006, *on-line*).

Com esse instrumento, as autoras oferecem uma forma de operacionalizar o processo de indexação, conectando a representação do documento com o sistema de recuperação de informação, o que é ilustrado através do QUADRO 3.

QUADRO 3 – Modelo de leitura Documentária de Fujita e Rubi (2006)

CONCEITO (ANÁLISE CONCEITUAL)	QUESTIONAMENTO (NORMA 12.676)	PARTE DA ESTRUTURA TEXTUAL
OBJETO	O documento possui em seu contexto um objeto sob efeito de uma atividade?	INTRODUÇÃO (OBJETIVOS)
AÇÃO	O assunto contém um conceito ativo (por exemplo, uma ação, uma operação, um processo etc.)?	INTRODUÇÃO (OBJETIVOS)
AGENTE	O documento possui um agente que praticou esta ação?	INTRODUÇÃO (OBJETIVOS)
MÉTODOS DO AGENTE	Este agente refere-se a modos específicos para realizar a ação (por exemplo, instrumentos especiais, técnicas ou métodos)?	METODOLOGIA
LOCAL OU AMBIÊNCIA	Todos estes fatores são considerados no contexto de um lugar específico ou ambiente?	METODOLOGIA
CAUSA E EFEITO	São identificadas algumas variáveis dependentes ou independentes?	RESULTADOS; DISCUSSÃO DE RESULTADOS
PONTO DE VISTA DO AUTOR; PERSPECTIVA	O assunto foi considerado de um ponto de vista, normalmente não associado com o campo de estudo (por exemplo, um estudo sociológico ou religioso)?	CONCLUSÕES

Fonte: Fujita e Rubi (2006, *on-line*). Disponível em: <[http://www.dgz.org.br/jun06/Art\\_04.htm](http://www.dgz.org.br/jun06/Art_04.htm)>.

Com base em teorias da área da biblioteconomia, esse modelo de leitura foi desenvolvido para a indexação de artigos científicos da área de biologia, “apoiado na combinação da exploração da estrutura textual para identificação de conceitos por meio de questionamento” (FUJITA; RUBI, 2006, *on-line*). Sua metodologia é apresentada por Fujita e Rubi (2006, *on-line*), em dois aspectos (combinação da estrutura textual com a identificação de conceitos e sistemática de identificação de conceitos) e segue os seguintes passos:

1) Identificação da estrutura textual de conceitos na literatura: conhecer a estrutura textual do documento auxilia na identificação dos conceitos na atividade de indexação. A título de argumentação,

no que diz respeito à estrutura do texto, afirma-se estar associada ao modo com o qual as idéias são organizadas com relação ao conteúdo, ao tema e aos conceitos tratados no texto. Como a estrutura do texto se articula ao seu conteúdo, o autor de um texto escolhe determinada estrutura textual que venha coincidir com o conteúdo que quer transmitir (FUJITA; RUBI 2006, *on-line*).

Assim, constata-se que as autoras partiram da concepção de estrutura e superestrutura, relacionando com a metodologia de indexação (KOBASHI, 1994) e a metodologia de Tálamo (1994), já citada anteriormente.

2) Identificação dos conceitos na leitura documentária (a abordagem sistemática e a análise conceitual): a abordagem sistemática consiste em usar as regras da Norma NBR-12.676, que sugerem questionamentos ao texto como forma de extrair conceitos dele, tais como:

#### QUADRO 4 – Sugestões de perguntas da Norma NBR-12.676

---

##### PERGUNTAS

---

- a) O documento possui em seu contexto um objeto sob efeito de uma atividade?
  - b) O assunto contém um conceito ativo (por exemplo, uma ação, uma operação, um processo, etc.)?
  - c) O objeto é influenciado pela atividade identificada?
  - d) O documento possui um agente que praticou esta ação?
  - e) Este agente refere-se a modos específicos para realizar a ação (por exemplo, instrumentos especiais, técnicas ou métodos)?
  - f) Todos estes fatores são considerados no contexto de um lugar específico ou ambiente?
  - g) São identificadas algumas variáveis dependentes ou independentes?
  - h) O assunto foi considerado de um ponto de vista, normalmente não associado com o campo de estudo (por exemplo, um estudo sociológico ou religioso)?
- 

Fonte: Norma NBR-12.676 (1992, p.2).

Em vista dessas colocações, é possível o entendimento de que, para cada item da primeira coluna, referente ao modelo apresentado no QUADRO 3, haverá um questionamento (QUADRO 4 e segunda coluna do QUADRO 3) e uma resposta apropriada.

No que diz respeito à análise conceitual, Fujita e Rubi (2006) trabalharam com as recomendações para a análise conceitual utilizados pelos princípios de indexação do sistema PRECIS (FUJITA, 2003). Segundo essa autora, esse sistema foi idealizado por Derek Austin, em 1964 (SILVA; FUJITA, 2004), para automatizar a indexação de documentos, cuja atividade visa preservar o contexto do serviço de informação. Além disso, alia o contexto de produção e de uso da informação, em uma abordagem sintática e

semântica, para a elaboração de índices de assunto. Esse sistema permite compatibilizar a linguagem natural dos textos e a linguagem natural do usuário com um vocabulário controlado que pode ser, e é preferível que seja assim, adaptado para o contexto, isto é, ele não é pré-determinado ou padronizado. No intuito de atender a esse critério, o sistema PRECIS indica o questionamento ao texto, buscando, justamente, identificar os conceitos, pois, para cada pergunta haverá um conceito em resposta:

- o que aconteceu? (ação)
- a que ou a quem isto aconteceu? (objeto da ação - sistema chave)
- o que ou quem fez isto? (agente da ação)
- onde aconteceu? (local)

A partir disso, os conceitos são identificados a partir da estrutura profunda do texto, ou seja, com a determinação dos conceitos nos quais ocorrem os significados mais generalizáveis. Em outras palavras, pode-se afirmar que esses conceitos mais gerais são as categorias (primeira coluna), que representam a estrutura textual do documento.

Ao término desses procedimentos, é possível indexar os documentos com completude e clareza, facilitando a tarefa do indexador. Entretanto, cabe salientar que diferentes tipos de textos podem possuir distintas estruturas de produção, o que impõe a adaptação do modelo proposto por Fujita e Rubi (2006).

Diante do exposto, verifica-se a necessidade de se identificar a estrutura dos documentos do *corpus* da presente dissertação, que são trabalhos acadêmicos do tipo teses e dissertações, de forma a estabelecer a adequação ou não do modelo de Fujita e Rubi. Dessa forma, esse é o tópico que se apresenta a seguir.

#### **4.9 Documentos acadêmicos: teses e dissertações**

Um “documento”, desde a concepção de Otlet, pode ser praticamente qualquer objeto de informação.

Para Paul Otlet, documento é o livro, a revista, o jornal; é a peça de arquivo, a estampa, a fotografia, a medalha, a música; é também, o filme, o disco e toda a parte documental que precede e sucede a emissão radiofônica. Ao lado dos textos e das imagens, existem objetos documentais por si mesmos (réalia). São as amostras, os espécimes, modelos, fac-símiles e, de maneira geral, tudo que tenha caráter representativo em três dimensões e, eventualmente, em movimento (CUNHA; CAVALCANTI, 2008, p.132).

Para Rabello (2009, p.12), o documento

faz parte do quadro conceitual da CI. Ele apresenta a materialidade e a susceptibilidade [sic] de inscrição ou registro informacional como características principais e apresenta, igualmente, uma estrutura e uma disposição que facilita e/ou propicia a identificação e a apreensão das informações nele registradas.

Esses dois conceitos de documento traduzem o que está expresso na norma NBR-12676:1992, que trata das regras para a prática normalizada da análise de documentos. A referida norma determina que documento é “qualquer unidade, impressa ou não, passível de catalogação ou indexação” e, ainda, acrescenta que:

Esta definição se refere não apenas a materiais escritos ou impressos em papel ou suas versões em microforma (p. ex.: livros, jornais, diagramas, mapas), mas também a suportes não-impressos (p. ex.: registros legíveis por máquinas, filmes, gravações sonoras), objetos tridimensionais e *realia* usadas como espécimens (NBR-12676, 1992, p.2).

Nesta dissertação, é adotada a concepção de “documento” como qualquer unidade informacional, impressa ou não, passível de catalogação ou indexação, seja qual for a sua natureza ou suporte. Além disso, a unidade informacional (informação) é considerada tangível, pois, segundo Buckland (1991), pode ser tratada e disseminada em SRIs, por meio de três usos: 1) informação como processo, com a transmissão e comunicação de uma informação; 2) informação enquanto conhecimento, capaz de reduzir incertezas através do uso da informação pelo indivíduo; 3) informação na perspectiva de objeto, que pode ser atribuído a qualquer unidade informacional, a qualquer “coisa”.

Partindo-se desses preceitos de “documento”, pode-se afirmar que os documentos acadêmicos constituem o conjunto de produções que formam a literatura científica. Entretanto, o escopo deste trabalho trata, exclusivamente, dos documentos acadêmicos do tipo teses e dissertações. Sendo assim, convém explicar que uma tese é um documento acadêmico que representa o resultado de pesquisa com tema único e escopo bem delimitado, é responsável por contribuição real para a questão levantada, com alto grau de profundidade de estudo. Nesses termos, a tese deve ser elaborada sob a supervisão de um orientador que possua a titulação de doutor. Com relação a isso, o Conselho Federal de Educação (Parecer 977/65) afirma que

a elaboração de uma tese constitui exigência para obtenção do grau de ‘Doutor’ [...] deverá ser elaborada com base em investigação original devendo representar trabalho de real contribuição para o tema escolhido.

A propósito da NBR 14724:2005, a tese é definida como

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 o especialista em questão. É feito sob a coordenação de um orientador (doutor) e visa à obtenção do título de doutor, ou similar.

Com relação à dissertação, ela configura um documento acadêmico de tema único, com um escopo bem delimitado. Seu objetivo é estudar e reunir subsídios para analisar e interpretar informações coletadas. Desse modo, deverá testar a capacidade e habilidade de sistematização do aluno, sob a supervisão de um orientador que também deve possuir a titulação de doutor. Sobre a dissertação, o Conselho Federal de Educação (Parecer 977/65) salienta que:

o preparo de uma dissertação será exigido para obtenção do grau de 'Mestre' [...] deverá evidenciar conhecimento da literatura existente e a capacidade de investigação do candidato, podendo ser baseada em trabalho experimental, projeto especial ou contribuição técnica [...].

A NBR 14724:2005, por sua vez, caracteriza a dissertação como um

documento que representa o resultado de um trabalho experimental ou exposição de um estudo científico retrospectivo, de tema único e bem delimitado em sua extensão, com 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 a coordenação de um orientador (doutor), visando à obtenção do título de mestre.

Nesse cenário, a pesquisa acadêmica, inscrita nas teses e dissertações, demanda conhecimento científico e metodológico, trazendo contribuições do autor ao tema escolhido, ainda que seja um pequeno aporte de conhecimento. Além disso, deve respeitar regras e normas técnicas para a normalização dos textos, de modo a possibilitar a comunicação e o compartilhamento da informação.

No Brasil, a Associação Brasileira de Normas Técnicas (ABNT<sup>67</sup>), é o órgão responsável pela normalização técnica e representa o país em entidades internacionais tais como a *International Organization for Standardization* (ISO) e *International Electrotechnical Commission* (IEC). Assim, a ABNT possui particular importância no âmbito de normalização de documentação, sobretudo na apresentação de documentos acadêmicos. Isso é confirmado pelo fato de que as diretrizes da Norma ABNT-NBR 14724:2011 são as que prevalecem, desde 17 de abril de 2011, para padronizar a elaboração e apresentação de documentos acadêmicos, dentre eles as teses e dissertações. Não obstante, outras normas complementam-na, quais sejam: 6023:2002 (Informação e documentação – Referência – Elaboração), NBR 6024:2003 (Numeração progressiva das seções de um documento escrito), NBR 6027:2003 (Sumário), NBR 6034:2004 (Índice), NBR 10520:2002 (Citações), NBR 12225:2004 (Títulos de lombada), NBR 6028:2003 (Resumo).

A respeito desta última norma, sobre os resumos, é preciso esclarecer que é parte constitutiva das teses e dissertações e, como tal, deve ser composto por uma descrição de todo o conteúdo textual do documento. Mesmo porque, a norma NBR-6028 (2003), recomenda que os trabalhos acadêmicos precisam, obrigatoriamente, apresentar um resumo informativo, como elemento pré-textual, com as seguintes características:

Resumo informativo: informa ao leitor finalidades, metodologia, resultados e conclusões do documento, de tal forma que este possa, inclusive, dispensar a consulta ao original. [...]

O resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. [...]

---

<sup>67</sup> Disponível em: <<http://www.abnt.org.br/>>. Acesso em: 2 out. 2010.

A primeira frase deve ser significativa, explicando o tema principal do documento. A seguir, deve-se indicar a informação sobre a categoria do tratamento (memória, estudo de caso, análise da situação etc.).

Quanto a sua extensão os resumos devem ter: a) de 150 a 500 palavras os de trabalhos acadêmicos (teses, dissertações e outros) (NBR-6028, 2003, p.1-2).

Nota-se a completude exigida para os resumos informativos, o que estabelece a importância desse componente que revelou-se a partir do século XIX, quando a comunicação científica passou a ser essencialmente obtida a partir de resumos, “os quais podiam substituir os artigos (resumos informativos) ou ser guias para eles (indicativos)” (MENDES, 2010, p.136). Infelizmente, problemas identificados na elaboração de resumos ainda são recorrentes nos trabalhos acadêmicos.

Retornando à estrutura dos documentos do tipo teses e dissertações, pode-se afirmar que há três partes fundamentais:

Parte textual: o texto é composto por uma parte introdutória, que apresenta os objetivos do trabalho e as razões de sua elaboração; o desenvolvimento, que detalha a pesquisa ou estudo realizado; e uma parte conclusiva (ABNT NBR 14724:2011, p.8).

Na parte de desenvolvimento da pesquisa, apresentar a investigação de forma detalhada pressupõe apresentar quatro elementos principais:

Elemento 1 – Revisão de literatura: contém uma explanação sobre os estudos realizados por outros autores acerca do tema da pesquisa. Restringe-se, geralmente, às contribuições mais contundentes, situando a evolução do assunto.

Elemento 2 – Metodologia: descreve, detalhadamente, os métodos, técnicas e processos seguidos na pesquisa, explicando as hipóteses ou pressupostos, população ou amostra, os instrumentos e a coleta de dados. Em especial, os métodos inéditos desenvolvidos pelo autor devem ser justificados, e devem ser apontadas suas vantagens em relação a outros autores.

Elemento 3 – Resultados: apresenta os dados obtidos durante o desenvolvimento da pesquisa, os quais podem ser oferecidos por meio de tabelas, quadros e figuras. Os diversos resultados alcançados devem ser agrupados e ordenados, sendo que, para maior clareza, podem ser acompanhados por gráficos com valores estatísticos.

Elemento 4 – Discussão: pode expor uma comparação dos resultados alcançados na pesquisa com aqueles estudos e já mencionados na revisão de literatura, ressaltando os aspectos que confirmem ou modifiquem, de modo significativo, as teorias estabelecidas.

Diante dessas verificações, percebe-se que o documento textual das teses e dissertações pode ser considerado padronizado em relação à sua estruturação. Então, de modo sintético, é possível a seguinte enumeração acerca da sua constituição: 1) seção inicial – com a introdução, as justificativas e os objetivos; 2) desenvolvimento – geralmente,

com a revisão da literatura, a metodologia, uma discussão (argumentação teórica, conceitual ou analítica), descrição dos dados, apresentação dos resultados; e 3) última seção – parte em que são apresentadas as conclusões.

Quanto ao tipo de conteúdo, as teses e dissertações têm caráter didático, apresentando metodologia e resultados oriundos da análise de dados de pesquisa científica (MARCONI; LAKATOS, 1990). Para tanto, utilizam linguagem formal e específica para a área ou domínio, através de uma terminologia própria. Entretanto, o documento textual das teses e dissertações é produzido em linguagem natural, que permite o uso de diferentes termos para um mesmo conceito (sinonímia<sup>68</sup>). Esse fato acaba por ocasionar uma dificuldade para a comunicação e compartilhamento do conhecimento registrado nesses documentos.

É preciso salientar, finalmente, que a estrutura textual das teses e dissertações foi utilizada para a criação da matriz categorial referente aos trabalhos acadêmicos, assim como para a criação da taxonomia facetada navegacional.

Agora, sugere-se o momento de descrever e analisar os resultados encontrados. Para isso, o próximo capítulo se apresenta.

---

<sup>68</sup> Palavra ou expressão que tem o mesmo significado ou conceito de outra.

## 5 DESCRIÇÃO DOS PROCEDIMENTOS E ANÁLISE DOS RESULTADOS

Neste capítulo, são apresentadas as etapas de desenvolvimento da taxonomia facetada navegacional, bem como as etapas da aplicação dos métodos e técnicas, as decisões tomadas durante as diferentes fases desta pesquisa, e a apresentação dos resultados. A taxonomia facetada navegacional, denominada de TAFNAVEGA, consiste em uma ferramenta que tem por objetivo auxiliar o usuário na busca por informações de sua necessidade, no banco de dados da BDTD do Programa de Pós-Graduação em Ciência da Informação da UFMG. Para tal, propõe estruturar esta taxonomia com base nas características específicas de seus usuários, evidenciando, no momento da busca, elementos tais como os temas, métodos, teorias e os resultados das pesquisas. Dessa forma, o usuário poderá acessar com maior facilidade não somente o título, autor e ano do documento, mas também essas outras informações que complementam e caracterizam a pesquisa acadêmica.

Na elaboração da TAFNAVEGA, primeiramente, considerou-se a delimitação temporal e o estudo do material empírico e, depois, fez-se um estudo dos documentos selecionados referentes ao *corpus*, com sua codificação. Posteriormente, houve a identificação do domínio, do usuário e de suas necessidades de informação, utilizando a teoria da análise de domínio. Na etapa seguinte, passou-se a determinar as categorias fundamentais temáticas da TAFNAVEGA, através do método da análise de conteúdo, com a técnica da análise categorial temática. Em síntese, a TAFNAVEGA é composta por dez categorias fundamentais temáticas e pelo algoritmo desenvolvido para o processo de extração de conceitos. Estes procedimentos serão detalhados a seguir.

### 5.1 Recorte temporal e levantamento do material empírico

A delimitação temporal desta pesquisa é o período compreendido entre 1998 e 2009. A escolha da data inicial teve estreita relação com a história do PPGCI da UFMG, pois esse programa foi criado em 1976, oferecendo mestrado em Administração de Bibliotecas, direcionado para a qualificação de docentes oriundos dos cursos de Biblioteconomia. Em 1990, houve uma reformulação do programa. Porém, somente em 1997, juntamente com a implantação do doutorado, o programa passou a se denominar Programa de Pós-Graduação em Ciência da Informação (PPGCI). Por esse motivo, foram computados apenas os trabalhos defendidos a partir de 1998, quando o Programa passou a acolher estudantes de mestrado e de doutorado, tendo seu foco direcionado a estudos na área da Ciência da Informação. Ademais, essa é a data mais antiga em que as teses e dissertações passaram a ser inseridas no banco de dados da BDTD da ECI. Em relação à data final, 2009, sua

escolha justificou-se pela necessidade de se estabelecer um limite para o *corpus*, uma vez que, a cada dia, novos trabalhos são defendidos, e novos documentos são inseridos na BDTD da ECI.

Em seguida, na etapa do levantamento do material empírico, houve consulta aos registros documentais da secretaria do PPGCI, ao catálogo disponibilizado pelo Sistema Pergamum, da Biblioteca Professora Etelvina Lima, da ECI e da Biblioteca Universitária (BU) da UFMG, além de busca física, nas prateleiras dessas duas bibliotecas.

Nesse arrolamento, contou-se com o apoio e a colaboração da secretária do PPGCI-UFMG, dando acesso aos registros dos trabalhos acadêmicos e linhas de pesquisa, e com a cooperação da chefia da biblioteca da ECI-UFMG para o cruzamento e validação dos dados levantados na secretaria do PPGCI-UFMG.

Entretanto, algumas dificuldades foram encontradas, como, por exemplo, trabalhos acadêmicos sem a identificação da linha de pesquisa à qual pertenciam; itens não registrados no catálogo da biblioteca; itens cuja catalogação apresentava pequenos erros, tais como grafia incorreta do nome dos autores; itens dispostos em locais errados nas prateleiras da biblioteca, o que demandou mais tempo de busca; e alguns *links* quebrados na BDTD, ocorrências que delongaram a coleta de dados.

Como resultado desse levantamento foram identificados 290 documentos, sendo 62 teses e 228 dissertações, referentes às linhas de pesquisa Gestão da Informação e do Conhecimento (GIC), Informação, Cultura e Sociedade (ICS) e Organização e Uso da Informação (OUI). Desse total, 62 trabalhos referiam-se à linha de pesquisa OUI.

Em seguida, com autorização da coordenação do PPGCI-UFMG, foi solicitado ao Sistema OPUS, que é o sistema oficial para disponibilização da produção intelectual da UFMG, a liberação dos registros do banco de dados referentes ao acervo da ECI-UFMG, para que fosse possível fazer o cruzamento de dados entre a lista dos 62 documentos e a lista do que efetivamente estava disponível no banco de dados.

A partir da análise dos registros dos dados, liberados em 7 de julho de 2010, constatou-se que, do total dos 290 documentos (universo), somente 146 documentos estavam disponibilizados no banco de dados, isto é, pouco mais de 50%.

Dentre os 146 documentos disponíveis, identificaram-se 41 documentos relativos à linha de pesquisa OUI, o que equivale a pouco mais de 66% do total. Esses 41 documentos representam o *corpus* desta pesquisa e são apresentados no QUADRO 5.

QUADRO 5 – Relação de documentos do *corpus* (por ano)\*

ANO	DOCUMENTOS		
	T	D	Total
1998	0	0	0
1999	0	0	0
2000	0	0	0
2001	0	0	0
2002	0	1	1
2003	0	1	1
2004	2	1	3
2005	4	7	11
2006	0	6	6
2007	1	5	6
2008	3	5	8
2009	2	3	5
<b>TOTAL</b>	<b>12</b>	<b>29</b>	<b>41</b>

\* Disponíveis no banco de dados da BDTD/ECl em julho/2010.  
Fonte: Elaborado pela autora.

Com a listagem do *corpus* pronta, passou-se à sua codificação.

### 5.1.1 Codificação do *corpus* da pesquisa

Para cada um dos documentos do *corpus* estabeleceu-se uma codificação, cuja listagem está disposta no APÊNDICE A. Os documentos foram identificados pela tipologia, sendo utilizada a letra “t” para as teses e a letra “d” para as dissertações. Junto às letras, foram atribuídos algarismos arábicos, de forma sequencial, iniciando em 001, até o último item, acompanhados do ano de defesa do documento, sendo apresentadas em ordem decrescente de defesa. Como exemplo, tem-se t001-2009, t002-2009 e d001-2009, d002-2009, e assim, sucessivamente.

Após a determinação do recorte temporal e do levantamento do material empírico (*corpus*), passou-se à identificação do domínio, do usuário e de suas necessidades, uma vez que antes de modelar um serviço de informação é necessário fazer essa identificação.

## 5.2 Identificação do domínio, do usuário e de suas necessidades

Nesta etapa houve a análise do *corpus* da pesquisa com a finalidade de identificar a qual domínio ele pertence, qual sua função dentro desse ambiente e as razões subjacentes para os fluxos de informação nesse campo (HJORLAND, 2002a). Além disso, identificou-se o usuário da BDTD, em relação ao tipo de informação contida no *corpus*, assim como de suas necessidades de informação. Para isso, foi utilizado o método da teoria da análise de domínio (TAD), com procedimentos realizados a partir das quatro etapas que se encontram descritas na fundamentação, no Capítulo Quatro, itens 4.5 e 4.5.1.

a) Identificação do domínio: a partir das indagações propostas para esta etapa da TAD, determinou-se que o domínio é o programa de pós-graduação *stricto sensu* da ECI-UFMG, limitado à linha de pesquisa OUI, cujos trabalhos acadêmicos do tipo teses e dissertações têm como característica comum apresentar resultados de pesquisa com tema único e escopo definido, analisando-o de forma sistemática e científica, por meio de fundamentos teórico-metodológicos.

b) Análise de dados: a caracterização do domínio foi obtida por meio dos questionamentos propostos para esta etapa, que indicou que o produtor do tipo de informação contido nas teses e dissertações é sempre um pesquisador de carreira ou de circunstância. Determinou, ainda, que todo pesquisador é um cientista e, como tal, é membro da comunidade científica. A finalidade dessa produção é ser aprovado perante uma banca de doutores, também da área acadêmica, buscando a aprovação de seus pares, dentro de uma mesma área de conhecimento ou entre áreas interdisciplinares. Em sentido *stricto sensu*, todo pesquisador tem como objetivo de ação empregar uma sistemática metodológica, apresentando seu tema, o problema de pesquisa, a proposta de solução, a fundamentação teórico-metodológica e os resultados alcançados. Essa sistemática faz parte da atividade profissional do pesquisador, e é exigida a sua observância para que a produção seja considerada e aceita como conhecimento científico. Para avançar em seu conhecimento, um pesquisador precisa saber o que já foi estudado sobre seu tema de pesquisa e sob quais condições, teóricas e metodológicas, esse tema foi abordado.

c) Modelagem do domínio: acolhendo ao recomendado para esta etapa e em atendimento ao objetivo proposto, considera-se que, em atividade profissional, o pesquisador necessita de informações sobre ideias, abordagens, teorias, métodos e técnicas que possam ser empregadas em sua própria produção, de modo que possa avançar a partir de resultados já atingidos. Seguindo a lógica indicada no método da TAD, tem-se: a) quem produz: a comunidade científica; b) quem comunica: as instituições (quando disponibilizam digitalmente); c) quem usa: a própria comunidade acadêmica. A BDTD é disponibilizada pela instituição (ECI-UFMG) e o serviço de informação atenderá prioritariamente aos pesquisadores da OUI.

d) Coleta e seleção de dados e validação: atendendo ao prescrito nesta etapa, foi realizada uma pesquisa especializada para encontrar fontes de informação que dessem respaldo à execução da etapa anterior (modelagem de domínio). Dessa maneira, para identificar e fundamentar a escolha do tipo de informação que seria disponibilizado ao usuário, foram utilizadas as fontes: 1) o estudo de Targino (2000); 2) as ideias de Kuhn (2006); 3) o dicionário da área de Biblioteconomia e Arquivologia de Cunha e Cavalcanti

(2008); 4) Alvarado e Oliveira (2008)<sup>69</sup>; 5) os três livros utilizados na etapa de desenvolvimento da TAFNAVEGA (descrita no item 5.3 deste capítulo), que são: Severino (2007), Marconi e Lakatos (2010) e Silva (2010). Nessas fontes, foram identificadas as seguintes necessidades de informação do pesquisador, em sua atividade profissional: a) temas que já foram investigados, com o intuito de entender o que já foi feito e para avançar em seu conhecimento; b) métodos e teorias utilizados nas pesquisas, com a finalidade de ampliar o conhecimento e aprimorar teorias e métodos científicos existentes; c) resultados alcançados nas pesquisas, para entender, ampliar e aplicar o conhecimento científico já conhecido, sem incorrer em retrabalhos. Além disso, para dar respaldo à identificação do produtor (autor) dos trabalhos acadêmicos do tipo teses e dissertações e dos responsáveis pela comunicação do conhecimento acumulado nesses documentos foram utilizadas as seguintes fontes: 1) a legislação brasileira<sup>70</sup> sobre programas de pós-graduação *stricto sensu*, incluindo seus decretos; 2) as resoluções e portarias do Ministério da Educação<sup>71</sup> (MEC); 3) as normas e resoluções da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior<sup>72</sup> (CAPES); 4) as resoluções internas do PPGCI-UFMG<sup>73</sup>. Todas essas quatro fontes estabelecem diretrizes para os programas de mestrado e doutorado, identificando quem pode ser produtor de teses e dissertações e sob quais condições, além de determinar as instituições que têm permissão para oferecer tais programas *stricto sensu*.

Tendo sido identificados o domínio, o usuário e suas necessidades de informação, passa-se ao desenvolvimento da taxonomia facetada navegacional (TAFNAVEGA), iniciando pela determinação das categorias fundamentais temáticas e depois passando à apresentação do algoritmo utilizado para a extração de conceitos dos documentos do *corpus* desta pesquisa.

### 5.3 Criação da taxonomia facetada navegacional

Para a criação de uma taxonomia facetada navegacional (TAFNAVEGA), visando uma representação sistematizada e semântica do conteúdo de teses e dissertações capaz de facilitar a busca e a recuperação das informações, primeiramente concebeu-se a modelagem das informações tendo em mente o usuário e suas necessidades de informação, obtidos a partir da etapa anterior, no subcapítulo 5.2, o que pode ser sintetizado conforme os dados a seguir:

<sup>69</sup> Estudo que discorre sobre “A Comunidade científica da Biblioteconomia e Ciência da Informação brasileira” apresentando suas características e necessidades de informação. Disponível em: <<http://revista.ibict.br/pbcib/index.php/pbcib/article/view/1051>>.

<sup>70</sup> Disponível em: <<http://www4.planalto.gov.br/legislacao>>.

<sup>71</sup> Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_content&view=article&id=12899:pos-graduacao-normativos&catid=323:orgaos-vinculados](http://portal.mec.gov.br/index.php?option=com_content&view=article&id=12899:pos-graduacao-normativos&catid=323:orgaos-vinculados)>.

<sup>72</sup> Disponível em: <<http://www.capes.gov.br/sobre-a-capes/legislacao/2341-resolucoes>>.

<sup>73</sup> Disponível em: <[http://ppgci.eci.ufmg.br/?O\\_Programa:Resolu%E7%F5es](http://ppgci.eci.ufmg.br/?O_Programa:Resolu%E7%F5es)>.

**Usuário:** pesquisadores

**Necessidades de informação:** temas, métodos, teorias e resultados das pesquisas.

Em seguida, no subcapítulo **5.3.1**, apresenta-se o desenvolvimento das categorias fundamentais temáticas da TAFNAVEGA e, posteriormente, no subcapítulo **5.3.2**, descreve-se e aponta-se o algoritmo para o processo de extração de conceitos dos documentos.

### **5.3.1 Desenvolvimento das categorias fundamentais temáticas**

Nesta etapa de desenvolvimento das categorias fundamentais temáticas para a TAFNAVEGA foi utilizado o método da análise de conteúdo, a partir da técnica da análise categorial temática (ACT), com uma abordagem orientada à demanda (ALBRECHTSEN, 1993).

A técnica da análise categorial temática (ACT) está descrita no Capítulo Quatro, no subcapítulo **4.6**, cujos procedimentos foram aplicados conforme descrito a seguir.

Tendo em vista a equivalência das metodologias dos autores Moraes (1999), Oliveira (2008) e Bardin (2009) para a aplicação da ACT, que foram apresentadas na fundamentação (subcapítulo **4.6**), utilizou-se a proposta de Moraes, porque ela possui etapas mais subdivididas, facilitando seu emprego. Dessa maneira, na metodologia proposta por Moraes, primeiramente identifica-se o objetivo pretendido com a análise e, em seguida, aplicam-se os procedimentos em cinco etapas distintas: 1) preparação das informações; 2) unitarização ou transformação do conteúdo em unidades; 3) categorização ou classificação das unidades em categorias; 4) descrição; 5) interpretação.

A seguir, apresenta-se a descrição do objetivo pretendido:

**Objetivo pretendido:** verificar qual conjunto de conceitos é necessário para representar a estrutura textual das teses e dissertações, de forma que possa ser utilizado como categorias fundamentais temáticas na TAFNAVEGA criada no escopo desta pesquisa.

Partindo disso, apresentam-se as cinco etapas de Moraes (1999), que possibilitou aplicar a metodologia da ACT.

#### **Etapa 1: preparação das informações**

O *corpus* de análise é o conjunto com as 41 teses e dissertações, apresentado no **Capítulo 2**, que foram formalmente analisados. A representatividade da amostra foi garantida porque houve a análise de todos os documentos disponíveis na base de dados da BDTD da ECI-UFMG, dentro do recorte temporal e da limitação do objeto empírico desta

pesquisa. Partiu-se da validade da estrutura textual para a representação do conteúdo dos documentos do *corpus* da pesquisa.

## **Etapa 2: unitarização ou transformação do conteúdo em unidades**

A codificação dos documentos do *corpus* da pesquisa está apresentada no subcapítulo **5.1.1**. Esclarece-se que o procedimento de codificação é contínuo, pois sempre que houve novas unidades para análise, as mesmas foram codificadas.

A transformação do conteúdo em unidades de conteúdo (UC) e em unidades de registro (UR) foi obtida a partir da estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações, que se baseou na fundamentação teórica do subcapítulo **4.8** desta pesquisa, na norma ABNT-14724:1992 e na consulta à literatura especializada sobre metodologias de pesquisa, que se baseou nas seguintes fontes: Severino (2007), Marconi e Lakatos (2010) e Silva (2010).

A escolha dessas três obras seguiu os seguintes critérios: uma obra sobre desenvolvimento de trabalhos acadêmicos para professores e discentes, uma obra tendo como autor um profissional respeitado e versado na área de metodologia de pesquisa, e uma obra sobre metodologia de pesquisa com foco na área das ciências humanas aplicadas. A qualificação dos autores e das obras está descrita no APÊNDICE B e a compilação da análise está apresentada no QUADRO 6.

QUADRO 6 – Consolidação da estrutura textual de documentos acadêmicos

<b>PARTE TEXTUAL (unidade de contexto)</b>	<b>SUBPARTES (unidade de registro)</b>
Introdução	Tema Tempo e espaço Problema Objetivos Justificativa Metodologia Referencial teórico
Desenvolvimento	Revisão literatura Hipóteses e variáveis Fundamentação teórica Métodos e técnicas Análise e interpretação
Conclusão	Conclusão (em relação às hipóteses e objetivos)

Fonte: Baseado na ABNT-14724:1992, Severino (2007), Marconi e Lakatos (2010) e Silva (2010).

A partir dessa consolidação, identificaram-se as unidades de contexto (UC) e as unidades de registro (UR), conforme descrito e codificado a seguir.

**Unidades de Contexto e codificação:** para esta codificação foram utilizadas as iniciais “UC”, que significa “unidade de contexto”, seguida da numeração sequencial, iniciada em “um”. Dessa forma, temos a seguinte codificação:

UC1: Introdução, UC2: Desenvolvimento e UC3: Conclusão.

Como exemplo de leitura, tem-se o código UC1, que significa “unidade de contexto referente à introdução”.

**Unidades de Registro e codificação:** para esta codificação foram adicionadas às “unidades de contexto (UC)”, as iniciais de “UR”, para representar as “unidades de registro”, seguidas das letras do alfabeto, em ordem sequencial, iniciada com a letra “a”. Dessa forma, tem-se a seguinte codificação:

UC1.ura: Tema;	UC2.urh: Revisão literatura;
UC1.urb: Tempo e espaço;	UC2.uri: Hipóteses e variáveis;
UC1.urc: Problema;	UC2.urj. : Fundamentação teórica;
UC1.urd: Objetivos;	UC2.urk: Métodos e técnicas;
UC1.ure: Justificativa;	UC2.url: Análise e interpretação;
UC1.urf: Metodologia;	UC3.urm: Conclusão.
UC1.urg: Referência teórica;	

Como exemplo de leitura, tem-se o código UC1.ura, que significa “unidade de contexto referente à introdução, com a unidade de registro representada pelo tema”.

### **Etapa 3: categorização ou classificação das unidades em categorias**

Nesta etapa, adotou-se o procedimento indicado pela técnica da análise categorial temática (ACT), que consiste em determinar o conjunto de categorias a partir da análise dos documentos do *corpus*, das unidades de registro e das fontes indicadas na **etapa 2** anterior. Seguindo o sugerido pela ACT para esta etapa, a determinação das categorias foi realizada *a priori*, em uma análise com abordagem semântica e dedutiva, resultando em um conjunto de categorias que é denominada pela ACT de “categorias iniciais”. Esse conjunto de “categorias iniciais” é apresentado no QUADRO 7.

QUADRO 7 – Etapa da categorização (Análise de Conteúdo)

UNIDADE DE REGISTRO	CATEGORIAS INICIAIS
Tema Tempo e espaço Problema Objetivos Justificativa Metodologia Referência teórica	TEMA DIMENSÃO  AÇÃO
Revisão literatura Hipóteses e variáveis Fundamentação teórica Métodos e técnicas Análise e interpretação	CAUSA E EFEITO FUNDAMENTO TEÓRICO MÉTODOS / COLETA DE DADOS PONTO DE VISTA DO AUTOR
Conclusão (em relação às hipóteses e objetivos)	PRINCIPAIS RESULTADOS ALCANÇADOS

Fonte: Baseado na ABNT-14724:1992, Severino (2007), Marconi e Lakatos (2010) e Silva (2010), nas unidades de registro e na análise nos documentos do *corpus*.

Para codificação do conjunto das categorias iniciais utilizou-se o seguinte esquema: a letra “C” para identificar todos os elementos como “categoria” e, em seguida, foi inserida uma numeração sequencial, iniciada em “um”. Dessa forma, tem-se a seguinte codificação:

**Conjunto de categorias iniciais:** C1.Tema; C2.Dimensão; C3.Ação; C4.Causa e efeito; C5.Fundamento teórico; C6.Métodos; C7.Coleta de dados; C8.Ponto de vista do autor/perspectiva; C9.Principais resultados alcançados.

Ainda nesta etapa (da categorização), passou-se para o procedimento de validação do conjunto de “categorias iniciais”. Essa validação corresponde, segundo os critérios da ACT, que as categorias devem atender às seguintes características: válidas, exaustivas e homogêneas; mutuamente exclusivas; consistentes. Essa validação é demonstrada no QUADRO 8.

QUADRO 8 - Validação das "categorias iniciais"

CRITÉRIOS	VALIDAÇÃO
Válidas	O conjunto de categorias é significativo em relação à estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações e adequado ao objetivo proposto de atender às necessidades de informação da comunidade científica.
Exaustivas	Cobre todo o conteúdo significativo da estrutura textual dos trabalhos acadêmicos. Ao completar os campos, nenhuma parte textual deixará de ser contemplada.
Homogêneas	O conjunto de categorias tem como único critério representar a estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações.
Mutuamente exclusivas	Todas as “unidades de conteúdo” se encaixam única e exclusivamente em uma categoria. A regra de classificação é acatar a estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações. O critério utilizado é o semântico (se atende ao critério “válida”). O critério de ordenação das “unidades de registro” é o alfabético.
Consistentes	Na classificação do mesmo conteúdo, visando ao mesmo objetivo proposto, outros pesquisadores chegarão a resultados semelhantes, ainda que estabeleçam conjuntos diferentes de categorias, pois os trabalhos acadêmicos possuem estrutura textual estável e persistente.

Fonte: Elaborado pela autora.

É importante salientar que a ACT sugere que seja efetuada a validação das categorias no percurso da aplicação da técnica até o término da etapa 4 (descrição), que será apresentada a seguir. O primeiro passo da etapa de descrição é a validação do conjunto de “categorias iniciais”, através das descrições destas, numa abordagem qualitativa, chegando-se ao conjunto de “categorias intermediárias”. Após essa análise, para validação dos critérios e adequação semântica e terminológica, foi feita também uma reformulação da ordem lógica das categorias, resultando no conjunto de “categorias finais”.

Ressalta-se que, em todos os momentos nos quais houve o procedimento de validação do conjunto de categorias, essa validação foi sempre com base nos critérios indicados no QUADRO 8 (anterior).

Passa-se, a seguir, à etapa 4, de descrição, na qual será apresentada a formulação do conjunto das “categorias intermediárias” e, depois, do conjunto das “categorias finais”, que será empregado na TAFNAVEGA.

#### **Etapa 4: descrição**

Nesta etapa, adotou-se uma abordagem qualitativa para a descrição das categorias, cujo detalhamento está apresentado no APÊNDICE C.

A seguir, apresenta-se uma síntese dessa descrição:

##### C1 Tema

É o assunto que a pesquisa investiga, sendo seu objeto de estudo.

##### C2 Dimensão

É a descrição do “onde” a pesquisa será desenvolvida, ou seja, o ambiente no qual determinado objeto ou ação ocorre.

##### C3 Ação

São os objetivos pretendidos pela pesquisa, incluindo os objetivos geral e os específicos.

##### C4 Causa e efeito

São enunciadas hipóteses ou pressuposições do tipo “se-então” (se variável independente, então variável dependente), advindas, geralmente, de uma teoria.

##### C5 Fundamento teórico

São os principais conceitos teóricos necessários ao desenvolvimento da pesquisa, provendo apoio teórico para os estudos, análise e reflexões sobre os dados e informações coletadas.

##### C6 Métodos

É a indicação do tipo de métodos e técnicas utilizados nos procedimentos metodológicos.

##### C7 Coleta de dados

É constituída pelos instrumentos utilizados como meio para coletar os dados para análise de resultados.

### C8 Ponto de vista do autor/perspectiva

Representa o olhar particular do pesquisador sobre o problema ou objeto estudado.

### C9 Principais resultados alcançados

É a apresentação da análise dos dados coletados durante o desenvolvimento da pesquisa, muitas vezes resultando na criação de um método, modelo ou na formulação de uma teoria.

Após a descrição das categorias, observou-se a necessidade de adequação semântica de alguns dos termos. Dessa forma, foram substituídos:

a) O termo da categoria C2, “dimensão”, foi substituído pelo termo “ambientação”, atendendo ao critério semântico, conforme verificado na literatura:

Dimensionar: [calcular a dimensão]: calcular ou preestabelecer as dimensões ou proporções (FERREIRA, 1988, p.222).

Dimensão: sentido em que se mede a extensão para avaliá-la. Tamanho (*ibidem*).

Ambientar: [ambientação é o ato de ambientar-se]: adaptar-se a um ambiente (*ibidem*, p.36).

Ambiente: [...] lugar, sítio, espaço, recinto (*ibidem*).

Nota-se que o significado de “ambientação” direciona para o entendimento de “lugar”, enquanto “dimensão” indica uma noção de “tamanho”, o que justificou a substituição, que torna o termo mais adequado na nomeação da categoria.

b) O termo da categoria C3, “ação”, foi substituído pelo termo “escopo”, atendendo ao critério semântico, conforme verificado na literatura:

Escopo: alvo, mira, intuito; intenção (FERREIRA, 1988, p.263).

Ação: ato ou efeito de atuar; atuação, ato, feito, obra (*ibidem*, p.8).

Nota-se que o significado da palavra “ação” denota “como fazer?” e que o significado da palavra “escopo” denota “para que?”, “qual a intenção?”, que é mais adequado à categoria dos objetivos;

c) O termo da categoria C9, “principais resultados alcançados”, foi substituído pelo termo “resultados”, pois qualquer pesquisa somente descreve os principais resultados alcançados e o uso do termo “resultados” já denota esse sentido completo.

Após as descrições apresentadas anteriormente, numa abordagem qualitativa, e adequações necessárias, chegou-se ao conjunto de “categorias intermediárias”, mantendo-se o mesmo esquema de codificação indicado na etapa anterior, descritos a seguir.

**Conjunto de categorias intermediárias:** C1.Tema; C2.Ambientação; C3.Escopo; C4.Causa e efeito; C5.Fundamento teórico; C6.Métodos; C7.Coleta de dados; C8.Ponto de vista do autor/perspectiva; C9.Resultados.

Em seguida, o conjunto das categorias intermediárias foi avaliado e analisado conforme os critérios indicados na etapa 2, descritos no QUADRO 8, resultando nas seguintes reformulações, que indicará o conjunto de “categorias finais”:

a) Categorias excluídas:

a1) a categoria C4, “causa e efeito”, foi excluída, porque esse tipo de elemento geralmente está relacionado a estudos muito específicos, muitas vezes sobre o comportamento de grupos sociais, em uma visão mais sociológica. Exige conhecimento profundo do texto, dificultando a identificação do seu conteúdo, não atendendo ao critério de características “válidas”.

a2) a categoria C8, “ponto de vista do autor/perspectiva”, foi excluída, porque exige análise profunda do texto do documento e geralmente refere-se a pesquisas de caráter teórico, em que se correlacionam conceitos e se formulam representações da realidade estudada. Não é pertinente manter uma categoria que raramente será preenchida e que não está alinhada às necessidades do usuário. Não atendeu ao critério de características “válidas”.

b) Categorias adicionadas (percebeu-se a carência de algumas categorias para atender aos critérios de características “válidas”, “exaustivas” e “mutuamente exclusivas”):

b1) Categoria “tipo de pesquisa”, para indicar a qualificação da pesquisa. A categoria é sinteticamente descrita da seguinte forma:

#### Categoria Tipo de pesquisa

São tipos de caracterização de pesquisa, tais como: natureza (básica, aplicada), abordagem (quantitativa, qualitativa), objetivos (exploratória, descritiva, explicativa) e procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, *expost-facto*, participante).

b2) Categoria “fundamento histórico/contextual”, para indicar a revisão de literatura que não esteja relacionada a teorias e métodos. A categoria é sinteticamente descrita da seguinte forma:

#### Categoria Fundamento histórico/contextual

É a apresentação das ideias presentes na literatura estudada, mostrando a relação que possuem com o tema pesquisado, de modo a construir e reconstruir os conceitos abrangidos na pesquisa.

b3) Categoria “objeto empírico”, para indicar o objeto utilizado para a análise e/ou coleta de dados. A categoria é sinteticamente descrita da seguinte forma:

#### Categoria Objeto Empírico

É o objeto de análise utilizado para investigar o tema (objeto de estudo), geralmente delimitado pelo *corpus* da pesquisa.

Após as validações fizeram-se reformulações na ordem lógica das categorias, respeitando a estrutura textual dos documentos do tipo teses e dissertações, resultando no conjunto de “categorias finais”, mantendo-se o mesmo esquema de codificação utilizado nas etapas anteriores, conforme segue:

**Conjunto de categorias finais:** C1.Tema; C2.Objeto empírico; C3.Escopo; C4.Ambientação; C5.Tipo de pesquisa; C6.Coleta de dados; C7.Métodos; C8.Fundamento teórico; C9.Fundamento histórico/contextual C10.Resultados.

Dando prosseguimento, a ACT determina que se faça a análise interpretativa de acordo com o “objetivo pretendido” ao aplicar a técnica, que é apresentada a seguir.

#### **Etapa 5: interpretação**

Neste procedimento a ACT recomenda que seja realizada uma análise interpretativa do resultado alcançado, com uma abordagem qualitativa, comparando esse resultado com a literatura especializada da área consultada. Em atendimento a esse preceito da ACT, as fontes utilizadas nesta etapa foram as mesmas da etapa 2, ou seja, a norma ABNT-14724:1992, Severino (2007), Marconi e Lakatos (2010), Silva (2010) e a estrutura textual dos documentos do tipo teses e dissertações.

Partindo da análise desse material, e comparando-a com o conjunto das “categorias finais”, percebe-se que esse conjunto de categorias é a “representação simplificada” do conteúdo dos documentos do *corpus* da pesquisa, tal como apresentado por Marconi e Lakatos (2010, p.93).

Neste cenário, notou-se que o uso da estrutura textual do *corpus* da pesquisa como parâmetro para a criação do conjunto das “categorias finais”, transformou esse conjunto em um modelo do documento. Percebeu-se, com isso, que o modelo reflete as partes mais estáveis do documento, pois as partes da estrutura textual mantêm relações consistentes com o conjunto de “categorias finais”.

Apresentar o conteúdo dos trabalhos acadêmicos a partir de sua representação simplificada faz prevalecer a ideia do todo do documento, deixando claro o conteúdo manifesto (explícito) da pesquisa realizada. Esse conteúdo é importante para o pesquisador em sua atividade profissional, como foi atestado pelas fontes da literatura especializada da área consultada.

Toda pesquisa científica necessita de aportes teóricos e do conhecimento sobre o “estado da arte” do tema que deseja investigar, além de ser importante conhecer os métodos e técnicas que são aplicáveis ao seu problema de pesquisa. Isso é confirmado pela ideia de que uma pesquisa científica é um processo contínuo e cíclico, o que remete à forma de operação de um sistema de informação, pois o pesquisador produz informação e também consome informação, alimentando e retro-alimentando esse sistema informacional.

Sendo assim, entende-se que a pesquisa científica é um processo sistêmico, no qual a pesquisa é o todo, composto pelas diferentes fases que são representadas por partes tais como a identificação do problema, a solução proposta e os métodos empregados para alcançar o resultado final. Mas, essas partes, ao se juntarem, extrapolam sua função primária. Assim, a pesquisa (objeto) não é percebida isoladamente, mas no seu contexto de origem (produção), de uso (demanda do usuário) e das relações que estabelecem entre si.

Essa ideia de processo sistêmico também remete ao pensamento de Dahlberg (1978), que, discursando sobre as relações paradigmáticas funcionais entre os conceitos de um sistema de classificação, aponta que elas ocorrem nos conceitos de relação do todo e suas partes (todo/parte), por exemplo. Nesse sentido, essas relações produzem a ideia de conexão entre as partes do texto, de produção (significado), com o uso (funcional) dos documentos acadêmicos pelos pesquisadores. Observa-se, assim, a relação funcional entre as partes da estrutura textual, pois, para o desenvolvimento de uma pesquisa, uma parte é dependente funcional de outra. Essa dependência pode ser percebida na literatura especializada da área, quando esta indica que, para alcançar o objetivo da pesquisa, devem-se levar em consideração os fundamentos teóricos e metodológicos, para manter consistência nos procedimentos adotados. Além disso, considera-se que, como é ressaltado

pela ACT, se outro pesquisador utilizar os mesmos critérios de avaliação aqui propostos e tendo em vista os mesmos objetivos dessa análise, chegará a resultados similares aos encontrados por esta pesquisa.

No próximo item, passa-se a descrever o desenvolvimento do algoritmo<sup>74</sup> para ser utilizado no processo de extração de conceitos do conteúdo de trabalhos acadêmicos do tipo teses e dissertações, para a criação da TAFNAVEGA. Segundo Dias e Naves (2007, p.69), “para definir em termos adequados o assunto de um texto é necessário que, primeiramente, se extraiam os conceitos que nele estão contidos” (grifo nosso). Nesse sentido, o algoritmo indica as instruções metodológicas a serem respeitadas no processo do tratamento temático documental<sup>75</sup> (análise conceitual) deste tipo de documento (*corpus*). Ao final, os termos representativos dos conceitos extraídos irão alimentar o conjunto de categorias fundamentais temáticas da TAFNAVEGA.

### 5.3.2 O algoritmo para o processo de extração de conceitos

Para a análise conceitual nos trabalhos acadêmicos, propõe-se uma sequência de tarefas com instruções bem definidas, com o objetivo de extrair os conceitos que estão constituídos nos textos, que irão alimentar as categorias fundamentais temáticas da TAFNAVEGA. Para isso, foi desenvolvido um algoritmo composto de uma tabela, com base no Modelo de Leitura Documentária (MLD) elaborado por Fujita e Rubi (2006), cujos procedimentos estão descritos no Capítulo 4, item 4.7 desta pesquisa. O MLD dessas autoras foi criado para a análise de artigos da área de ciências biológicas e, neste trabalho, ele foi adaptado para a análise de trabalhos acadêmicos do tipo teses e dissertações, especialmente para a área da CI.

Na adaptação da tabela, denominada Matriz Categorical para trabalhos acadêmicos (MCTCA), aplicou-se a mesma metodologia empregada no MLD, que consiste no desenvolvimento de três colunas: 1) a primeira coluna é composta pelos conceitos que representam a estrutura textual dos documentos; 2) a segunda coluna indica a abordagem sistemática, que é realizada através de questionamentos, a partir das recomendações da norma 12.676 (1992) e do sistema de indexação PRECIS; 3) a terceira coluna determina as partes da estrutura textual dos documentos, nas quais os conceitos da primeira coluna devem ser buscados.

---

<sup>74</sup> Um algoritmo mostra os procedimentos (passo a passo) necessários para completar uma tarefa, respondendo à pergunta “como fazer?”. Ele é uma sequência lógica, finita e bem definida de instruções para executar uma tarefa. Nem sempre um algoritmo representa um programa de computador, pois ele pode determinar também as ações de um ser humano.

<sup>75</sup> O tratamento temático documental é a aplicação de métodos e técnicas funcionais, contextualizada na necessidade informacional do usuário, com vistas ao acesso à informação, potencializando o desenvolvimento do conhecimento e aprimorando a recuperação da informação. É a ênfase dada a “o que” tratar, “para quem” tratar e “como” tratar (GUIMARÃES, 2001).

Em seguida, passa-se a descrever os procedimentos adotados para o desenvolvimento de cada uma das três colunas do MCTCA.

### **Primeira coluna – conceitos da estrutura textual de trabalhos acadêmicos**

Esta primeira coluna foi composta pelos conceitos que representam a estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações. Esses conceitos se referem ao conjunto de categorias fundamentais temáticas (CAFTE), da taxonomia facetada navegacional (TAFNAVEGA), já apontado na etapa anterior, no subcapítulo **5.3.1**, que são apresentados a seguir e expostos no QUADRO 9.

**Conjunto de categorias fundamentais temáticas:** C1.Tema; C2.Objeto empírico; C3.Escopo; C4.Ambientação; C5.Coleta de dados; C6.Tipo de pesquisa; C7.Métodos; C8.Fundamento teórico; C9.Fundamento histórico/contextual; C10.Resultados.

QUADRO 9 – Primeira coluna do MCTCA

<b>TERMOS CAFTE</b>
C1. TEMA
C2. OBJETO EMPÍRICO
C3. ESCOPO
C4. AMBIENTAÇÃO
C5. TIPO DE PESQUISA
C6. COLETA DE DADOS
C7. MÉTODOS
C8. FUNDAMENTO TEÓRICO
C9. FUNDAMENTO HISTÓRICO/ CONTEXTUAL
C10. RESULTADOS

Fonte: Elaborado pela autora.

Com a apresentação do conjunto de termos (representando os conceitos) que compõem a primeira coluna do MCTCA (QUADRO 9), passou-se aos procedimentos adotados no desenvolvimento da segunda coluna.

### **Segunda coluna – abordagem sistemática**

Esta segunda coluna do MCTCA foi composta pelos “questionamentos” sugeridos na norma ABNT-12.676:1992 e pelos princípios de análise do sistema de indexação PRECIS (FUJITA; RUBI, 2006). Além disso, como é recomendado pela mesma

norma, foram acrescentados outros questionamentos aos já sugeridos, tendo como objetivo tornar a abordagem sistemática mais detalhada.

Seguindo as indicações descritas acima, foram definidos os “questionamentos” para cada termo do conjunto de termos/categorias da primeira coluna, que estão descritos a seguir e apresentados no QUADRO 10.

#### C1. Tema

“Qual o assunto de que trata o documento?”.

#### C2. Objeto empírico

“Qual o objeto empírico do estudo em questão? Qual objeto foi analisado e/ou utilizado na pesquisa?”.

#### C3. Escopo

“O que pretende a pesquisa, de forma geral e específica, que seja relevante determinar? A que objetivos a pesquisa tem intenção de atender (aprimorar, avaliar, analisar, identificar, contribuir, etc.)?”.

#### C4. Ambientação

“O tema, o objeto empírico e/ou a ação são considerados no contexto de um lugar específico ou ambiente?”.

#### C5. Tipo de pesquisa

“Tendo em vista a natureza (básica, aplicada), a abordagem (quantitativa, qualitativa), os objetivos (exploratória, descritiva, explicativa) e os procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, *expost-facto*, participante), quais classificações podem tipificar a pesquisa realizada?”.

#### C6. Coleta de dados

“Quais instrumentos específicos (questionários, entrevistas, registros áudios-visuais, coleta de documentos, etc.) foram utilizados para realizar a ação?”.

#### C7. Métodos

“Quais modos específicos foram utilizados para realizar a ação (por exemplo, técnicas ou métodos para tratamento dos dados, que podem ser do tipo: modelagem

estatística, análise estrutural, codificação, análise de conteúdo, indexação, análise semiótica, retórica ou de discurso)?”.

#### C8. Fundamento teórico

“Houve alguma corrente ou abordagem teórica específica (teorias, hipóteses, pressupostos, etc.) utilizada em função da natureza do objeto a ser pesquisado e dos objetivos pretendidos, que foram descritos na pesquisa?”.

#### C9. Fundamento histórico/contextual

“Quais temas foram tratados e revisados, a partir de pesquisa bibliográfica, para contextualizar o tema pesquisado de forma profunda e consistente?”.

#### C10. Resultados

“Quais pontos a pesquisa alcançou, levando em consideração os objetivos propostos? Houve formulação ou reformulação de teoria, criação de um método ou de um produto?”.

QUADRO 10 – Primeira e segunda coluna do MCTCA

TERMOS CAFTE	QUESTIONAMENTOS (NORMA 12.676 e PRECIS)
C1. TEMA	Qual o assunto de que trata o documento?
C2. OBJETO EMPÍRICO	Qual o objeto empírico do estudo em questão? Qual objeto foi utilizado e/ou analisado na pesquisa?
C3. ESCOPO	O que pretende a pesquisa, de forma geral e específica, que seja relevante determinar? A que objetos a pesquisa tem intenção de atender (aprimorar, avaliar, analisar, identificar, contribuir, etc.)?
C4. AMBIENTAÇÃO	O tema, objeto empírico e/ou ação são considerados no contexto de um lugar específico ou ambiente?
C5. TIPO DE PESQUISA	Tendo em vista a natureza (básica, aplicada), a abordagem (quantitativa, qualitativa), os objetivos (exploratória, descritiva, explicativa) ou os procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, <i>ex post-facto</i> , participante), quais as classificações podem tipificar a pesquisa realizada?
C6. COLETA DE DADOS	Quais instrumentos específicos (questionários, entrevistas, registros áudios-visuais, coleta de documentos, etc.) foram utilizados para realizar a ação?
C7. MÉTODOS	Quais modos específicos foram utilizados para realizar a ação (por exemplo, técnicas ou métodos para tratamento dos dados, que podem ser do tipo: modelagem estatística, análise estrutural, codificação,

TERMOS CAFTE	QUESTIONAMENTOS (NORMA 12.676 e PRECIS)
	análise de conteúdo, indexação, análise semiótica, retórica ou de discurso)?
C8. FUNDAMENTO TEÓRICO	Houve alguma corrente ou abordagem teórica específica (teorias, hipóteses, pressupostos, etc.) utilizada em função da natureza do objeto a ser pesquisado e dos objetivos pretendidos, que foram descritos na pesquisa?
C9. FUNDAMENTO HISTÓRICO/ CONTEXTUAL	Quais temas foram tratados e revisados, a partir de pesquisa bibliográfica, para contextualizar o tema pesquisado de forma profunda e consistente?
C10. RESULTADOS	Quais pontos a pesquisa alcançou, levando em consideração os objetivos propostos? Houve formulação ou reformulação de teoria, criação de um método ou de um produto?

Fonte: Adaptado do MLD de Fujita e Rubi (2006).

Com a determinação da primeira e segunda colunas do MCTCA, passou-se ao procedimento do desenvolvimento da terceira coluna.

### Terceira coluna – partes da estrutura textual dos documentos

Esta coluna do MCTCA foi composta com a indicação das partes da estrutura textual dos trabalhos acadêmicos, nas quais se devem procurar as respostas aos questionamentos da segunda coluna. Esse procedimento foi realizado com base na norma NBR-14724:2011 e nas fontes da literatura especializada da área, as mesmas utilizadas na etapa 2, do subcapítulo **5.3.1**.

Partindo dessas premissas, a seguir indicam-se as partes da estrutura textual dos documentos, nas quais se encontram as informações referentes a cada conceito da primeira coluna.

#### C1. Tema

Resumo / Introdução (problema, justificativa, objetivos).

#### C2. Objeto empírico

Resumo / Introdução (problema, justificativa, objetivos) / Metodologia.

#### C3. Escopo

Resumo / Introdução (problema, justificativa, objetivos).

#### C4. Ambientação

Resumo / Introdução (problema, justificativa, objetivos) / Metodologia.

#### C5. Tipo de pesquisa

Resumo / Metodologia / Resultados / Discussão de resultados.

#### C6. Coleta de dados

Resumo / Metodologia / Resultados / Discussão de resultados.

#### C7. Métodos

Resumo / Metodologia / Resultados / Discussão de resultados.

#### C8. Fundamento teórico

Resumo / Revisão de literatura / Fundamentação.

#### C9. Fundamento histórico/contextual

Resumo / Revisão de literatura / Fundamentação.

#### C10. Resultados

Resumo / Resultados / Discussão de resultados / Conclusão.

Finalizado o desenvolvimento da terceira coluna, apresenta-se o algoritmo do modelo MCTCA completo, conforme o QUADRO 11.

QUADRO 11 – Matriz categorial para trabalhos acadêmicos (teses e dissertações)

<b>TERMOS CAFTE</b>	<b>QUESTIONAMENTOS (NORMA 12.676) e PRECIS</b>	<b>PARTE DA ESTRUTURA TEXTUAL</b>
C1. TEMA	Qual o assunto de que trata o documento?	RESUMO / INTRODUÇÃO (problema, justificativa, objetivos)
C2. OBJETO EMPÍRICO	Qual o objeto empírico do estudo em questão? Qual objeto foi utilizado e/ou analisado na pesquisa?	RESUMO / INTRODUÇÃO (problema, justificativa, objetivos) / METODOLOGIA

TERMOS CAFTE	QUESTIONAMENTOS (NORMA 12.676) e PRECIS	PARTE DA ESTRUTURA TEXTUAL
C3. ESCOPO	O que pretende a pesquisa, de forma geral e específica, que seja relevante determinar? A que objetos a pesquisa tem intenção de atender (aprimorar, avaliar, analisar, identificar, contribuir, etc.)?	RESUMO / INTRODUÇÃO (problema, justificativa, objetivos)
C4. AMBIENTAÇÃO	O tema, objeto empírico e/ou ação são considerados no contexto de um lugar específico ou ambiente?	RESUMO / INTRODUÇÃO (problema, justificativa, objetivos) / METODOLOGIA
C5. TIPO DE PESQUISA	Tendo em vista a natureza (básica, aplicada), a abordagem (quantitativa, qualitativa), os objetivos (exploratória, descritiva, explicativa) ou os procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, <i>expost-facto</i> , participante), quais as classificações podem tipificar a pesquisa realizada?	RESUMO / METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS
C6. COLETA DE DADOS	Quais instrumentos específicos (questionários, entrevistas, registros áudios-visuais, coleta de documentos, etc.) foram utilizados para realizar a ação?	RESUMO / METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS
C7. MÉTODOS	Quais modos específicos foram utilizados para realizar a ação (por exemplo, técnicas ou métodos para tratamento dos dados, que podem ser do tipo: modelagem estatística, análise estrutural, codificação, análise de conteúdo, indexação, análise semiótica, retórica ou de discurso)?	RESUMO / METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS
C8. FUNDAMENTO TEÓRICO	Houve alguma corrente ou abordagem teórica específica (teorias, hipóteses, pressupostos, etc.) utilizada em função da natureza do objeto a ser pesquisado e dos objetivos pretendidos, que foram descritos na pesquisa?	RESUMO / REVISÃO DE LITERATURA / FUNDAMENTAÇÃO
C9. FUNDAMENTO HISTÓRICO/ CONTEXTUAL	Quais temas foram tratados e revisados, a partir de pesquisa bibliográfica, para contextualizar o tema pesquisado de forma profunda e consistente?	RESUMO / REVISÃO DE LITERATURA / FUNDAMENTAÇÃO
C10. RESULTADOS	Quais pontos a pesquisa alcançou, levando em consideração os objetivos propostos? Houve formulação ou reformulação de teoria, criação de um método ou de um produto?	RESUMO / RESULTADOS / DISCUSSÃO DE RESULTADOS / CONCLUSÕES

Fonte: Adaptado do MLD de Fujita e Rubi (2006).

O objetivo do modelo MCTCA é orientar e sistematizar a análise conceitual nos documentos do tipo teses e dissertações, que, ao mesmo tempo, facilita a extração de conceitos e alimenta a estrutura da TAFNAVEGA, tornando-a um mecanismo para a navegação facetada.

Para demonstrar a utilização desse algoritmo na extração de conceitos, e a funcionalidade da TAFNAVEGA, foi preparada uma experiência<sup>76</sup> composta por duas etapas: 1) na primeira etapa, no subcapítulo 5.4, a seguir, apresenta-se o uso do MCTCA na extração dos conceitos, com a análise nos resumos informativos dos documentos do *corpus* desta pesquisa; 2) na segunda etapa, descrita no subcapítulo 5.5, foi realizado uma apresentação da implementação tecnológica, com o objetivo de demonstrar a visualização da navegação pela “taxonomia facetada navegacional” (TAFNAVEGA), em meio digital.

Passa-se, então, a apresentar os procedimentos da primeira etapa.

#### 5.4 O experiência do MCTCA na análise do *corpus* da pesquisa

Nesta etapa da experiência, foi realizada uma análise conceitual, utilizando o MCTCA, para a extração dos conceitos referentes aos resumos informativos dos documentos do *corpus* (41 documentos) desta pesquisa.

Na aplicação do MCTCA, sentiu-se a necessidade de acrescentar uma quarta coluna à tabela, com a finalidade de registrar os termos (representando os conceitos) extraídos do documento, referentes aos termos/categorias da primeira coluna, conforme o QUADRO 12.

QUADRO 12 – Matriz categorial para extração de conceitos

Primeira coluna	Segunda coluna	Terceira coluna	Quarta coluna
TERMOS CAFTE	QUESTIONAMENTOS (NORMA 12.676) e PRECIS	PARTE DA ESTRUTURA TEXTUAL	TERMOS DE INDEXAÇÃO

Fonte: Adaptado do modelo MCTCA.

Descreve-se, a seguir, a sequência dos procedimentos da aplicação do algoritmo do MCTCA nos resumos informativos do *corpus* desta pesquisa: 1) selecione um conceito CAFTE na primeira coluna; 2) procure, no resumo informativo do documento analisado, as partes da estrutura textual indicadas na terceira coluna; 3) analise o resumo informativo do documento, fazendo os questionamentos indicados na segunda coluna; 4) extraia e selecione os conceitos dos resumos informativos.

Contudo, para concluir a experiência da extração de conceitos e registrar os conceitos extraídos através do algoritmo, foi preciso adicionar, aos procedimentos

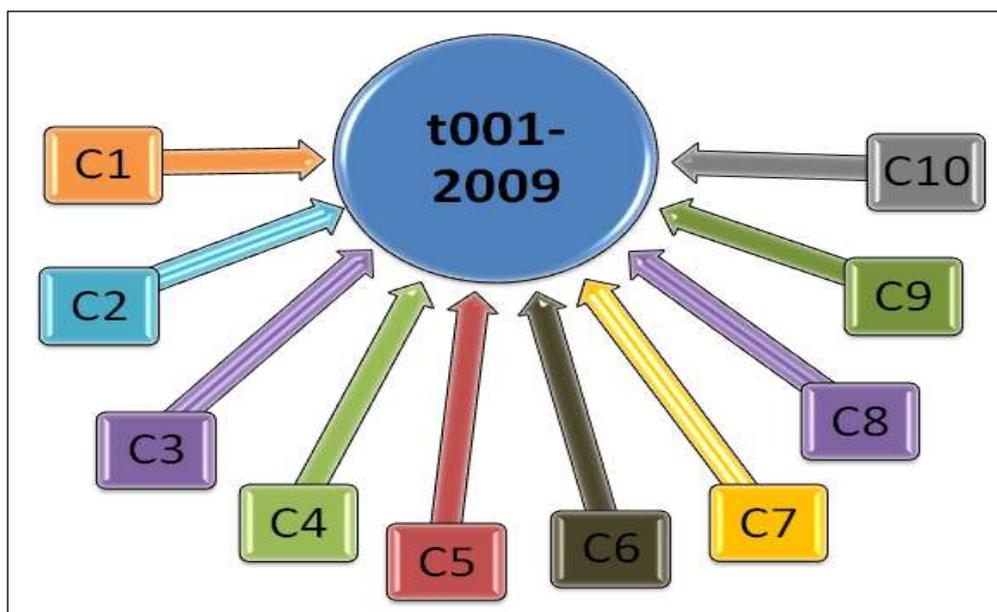
<sup>76</sup> Na presente pesquisa, o vocábulo “experiência” significa o ato ou efeito de conhecer a aplicação de um instrumento ou objeto, verificando a sua qualidade de uso (FERREIRA, 1988).

anteriores, a seguinte instrução: 5) faça a tradução<sup>77</sup> dos conceitos, utilizando um vocabulário controlado, e o insira na quarta coluna.

Para essa tradução dos conceitos selecionados, foi utilizado o vocabulário controlado “Taxonomia da Ciência da Informação”, desenvolvido por Oddone e Gomes (2003). É importante novamente esclarecer que o escopo desta pesquisa não inclui a avaliação de vocabulários controlados quando empregados na atividade de tradução. Contudo, salienta-se que essa validação é necessária e pretende-se investigá-la em um trabalho posterior. A escolha pelo uso do vocabulário controlado (VC) de Oddone e Gomes, nessa experiência, deve-se à preferência pela utilização de um instrumento (a taxonomia) que, originalmente, fosse de mesmo tipo e princípio do mecanismo proposto nesta pesquisa.

Prosseguindo-se na análise conceitual, cada resumo informativo do *corpus* desta pesquisa foi indexado pelo conjunto de termos (representando os conceitos) da primeira coluna do MCTCA, conforme a FIG. 6.

FIGURA 6 – Processo de extração de conceitos dos resumos informativos do *corpus*



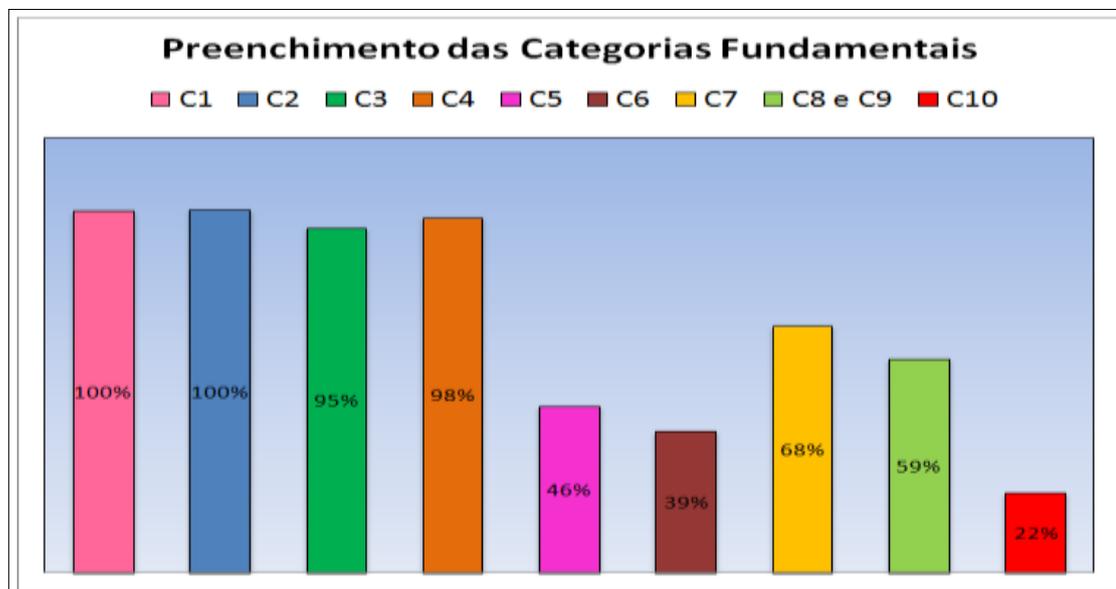
Fonte: Elaborado pela autora.

Ao final do procedimento de análise conceitual, foi coletado um total de 407 termos indexadores. Esses termos foram refinados somente pelo critério de exclusão de termos idênticos, uma vez que, com o uso do vocabulário controlado na tradução, não houve problema de sinonímia. Após esse refinamento, totalizaram 168 termos indexadores (APÊNDICE H).

<sup>77</sup> A tradução é uma das etapas da indexação na qual a linguagem natural dos autores é convertida para a linguagem controlada utilizada no sistema, fazendo a compatibilização das duas linguagens.

Como resultado dessa etapa, apresenta-se o percentual de preenchimento do conjunto dos termos CAFTE no GRÁF. 1.

GRÁFICO 1 – Percentual de preenchimento das CAFTE (resumos do *corpus*)



Fonte: Síntese dos dados do APÊNDICE G.

Observando o GRÁF. 1, nota-se, principalmente, o baixo percentual da categoria C10 (resultados). Entretanto, esse percentual pode representar a totalidade de trabalhos que tiveram como resultado de pesquisa um produto (protótipo), modelo, método ou teoria desenvolvida. Porém, ao longo desse procedimento, alguns problemas foram detectados, e o primeiro deles é a incompletude dos resumos informativos do *corpus* analisado. Esse problema também pode explicar o baixo percentual do resultado da categoria C10, pois nem sempre esses resumos incluíram todos os resultados alcançados pelas pesquisas, conforme apresentado no resultado da análise, que são descritos a seguir:

a) Resumos incompletos: quase 95% do total, pois apenas dois documentos (d021-2006; d023-2006) apresentaram informações suficientes para o preenchimento de todos os campos do MCTCA;

b) Inexistência de palavras-chave: quase 25% do total dos resumos informativos não possuíam palavras-chave (d037-2003; d025-2005; d020-2006; d036-2004; d030-2005; t012-2005; d017-2007; t009-2007; d013-2008; t002-2009);

c) Palavras-chave insuficientes para representar o conteúdo do documento: 100% dos resumos informativos não contêm palavras-chave suficientes para representar todo o conteúdo das pesquisas.

Esse resultado está tabulado no GRÁF. 2.

GRÁFICO 2 – Problemas em resumos informativos do *corpus* da pesquisa

Fonte: Dados extraídos da análise conceitual nos resumos do *corpus*.

Observa-se que o problema mais grave é mesmo o de resumos incompletos, pois esse problema traz prejuízos à qualidade da recuperação do conteúdo dos documentos. Decerto, um pesquisador geralmente recorre aos resumos informativos dos trabalhos acadêmicos quando busca por uma informação sobre pesquisas já concluídas, uma vez que a consulta aos textos completos demandaria um tempo demasiado, e poderia não ser viável. Dessa forma, a qualidade dos resumos informativos das teses e dissertações tem influência direta na melhora da recuperação de informações.

Outros problemas foram verificados nos resumos informativos, tais como: sinonímia na atribuição de palavras-chave (ex. d023-2006), falta de objetividade (ex. d029-2005) e itens pré-textuais inseridos após o sumário (ex. d014-2007).

Prosseguindo com a experiência do uso do MCTCA na análise dos resumos informativos do *corpus* da pesquisa, foram selecionadas cinco, das dez categorias possíveis, como amostra para a demonstração e análise dos resultados. A escolha da amostra foi obtida aplicando-se o critério de amostra probabilística<sup>78</sup>, com uma seleção aleatória simples (por sorteio), chegando-se à seguinte lista de categorias: C1. Tema, C2. Objeto empírico, C5. Tipo de pesquisa, C6. Coleta de dados e C7. Métodos.

O objetivo dessa demonstração e análise dos resultados foi indicar que, utilizando a TAFNAVEGA em uma situação empírica, quais dados podem ser coletados e, a partir dos resultados encontrados, quais tipos de análises podem ser realizados.

Dessa maneira, em seguida apresentam-se os resultados da análise nas categorias da amostra (gráficos no APÊNDICE I), respeitando-se a ordem decrescente do

<sup>78</sup> Quando a população é finita e pequena pode-se aplicar a amostra probabilística simples escrevendo cada elemento da população em um cartão, misturando-os em uma urna e sorteando tantos cartões quantos se deseja na amostra.

percentual de preenchimento das mesmas (GRÁF. 1), que será a seguinte: C1. Tema, C2. Objeto empírico, C7. Métodos, C5. Tipo de pesquisa e C6. Coleta de dados.

**Categoria C1 – Tema:** houve 100% de preenchimento dessa categoria. Observou-se que há, basicamente, duas vertentes na linha de pesquisa OUI da ECI-UFMG: uma vertente “organização da informação” e outra vertente “estudos de uso”. A vertente “organização” totalizou 66% dos temas pesquisados, e a vertente “uso” totalizou 34%. Esse resultado demonstra que as pesquisas na linha OUI priorizam as questões relacionadas com a organização da informação. Entretanto, ressalta-se que outra linha de pesquisa da ECI-UFMG, “Informação, Cultura e Sociedade” (ICS), também investiga o “uso de informações”, o que pode explicar a diferença percentual. Entretanto, o escopo desta pesquisa não abarcou a análise das outras linhas de pesquisas, e essa informação representa somente um indício para futuras análises que envolvam as três linhas de pesquisa do PPGCI-UFMG.

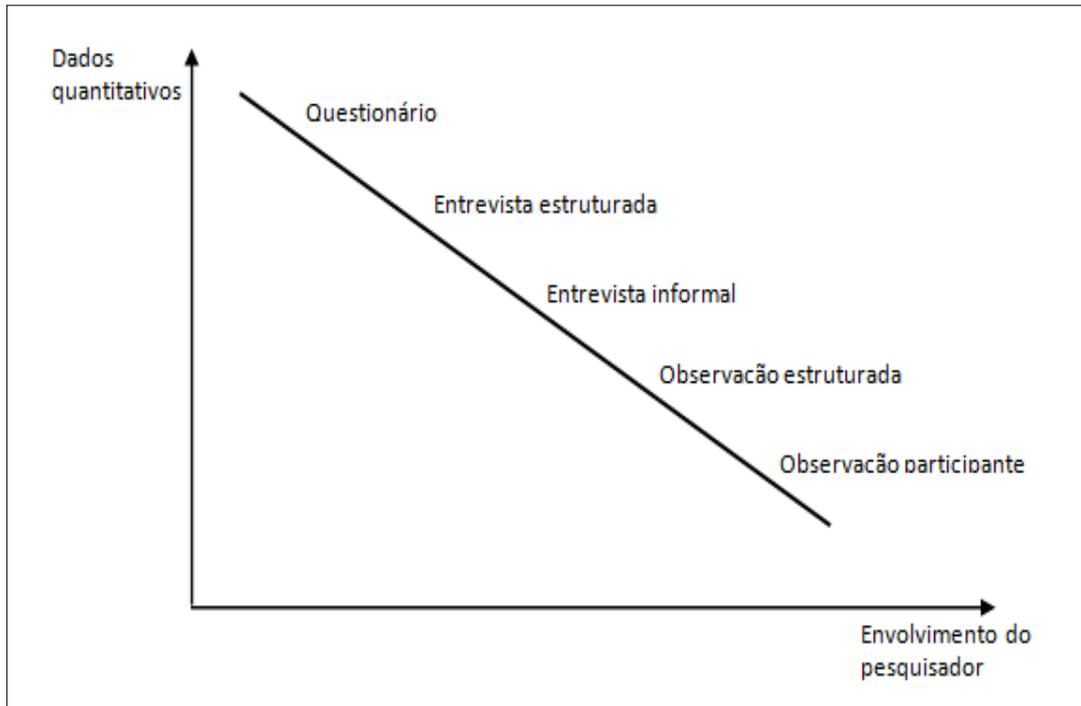
**Categoria C2 – Objeto empírico:** houve 100% de preenchimento dessa categoria. Percebeu-se bastante heterogeneidade nesse item. Uma particularidade observada foi em relação ao uso de prontuários médicos (três trabalhos, quase 8% do total). Esse resultado indica um interesse especial das pesquisas na linha OUI com a organização e o acesso de informações médicas, provavelmente devido à ampliação no número de hospitais e clínicas que estão adotando os prontuários eletrônicos, nos últimos anos.

**Categoria C7 – Métodos:** houve 68% de preenchimento dessa categoria, totalizando 28 documentos do *corpus* da pesquisa. Observou-se que um mesmo documento, muitas vezes, indicou a utilização de mais de um método de pesquisa, o que é uma característica comum. Verificou-se que os métodos mais utilizados na vertente “organização da informação” foram os métodos da análise documental, da classificação (análise facetada) e do analítico-sintético, revelando a recorrência do uso dos princípios dos estudos de Ranganathan na linha de pesquisa OUI. Já na vertente “uso da informação”, os métodos mais utilizados foram os de estatística (bibliometria), geralmente combinando a abordagem quantitativa com uma análise qualitativa dos dados, e o método da análise de citações. Esse resultado indica uma visão mais positivista da vertente “uso da informação”.

**Categoria C5 – Tipo de pesquisa:** houve 46% de preenchimento dessa categoria, totalizando 19 documentos do *corpus* da pesquisa. Da mesma forma como na categoria C7 (Métodos), nessa categoria houve a indicação de mais de um tipo de qualificação para uma pesquisa, característica também comum. Os resultados revelaram que prevalecem as pesquisas qualitativas e comparativas, o que parece indicar o interesse em investigar melhores formas de organização de informações para atender às necessidades do usuário final. Além disso, prevalecem também as pesquisas do tipo “estudos de caso”, dificultando as generalizações dos resultados encontrados nessas pesquisas.

**Categoria C6 – Coleta de dados:** houve 39% de preenchimento dessa categoria, totalizando 16 documentos do *corpus* da pesquisa. As entrevistas e os questionários são os instrumentos de coleta de dados mais utilizados. Isso demonstra o grau de envolvimento dos pesquisadores com os respondentes, conforme FIG. 7.

FIGURA 7 – Envolvimento do pesquisador com respondentes



Fonte: Cação (2008)<sup>79</sup>

Observa-se que o pesquisador da linha de pesquisa OUI, geralmente, tem um grau de envolvimento com seus respondentes que fica entre médio e baixo, denotando uma abordagem mais positivista, ainda que haja uma análise qualitativa dos dados coletados.

Refletindo sobre os dados coletados, principalmente, das últimas categorias analisadas acima (C5 e C6), que tiveram um percentual de preenchimento considerado baixo, percebe-se que o problema da incompletude dos resumos informativos atrapalha a recuperação de importantes informações das pesquisas.

Ademais, verificou-se que o MCTCA pode, também, ser utilizado para auxiliar na sistematização de resumos informativos para trabalhos acadêmicos (teses e dissertações). Para isso, é necessário acrescentar uma coluna ao MCTCA, para que as informações referentes ao documento sejam inseridas, conforme mostra a FIG. 8.

<sup>79</sup> CAÇÃO, Rosário. Análise Qualitativa de Dados com o NVIVO. 31 de out. 2008. Disponível em: <<http://slidesha.re/11XyZq>>. Acesso em: 12 mar. 2011.

FIGURA 8 – Tabela para elaboração de resumo informativo

TIPO DOCUMENTO: ( ) TESE ( ) DISSERTAÇÃO TÍTULO: _____				
CONCEITO (ANÁLISE CONCEITUAL)	QUESTIONAMENTO (NORMA 12.676) e PRECIS	PARTE DA ESTRUTURA TEXTUAL	INFORMAÇÕES DO DOCUMENTO	TERMOS DE INDEXAÇÃO
TEMA	Qual o assunto de que trata o documento?	INTRODUÇÃO (problema, justificativa, objetivos)		
OBJETO EMPÍRICO	Qual o objeto de estudo em questão? O que está sob estudo da pesquisa?	INTRODUÇÃO (problema, justificativa, objetivos) / METODOLOGIA		
ESCOPO	O que pretende a pesquisa, de forma geral e específica, que seja relevante determinar? A que objetos a pesquisa tem intenção de atender (aprimorar, avaliar, analisar, identificar, contribuir, etc.)?	INTRODUÇÃO (problema, justificativa, objetivos)		
AMBIENTAÇÃO	O tema, objeto empírico e/ou ação são considerados no contexto de um lugar específico ou ambiente?	INTRODUÇÃO (problema, justificativa, objetivos) / METODOLOGIA		
TIPO DE PESQUISA	Tendo em vista a natureza (básica, aplicada), a abordagem (quantitativa, qualitativa), os objetivos (exploratória, descritiva, explicativa) ou os procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, <i>ex-post-facto</i> , participante), quais as classificações podem tipificar a pesquisa realizada?	METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS		
COLETA DE DADOS	Quais instrumentos específicos (questionários, entrevistas, registros áudio-visuais, coleta de documentos, etc.) foram utilizados para realizar a ação?	METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS		
MÉTODOS	Quais modos específicos foram utilizados para realizar a ação (por exemplo, técnicas ou métodos, para tratamento dos dados, que podem ser do tipo: modelagem estatística, análise estrutural, codificação, análise de conteúdo, indexação, análise semiótica, retórica ou de discurso)?	METODOLOGIA / RESULTADOS / DISCUSSÃO DE RESULTADOS		
FUNDAMENTO TEÓRICO	Houve alguma corrente ou abordagem teórica específica (teorias, hipóteses, pressupostos, etc.) utilizada em função da natureza do objeto a ser pesquisado e dos objetivos pretendidos, que foram descritos na pesquisa?	REVISÃO DE LITERATURA / FUNDAMENTAÇÃO		
FUNDAMENTO HISTÓRICO / CONTEXTUAL	Quais temas foram tratados e revisados, a partir de pesquisa bibliográfica, para contextualizar o tema pesquisado de forma profunda e consistente?	REVISÃO DE LITERATURA / FUNDAMENTAÇÃO		
RESULTADOS	Quais pontos a pesquisa alcançou, levando em consideração os objetivos propostos? Houve formulação	RESULTADOS / DISCUSSÃO DE RESULTADOS / CONCLUSÕES		

Fonte: Adaptada do MCTCA.

A partir das informações inseridas na coluna “informações do documento”, elabora-se um texto corrido (texto sem tópicos e em parágrafo único), de acordo com as regras para redação de resumos informativos da norma NBR-6028:2003. Dessa forma, apesar de não fazer parte dos objetivos desta pesquisa, essa tabela é, também, um resultado obtido por esta pesquisa. Vislumbra-se que ela poderá auxiliar os autores e/ou indexadores na preparação dos resumos informativos, tornando-os mais completos e objetivos, podendo, ainda, impactar positivamente a recuperação de informações.

### **Análise geral**

O conteúdo textual dos documentos do tipo teses e dissertações é semi-estruturado e, dessa maneira, muitas vezes não é facilmente recuperado pelo pesquisador. A partir do exposto nesta pesquisa sobre as necessidades informacionais de um pesquisador, verificou-se que, em atividade profissional, ele precisa obter informações sobre os resultados de pesquisa, tais como teorias, métodos, instrumentos de coleta de dados e tipos de pesquisa. Dessa maneira, percebeu-se que essas informações estão armazenadas e disponibilizadas, mas não, necessariamente, acessíveis ao usuário.

Acredita-se que o pesquisador somente conseguirá obter facilmente uma informação, sobre o método utilizado em uma pesquisa, por exemplo, se ele souber exatamente onde essa informação necessitada está. Ou seja, será fácil recuperar a informação se o pesquisador souber o título ou o autor do documento. Ou, talvez, se essa informação estiver descrita no resumo informativo e, ainda, se o usuário for hábil no uso das expressões de busca *booleanas*. Caso contrário, recuperar a informação de que necessita será mais difícil para o usuário.

Ao realizar a implantação de bibliotecas digitais de teses e dissertações nas instituições de ensino superior, com o objetivo de evidenciar o conhecimento acumulado nas pesquisas concluídas, não basta disponibilizá-las digitalmente. É preciso agregar valor à informação disponibilizada e criar condições para que esse conhecimento possa ser acessado, possibilitando a sua recuperação. Atualmente, ao utilizar a BDTD, o usuário se depara com grande quantidade de informações que podem dar subsídio para sua pesquisa, porém, elas nem sempre são fáceis de explorar, localizar e recuperar. Propõe-se a TAFNAVEGA como mecanismo para a estruturação das informações contidas nos trabalhos acadêmicos, capaz de permitir a exploração do recurso informacional da BDTD, através da navegação por suas categorias e facetas, facilitando a busca e, conseqüentemente, aprimorando a recuperação da informação.

A TAFNAVEGA foi criada a partir da ordenação lógica da estrutura textual de trabalhos acadêmicos, permitindo mapear e organizar o conteúdo considerado relevante

para o pesquisador, em sua atividade profissional. O mecanismo possui um conjunto de categorias fundamentais temáticas, que agrupam e interrelacionam os diferentes conteúdos dos documentos. Cada categoria pode ser composta por uma hierarquia (taxonomia) de facetas e subfacetadas, que descreve os documentos sob essa dimensão específica. Essa característica dá, ao instrumento, alto grau de flexibilidade e, ao usuário, a possibilidade de combinar diferentes aspectos e documentos.

Com o algoritmo proposto para a extração dos conceitos contidos nos documentos, cujos termos representativos que irão alimentar as categorias fundamentais temáticas da TAFNAVEGA, a atividade de análise conceitual do processo de indexação do recurso informacional será objetivado. Isto é, a atividade será sistematizada, minimizando a subjetividade envolvida pois usa parâmetros objetivos relativos à análise conceitual.

Neste subcapítulo 5.4, foi descrita e analisada a primeira etapa da experiência, composta por duas etapas, para demonstrar (1) a utilização do algoritmo na extração de conceitos e (2) a funcionalidade da TAFNAVEGA.

Dando prosseguimento, os conceitos obtidos nessa primeira etapa (407 termos indexadores) foram utilizados para realizar a segunda parte dessa experiência, que será descrita no subcapítulo 5.5.

## **5.5 Apresentação da implementação tecnológica da TAFNAVEGA**

O desenvolvimento desta segunda etapa da experiência tecnológica teve por objetivo a visualização da interface de navegação (digital) pela estrutura facetada da TAFNAVEGA, criada para a BDTD da ECI-UFMG. Essa segunda etapa foi desenvolvida a partir dos 407 termos indexadores, obtidos na primeira etapa que está descrita no subcapítulo 5.4, anterior.

Esse foi um trabalho conjunto com Pontes<sup>80</sup>, também integrante do grupo de pesquisa “Protótipo Mapa Hipertextual” (MHTX), realizado dentro do laboratório experimental do protótipo de mesmo nome (LIMA, 2004a), buscando desenvolver novos mecanismos a fim de melhorar a recuperação da informação na biblioteca digital experimental. Esse grupo de pesquisa é coordenado pela professora Gercina Lima, orientadora desta pesquisa.

Esta etapa da experiência foi desenvolvida em uma base de dados relacional, o *Microsoft Office Access*, também conhecido por *MSAccess*, foi criado pelo fabricante

---

<sup>80</sup> Flávio Pontes é estudante de doutorado do PPGCI-ECI/UFMG e desenvolveu, em seu projeto de qualificação intitulado “Organização do conhecimento em bibliotecas digitais de teses e dissertações: uma abordagem baseada na classificação facetada”, um aplicativo para apresentar sua proposta que tem por objetivo implementar e avaliar mecanismos para a representação e organização do acervo de uma biblioteca digital de teses e dissertações, e a criação de um protótipo, com base na taxonomia facetada criada nesta presente pesquisa.

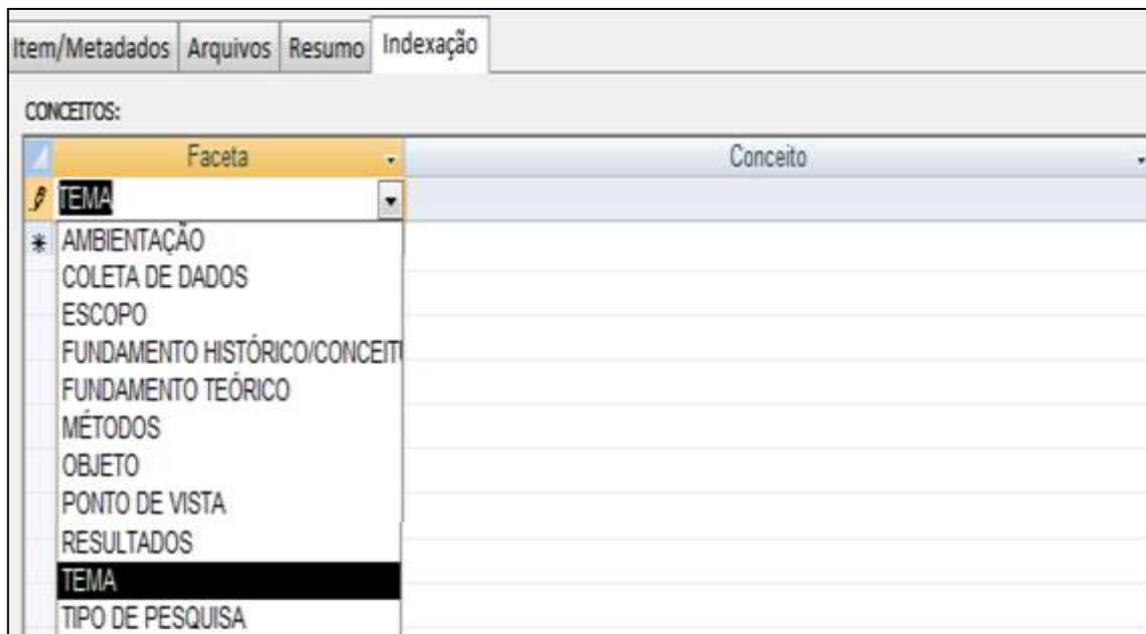
*Microsoft*. Ele é um *software* de gerenciamento de banco de dados, que está incluído no pacote do *Microsoft Office Professional*. É importante destacar que essa experiência se baseia no princípio de sistemas facetados apresentados por Sacco e Tzitzicas (2009), cujo modelo se fundamenta em um padrão genérico de conhecimento e em uma classificação multidimensional de objetos (itens de informação), usado para a navegação em bases de dados, com foco no acesso e nas necessidades do usuário. Dessa maneira, poderia ter sido desenvolvida em qualquer outra ferramenta ou *software* que permite criar uma base relacional.

O *MSAccess* é um programa que oferece um modelo relacional bastante simples, que consiste em diferentes tabelas, estruturadas por campos (colunas), e, em cada tabela, há distintos registros (linhas) de informação, que podem relacionar-se, através de campos-chave, com as diferentes tabelas.

No desenvolvimento dessa experiência, foram criadas tabelas diferentes para cada tipo de informação, para armazenar os dados. Os procedimentos para a estruturação do banco de dados seguiram as recomendações do fabricante, com as seguintes etapas:

1) Determinação das tabelas: foi criada uma tabela para cada uma das dez categorias fundamentais temáticas (FIG. 9);

FIGURA 9 – Tabela de categorias fundamentais



Fonte: Tela da implementação tecnológica no *MSAccess*.

2) Determinação dos campos necessários: foram atribuídos diferentes campos, independentes, para cada faceta encontrada (setas verdes na FIG. 10);

3) Determinação dos relacionamentos: para efeito desta experiência, todos os relacionamentos se basearam no relacionamento do tipo “é um” (setas vermelhas na FIG.

10), uma vez que, no escopo desta pesquisa, esse tipo de atribuição foi suficiente para demonstrar a experiência.

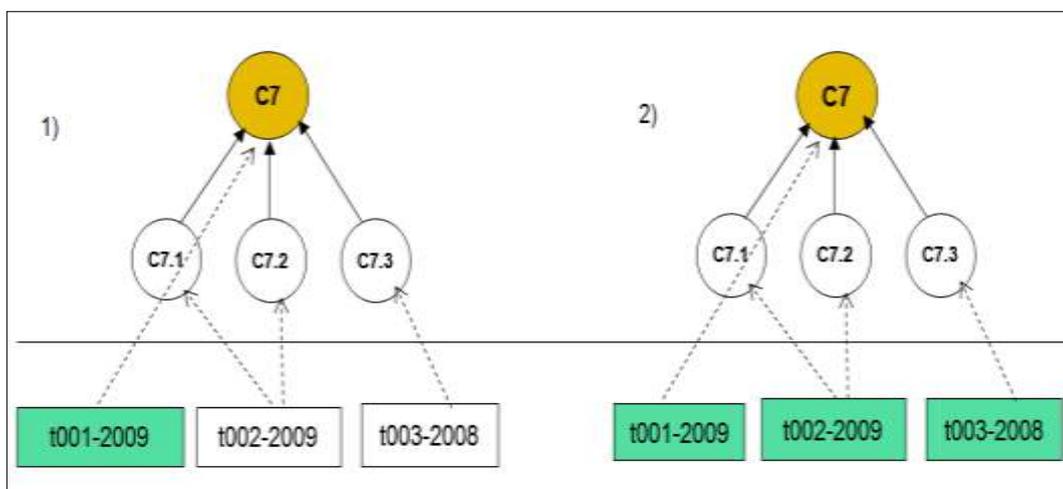
FIGURA 10 – Campos e relacionamentos entre facetas

The screenshot displays the 'TAXONOMIA: Menu Principal' interface. At the top, there are menu options: 'Finalizar o Sistema', 'Taxonomia', 'Itens', 'Relatórios', and 'Sobre o Sistema'. Below these are sub-menus: 'Sistema', 'Taxonomia', 'Itens Bibliográficos', 'Relatórios', and 'Ajuda'. The main area is titled 'Itens Bibliográficos' and 'Cadastro de Conceitos'. It features two relationship tables: 'RELAÇÕES (Botton-Up):' and 'RELAÇÕES (Top-Down):'. The 'Botton-Up' table has columns for 'Relação', 'Conceito', and 'Descrição', with a dropdown menu showing 'É um(a)'. The 'Top-Down' table has columns for 'Conceito', 'Relação', and 'Descrição'. A search panel on the right, titled 'Busca:', contains a list of concepts, with 'Análise documentária' highlighted. Below the search panel are buttons for 'Adicionar', 'Excluir', and 'Fechar'. The interface also shows a 'Conceito' field with the value 'Análise documentária' and an ID of '489'. A 'TERMOS:' section is also visible.

Fonte: Tela da implementação tecnológica no MSAccess.

4) Refinamento da estrutura do banco de dados: os dados das tabelas foram inseridos aos poucos, possibilitando verificar a estabilidade e o funcionamento da estrutura que estava sendo desenvolvida, com constantes averiguações. Esse refinamento também foi conseguido a partir da extração de relatórios, que permitiram verificar a condição de cada tabela e de cada elemento dentro dela.

Em seguida, descreve-se a modelagem informacional criada e, logo depois, a visualização de uma busca através da TAFNAVEGA, por uma sequência de figuras.

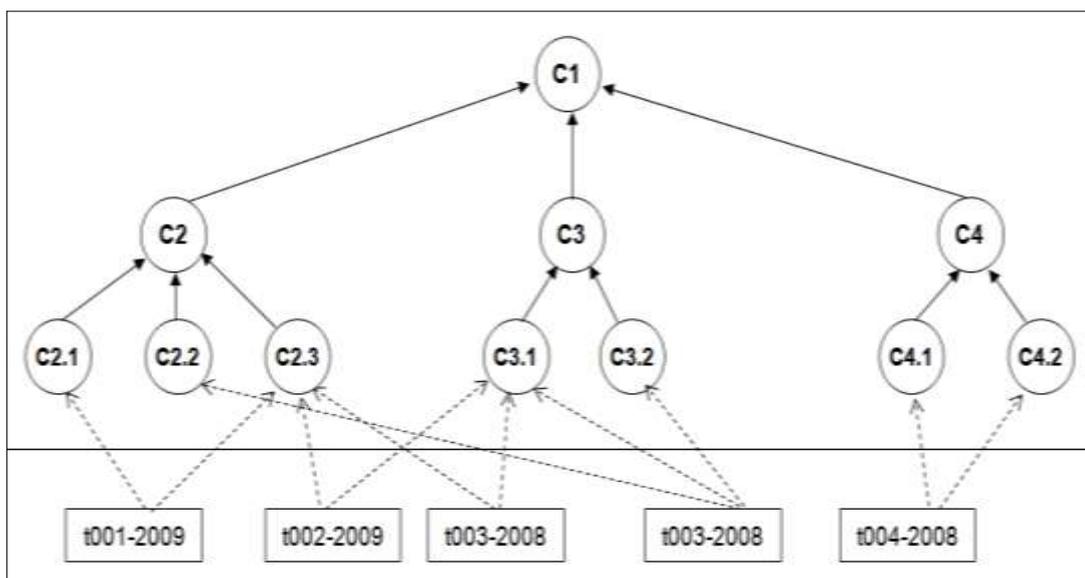
FIGURA 11 – Relações entre documentos do *corpus*

Fonte: Adaptado de Sacco e Tzitzicas (2009, p. 22 e 32).

Observando a FIG. 11, à esquerda (n.1) tem-se a relação isolada do documento t001-2009. Isto é, o documento foi representado sob a categoria C7. Já à direita (2), tem-se a relação do documento t001-2009 com os diferentes documentos na BDTD que também foram representados sob a categoria C7 em diferentes facetas (C7.1, C7.2 e C7.3). Dessa forma, ao escolher a categoria C7, o usuário terá à disposição todos os documentos representados sob essa categoria.

Com a utilização da taxonomia facetada para a representação do teor dos documentos do *corpus* (listados no APÊNDICE A, p.172), esses conteúdos tornam-se interrelacionados dentro da estrutura facetada, de forma multidimensional. A FIG. 12 mostra uma interrelação hipotética entre diferentes categorias.

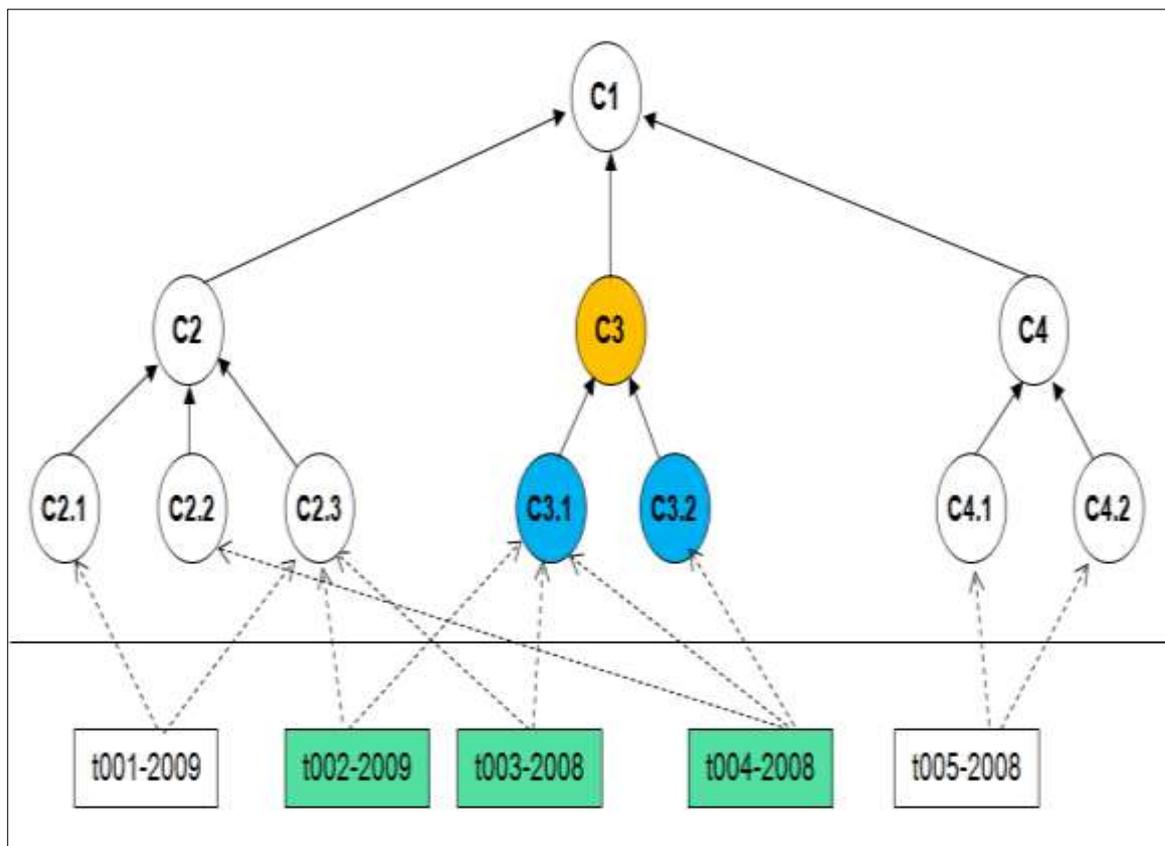
FIGURA 12 – Relação entre categorias



Fonte: Adaptado de Sacco e Tzitzicas (2009, p. 22 e 32).

Conforme a FIG. 12, através da navegação na taxonomia facetada, o usuário pode explorar todo o conjunto de categorias. Em uma situação, por exemplo, em que o usuário selecione a categoria C3 para a busca, a FIG. 13 mostra o que o sistema apresenta como resultado para essa busca.

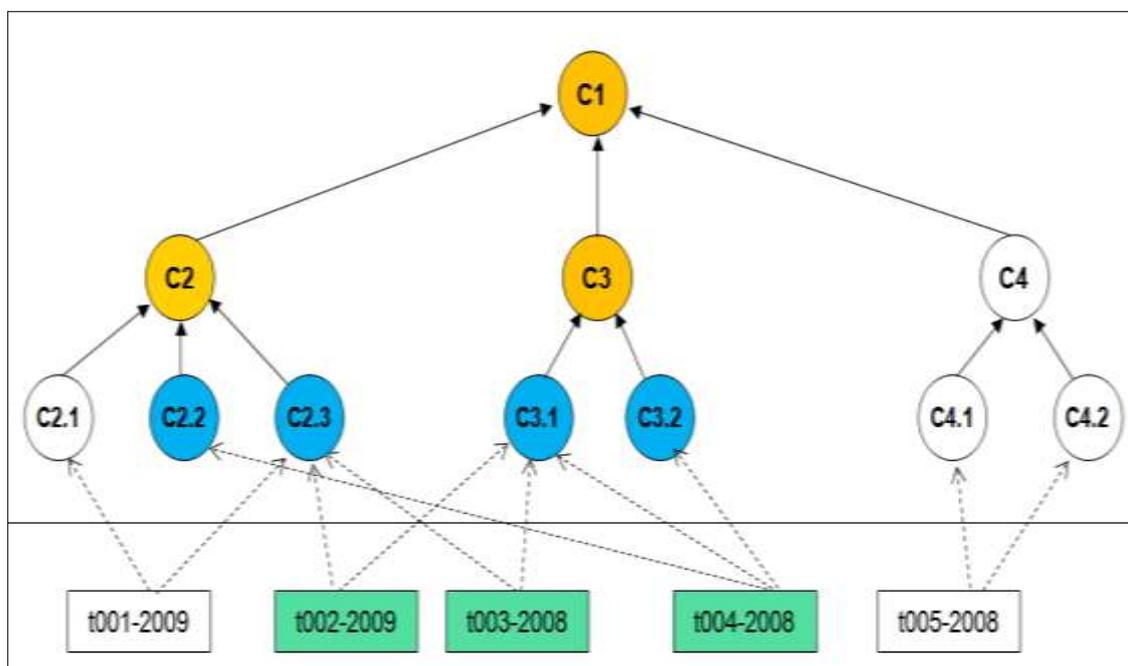
FIGURA 13 – Relação direta entre categorias e documentos



Fonte: Adaptado de Sacco e Tzitzicas (2009, p. 22 e 32).

A FIG. 13 mostra que a categoria C3 possui duas facetas, que estão diretamente relacionadas com os documentos t002-2009, t003-2008 e t004-2008. Entretanto, como os documentos são representados sob todas as categorias da taxonomia facetada, o resultado para o usuário é expandido, permitindo que esse usuário possa explorar um conjunto maior de temas relacionados com os documentos recuperados, como é mostrado na FIG. 14.

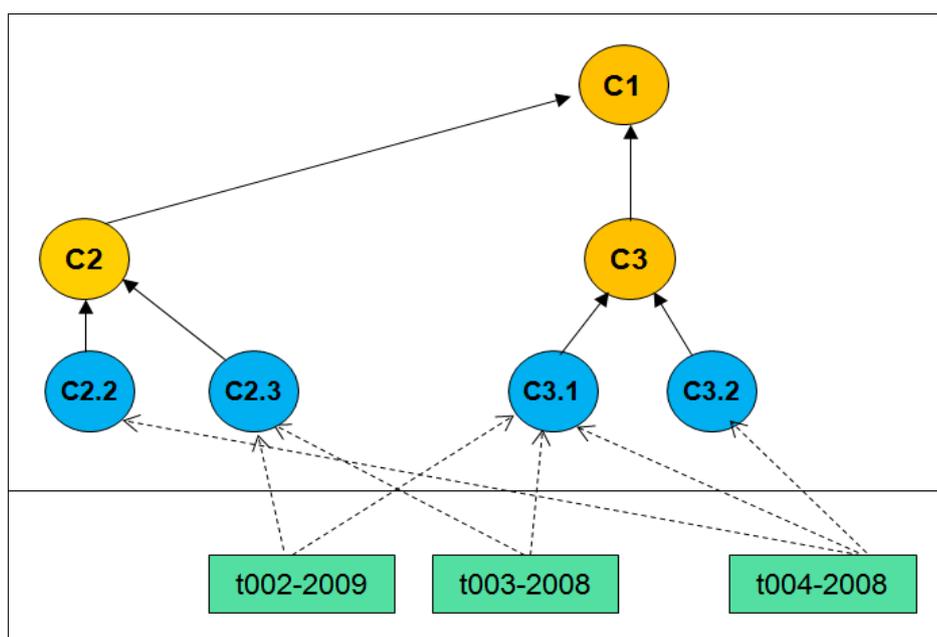
FIGURA 14 – Relação expandida entre categorias e documentos



Fonte: Adaptado de Sacco e Tzitzicas (2009, p. 22 e 32).

Observando a FIG. 14, nota-se que a taxonomia facetada permite que o usuário explore todas as categorias e facetas sob as quais os documentos (t002-2009, t003-2008 e t004-2008) foram representados. A FIG. 15 mostra como é apresentado o resultado ao usuário.

FIGURA 15 – Resultado das interrelações disponíveis para recuperação



Fonte: Adaptado de Sacco e Tzitzicas (2009, p. 22 e 32).

Nota-se que somente é apresentado ao usuário o que efetivamente está interrelacionado e disponibilizado no banco de dados. Com isso, o usuário tem à sua disposição a possibilidade de combinar suas buscas, sem encontrar como resultado uma resposta vazia, pois somente são apresentadas a ele as opções que estão disponíveis na BDTD.

Concluída a demonstração da modelagem informacional criada, a seguir apresenta-se a visualização de uma busca através da TAFNAVEGA.

### Demonstração empírica da implementação tecnológica

Abaixo se apresenta um exemplo hipotético de busca a partir da “taxonomia facetada navegacional” (TAFNAVEGA).

FIGURA 16 – Busca usando a TAFNAVEGA (1)

1) Mostra o “universo OUI”, referente ao *corpus* (41 documentos);

2) Mostra as categorias fundamentais e o número de seus respectivos elementos;

3) Escolhi a categoria “tipo de pesquisa”, que contém 19 documentos, que estão qualificados pelos 10 tipos de pesquisa disponíveis para busca (comparativa, descritiva, documental, empírica, estudo de caso, experimental, qualitativa, quantitativa, exploratória e teórica), assim como o número de documentos.

Fonte: Tela da implementação tecnológica no MSAccess.

Nessa busca, conforme FIG. 16, o pesquisador tem 41 documentos disponíveis, que são todo o universo referente ao *corpus* (1). O universo está classificado sob as 10 categorias da TAFNAVEGA (2), estabelecidas a partir da estrutura textual dos trabalhos

acadêmicos do tipo teses e dissertações. Dentre as diferentes opções, o pesquisador escolhe a categoria “tipo de pesquisa”, que lhe oferece 10 qualificações diferentes para escolha. Se o pesquisador optar pela faceta “comparativa”, o resultado será como mostra a FIG.17.

FIGURA 17 – Busca usando a TAFNAVEGA (2)

The screenshot displays the TAFNAVEGA search interface. On the left, a hierarchical tree of search options is shown, with a green arrow pointing to 'UNIVERSO (5)' and a red arrow pointing to 'Comparativa (5)'. A bracket on the right side of the tree is labeled 'Opções de busca'. The right side of the interface shows a list of search results under the heading 'Resultados:'. The results are as follows:

Resultado	Descrição
2112	A interação dos usuários da UFMG com o catálogo on-line do Sistema Pergam
2101	Comparação entre termos de indexação e palavras dos títulos dos artigos do
1282	Estudo comparativo entre interfaces hipertextuais de softwares para a repre
18	Tesouros e ontologias: estudo de definições presentes na literatura das área
4424	Um estudo sobre o uso de periódicos eletrônicos: O Portal De Periódicos CAPI

Below the results list, there is a large white arrow pointing upwards with the text 'Resultado da busca' inside it. At the bottom of the interface, there is a 'Histórico de Navegação:' section showing '(TIPO DE PESQUISA=>Comparativa)' and a '5 documento(s)' indicator.

Fonte: Tela da implementação tecnológica no MSAccess.

Conforme se observa na FIG. 17, o universo (seta verde) agora são os cinco documentos que foram classificados sob a faceta “comparativa” (seta vermelha). O sistema recupera e apresenta os cinco documentos disponíveis (resultado da busca) e o pesquisador ainda pode associar sua busca a outros “tipos de pesquisa” (descritiva, qualitativa e quantitativa). Além disso, há ainda a possibilidade de escolher entre todas as opções de busca dentro da chave.

Vale ressaltar, que todas as 10 categorias deveriam oferecer pelo menos cinco opções de documentos, uma vez que é o número de documentos disponível. Entretanto,

algumas categorias apresentam opções menores (ex. ambientação, coleta de dados e métodos) devido à incompletude dos resumos informativos utilizados na extração de conceitos nesta experiência (mencionada no item 5.3.4). Contudo, numa situação real de uso deste protótipo, todos os documentos são classificados sob as 10 categorias. Com isso, é possível buscar os conteúdos dos documentos através da navegação, facilitando o acesso aos resultados detalhados das pesquisas.

Ainda no caso hipotético da busca acima, se o pesquisador quiser recuperar apenas os documentos classificados sob a faceta “comparativa”, mas somente aqueles nos quais a pesquisa tenha sido realizada em uma abordagem qualitativa, isso seria facilmente recuperado, como mostra a FIG. 18.

FIGURA 18 – Busca usando a TAFNAVEGA (3)

The screenshot displays the TAFNAVEGA search interface. On the left, a hierarchical tree of search options is shown, with a green arrow pointing to 'UNIVERSO (2)' and a red arrow pointing to 'Qualitativa (2)'. A bracket on the right side of the tree is labeled 'Opções de busca'. The search results on the right are listed under 'Resultados:' and include two entries: '2112 A interação dos usuários da UFMG com o catálogo on-line do Sistema Pergam' and '18 Tesouros e ontologias: estudo de definições presentes na literatura das áreas'. A large white arrow points to a box labeled 'Resultado da busca'. At the bottom, the 'Histórico de Navegação:' shows '(TIPO DE PESQUISA=>Comparativa)' and '(TIPO DE PESQUISA=>Qualitativa)'. The bottom right corner indicates '2 documento(s)'. Navigation buttons like '+', '-', '<=>', 'Or', 'And', and 'Cancelar' are visible.

Fonte: Tela da implementação tecnológica no MSAccess.

Nota-se que o universo, agora, possui somente dois documentos (seta verde). Analisando o conjunto percebe-se, por exemplo, que esses dois documentos (resultado da busca) possuem “tipo de pesquisa” “comparativa”, que também são “qualitativa” (seta vermelha). Nota-se que, novamente, que o pesquisador tem à sua disposição outras opções

de busca (dentro da chave), podendo restringir sua busca (selecionando a faceta “descritiva” ou “quantitativa”) ou, ainda, associá-la com as outras categorias disponíveis.

No presente capítulo foi relatado o desenvolvimento da taxonomia facetada navegacional (TAFNAVEGA), criada com base na estrutura textual de documentos do tipo teses e dissertações. Esse relato incluiu a identificação do domínio, do usuário e de suas necessidades de informação, assim como as descrições dos procedimentos utilizados, a apresentação e a análise dos resultados, que demonstraram como esse mecanismo auxilia o usuário na busca por informações sobre os resultados das pesquisas disponibilizadas na BDTD do PPGCI-UFMG.

A seguir, será apresentado um breve relato sobre o resultado encontrado na comparação entre a aplicação da técnica da análise categorial temática (ACT) e a análise facetada (AF), discutida no Capítulo 4, subcapítulo 4.6.1, mostrando a relação existente entre as categorias fundamentais de Ranganathan, do CRG e as dez categorias fundamentais temáticas (CAFTE) determinadas pela aplicação da ACT neste Capítulo 5.

## 5.6 Paridade entre as categorias fundamentais de Ranganathan, CRG e da TAFNAVEGA

Neste subcapítulo analisam-se as similaridades e correspondências ressaltadas entre as categorias fundamentais temáticas (CAFTE) da taxonomia facetada navegacional (TAFNAVEGA), encontradas com a aplicação da técnica de análise categorial temática e as categorias fundamentais de Ranganathan e do CRG. Adverte-se, entretanto, que esta é uma visão particular da preponente, conforme QUADRO 13.

QUADRO 13 – Paridade entre as categorias de Ranganathan, CRG e CAFTE

RANGANATHAN	CRG	ACT (CAFTE)
Personalidade	Tipos de Produto Final	Objeto Resultados Tema
Matéria	Partes Materiais Propriedades	Escopo Fundamento histórico/contextual Tipo de pesquisa
Energia	Processos Operações Agentes	Coleta de dados Fundamento teórico Métodos
Espaço	Espaço	Ambientação
Tempo	Tempo	(já há busca por data na BDTD)

Fonte: Elaborado pela autora.

Em seus estudos, Ranganathan descreve que a categoria mais difícil de identificar é a categoria Personalidade. No QUADRO 13, nota-se que as CAFTE Objeto, Resultados e Tema podem ser considerados Personalidade, pois a primeira configura o objeto empírico utilizado, a segunda, o assunto tratado e, a terceira, um produto final criado a partir dos procedimentos oriundos do campo “Energia”. É possível perceber, ainda, que o Objeto e o Tema podem ser considerados “tipos de produtos final”, pois exercem influência no modo de conduzir a pesquisa, e o item “Resultados” refere-se a um produto obtido em decorrência de tudo que foi desenvolvido na investigação, tornando-se uma Personalidade do estudo. Quanto aos outros, não é difícil acomodar, porque no item “Matéria” são inseridos os elementos que caracterizam a pesquisa e o campo “Energia” é composto por tudo que é utilizado como “ações” no decorrer da pesquisa. O Espaço (lugar), pela sua obviedade, dispensa explicações e o Tempo está coberto pelas opções de busca já existentes na BDTD.

Por ocasião desses apontamentos, é concebível a afirmação de que as categorias não surgem por encanto, pois são consequência de um olhar específico do pesquisador e pela abordagem ao texto. Assim, determinar o objetivo (o que) e a finalidade (para quem) é essencial em um trabalho de categorização de um domínio, assim como para o fornecimento de um serviço de informação. Por esse motivo, acredita-se que o entendimento do texto, o que o texto “fala”, está induzido pelo olhar lançado a ele. Nesse sentido, o olhar desta pesquisa recaiu sobre a estrutura textual dos trabalhos acadêmicos do tipo teses e dissertações.

Nesse percurso, ainda é significativo destacar que a sequência dos termos dentro de cada categoria respeitará a ordem alfabética, uma vez que a ordem não tem qualquer influência nos resultados da recuperação da informação em ambiente digital. A listagem alfabética facilitará a busca e a navegação pelo usuário, pois padroniza a ordem de entrada e, assim, a forma apresentada na estrutura. Além disso, auxiliará aquele usuário que tenha menor conhecimento sobre o tema buscado. Com isso, atende-se ao requisito de sequência relevante e útil do CRG e de Ranganathan. Além disso, obedece também à lei da imparcialidade (RANGANATHAN, 1967, p.125), não beneficiando, arbitrariamente, qualquer outra ordem de importância.

Finalizando este capítulo de descrição de procedimentos e de análise dos resultados, passa-se à apresentação das considerações finais desta pesquisa e as sugestões de trabalhos futuros.

## 6 CONSIDERAÇÕES FINAIS

A presente pesquisa iniciou-se com a perspectiva de priorizar o atendimento às necessidades de informação do usuário da biblioteca digital de teses e dissertações (BDTD), com recorte nos trabalhos defendidos na linha de pesquisa de Organização e Uso da Informação (OUI), no Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação (ECI) da UFMG. Para tanto, foi necessário identificar as características do domínio, do usuário e do que precisa em termos informacionais. Nesse sentido, o uso da teoria da análise de domínio (TAD) mostrou-se adequado para indicar a comunidade de discurso que produz e usa o tipo de informação contida nesses trabalhos acadêmicos. A abordagem sócio-cognitiva da TAD, também utilizada para expor o usuário e suas necessidades de informação, levou ao entendimento pragmático de que o fluxo da informação produzida nessa comunidade de discurso, que é a comunidade científica, tem um sentido prático de produção e de uso. A partir da aplicação da TAD, definiu-se que o produtor e o usuário da BDTD são os pesquisadores (ou cientistas) em exercício de atividade profissional.

Dentro desse contexto, entende-se que todo pesquisador é membro da comunidade científica. Ele busca a aprovação de seus pares e segue procedimentos metodológicos ordenados e sistemáticos na produção de seus trabalhos acadêmicos. Toda pesquisa é construída a partir de um conhecimento anterior e, como consequência, a facilidade de acesso, a disponibilidade e a visibilidade desse conhecimento anterior torna a produção científica mais eficiente e menos redundante. Acredita-se que a ciência avança na medida em que há um processo de construção de conhecimento, maximizado pela comunicação do conhecimento já edificado. Sendo assim, o texto produzido pelo cientista expressa um significado e um objetivo comunicativo dentro de sua comunidade discursiva, que, por sua vez, exerce certo controle sobre os procedimentos que são seguidos por esse cientista na produção do seu texto científico.

A revisão de literatura realizada neste estudo demonstrou que a abordagem da navegação facetada é empregada desde o final da década de 90, como solução de exploração, busca e recuperação de informações na *Web* utilizando, principalmente, a determinação de categorias de forma automatizada. Essa característica muitas vezes se deve ao fato de serem projetos oriundos de áreas como a Ciência da Computação, Engenharia de *Software* e Informática, campos nos quais há forte apelo para soluções tecnológicas automáticas. Entretanto, encontram-se propostas, como a do FLAMENCO, cujos metadados categoriais são atribuídos de forma não automática (manual), uma vez que é uma iniciativa para navegação em ambiente *Web*, porém, para coleções mais restritas.

No âmbito da Ciência da Informação (CI) brasileira, a literatura especializada aborda a taxonomia como instrumento de representação para mapear uma área de conhecimento, geralmente empregado na entrada de dados nos sistemas de informação, para a organização e a classificação de registros do conhecimento. A revisão de literatura indicou que os estudos realizados a partir dos princípios do método da análise facetada, na elaboração de taxonomias ou classificações facetadas, enfatizam seu uso mais como modelo da lógica analítico-sintética para a descrição dos conteúdos, visando à recuperação da informação, do que como um mecanismo de busca. Ademais, a taxonomia facetada pode ser considerada uma cadeia de relações conceituais, que tem também potencial valor para uso na saída dos sistemas de informação, como um mecanismo para navegação, facilitando a busca e a recuperação de informações em bibliotecas digitais

Desde a década de 90 nota-se o avanço nos serviços oferecidos pelas bibliotecas digitais e nas formas de gerenciamento do recurso informacional, tentando se antecipar às necessidades dos usuários. Nas instituições de ensino superior, o aumento no número de bibliotecas digitais de teses e dissertações (BDTD) contribuiu para a disponibilização e o acesso a informações acadêmicas em formato digital. Como consequência dessas implementações, espera-se a divulgação de resultados de pesquisas já concluídas, evitando-se esforços replicados e a perda de recursos intelectuais. Todavia, no Brasil ainda não se notam esforços no sentido de fazer com que a busca e a recuperação de informações em BDTDs extrapolem as opções oferecidas por uma biblioteca tradicional.

Geralmente, as BDTDs oferecem dois tipos de opção de buscas: a “busca simples” e a “busca avançada”. A opção da “busca simples” apresenta problemas relativos a resultados com ambiguidades. Por outro lado, na opção de “busca avançada” os usuários se deparam com barreiras tecnológicas e psicológicas, pois consideram esse tipo de busca de difícil compreensão. Além disso, o texto completo das teses e dissertações é disponibilizado em formato *Portable Document Format* (PDF), o que impede que o sistema varra o seu conteúdo durante as buscas e recupere seu conteúdo.

Ademais, a literatura especializada e os resultados da presente pesquisa indicam problemas na atribuição insuficiente de palavras-chave em trabalhos acadêmicos e da elaboração incompleta de resumos informativos. Isso acarreta perda de informações, prejudicando a qualidade da recuperação de informações.

Partindo dos problemas verificados nos resumos, o primeiro resultado computado na presente pesquisa não fazia parte do escopo de objetivos propostos. Trata-se de uma tabela para elaboração de resumos informativos, que pode ser utilizada para orientar e auxiliar os autores e/ou indexadores na preparação desses resumos. Para isso, basta que se extraiam as informações a partir das instruções da tabela e, depois, constroi-se um texto sem tópicos e em parágrafo único, como exigido para um resumo informativo. O

uso dessa tabela pode vir a melhorar a qualidade dos resumos disponibilizados na BDTD e, conseqüentemente, aprimorar a recuperação das informações referentes aos documentos.

Para o desenvolvimento desta dissertação partiu-se do pressuposto de que era possível oferecer ao usuário da BDTD uma opção de busca fácil de usar e interpretar, com informações organizadas. Além disso, a expectativa recaiu sobre o objetivo de dar maior visibilidade ao conhecimento científico acumulado nas teses e dissertações, oferecendo a possibilidade de refinamento da busca do usuário, sem que ele fosse sobrecarregado com informações.

A taxonomia facetada para navegação (TAFNAVEGA) apresentada aqui é constituída por um conjunto de dez categorias fundamentais temáticas (CAFTE) e pelos conceitos (representados pelos termos) extraídos pela análise conceitual nos documentos do *corpus* desta pesquisa, que compuseram as facetas da estrutura facetada da taxonomia. A TAFNAVEGA foi desenvolvida dentro dos princípios da teoria da análise facetada (TAF) dos estudos de Ranganathan (1957) e das recomendações do *Classification Research Group* (CRG), que estão indicadas na simplificação apresentada por Spiteri (1998). A validação do conjunto CAFTE foi realizada a partir do método da Análise de Conteúdo, com a aplicação da técnica da análise categorial temática (ACT). O uso da ACT mostrou-se apropriada nesse procedimento, uma vez que o confronto realizado entre essa técnica e os fundamentos da TAF determinou o alto grau de similaridade entre elas. Essa similaridade é o que se denomina de “similaridade funcional e de execução”. Na similaridade funcional a TAF e a ACT comungam a função de potencializar a estruturação e ordenação das informações do conteúdo das teses e dissertações. Já a similaridade de execução está relacionada aos processos metodológicos lógicos executados para se chegar às categorias fundamentais, conforme indicado no QUADRO 13.

As teses e dissertações são produzidas em linguagem natural e de forma semiestruturada, o que dificulta sua análise conceitual. A representação desse tipo de documento por meio do conjunto das CAFTE organizou e ordenou o conteúdo textual dos trabalhos acadêmicos, uma vez que estruturou esse tipo de documento em categorias, respeitando-se a sua estrutura textual de produção. Dessa maneira, o conjunto das CAFTE mostrou-se adequada, pois possibilitou o acesso a resultados detalhados das pesquisas tais como temas, métodos, instrumentos de coleta de dados e fundamentos teórico-metodológicos. A estrutura facetada taxonômica agrupou os termos indexadores (extraídos dos documentos através do algoritmo) que compartilham características semelhantes, transformando-os em facetas e subfacetas navegáveis. Com isso, alcançou-se o objetivo de oferecer informações organizadas, pois elas foram arranjadas de uma forma fácil de apreender.

Além disso, o uso do vocabulário controlado para a tradução dos termos extraídos dos documentos, ainda que não tenha sido criado especificamente para o contexto desta pesquisa, permitiu compatibilizar as distintas “linguagens naturais” (o vocabulário de cada autor) utilizadas nos documentos do *corpus* da pesquisa. Essa compatibilidade é importante para oferecer uma estrutura temática única para o domínio, o que permitiu interligar os diferentes documentos no sistema.

A visibilidade do conteúdo dos documentos foi obtida através do mapeamento do conteúdo dos trabalhos acadêmicos, o que foi demonstrado pela experiência realizada com a TAFNAVEGA, no processo de extração de conceitos dos documentos. Nessa experiência, pela análise dos resultados, foi possível inferir, ainda que parcialmente, que as pesquisas na linha OUI priorizam os temas relacionados com a “organização da informação”, em detrimento aos temas referentes ao “uso da informação”. Foi possível, também, descobrir que a maior parte das investigações, nessa linha de pesquisa, baseia-se no método analítico-sintético e na classificação facetada. Esses apontamentos evidenciaram, dentro dos limites estabelecidos neste estudo, o “estado da arte” dos trabalhos acadêmicos da linha de pesquisa OUI. Acredita-se que o uso da TAFNAVEGA pode ser estendido a outras linhas de pesquisas e a outras áreas do conhecimento, ampliando o campo de aplicação desse mecanismo.

A matriz categorial para trabalhos acadêmicos (MCTCA) é um produto que foi desenvolvido neste estudo, a partir do Modelo de Leitura Documentária de Fujita e Rubi (2006). O MCTCA é um algoritmo de alto nível, para uso humano, composto pelo conjunto das CAFTE (primeira coluna) e por uma sequência lógica e sistemática de procedimentos. A aplicação desse algoritmo culmina na obtenção de metadados, que são os termos indexadores obtidos pela análise conceitual nos documentos, cujos termos alimentam a TAFNAVEGA. Com sua utilização na análise dos documentos, o usuário terá acesso ao documento que efetivamente possui o conteúdo que o termo indexado representa e não simplesmente a um documento que contém (cita) tal termo.

A abordagem da navegação facetada criada na presente pesquisa segue o princípio da árvore *baniana*, como um organismo vivo, pois tem uma perspectiva multidimensional, quando atribui a um mesmo documento mais de uma dimensão (por ex.: tema, ambientação, tipo de pesquisa), e outro sentido multidimensional, quando um mesmo termo na estrutura leva a mais de um documento (por ex.: o termo “comparativa” leva a todos os documentos que utilizaram esse tipo de pesquisa). Essa característica torna a busca facetada dinâmica, com uma interface interativa, na qual o usuário pode visualizar e explorar recursos informacionais, descobrindo o universo do conteúdo em questão e refinando sua busca progressivamente, podendo, ainda, combinar uma ou mais dimensões. Desse modo, o usuário sabe, exatamente, quais informações estão disponíveis no banco de

dados, e isso evita um resultado de busca vazio, minimiza a sobrecarga de informação e, sobretudo, melhora a recuperação de informações. Essa melhora é obtida, porque o usuário aproveitará os recursos de uma busca avançada, sem a necessidade de indicar as combinações lógicas, pois isso já se apresenta, semanticamente, implícito na estrutura facetada.

Ainda é importante destacar que esse mecanismo suporta novos termos sem que perca a característica de ser uma estrutura consistente, pois o sistema facetado é hospitaleiro e pode, e deve, ser sempre atualizado.

Ao longo desta pesquisa de mestrado, algumas questões enfrentadas mostraram que há possibilidade de estudos futuros, tal como a realização de cruzamento de dados entre metodologias e teorias utilizadas nas pesquisas, incluindo-se, também, as outras duas linhas de pesquisa da ECI-UFMG, enriquecendo a percepção sobre a natureza das práticas de pesquisa realizadas nessa Escola. Outro ponto sugerido é o estudo do uso de um vocabulário controlado na indexação de documentos acadêmicos, visando determinar se sua aplicação pode minimizar o problema de ambiguidade terminológica percebida entre o sistema e os diferentes atores, produtores de informação. Evidenciou-se, ainda, a necessidade da criação de uma política de indexação que possa determinar o grau de exaustividade e especificidade para a indexação de documentos do tipo teses e dissertações, como forma de padronizar os termos indexadores utilizados em sua descrição temática. Além disso, indica-se verificar a viabilidade da criação de um aplicativo tecnológico capaz de padronizar a entrada de dados (metadados) conforme a matriz categorial temática para trabalhos acadêmicos desenvolvida nesta pesquisa.

Finalmente, conclui-se que os pressupostos foram validados e os objetivos alcançados, ao mesmo tempo que considera-se importante trazer o tema da abordagem da navegação facetada, que é ancorada nos princípios da classificação facetada, para o seio da Ciência da Informação brasileira, especificamente para o campo da Organização e Uso da Informação (OUI), tendo em vista que o objeto já é bastante investigado na literatura estrangeira. Na sociedade contemporânea, a informação é primordial para a criação de novos conhecimentos, e para isso é preciso estabelecer a transferência dessa informação para a comunidade de usuários que a busca. Ao propor o mecanismo de acesso – a taxonomia facetada navegacional – auxilia-se nessa busca e recuperação, pois agrega-se valor à informação ao oferecê-la de forma estruturada e ordenada. Avalia-se que essa é a maior contribuição da presente pesquisa: a demonstração do papel da taxonomia facetada para a visualização, exploração e obtenção de informações relevantes dos resultados das pesquisas à comunidade acadêmica. Nesse sentido, o recurso informacional que está obscuro e pulverizado no conteúdo textual dos documentos, tais como teorias, ferramentas, métodos e técnicas, pode ser mais facilmente comunicado, o que permitirá um avanço na

área da Ciência da Informação, sobretudo no sentido de ajudar a perceber “o estado da arte” de um domínio investigado. Atualmente, as teses e dissertações já são inseridas no sistema de forma manual, e acoplar o aplicativo ao sistema não representa problema, pois ele está sendo desenvolvido em linguagem compatível com a linguagem empregada pela BDTD. Dessa forma, ao inserir os documentos na base de dados, todos os campos do conjunto das dez categorias da TAFNAVEGA podem ser preenchidos, o que permitirá interrelacionar documentos e informações, auxiliando o acesso, a recuperação e a reutilização dos resultados detalhados das pesquisas.

À luz dessas considerações, entende-se que os resultados encontrados neste estudo podem impactar positivamente o processo de comunicação científica a partir do acesso às informações contidas nas teses e dissertações e disponibilizadas na BDTD.

## REFERÊNCIAS

AAKER, D.; KUMAR, V.; DAY, G. S. *Marketing research*. Hoboken, NJ: Wiley, 1995.

AHLBERG, C.; SHNEIDERMAN, B. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS: Celebrating Interdependence, Boston, MA, USA, Apr. 24-28, 1994. *Proceedings...* New York: ACM, 1994. p. 313-317. Disponível em: <<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.123.805>>. Acesso em: 12 mar. 2011.

ALBRETCHTSEN, H. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, London, v.18, n. 4, p. 219-24, 1993.

ALVARADO, Rubén U.; OLIVEIRA, Marlene. A Comunidade científica da Biblioteconomia e Ciência da Informação brasileira. *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia*, v. 3, n. 2, 2008. Disponível em: <<http://revista.ibict.br/pbcib/index.php/pbcib/article/view/1051>>. Acesso em: 12 jan. 2011.

ALVES, Rachel C. V. *et al.* Ciência da Informação, Ciência da Computação e Recuperação da Informação: algumas considerações sobre os métodos e tecnologias da informação utilizados ao longo do tempo. *Revista Eletrônica Informação e Cognição*, Marília, SP, v.6, n.1, p.28-40, 2007.

ANDERSON, J. D.; HOFMANN, M. A. A fully faceted syntax for Library of Congress subject headings. *Cataloging & Classification Quarterly*, v. 43, n. 1, p. 7-38, 2006.

ARAÚJO, Carlos Alberto Ávila. *Análise temática da produção científica em comunicação no Brasil baseada em um sistema classificatório facetado*. 2005. 427f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

ARAÚJO, Samur F. C. de. *Explorator*: ferramenta para exploração de dados RDF baseado em uma interface de manipulação direta. 2009. 127f. Dissertação (Mestrado em Informática) – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009. Disponível em: <[http://www.maxwell.lambda.ele.puc-rio.br/Busca\\_etds.php?strSecao=resultado&nrSeq=13792@1](http://www.maxwell.lambda.ele.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=13792@1)>. Acesso em: 12 mar. 2011.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676. Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação*. Rio de Janeiro, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR10520. Informação e documentação: citações em documentos: apresentação*. Rio de Janeiro, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR12225. Informação e documentação: lombada: apresentação*. Rio de Janeiro, 2004.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR14724. Informação e documentação: trabalhos acadêmicos: apresentação*. Rio de Janeiro, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR14724. Informação e documentação: trabalhos acadêmicos: apresentação*. Rio de Janeiro, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR14724. Informação e documentação: trabalhos acadêmicos: apresentação*. Rio de Janeiro, 2011.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6023. Informação e documentação: referências: elaboração*. Rio de Janeiro, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6024. Informação e documentação: numeração progressiva das seções de um documento escrito: apresentação*. Rio de Janeiro, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6027. Informação e documentação: sumário: apresentação*. Rio de Janeiro, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6028. Informação e documentação: resumo: apresentação*. Rio de Janeiro, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR6034. Informação e documentação: índice: apresentação*. Rio de Janeiro, 2004.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. *Modern Information retrieval*. New York: Addison-Wesley, 1999.

BARDIN, Laurence. *Análise de conteúdo*. 4. ed. rev. e atual. Lisboa: Edições 70, 2009.

BARQUÍN, Beatriz A. R.; MOREIRO GONZÁLEZ, José Antonio; PINTO Adilson L. Construção de uma ontologia para sistemas de informação empresarial para a área de telecomunicações. *DataGramaZero – Rev. Ci. Inf.* Rio de Janeiro, v. 7, n. 2, abr. 2006.

BONILLA, Maeia Helena Silveira. *Escola aprendente: para além da sociedade da informação*. Rio de Janeiro: Quartet, 2005. p. 139.

BORKO, H. Toward a theory of indexing. *Information Processing and Management*, v. 13, p. 355-365, 1977.

BORKO, H.; CHATMAN, S. Criteria for acceptable abstracts: a survey of abstractor's instructions. *American Documentation*, v.14, p.149-160, 1963.

BORKO, Harold. Information science: what is it? *American Documentation*, Washington, v. 19, n. 1, p. 3-5, jan. 1968. Disponível em: <<http://www.cas.usf.edu/lis/lis6260/lectures/infosci.htm>>. Acesso em: 2 jul. 2007.

BRASIL. Ministério da Educação. CAPES. *Portaria 13*. 2006. Disponível em: <<http://www.abmes.org.br/legislacao/2006/portarias.asp>>. Acesso em: 22 mar. 2010.

BREEDING, Marshall. The birth of a new generation of library interfaces. *Computers in Libraries*, v. 27, n. 9, p. 34-37, Oct. 2007.

BROUGHTON, Vanda. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives*, v. 58, n. 1-2, p. 49-72, 2006.

BUCKLAND, Michel. Information as thing. *Journal of American Society of Information Science*, v. 42, n. 5, p. 351-360, 1991.

CAÇÃO, Rosário. *Análise Qualitativa de Dados com o NVIVO*. 31 de out. 2008. Disponível em: <<http://slidesha.re/11XyZq>>. Acesso em: 12 mar. 2011.

CAFÉ, Lígia Maria A.; BRATFISCH, Aline. Classificação analítico-sintética: reflexões teóricas e aplicações. *TransInformação*, Campinas, v. 19, n. 3, p. 237-250, set./dez. 2007.

CALHOUN, K. *The changing nature of the catalog and its integration with other discovery tools*. 2006. Disponível em: <[www.loc.gov/catdir/calhoun-report-final.pdf](http://www.loc.gov/catdir/calhoun-report-final.pdf)>. Acesso em: 12 mar. 2011.

CAMPOS, Claudinei José G. Método de análise de conteúdo: ferramenta para a análise de dados qualitativos no campo da saúde. *Rev. Bras. Enferm.*, Brasília, DF, v. 57, n. 5, p. 611-614, set./out. 2004.

CAMPOS, Maria Luiza de A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ci. Inf.* Brasília, DF, v. 33, n. 1, 2004. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/56/52>>. Acesso em: 22 abr. 2010.

CAMPOS, Maria Luiza de A.; GOMES, Hagar Espanha. Taxonomia e classificação: o princípio de categorização. *DataGramaZero – Rev. Ci. Inf.* Rio de Janeiro, v. 9, n. 4, ago. 2008. Disponível em: <[http://www.datagramazero.org.br/ago08/Art\\_01.htm](http://www.datagramazero.org.br/ago08/Art_01.htm)>. Acesso em: 21 jun. 2009.

CAPRA, R.; MARCHIONINI, G. The relation browser tool for faceted exploratory search. In: CONFERENCE ON DIGITAL LIBRARIES – JCDL'08, Pittsburg, Pennsylvania, 2008. *Proceedings...* Pittsburg, Pennsylvania, jun. 2008. p. 16-20. Disponível em: <<http://portal.acm.org/citation.cfm?id=1378967>>. Acesso em: 15 mar. 2011.

CAPURRO, R. Epistemologia e Ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., Belo Horizonte, 2003. *Anais...* Belo Horizonte: Escola de Ciência da Informação da UFMG, 2003.

CASE, D. O. *Looking for information: a survey of research on information seeking, needs and behavior*. San Diego, CA: Elsevier Academic, 2002. Disponível em: <<http://books.google.com.br/>>. Acesso em: 18 mar. 2011.

CONKLIN, Jeff. Hipertext: an introduction and survey. In: GREIF, Irene (Ed.). *Computer-supported cooperative work: a book of reading*. San Mateo, CA: Morgan Kaufmann, 1988. p. 423-476. Disponível em: <<http://books.google.com.br/books?id=4oVMJ1vi8IkC&printsec=frontcover#v=onepage&q&f=false>>. Acesso em: 14 mar. 2011.

CONWAY, Susan; SLIGAR, Char. Building taxonomies. In: \_\_\_\_\_. *Unlocking knowledge assets*. Redmont: Microsoft Press, 2002. cap. 6.

COSTA, Fernanda C. da; RAMOS, Luciene B. Análise facetada: em busca de uma classificação para o teatro. *Ponto de Acesso*, Salvador, v. 2, n. 3, p. 30-42, dez. 2008.

CUNHA, M. B. da; CAVALCANTI, Cordélia R. de O. *Dicionário de Biblioteconomia e Arquivologia*. Brasília, DF: Briquet de Lemos, 2008.

CUNHA, Murilo Bastos. Desafios na construção de uma biblioteca digital. *Ci. Inf.*, Brasília, v. 28, n. 3, p. 255-266, set./dez 1999.

CURRÁS, Emilia. *Tesouros: linguagens terminológicas*. Brasília: IBICT, 1995.

DAHLBERG, I. Teoria da classificação: ontem e hoje. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, Rio de Janeiro, 12-17 de setembro de 1972, Brasília. *Anais...* Brasília: IBICT/ABDF, v. 1, 1979. Disponível em:

<[http://www.conexaorio.com/bit/dahlbergteoria/dahlberg\\_teoriam.htm](http://www.conexaorio.com/bit/dahlbergteoria/dahlberg_teoriam.htm)>. Acesso em: 12 abr. 2010.

DAHLBERG, Ingetraut. Fundamentos teórico-conceituais da classificação. *R. Bibliotecon.*, Brasília, v. 6, n. 1, jan./jun. 1978. Disponível em: <<http://www.tempusactas.unb.br/index.php/RBB/article/viewFile/206/187>>. Acesso em: 22 mar. 2010.

DEMO, Pedro. *Metodologia do conhecimento científico*. São Paulo: Atlas, 2000.

DIAS, Célia da Consolação. *Análise do domínio organizacional na perspectiva arquivística: potencialidade no uso da Metodologia DIRKS – Designing and Implementing Redordkeeping Systems*. 2010. 333f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

DIAS, Eduardo W; NAVES, Madalena M. L. *Análise de assunto: teoria e prática*. Brasília: Thesaurus, 2007.

DOAN, K.; PLAISANT, C.; SHNEIDERMAN, B. Query previous in networked information systems. In: INTERNATIONAL FORUM ON RESEARCH AND TECHNOLOGY ADVANCES IN DIGITAL LIBRARIES, 3., 1995, Washington. *Proceedings...* Washington, DC: IEEE Computer Society, 1995. Disponível em: <<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.9728>>. Acesso em: 12 mar. 2011.

DODEBEI, Vera Lúcia D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto, 2002.

ELIEL, Regiane Alcântara. Institucionalização da ciência da informação no Brasil: estudo da convergência entre a produção científica e os marcos regulatórios da área. *TransInformação*, Campinas, v. 20, n. 3, p. 207-224, set./dez. 2008. Disponível em: <[http://biblioteca.ricesu.com.br/art\\_link.php?art\\_cod=6347](http://biblioteca.ricesu.com.br/art_link.php?art_cod=6347)>. Acesso em: 12 abr. 2010.

ENGLISH, Jennifer *et al.* *Flexible search and navigation using faceted metadata*. The Flamenco Search Interface Project, National Science Foundation. [Paper]. 2002. Disponível em: <[flamenco.berkeley.edu/papers/flamenco02.pdf](http://flamenco.berkeley.edu/papers/flamenco02.pdf)>. Acesso em: 12 mar. 2011.

FAGAN, Jody C. Usability studies of faceted browsing: a literature review. *Information Technology and Libraries*, v. 29, n. 2, p. 558-66, Jun. 2010.

FERNEDA, Edberto. *Recuperação da informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação*. 2003.147f. Tese (Doutorado em Ciência da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003.

FERREIRA, Aurélio B. de H. *Novo dicionário Aurélio da língua portuguesa*. Rio de Janeiro: Nova Fronteira, 1988.

FERREIRA, Aurélio B. de H. *Novo dicionário Aurélio da língua portuguesa*. 4. ed. Curitiba: Positivo, 2009.

FIGUEIRA NETTO, Silvino C. *A comunicação científica através de rede de computadores; a experiência de pesquisadores brasileiros*. 1994. 161f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio de Janeiro. Rio de Janeiro, 1994.

FIGUEIREDO, Messias B. *et al.* Biblioteca híbrida especializada em gestão social: projeto piloto. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2., Campinas, 2004. *Anais...* Campinas: 2004. Disponível em: <<http://libdigi.unicamp.br/document/?view=8294>>. Acesso em: 14 abr. 2010.

FIGUEIREDO, N. M. Da necessidade de promover o uso da informação. *Ci. Inf.*, v.6, n.1, p.75-79, 1987.

FIGUEIREDO, Saulo. *A taxonomia na gestão estratégicas das empresas*. [S.l.]: Webinsider, 7 dez. 2006. Disponível em: <<http://webinsider.uol.com.br/index.php/2006/12/07/a-taxonomia-na-gestao-estrategica-das-empresas/>>. Acesso em: 19 abr. 2010.

FRANCO, Maria Laura P. B. *Análise de conteúdo*. 3. ed. Brasília: Liber Livro, 2008. (Série Pesquisa, 6).

FUJITA Mariângela S. L.; RUBI, Milena P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. *DataGramaZero – Rev Ci. Inf.*, Rio de Janeiro, v. 7, n. 3, jun. 2006. Disponível em: <[http://www.dgz.org.br/jun06/Art\\_04.htm](http://www.dgz.org.br/jun06/Art_04.htm)>. Acesso em: 22 abr. 2010.

FUJITA, Mariângela S. L. *A leitura documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional*. 2003. 321f. Tese (Livre-Docência nas disciplinas Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2003.

FURTADO, José Afonso. Hipertexto revisited. *Letras de Hoje*, Porto Alegre, v. 45, n. 2, p. 31-55, abr./jun. 2010.

GIL LEIVA, Isidoro; FUJITA, Mariângela S. L.; RUBI, Milena P. Consistência na indexação em bibliotecas universitárias. *TransInformação*, Campinas, v. 20, n. 3, p. 233-253, set./dez. 2008.

GIL, A. C. *Métodos e técnicas de pesquisa social*. 4. ed. São Paulo: Atlas, 1994. 207p.

GROLIER, Eric de. *Étude sur les catégories générales applicables aux classifications et codifications documentaires*. Paris: UNESCO, 1962.

GRUBER, T. R. *What is an Ontology?*. 1996. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 18 jun. 2010.

GUIMARAES, J. A. C. Perspectiva de ensino e pesquisa em organização do conhecimento em cursos de Biblioteconomia: uma reflexão. In: CARRARA, K. (Org.). *Educação, Universidade e Pesquisa*. Marília: UNESP Marília, FAPESP, 2001, p. 61-74.

GUIZZARDI, G. *Ontological foundations for structural conceptual models*. Twente, Netherlands: University of Twente, Netherlands, 2005.

HANSEN, M. H.; HURWITZ, W. N.; MADOW, W. G. *Sample survey methods and theory*. New York; Chichester: Wiley, 1996. v.1.

HEARST, M. A. Next generation web search: setting our sites. *IEEE Data Engineering Bulletin*, v. 23, n. 3, p. 38-48, 2000. Disponível em: <<http://people.ischool.berkeley.edu/~hearst/papers/data-engineering/>>. Acesso em: 15 mar. 2011.

HEUSELER, Fábio M. *Uma abordagem multifacetada para exploração integrada de dados estruturados e não-estruturados em ambientes OLAP*. 2010. 131f. Dissertação (Mestrado em Informática) – Instituto de Matemática, Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

HJORLAND, B. Domain analysis in information science: eleven approaches: traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p. 422-462, 2002a.

HJORLAND, B. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology - JASIST*, v. 53, n. 4, p. 257-270, 2002b.

HJORLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science - JASIS*, v. 46, n. 6, p. 400-425, 1995.

HYVÖNEN, Eero *et al.* Finnish museums on the semantic web: the user's perspective on MuseumFinland. In: MUSEUMS AND THE WEB, 2004. *Papers...* Helsinki, Institute for Information Technology, University of Helsinki. Finland: HIIT, 2004. Disponível em: <<http://www.archimuse.com/mw2004/papers/hyvonen/hyvonen.html#ixzz1Lp3Tut8V>>. Acesso em: 20 mar. 2011.

JUGEND, D.; SILVA, S. L. da. Gestão do processo de desenvolvimento de produto: um estudo comparativo entre empresas de base tecnológica do setor de automação industrial. In: Encontro Nacional de Engenharia de Produção – ENEGEP, 25., 2005, Porto Alegre. *Anais...* Porto Alegre: ABEPRO, 2005.

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

KRUK, S. R.; DECKER, S.; ZIEBORAK, L. *JeromeDL: reconnecting digital libraries and the semantic web*. Chiba, Japan: [s. n.], 2005. Disponível em: <[http://www.marcont.org/marcont/pdf/www2005\\_jeromedl.pdf](http://www.marcont.org/marcont/pdf/www2005_jeromedl.pdf)>. Acesso em: 14 jan. 2010.

KUHN, Thomas S. *A tensão essencial*. 9. ed. Lisboa: Edições 70, 2006.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. *Ci. Inf.*, Brasília, v. 25, n. 2, 1995.

KWASNICK, Barbara H. The role of classification in knowledge representation and discovery. *Library Trends*, v. 48, n. 1, p. 22-47, 1999. Disponível em: [http://findarticles.com/p/articles/mi\\_m1387/is\\_1\\_48/ai\\_57046525](http://findarticles.com/p/articles/mi_m1387/is_1_48/ai_57046525)>. Acesso em: 16 ago. 2009.

LACHTIM, Fernanda Aparecida. Organização e instanciação automática de conteúdos em portais Semânticos. 2008. 192f. Dissertação (Mestrado em Ciência da Computação) – Instituto Militar de Engenharia do Rio de Janeiro, Rio de Janeiro, 2008.

LAMBE, Patrick. *Organising Knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford, England: Chandos, 2007. 277p.

LANA, Sergio V. de. *Planejamento de pesquisa: uma introdução – elementos para uma análise metodológica*. 2. ed. São Paulo: EDUC, 2009.

LANCASTER, F. W. Princípios da indexação In: \_\_\_\_\_. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 1993, cap. 2.

LANCASTER, F. W.; WARNER, A. J. *Information retrieval today*. Arlington, Va: Information Resources, 1993.

LEMOS, Antônio A. B. de. Bibliotecas. In: CAMPELLO, Bernadete S.; CALDEIRA, Paulo da T.; MACEDO, Vera Amália A.. *Formas e expressões do conhecimento: introdução às fontes de informação*. Belo Horizonte: Escola de Biblioteconomia da UFMG, 1998. p. 345-366.

LIMA, G. A. B. O. *Mapa hipertextual (MHTX) um modelo para a organização hipertextual de documentos*. 2004. 199f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004a.

LIMA, Gercina Ângela B. de O. O modelo simplificado para análise facetada de Spiteri a partir de Ranganathan e do Classification Research Group (CRG). *Inf. Cult. Soc.*, Ciudad Autónoma de Buenos Aires, n. 11, jul./dic. 2004b.

LIMA, Vânia Mara A. de *et al.* Estudos para implantação de ferramenta de apoio à gestão de linguagens documentárias: vocabulário controlado da USP. *TransInformação*, Campinas, v. 18, n. 1, p. 17-25, jan./abr., 2006.

LIMA, Vânia Maria A.; BOCCATO, Vera Regina C. O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática. *Perspect. Ciênc. Inf.*, v. 14, n. 1, p. 131-151, jan./abr. 2009.

LLORÉNS, Juan *et al.* Automatic generation of domain representations using thesaurus structures. *Journal of the American Society for Information Science and Technology*, v. 55, n. 10, p. 846-858, 2004.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ci. Inf.*, v. 31, n. 1, 2002. Disponível em:  
<[http://www.scielo.br/scielo.php?pid=S0100-19652002000100005&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0100-19652002000100005&script=sci_arttext)>.  
Acesso em: 22 jun. 2010.

MACEDO, Neusa D. de; MODESTO, Fernando. Equivalências: do serviço de referência convencional a novos ambientes de redes digitais em bibliotecas. *Revista Brasileira de Biblioteconomia e Documentação: Nova Série*, São Paulo, v. 1, n. 1, p. 38-54, 1999.

MACHADO, Raymundo das Neves *et al.* Biblioteca do futuro na percepção de profissionais da informação. *TransInformação*, v. 11, n. 3, p. 215-222, set./dez. 1999.

MALHOTRA, N. K. *Pesquisa de Marketing: uma orientação aplicada*. 3. ed. Porto Alegre: Bookman, 2001.

MARCHIONINI, Gary. Designing hypertext; start with an index. In: FIDEL, Raya *et al.* *Challenges on indexing electronic text and images*. Medford, NJ: ASIS, 1994. cap. 4, p. 77-89.

MARCHIORI, Patrícia Zeni. Ciberteca ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação. *Ci. Inf.*, Brasília, v. 26, n. 2, p. 115 - 124, maio/ago. 1997.

MARCONI, M. A.; LAKATOS, Eva Maria. *Fundamentos de metodologia científica*. 7. ed. São Paulo: Atlas, 2010.

MARCONI, M. A.; LAKATOS, Eva Maria. *Técnicas de pesquisa*. 2. ed. São Paulo: Atlas, 1990.

MARTINEZ, A. *et al.* Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios Web argentinos. *Ci. Inf.*, Brasília, v. 33, n. 2, p. 106-111, maio/ago. 2004.

MELTON, Jessica. *A use for the techniques of structural linguistics in documentation research*. Cleveland: Western Reserve University, 1964.

MENDES, Renata dos Santos. A importância da adequada estruturação de resumo e resenha. *Revista Espaço Acadêmico*, Maringá, PR, n. 114, p. 135-140, nov. 2010. Disponível em: <<http://periodicos.uem.br/ojs/index.php/EspacoAcademico/index>>. Acesso em: 2 fev. 2011.

MORAES, Roque. Análise de Conteúdo. *Revista Educação*, Porto Alegre, v. 22, n. 37, p. 7-32, 1999.

MOREIRA, W. Biblioteca tradicional x biblioteca virtual: modelos de recuperação da informação. 1998. 90f. Dissertação (Mestrado em Biblioteconomia), Pontifícia Universidade Católica de Campinas, Campinas, 1998.

MORVILLE, Peter; ROSENFELD, Louis. *Information architecture for the World Wide Web*. 3th ed. Sebastopol, CA: O'Reilly Media, 2006.

MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeanette Marguerite. (Orgs.). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: Editora UFMG, 2000. p. 21-34.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual thesauri*. 2005.

NELSON, Theodor Holm. Xanalogical structure, needed now more than ever: parallel documents, deep links to content, deep versioning, and deep re-use. *ACM Computing Surveys*, v. 31, n. 4, Dec. 1999. Disponível em: <<http://www.acm.org/surveys/Formatting.html>>. Acesso em: 4 mar. 2010.

NOGUEIRA, Maria do Carmo de C. *Artigos de periódicos: produção de pesquisadores de Ciência Espacial do INPE*. 1997. 149f. Dissertação (Mestrado em Biblioteconomia – Planejamento e Administração de Sistemas) – Pontifícia Universidade Católica de Campinas, Campinas, 1997.

NONATO, R. dos S. *et al.* Arquitetura da Informação em Bibliotecas Digitais: uma abordagem da Ciência da Informação e da Biblioteconomia. *Inf. Inf.*, Londrina, v. 13, n. 2, p. 125-141, jul./dez. 2008. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1812/1682>> Acesso em: 7 fev. 2010.

NOVO, Hildenise Ferreira. *A elaboração de taxonomia: princípios classificatórios para domínios interdisciplinares*. 2007. 172 f. Dissertação (Mestre em Ciência da Informação) – Universidade Federal Fluminense, Instituto de Arte e Comunicação Social, Niterói, Rio de Janeiro, 2007.

ODDONE, N. E.; GOMES, M. Y. F. S. F. Os temas de pesquisa em Ciência da Informação e suas implicações político-epistemológicas. In: ENCONTRO NACIONAL DE CIÊNCIA DA

INFORMAÇÃO, 5., 2004, Salvador. *Anais...* Salvador: CIFORM, 2004. Disponível em: <[http://dici.ibict.br/archive/00000558/01/temas\\_de\\_pesquisa.pdf](http://dici.ibict.br/archive/00000558/01/temas_de_pesquisa.pdf)> Acesso em: 13 mar. 2010.

ODDONE, Nanci; GOMES, Maria Yêda F. S. de F. Uma nova taxonomia para a ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., Belo Horizonte, 2003. *Anais...* Belo Horizonte: ECI/UFMG, 2003.

OHIRA, Maria Lourdes B.; PRADO, Noêmia S. Bibliotecas virtuais e digitais: análise de artigos de periódicos brasileiros: 1995/2000. *Ci. Inf.*, Brasília, v. 31, n.1, jan. 2002. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0100-19652002000100007&script=sci\\_arttext&tlng=es](http://www.scielo.br/scielo.php?pid=S0100-19652002000100007&script=sci_arttext&tlng=es)>. Acesso em: 12 abr. 2010.

OLIVEIRA, Denize Cristina de. Análise de conteúdo temático-categorial: uma proposta de sistematização. *Rev. Enferm.*, Rio de Janeiro, v. 16, n. 4, p. 569-576, out./dez. 2008.

OLSON, Tod A. Utility of a faceted catalog for scholarly research. *Library Hi Tech*, v. 25, n. 4, p. 550-561, 2007.

PEREIRA, Edmeire Cristina; RUTINA, Raquel. O século XXI e o sonho da biblioteca universal: quase seis mil anos de evolução na produção, registro e socialização do conhecimento. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 4, n. 1, p. 5-19, jan./jun. 1999.

PINHO, Carolina M. D. de; KUX, Hermann H.; ALMEIDA, Cláudia Maria. Elaboração de rede semântica para classificação de cobertura do solo de ambientes intra-urbanos: o Caso de São José dos Campos – SP. In: SIMPOSIO BRASILEIRO DE SENSORIMANTO REMOTO, 13., Florianópolis, 21-26 abr. 2007. *Anais...* Florianópolis: SBSR, 2007. p. 637-644.

POLLITT, A. S. *et al.* View-based searching systems: progress towards effective disintermediation. In: ONLINE INFORMATION, 96.; INTERNATIONAL ONLINE INFORMATION MEETING, 20., 1996. London, England. *Proceedings...* London, England: [s.n], 1996. p. 433-441. Disponível em: <[http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?\\_nfpb=true&\\_ERICExtSearch\\_SearchValue\\_0=ED411866&ERICExtSearch\\_SearchType\\_0=no&accno=ED411866](http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED411866&ERICExtSearch_SearchType_0=no&accno=ED411866)>. Acesso em: 14 mar. 2011.

PRATT, Wanda; HEARST, Marti A.; FAGAN, Lawrence M. A knowledge-based approach to organizing retrieved documents. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16., July 18-22, 1999. *Proceedings...* Orlando, Florida, 1999. p. 80-85.

PRAZERES, C. V. S. *et al.* Uma proposta de navegação na web utilizando facetas. In: ERBD, 6.; ESCOLA REGIONAL DE BANCO DE DADOS, 2., 2006, Passo Fundo, RS. *Anais...* Passo Fundo,RS: [s.n.], 2006. v. 1, p. 1-6.

RABELLO, Rodrigo. *A face oculta do documento: tradição e inovação no limiar da Ciência da Informação*. 2009. 331f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Federal Paulista, São Paulo, 2009.

RANGANATHAN, Shiyali R. *Prolegomena to library classification*. 3. ed. London: Asia Publishing House, 1967. Disponível em: <<http://dlist.sir.arizona.edu/arizona/handle/10150/106370>>. Acesso em: 12 fev. 2011.

ROCHA, Décio; DEUSDARÁ, Bruno. Análise de Conteúdo e Análise do Discurso: aproximações e afastamentos na (re)construção de uma trajetória. *Alea*, Rio de Janeiro, v. 7, n. 2, p. 305-322, jul./dez. 2005.

ROSETTO, Márcia. Bibliotecas digitais: cenários e perspectivas. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v.4, n.1, p. 101-130, jan./jun. 2008.

ROWLEY, Jennifer. *Informática para bibliotecas*. Brasília: Brique de Lemos, 1994.

ROZENFELD, H. et al. *Gestão de desenvolvimento de produtos: uma referência para a melhoria de processos*. São Paulo: Saraiva, 2006.

SACCO, Giovanni Maria; TZITZIKAS, Yannis (Ed.). *Dynamic taxonomies and faceted search: theory, practice and experience*. Berlin: Springer, 2009.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, v. 24, n. 5, p. 513-23, 1988.

SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, p. 1051-1063, 1999.

SAYÃO, Luís Fernando. Padrões para bibliotecas digitais abertas e interoperáveis. *Enc. Bibli.: R. Eletr. Bibliotecon. Ci. Inf.*, Florianópolis, n. esp., 1. sem. 2007.

SAYÃO, Luiz Fernando. Bibliotecas digitais e suas utopias. *Ponto de Acesso*, Salvador, v.2, n.2, p. 2-36, ago. /set. 2008.

SETZER, Valdemar W. Dado, Informação, Conhecimento e Competência. *DataGramaZero - Revista de Ciência da Informação*, n. zero, dez. 1999. Disponível em: <[http://www.dgz.org.br/dez99/Art\\_01.htm](http://www.dgz.org.br/dez99/Art_01.htm)>. Acesso em: 12 abr. 2011.

SEVERINO, Antônio Joaquim. *Metodologia do trabalho científico*. 23. ed. rev. atual. São Paulo: Cortez, 2007.

SIDOROFF, T.; HYVÖNEN, E. *Semantic e-government portals: a case study*. 2005. Disponível em: <<http://www.cs.helsinki.fi/u/eahyvone/publications/2005/suomifi.pdf>>. Acesso em: 12 abr. 2010.

SILVA, Antônio Carlos R. da. *Metodologia da pesquisa aplicada à contabilidade: orientações de estudos, projetos, artigos, relatórios, monografias, dissertações, teses*. 3. ed. São Paulo: Atlas, 2010.

SILVA, Luiz Antônio G. da; MÁRDERO, Miguel Á.; CLAUDIO, Silvana. Acompanhamento das bibliotecas brasileiras na Internet. *Ci. Inf.*, Brasília, v. 26, n. 2, maio/ago. 1997.

SILVA, Maria dos Remédios da; FUJITA, Mariângela S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. *TransInformação*, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004.

SIMON, H. A.; VALDES-PEREZ, R. E.; SLEEMAN, Derek H. Scientific discovery and simplicity of method. *Artificial Intelligence*, v. 91, n. 2, p. 177-181, Apr. 1997.

SMIT, Johanna W.; KOBASHI, Nair Y. Como *elaborar vocabulário controlado para aplicação em arquivos*. São Paulo: Arquivo do Estado, Imprensa Oficial, 2003. (Projeto Como Fazer, 10).

SMITH, Greg et al. FacetMap: a scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics*, v. 12, n. 5, Sep./Oct. 2006. Disponível em: <<http://research.microsoft.com/apps/pubs/default.aspx?id=64297>>. Acesso em: 11 abr. 2011.

SOERGEL, Dagobert. *Functions of a thesaurus, classification, ontological knowledge base*. [S.l.]: College of Library and Information Services, University of Maryland, 1997.

SOUZA, Allan Rocha de. A Portaria 13 de 2006 da CAPES e os direitos autorais. *Revista da Faculdade de Direito de Campos*, ano 7, n. 8, 15 jun. 2006.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ci. Inf*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

SOUZA, Renato R. Sistemas de Recuperação de Informações e Mecanismos de Busca na web: panorama atual e tendências. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 11, n. 2, p. 161 - 173, maio/ago. 2006.

SOUZA, Vanessa I. de. *Indexação: teorias e práticas do corpo indexador da Universidade Federal do Rio Grande do Sul*. 2009. 36f. Monografia (Especialização em Gestão de Bibliotecas Universitárias) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/18885/000730118.pdf?sequence=>>>. Acesso em: 22 abr. 2010.

SPITERI, Louise. A Simplified model for facet analysis: Ranganathan 101. *Canadian Journal of Information and Library Science*, v. 23, p. 1-30, Apr./Jun. 1998. Disponível em: <[http://iaainstitute.org/en/learn/research/a\\_simplified\\_model\\_for\\_facet\\_analysis.php](http://iaainstitute.org/en/learn/research/a_simplified_model_for_facet_analysis.php)>. Acesso em: 21 jun. 2010.

STEVENSON, Leslie; BYERLY, Henry. *The many faces of science: and introduction to scientists, values, and society*. Boulder: Westview, 1995.

STOICA, E.; HEARST, M. *Nearly-automated metadata hierarchy creation*. 2004. Disponível em: <<http://biotext.berkeley.edu/papers/hlt-naacl04.pdf>>. Acesso em: 19 abr. 2010.

STOICA, E.; HEARST, M.; RICHARDSON, M. *Automating creation of hierarchical faceted metadata structure*. 2007. Disponível em: <<http://flamenco.berkeley.edu/papers/castanet.pdf>>. Acesso em: 29 mar. 2010.

SUOMINEN, O.; VILJANEN, K.; HYVÖNEN, E. User-centric faceted search for semantic portals. In: EUROPEAN SEMANTIC WEB CONFERENCE – ESWC, 4., 2007, Innsbruck, Austria, *Proceedings...* Innsbruck, Austria: [s.n], 2007. p. 356–370, 2007. Disponível em: <<http://www.seco.tkk.fi/publications/2007/suominen-et-al-user-centric-faceted-search-2007.pdf>>. Acesso em: 29 mar. 2010.

SUOMINEN, Osmo *et al.* HealthFinland-a national semantic publishing network and portal for health information. *Journal of Web Semantics*, v. 7, n. 4, p. 271-376, dec. 2009. Disponível em: <<http://www.seco.tkk.fi/publications/2009/suominen-et-al-healthfinland-2009.pdf>>. Acesso em: 23 mar. 2011.

SWED. Semantic web environmental directory. Disponível em: <<http://www.swed.org.uk/swed/index.html>>. 2004.

TÁLAMO, M. F. G. M. Elaboração de resumos. *Cadernos de análise documentária*, São Paulo, n.1, p. 23-31, maio, 1994.

TÁLAMO, Maria de Fátima G. M. *et al.* A interface análise documentária, lingüística documental e terminologia. In: SIMPÓSIO DE TERMINOLOGIA, 3., 1992, Millán de la Cogolla. *Anais...* San Millán de la Cogolla: RITerm, 1992. Disponível em: <<http://www.riterm.net/actes/3simposio/talamo.htm>>. Acesso em: 12 mar. 2011.

TARGINO, M. G. Comunicação científica: uma revisão de seus elementos básicos. *Revista Informação & Sociedade: Estudos*, João Pessoa, v. 10, n. 2, 2000.

TERRA, J. C. C. *et al.* *Taxonomia*: elemento fundamental para a gestão do conhecimento. 2005. Disponível em: <<http://www.terraforum.com.br>>. Acesso em: 12 ago. 2008.

THELLEFSEN, Torkild L.; THELLEFSEN, Martin M. Pragmatic semiotics and knowledge organization. *Knowledge Organization*, v. 31, n. 3, p. 177-187, 2004.

TUNKELANG, Daniel. *Faceted search*. North Carolina: Morgan e Claypool, 2009.

TVAROŽEK, Michal *et al.* Improving semantic search via integrated personalized faceted and visual graph navigation. In: 34<sup>TH</sup> CONFERENCE ON CURRENT TRENDS IN THEORY AND PRACTICE OF COMPUTER SCIENCE, 34., Jan. 19-25 2008, Berlin. *Proceedings...* Berlin: Springer, 2008. p. 778-789.

TVAROŽEK, Michal; BIELIKOVÁ, Mária. Personalized faceted browsing for digital libraries. In: 11<sup>th</sup> EUROPEAN CONFERENCE, 11., Sep. 16-21 2007, Budapest. *Proceedings...* Budapest, Hungary: ECDL, 2007. p. 485-488.

TZITZIKAS, Y *et al.* An algebraic approach for specifying compound terms in faceted taxonomies. In: INFORMATION MODELLING AND KNOWLEDGE BASES, 15.; EUROPEAN-JAPANESE CONFERENCE ON INFORMATION MODELLING AND KNOWLEDGE BASES, 13., [2003], Kytakyushu. *Proceedings...* Amsterdam: IOS Press, 2004.

TZITZIKAS, Y.; ANALYTI, A. Faceted taxonomy-based information management. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, 18., 2007, Regensburg. *Proceedings...* Linz: DEXA Society, 2007.

VICKERY, Brian C. *Classification and Indexing in Science*. 3. ed. London: Butterworths, 1975.

VICKERY, Brian C. *Technique of information retrieval*. London, Butterworth, 1970.

VIEIRA, Simone B. Indexação automática e manual: revisão de literatura. *Ci. Inf*, Brasília, DF, v. 17, n. 1, 1988.

WANDERLEY, Manoel A. Linguagem documentária: acesso à informação - aspectos do problema. *Ci. Inf.*, Brasília, v. 2, n. 2, p. 175-217, 1973. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/1699/1302>>. Acesso em: 22 jun. 2010.

WHORF, Benjamin L. On the connection of ideas. Letter to Horace B. English. In: \_\_\_\_\_. *Language, thought, and reality*. Cambridge, MA: MIT, 1956. p. 35-39.

ZAFALON, Zaira Regina. Biblioteca em tempo real: o acesso em foco: proposta crítica do modelo de organização da informação na contemporaneidade. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 6, n. 1, p. 61-83, jul./dez. 2008.

#### BIBLIOGRAFIA COMPLEMENTAR

ALLEN, R. B. Retrieval from facet spaces. *Electronic Publishing*, v. 8, p. 247-257, 1996. Disponível em:

<<http://www.ischool.drexel.edu/faculty/ballen/PAPERS/FACETS/facets.html>>. Acesso em: 12 mar. 2011.

ALVARENGA, L. Representação do conhecimento em tempo e espaço digitais. *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, Florianópolis, n.15, p.1-23, 2003.

ALVES, R. C. V. *Web semântica: uma análise focada no uso de metadados*. 2005. 180f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2005.

ALVES, Renato M. B. *Uma análise empírica sobre o uso combinado de técnicas de busca e de navegação na recuperação de informação*. 122f. 2010. Dissertação (Mestrado em Informática) – Instituto de Matemática, Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

ANDER-EGG, Ezequiel. *Introducción a las técnicas de investigación social: para trabajadores sociales*. 7. ed. Buenos Aires: Humanitas, 1978.

AZEVEDO NETTO, C. X. de. A abordagem do conceito como uma estrutura semiótica. *TransInformação*, Campinas, v. 20, n. 1, p. 47-58, jan./abr. 2008.

BARBOSA, Alice. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.

BARDIN, Laurence. *Análise de conteúdo*. 3. ed. Lisboa: Edições 70, 2004.

BARITÉ, Mario G. *et al.* Garantia literária: elementos para uma revisão crítica após um século. *TransInformação*, Campinas, v. 22, n. 2, p. 123-138, maio/ago. 2010.

BARITÉ, Mario G. The notion of category: its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge Organization*, v. 27, n. 1-2, p. 5-10, 2000.

BATISTA, Edinelson A. *Uma Taxonomia facetada para técnicas de elicitação de requisitos*. 2003. 150 f. Dissertação (Mestrado em Ciência da Computação) – Universidade de Campinas, Instituto de Computação, Campinas, 2003.

BELLUZZO, Regina Célia B. *et al.* Indexação automática de documentos com "PRECIS software" no Serviço de biblioteca e Documentação da Faculdade de Odontologia de Bauru (USP). In: SIMPÓSIO LATINO-AMERICANO DE TERMINOLOGIA TÉCNICO-CIENTÍFICA (RITerm), 2., 1990, Brasília. *Actas...* Brasília: Secretaria da Ciência e Tecnologia da Presidência da República/Conselho Nacional de Desenvolvimento Científico e Tecnológico / Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 1990. Disponível em: <<http://www.riterm.net/actes/2simposio/belluzzo.htm>>. Acesso em: 12 mar. 2011.

BONINI, Adair. O conhecimento de jornalistas sobre gêneros textuais: um estudo introdutório. *Revista Linguagem em (Dis)curso*, Palhoça, SC, v. 2, n. 1, jul./dez. 2001. Disponível em: <<http://www3.unisul.br/paginas/ensino/pos/linguagem/0201/01.htm>>. Acesso em: 12 mar. 2010.

BORBINHA, José. Bibliotecas, arquivos e outras coisas digitais. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 9., 2007, Açores. *Anais...* Universidade dos Açores de Ponta Delgada. Açores, Lisboa: Congresso BAD, 2007. Disponível em: <<http://badinfo.apbad.pt/congresso9/COM68.pdf>>. Acesso em: 2 abr. 2010.

- BRASIL, Maria Irene *et al.* *Vocabulário sistematizado: a experiência da Fundação Casa de Rui Barbosa*. São Paulo: INTEGRAR. São Paulo, 2002. Disponível em: <[http://www.casaruibarbosa.gov.br/dados/DOC/artigos/a-j/FCRB\\_Vocabulario\\_sistematizado\\_a\\_experiencia\\_da\\_Fundacao\\_Casa\\_de\\_Rui\\_Barbosa.pdf](http://www.casaruibarbosa.gov.br/dados/DOC/artigos/a-j/FCRB_Vocabulario_sistematizado_a_experiencia_da_Fundacao_Casa_de_Rui_Barbosa.pdf)>. Acesso em: 23 mar. 2010.
- BRIGHT, M.; HURSON, A.; PAKZAD, S. A taxonomy and current issues in multidatabase systems. *Computer*, v. 25, n. 3, p. 50-60, 1992.
- BUSH, V. As we may think. *The Atlantic Monthly*, Boston, v. 176, n. 1, p. 101-108, 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: 28 fev. 2010.
- CAMARGO, A. C. *O desafio da informação: no marketing da comunicação*. Rio de Janeiro: Expressão e Cultura, 1983.
- CAMPOS, C. M. S.; BATAIERO, M. O. Necessidades de saúde: uma análise da produção científica brasileira de 1990 a 2004. *Interface - Comunic., Saúde, Educ.*, v.11, n. 23, p. 605-18, set./dez. 2007.
- CAMPOS, M. L. de A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: EDUFF, 2001.
- CAREGNATO, Rita Catalina A.; MUTTI, Regina. Pesquisa qualitativa: análise de discurso *versus* análise de conteúdo. *Texto Contexto Enferm.*, Florianópolis, v. 15, n. 4, p. 679-84, out./dez. 2006.
- CASTELLS, Manuel. *A sociedade em rede*. São Paulo: Paz e Terra, 1999. (A era da informação: economia, sociedade e cultura, 1).
- CASTRO, Fabiano F. de; SANTOS, Plácida L. V. A. da C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da *web* semântica. *Inf. & Soc.: Est.*, João Pessoa, v. 17, n. 2, p.13-19, maio/ago. 2007. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/840/1442>>. Acesso em: 22 mar. 2010.
- CHANG, Shan-Ju, RICE, R. E. Browsing: a multidimensional framework. *ARIST*, v. 28, p. 231-276, 1993.
- COELHO, Maria de Fátima P. *et al.* A aprendizagem organizacional e gestão estratégica da informação na UFMG: a implantação da Biblioteca Digital de Teses e Dissertações. In: SEMINÁRIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS BRASIL – SIBDB, 2007, São Paulo. *Anais...* São Paulo: UNICAMP, 2007. Disponível em: <<http://libdigi.unicamp.br/document/?code=23462>>. Acesso em: 24 mar. 2010.
- CONWAY, Susan; SLIGAR, Char. *Unlocking knowledge assets*. Redmon: Microsoft, 2002. cap. 6. [on-line]. Disponível em: <<http://www.microsoft.com/mspress/books/sampchap/5516a.aspx#104>>. Acesso em: 13 mar. 2011.
- CUNHA, I. M. R. F. (Coord.). *Análise documentária: considerações teóricas e experimentações*. São Paulo: FEBAB, 1989.
- CUNHA, Murilo B. da. Biblioteca digital: bibliografia internacional anotada. *Ci. Inf.*, Brasília, v. 26, n. 2, p. 195-213, 1997.

CUNHA, Murilo B. da. Construindo o futuro: a biblioteca universitária brasileira em 2010. *Ci. Inf.*, Brasília, v. 29, n. 1, p. 71-89, jan./abr. 2000.

CUNHA, Murilo Bastos da. Das bibliotecas convencionais às digitais: diferenças e convergências. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 13, n. 1, p. 2-17, jan./abr. 2008.

DAHLBERG, Ingetraut. Knowledge organization and terminology: philosophical and linguistic bases. *Int. Classif.*, v. 19, n. 2, p. 65-71, 1992.

DAHLBERG, Ingetraut. Knowledge organization: a new science?. *Knowledge Organization*, v. 33, n. 1, 2006.

DAHLBERG, Ingetraut. Teoria do conceito. *Ci. Inf.* Brasília, DF, v. 7, n. 2, 1978. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/1680>>. Acesso em: 14 ago. 2009.

DAHLBERG, Ingetraut. Uma teoria para o interconcept: teoria analítica do conceito voltada para o referente. Título original: A referent-oriented analytical concept theory of interconcept. [Publicado originalmente na revista] *International Classification*, v. 5, n. 3, p. 142-151, 1978. [Traduzido por Vânia Teixeira Gonçalves, Bolsista do CNPq, da equipe de Hagar. E. Gomes, Rio de Janeiro, 1990. 34p.].

DENTON, William. *How to make a faceted classification and put it on the web*. Disponível em: <<http://www.miskatonic.org/library/facet-web-howto.html>>. Acesso em: 10 jan. 2010.

DIAS, Aldo A.; NARVAI, Paulo C.; REGO, Delane Maria. Tendências da produção científica em odontologia no Brasil. *Ver. Panam. Salud. Publica [on-line]*, v. 24, n. 1, p. 54-60, 2008. Disponível em: <<http://www.scielosp.org/>>. Acesso em: 12 mar. 2011.

DIAS, C. A. Portal corporativo: conceitos e características. *Ci. Inf.* Brasília, DF, v. 30, p. 50-60, 2001.

DIAS, Eduardo W. Contexto digital e tratamento da informação. *DataGramaZero - Rev Ci. Inf.*, Rio de Janeiro, v. 2, n. 5, out. 2001. Disponível em: <[http://www.dgz.org.br/out01/Art\\_01.htm](http://www.dgz.org.br/out01/Art_01.htm)>. Acesso em: 20 out. 2009.

DIAS, S. A. *Integração semântica de dados através de federação de ontologias*. 2006. 83f. Dissertação (Mestrado em Informática) – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

DOSI, G. Sources, procedures and microeconomic effects of innovation. *Journal of Economic Literature*, v. 27, p. 1126-1171, 1988.

DUARTE Elizabeth A.; CERQUEIRA, Roberto F. P. de. Análise facetada: um olhar face a modelagem conceitual. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 4, n. 2, p. 39-52, jan./jun. 2007.

ENGELMANN, Jaqueline. *Teoria da definição: das definições reais às definições predicativas*. 2006. 140f. Tese (Doutorado em Filosofia) – Departamento de Filosofia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006. Disponível em: <[http://www.maxwell.lambda.ele.puc-rio.br/Busca\\_etds.php?%20strSecao=resultado&nrSeq=9106@1](http://www.maxwell.lambda.ele.puc-rio.br/Busca_etds.php?%20strSecao=resultado&nrSeq=9106@1)>. Acesso em: 12 dez. 2009.

ESTEBAN NAVARRO, Miguel A. El marco disciplinar de los lenguajes documentales: la Organización del Conocimiento y las ciencias sociales. *Scire*, Zaragoza, v. 2, n. 1, p. 93-107, ene./jun. 1996.

FAULSTICH, Enilde. A socioterminologia na comunicação científica e técnica. *Cienc. Cult.*, v. 58, n. 2, p. 27-31, jun. 2006. Disponível em: <<http://cienciaecultura.bvs.br/pdf/cic/v58n2/a12v58n2.pdf>>. Acesso em: 18 fev. 2010.

FAZENDA, Ivani. *Integração e interdisciplinaridade no ensino brasileiro: efetividade ou ideologia?*. São Paulo: Loyola, 1979.

FELBER, H. *Terminology manual*. Paris: UNESCO, 1984.

FERREIRA, Aloizio S.; ABREU, Márvio L. T. Desconstruindo um artigo científico. *Revista Brasileira de Zootecnia*, Piauí, v. 36, 2007.

FERREIRA, Luiz G. R. *Redação Científica: como escrever artigos, monografias, dissertações e teses*. Fortaleza: UFC, 1994.

FINATTO, Maria José B. *Definição terminológica: fundamentos teórico-metodológicos para sua descrição e explicação*. 2001. 395f. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2001.

FOSKETT, A.C. *The subject approach to information*. London: C. Bingley, 1996.

FOUCAULT, M. *A arqueologia do saber*. Rio de Janeiro: Forense Universitária, 1987.

FREEMAN, C.; SOETE, L. *The economics of industrial innovation*. London: Pinter, 1994.

FUJITA, Mariângela S. L.. Organização do conhecimento: algumas considerações para o tratamento temático da informação. In: CARRARA, Kester. (Org.). *Educação, Universidade e Pesquisa*. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p. 29-34.

FURGERI, Sérgio. *Representação de informação e conhecimento: estudo das diferentes abordagens entre a Ciência da Informação e a Ciência da Computação*. 2006. 161f. Dissertação (Mestre em Ciência da Informação) – Centro de Ciências Sociais Aplicadas, Pontifícia Universidade Católica de Campinas, Campinas, 2006.

GILLILAND-SWETLAND, Anne J.; KAFAL, Yasmin; LANDIS, William E. Metadata applications in the creation and description of digital primary sources for elementary school classrooms. *Journal of the American Society for Information Science*, v. 50, n. 14, p. 193-201, 1999.

GINEZ DE LARA, M. L. *A representação documentária: em jogo a significação*. São Paulo, 1993. 133f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.

GRÁCIO, José Carlos Abbud. *Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade*. 2002. 127f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

GRUBER, T. R., A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, p. 199-220, 1993.

GUIMARÃES, José Augusto Chaves. A análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais. In: RODRIGUES, Georgete Medleg; LOPES, Ilza Leite. *Organização e representação do conhecimento na perspectiva da Ciência da*

*Informação*. Brasília: Thesaurus, 2003. p. 100-117. (Estudos avançados em Ciência da informação, 2).

GUINCHAT, C.; MENO, M. *Introdução geral às ciências e técnicas da informação e documentação*. Brasília: IBICT, 1994.

HEARST, M. A. *Design recommendations for hierarchical faceted search interfaces*. 2007. Disponível em: <[flamenco.berkeley.edu/papers/faceted-workshop06.pdf](http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf)>. Acesso em: 12 mar. 2011.

HEARST, M. A. Clustering versus faceted categories for information exploration. *Communications of the ACM*, v. 40, n. 4, p. 56-61, 2006.

HINTAKU, Milton; BRÄSCHER, Marisa. DSpace versão 1.4: uma análise das facilidades relacionadas ao assunto. In: SEMINÁRIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS BRASIL, 2007, São Paulo. *Anais...* São Paulo: SIBDB, 2007. Disponível em: <[libdigi.unicamp.br/document/?down=23471](http://libdigi.unicamp.br/document/?down=23471)>. Acesso em: 23 mar. 2010.

HJORLAND, B. Documents, memory institutions, and information science. *Journal of Documentation*, v. 56, p. 27-41, 2000.

HOUAISS, Antônio et al. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2001.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Classification of information in the construction industry*. Geneva, 1994. 1 v. (ISO technical report, 14177).

JACOB, Elin K.; ALBRECHTSEN, Hanne. When essence becomes function: post-structuralist implications for an ecological theory of organizational classification systems source. In: MCCASKIE, Russell. *Exploring the contexts of information behavior*. London, UK: Taylor Graham Publishing, 1999. p. 519-534.

JAPIASSU, H. A. *Interdisciplinaridade: algumas conclusões*. Rio de Janeiro: Ed. PUC/RJ, 1998.

JESUS, Leandro Martins de. *Interdisciplinaridade*. [S.l.]: Recanto das Letras, 18 nov. 2006. [On-line]. Disponível em: <<http://recantodasletras.uol.com.br/artigos/294810>>. Acesso em: 5 ago. 2008.

KLEIN, Julie. Ensino interdisciplinar: didática e teoria. In: FAZENDA, Ivani (Org.). *Didática e interdisciplinaridade*. Campinas: Papirus, 1998. p. 109-132.

KOBASHI, N. Y. Análise documentária e representação da informação. *Informare*, São Paulo, v. 2, n. 2, p. 5-27, 1996.

KOBASHI, N. Y. Terminologia e operações documentárias. In: SIMPOSIO IBEROAMERICANO DE TERMINOLOGIA: RITerm 1988-2002, 4., 1994, Buenos Aires. *Atas...* Buenos Aires: RITERM, 1994. Disponível em: <<http://www.riterm.net/actes/4simposio/kobashi.htm>>. Acesso em: 12 abr. 2010.

KUHN, Thomas S. *Estrutura das revoluções científicas*. 3. ed. São Paulo: Perspectiva, 1989.

LAKATOS, Eva Maria; MARCONI, Marina de A. *Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos*. 6. ed. São Paulo: Atlas, 2001.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2004.

LAUSEN, H. *et al.* Semantic web portals: state-of-the-art survey. *Journal of Knowledge Management*, v. 9, n. 5, p. 40-49, may 2005. Disponível em: <<http://dx.doi.org/10.1108/13673270510622447>>. Acesso em: 18 mar. 2011.

LAZZAROTO, Juliana F.; SANTIAGO, Leilah B.; FAUSTINO, Rene G. Junior. Proposta de metodologia para a recuperação da produção científica em ciência da informação na base BRAPCI. *Ponto de Acesso*, Salvador, v. 4, n.3, p. 45-67, 2011. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/4629/3564>>. Acesso em: 20 mar. 2011

LEI, Y., UREN, V.; MOTTA, E. Semsearch: a search engine for the semantic web. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT MANAGING KNOWLEDGE IN A WORLD OF NETWORKS, 5., 2006, Podebrady. *Proceedings...* Podebrady, Czech Republic: EKAW-2006, 2006. p. 238-245.

LEITE, Fernando C. L.; MÁRDERO-ARELLANO, Miguel Á.; MORENO, Fernanda P. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil. *Perspect. Ciênc. Inf.*, Belo Horizonte, v.11, n.1, p. 82-94, jan./abr. 2006.

LI, Y.; BELKIN, N. J. A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, Elmsford, v. 44, n. 6, p. 1822-1837, nov. 2008.

LITWIN, W.; MARK, L.; ROUSSOPOULOS, N. Interoperability of multiple autonomous databases. *ACM Computing Surveys (CSUR)*, v. 22, n. 3, p. 267-293, 1990.

LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. *Ci. Inf.*, Brasília, v. 31, n. 2, p. 60-71, maio/ago. 2002.

LOPES, Jorge Expedito de G. *et al.* Procedimentos metodológicos para elaboração de projetos pesquisa relacionados a dissertações de mestrado em ciências contábeis. *Revista Contabilidade & Finanças*, São Paulo, v. 4, n. 36, p. 97-104, set.-dez./2004.

LYRA, Tania Maria de Paula; GUIMARÃES, Jorge A. Produção científica brasileira em comparação com o desempenho mundial em ciências agrárias. *Planejamento e Políticas Públicas*, Brasília, n. 30, 2007.

MAHEU, Cristina d'Ávila. *Interdisciplinaridade e mediação pedagógica*. [Salvador]: UNIFACS: NUPPEAD/NPP, UFBA (FACED) e UNEB, 2000. Disponível em: <[www.nuppead.unifacs.br/artigos/Interdisciplinaridade.pdf](http://www.nuppead.unifacs.br/artigos/Interdisciplinaridade.pdf)>. Acesso em: 15 ago. 2009.

MÄKELÄ, E. *View-based search interfaces for the semantic web*. 2006. 86f. Dissertation (Doctor of Science in Technology) – Faculty of Science, University of Helsinki, Department of Computer Science, 2006.

MÄKELÄ, E; HYVÖNEN .E.; SIDOROFF. *View-based user interfaces for information retrieval on the semantic web*. 2005. Disponível em: <<http://www.seco.tkk.fi/publications/2005/makela-hyvonen-et-al-view-based-user-2005.pdf>>. Acesso em: 20 jun. 2010.

MARCHIONINI, G.; BRUNK, B. Towards a general relation browser: a GUI for information architects. *Journal of Digital Information*, v. 4, n. 1, 2003.

- MARCONDES, C. H. Representação e economia da informação. *Ci. Inf.*, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001.
- MAYVILLE, W. V. Interdisciplinarity: The mutable paradigm. *Education Research Report, AAHE/ERIC*, San Francisco, CA, v. 7, n. 9, 1978.
- MCGARRY, K. *O contexto dinâmico da informação*. Brasília: Briquet de Lemos, 1999.
- MICHAELIS Língua Portuguesa. *Dicionário escolar: nova ortografia*. São Paulo: Melhoramentos, 2009.
- MORAES, Roque. Uma tempestade de luz... *Ciênc. educ. (Bauru)*, Bauru, v. 9, n. 2, p. 191-211, 2003.
- MOTTA, Dilza Fonseca da. *Método relacional como nova abordagem para a construção de tesouros*. 1987. 89 f. Dissertação (Mestrado em Ciência da Informação) – Centro de Ciências Jurídicas e Econômicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1987.
- MUELLER, Suzana. O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. *Revista da Escola de Biblioteconomia da UFMG*, v. 24, n. 1, p. 63-84, jan./jun. 1995.
- MURRAY, B. Weaving an interdisciplinary education. *APA Monitor*, v. 30, n. 5, 1999. Disponível em: <www.apa.org>. Acesso em: 20 out. 2009.
- NAVES, Madalena M. L. Estudo dos fatores interferentes no processo de análise de assunto. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 6, n. 2, p. 189-203, jul./dez. 2001.
- NAVES, Madalena M. L. *Fatores intervenientes no processo de análise de assunto: estudo de caso de indexadores*. Belo Horizonte, 2000. 253f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.
- NISSANI, M. Ten cheers for interdisciplinarity: the case for interdisciplinary knowledge and research. *Social Science Journal*, v. 34, n. 2, p. 202-217, 1997.
- NORONHA, Ana Paula P.; FERNANDES, Dario Cecílio. Estresse laboral: análise da produção científica brasileira na Scielo e BVS-PSI. *Fractal, Rev. Psicol.*, Niterói, RJ, v. 20, n. 2, p. 491-502, jul./dez. 2008.
- NORONHA, Daisy Pires; FUJINO, Asa. Teses e dissertações em ciências da informação: a multidisciplinaridade não revelada na avaliação da produção científica. *TransInformação*, Campinas, v. 21, n. 2, p. 123-132, maio/ago. 2009. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=283&locale=>>>. Acesso em: 12 mar. 2011.
- OLIVEIRA, Marcos B. de; OLIVERIA, Marta Kohl. *Investigações cognitivas: conceitos, linguagem e cultura*. Porto Alegre: Artmed, 1999.
- OREN, E.; BRESLIN, J. G.; DECKER, S.. How semantics make better wikis. In: INTERNATIONAL WORLD-WIDE WEB CONFERENCE, 15., 2006, Edinburgh, Scotland. *Proceedings...* New York : ACM Press, 2006. p. 1071-1072
- OREN, E. An overview of information management and knowledge work studies: lessons for the semantic desktop. In: INTERNATIONAL WORLD-WIDE WEB CONFERENCE, 5.;

Workshop on the Semantic Desktop, 2006, Athens. *Proceedings...* Athens, GA, USA: ISWC, 2006.

PANSANATO, Luciano T. E. Um modelo de navegação exploratória para a infra-estrutura da web semântica. 2007.166f. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, São Carlos, 2007

PASSARELLI, Brasilina. Do Mundaneum à web semântica: discussão sobre a revolução nos conceitos de autor e autoridade das fontes de informação. *DataGramaZero - Rev Ci. Inf.*, Rio de Janeiro, v. 9, n. 5, out. 2008.

PIEIDADE, Maria Antonietta. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977.

PINHEIRO, Lêna Vânia Ribeiro. Processo evolutivo e tendências contemporâneas da ciência da informação. *Inf. & Soc.: Est.*, João Pessoa, v. 15, n. 1, p. 13-48, jan./jun. 2005.

PINHEIRO, W. A. *Busca em portais semânticos: uma abordagem baseada em ontologias*. 2004. 170f. Dissertação (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 2004.

PINTO, Maria; DOUCET, Anne-Vinciane; FERNÁNDEZ-RAMOS, Andrés. The role of information competencies and skills in learning to abstract. *Journal of Information Science*, v. 34, n. 6, p.799-815, 2008. Disponível em: <<http://jis.sagepub.com/content/34/6/799>>. Acesso em: 12 fev. 2011.

PRESTES, Maria Luci de Mesquita. *A pesquisa e a construção do conhecimento científico: do planejamento aos textos, da escola à academia*. São Paulo: Rêspel, 2002.

RAFFERTY, Pauline; HIDDENLEY, Rob. Flickr and democratic indexing: dialogic approaches to indexing. *Aslib Proceedings*, v. 59, n. 4-5, p. 397-410, 2007. Disponível em: <<http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2760590407.pdf>>. Acesso em: 22 mar. 2010.

ROCKMORE, Marlene. The 5 types of taxonomies: from lists to ontologies. *The taxonomy blog*, 26 jun. 2009. Disponível em: <<http://thetaxonomyblog.wordpress.com/2009/06/26/5-types-of-taxonomies-from-lists-to-ontologies/>>. Acesso em: 12 mar. 2011.

ROSA, Malvina Vianna. O papel da classificação na recuperação da informação: classificação facetada. *R. Bibliotecon.*, Brasília, v. 1, n. 2, p. 207-210, jul./dez. 1973. Disponível em: <<http://www.tempusactas.unb.br/index.php/RBB/article/viewFile/33/25>>. Acesso em: 2 abr. 2010.

ROSAS, P. Introduções redatoriais e a indexação em publicação periódica. In: GOLDENBERG, S; GUIMARÃES C, A; Castro A, A. *Elaboração e apresentação de comunicação científica*. São Paulo: Metodologia.org, 2001. Disponível em: <[www.metodologia.org](http://www.metodologia.org)>. Acesso em: 2 ago. 2010.

RUSSO, Iêda C. P. Editorial II: a relevância da pesquisa científica na audiologia brasileira. *Revista CEFAC*, São Paulo, 2011. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-18462009000500002&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-18462009000500002&lng=en&nrm=iso)>. Acesso em: 11 abr. 2011.

SACCO, Giovanni Maria. Research results in dynamic taxonomy and faceted search systems. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS

APPLICATIONS; *Database and Expert Systems Applications, International Workshop*, 18., 2007, Regensburg. *Proceedings...* Regensburg, Germany: [s.n], 2007. p. 201-206.

SALES, Rodrigo de. Teoria comunicativa da terminologia (TCT) como aporte teórico para a representação do conhecimento especializado. In: ENANCIB Encontro Nacional 2007 de Pesquisa em Ciência da Informação, 7., 2007, Salvador. *Anais...* Salvador: ENANCIB, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--036.pdf>>. Acesso em: jun. 2009.

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. *TransInformação*, Campinas, v. 20, n. 2, p. 133-148, maio/ago., 2008. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=215>>. Acesso em 12 ago. 2009.

SHETH, A.; LARSON, J. A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.*, v. 22, n. 3, p. 183-236, 1990.

SHNEIDERMAN, Ben. Clarifying search: a user-interface framework for text searches. *DLIB Magazine*, [on-line]. Disponível em: <<http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>>. Acesso em: 12 abr. 2010.

SILVA, Antônio Carlos R. de. *Metodologia da pesquisa aplicada à contabilidade*: orientações de estudos, projetos, relatórios, monografias, dissertações, teses. São Paulo: Atlas, 2003.

SILVEIRA, José Maria J. da. Kenneth J. Arrow: economic welfare and the allocation of resources for invention. *Revista Brasileira de Inovação*, Rio de Janeiro, v. 7, n. 2, p. 268-285, jul./dez. 2008. Disponível em: <[http://www.finep.gov.br/revista\\_brasileira\\_inovacao/decima\\_quarta\\_edicao/Ideias%20Fundadoras.pdf](http://www.finep.gov.br/revista_brasileira_inovacao/decima_quarta_edicao/Ideias%20Fundadoras.pdf)>. Acesso em: 20 mar. 2010.

SILVEIRA, Murilo A. A. da; SILVA, Fábio M. e; CORRÊA, Renato F. Os periódicos em Ciência da Informação no Brasil: estudo cienciométrico da frente de pesquisa. In: ENCONTRO NACIONAL DE PESQUISA EM CIENCIA DA INFORMAÇÃO, 11., 2010, Rio de Janeiro. *Anais...* Rio de Janeiro: ENANCIB, 2010.

SMIT, J. W. (Coord.). *Análise documentária*: a análise da síntese. 2. ed. Brasília: IBICT, 1987.

SOERGEL, Dagobert. The rise of ontologies or the reinvention of classification. *Journal of the American Society of Information Science*, v. 50, n. 12, p. 1119-1120, 1999.

SWANSON, Don R.; SMALHEISER, Neil R. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence*, v. 91, n. 2, p. 183-203, Apr. 1997.

TÁLAMO, M. F. G. M. A compreensão literal de textos. *Cadernos de análise documentária*, São Paulo, n. 1, p. 13-22, maio 1994.

TAMMARO, Anna Maria; SALARELLI, Alberto. *A biblioteca digital*. Brasília: Briquet de Lemos, 2008.

TANG, M. *Browsing and searching in a faceted information space*: a longitudinal study of PubMed users' interactions with a novel display tool. New Brunswick, NJ: Rutgers, The State University of New Jersey, 2005.

TOLEDO, A. M. *Portais corporativos: uma ferramenta estratégica de apoio a gestão do conhecimento*. 2002. 126f. Monografia (Especialização em Sistemas de Negócios Integrados) – Escola de Engenharia, Universidade do Rio de Janeiro, Rio de Janeiro, 2002.

TRISTÃO, Ana Maria D.; FACHIN, Gleisy Regina B.; ALARCON, Orestes E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. *Ci. Inf.*, Brasília, v. 33, n. 2, maio/ago. 2004. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652004000200017](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000200017)>. Acesso em: 12 mar. 2010.

TUDHOPE, Douglas; CUNLIFFE, Daniel. Semantically indexed hypermedia: linking information disciplines. *ACM Computing Surveys*, v. 31, n. 4, dez. 1999.

UDDIN, M. N.; JANECEK, P. Faceted classification in web information architecture. *The Electronic Library*, v. 25, n.2, p. 219 - 233, 2007.

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris: The Organization, [1973]. 37p.

VAN DIJCK, Peter. Introduction to XFML: core exchangeable faceted metadata language. *O'Reilly Media.*, 22 jan. 2003. Disponível em: <<http://www.xml.com/pub/a/2003/01/22/xfml.html>>. Acesso em: 20 nov. 2009.

VASCONCELOS, Celso dos S. *Construção do conhecimento em sala de aula*. São Paulo: Salesiana Dom Bosco, 1993.

VICKERY, Brian C. *Faceted classification: a guide to construction and use of special schemes*. London: Aslib, 1960.

VILJANEN, K.; *et al.* Ontodella: a projection and linking service for semantic web applications. In: International Conference on Database and Expert Systems Applications, 17., 2006, Krakow. *Proceedings...* Krakow, Poland: IEEE. p. 370–376.

VITAL, Luciane Paula. *Recomendações para construção de taxonomia em portais corporativos*. 2007. 113 f. Dissertação (Mestrado em Ciência da Informação) – Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, SC, 2007. Disponível em: <<http://www.cin.ufsc.br/pgcin/Vital,%20Luciane.pdf>>. Acesso em: 16 ago. 2008.

WÜSTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l' ontologie, L' informatique et les sciences des chose. In: RONDEAU, G.; FELBER, E. (Org.). *Textes choisis de terminologie*. Québec: GIRSTERM, 1981.

## APÊNDICES

### APÊNDICE A – Lista do *corpus* da pesquisa

CODIGO DOCUMENTO	CÓDIGO DO DADO	TÍTULO DO DOCUMENTO
t001-2009	3454	Registro de informação no Sistema de Informação em Saúde: um estudo das bases SINASC, SIAB e SIM no Estado de Alagoas
t002-2009	4856	As relações interdisciplinares refletidas na Ciência da Informação
t003-2008	3962	Uso de sintagmas nominais na classificação automática de documentos eletrônicos
t004-2008	4857	Fatores de influência no uso de sistemas de informação via internet: proposta de um modelo integrativo
t006-2008	2134	Estudo das características de software e implementação de um software livre para o sistema de gerenciamento de Bibliotecas Universitárias Federais Brasileiras
t009-2007	2095	Gestão arquivística na era do cinema digital: formação de acervos de documentos digitais provindos da prática cinematográfica
t010-2005	79	Ambiente para geração e manutenção semiautomática de tesouros
t011-2005	550	Análise do padrão brasileiro de metadados de teses e dissertações segundo o modelo entidade-relacionamento
t012-2005	1265	Análise temática da produção científica em Comunicação no Brasil baseada em um sistema classificatório facetado
t013-2005	48	Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais
t014-2004	695	Aspectos metacognitivos na leitura do indexador
t015-2004	1261	Mapa Hipertextual (MHTX): um modelo para organização hipertextual de documentos
d004-2009	4054	Teoria do conceito e hipertextos: uma proposta para determinação de relacionamentos em links conceituais
d007-2009	3773	Análise conceitual sobre as relações semânticas em ciência da informação: contribuições para o desenvolvimento de ontologias
d008-2009	3844	Uso de bibliotecas digitais de periódicos: um estudo comparativo no Portal de Periódicos CAPES entre as áreas do conhecimento
d009-2008	3446	Análise de citações de teses como apoio à tomada de decisão no desenvolvimento de coleções de bibliotecas universitárias
d010-2008	2112	A interação dos usuários da UFMG com o catálogo on-line do Sistema Pergamum
d011-2008	4333	A utilização de mapas de tópicos na compatibilização de conteúdos hipertextuais semanticamente estruturados
d012-2008	4342	Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação
d013-2008	4520	Padrões de disciplinaridade no campo de pesquisa sobre a AIDS: uma prospecção a partir de publicações periódicas e pesquisadores
d014-2007	4059	O uso da internet como instrumento na prestação de serviços públicos na Prefeitura Municipal de Belo Horizonte através do Sistema de Atendimento ao Cidadão
d015-2007	1282	Estudo comparativo entre interfaces hipertextuais de softwares para a representação do conhecimento
d016-2007	2101	Comparação entre termos de indexação e palavras dos títulos dos artigos do periódico "Cadernos de Saúde Pública (2000/2005)"
d017-2007	1492	Centrais de <i>Help-Desk</i> : avaliação do serviço de atendimento ao cliente de Empresas Desenvolvedoras de Sistemas de Gestão Empresarial
d018-2007	747	Sistemas eletrônicos de gerenciamento de documentos arquivísticos: um questionário para escolha, aplicação e avaliação
d019-2006	3483	A utilização do prontuário eletrônico do paciente pelos hospitais de Belo Horizonte
d020-2006	144	Estudo de uso de informação externa para tomada de decisão: panorama geral das

CODIGO DOCUMENTO	CÓDIGO DO DADO	TÍTULO DO DOCUMENTO
		empresas de informática
d021-2006	118	Metadados para descrição de documentos remanescentes de fundo eclesiástico: uma proposta de metadados e biblioteca digital para os assentos de batismo, casamento e óbito da Matriz do Pilar de Ouro Preto
d022-2006	116	Educação Física no PPGMH/UFRGS: uma visão a partir da análise de citações e perfil dos pesquisadores
d023-2006	119	Organização e uso das bases de informação para o atendimento a clientes em <i>call centers</i>
d024-2006	2199	A análise de assunto na literatura ficcional infantil: categorias para ler o que você tem
d025-2005	64	Uma metodologia para avaliação automática de sites e sua aplicação em sites do Governo do Estado de Minas Gerais
d026-2005	78	Análise de citações da produção científica de uma comunidade: a construção de uma ferramenta e sua aplicação em um acervo de teses e dissertações do PPGCI-UFGM
d027-2005	4424	Um estudo sobre o uso de periódicos eletrônicos: O Portal De Periódicos CAPES na Universidade Federal de Minas Gerais
d028-2005	955	Método para produção de representações conceituais a partir de literatura especializada utilizando a abordagem Analítico-Sintética
d029-2005	90	Estratégias de produção e organização de informações na WWW: uma análise de sites turísticos
d030-2005	952	Prontuário Eletrônico do Paciente: estudo de uso pela equipe de saúde do Centro de Saúde Vista Alegre
d031-2005	694	A indexação temática de recursos fundamentada por estrutura profunda e abordagem objeto-relacionamento
d036-2004	949	O uso de sistemas de informações gerenciais em empresas: estudo de caso na Andrade Gutierrez Concessões - AGC
d037-2003	18	Tesaurus e ontologias: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se método analítico-sintético
d038-2002	901	O usuário do sistema de informação hospitalar: necessidades e usos no contexto da informação

## **APÊNDICE B – Qualificação dos autores das fontes especializadas**

a) Severino (2007): é atualmente professor titular de Filosofia da Educação na Faculdade de Educação da USP; sendo bacharel em Filosofia pela Universidade Católica de *Louvain*, Bélgica, em 1964; doutor pela Pontifícia Universidade Católica de São Paulo, em 1972. Suas publicações e interesses científicos atuais têm foco em questões de epistemologia da educação. O livro em questão versa sobre subsídios teóricos e metodológicos para a prática do desenvolvimento de trabalho científico, tanto para discentes quanto para docentes.

b) Marconi e Lakatos (2010): Marconi é conhecida na área da metodologia científica, sendo doutora pela UNESP/Franca; Pós-graduada pela Escola de Sociologia e Política; graduada em História pela UNESP/Franca, em Pedagogia pelas Faculdades Claretianas de Batatais, e em Estudos Sociais pela Faculdade de Pedagogia de Ituverava; em. Entre outras atuações, foi professora de Antropologia da UNESP de Franca, por 16 anos, e professora de Sociologia das Faculdades Claretianas de Batatais, por 10 anos.

Lakatos é reconhecida na área por sua especialidade no campo disciplinar da metodologia científica, sendo bacharel em Jornalismo pela Faculdade de Comunicação Social Cásper Libero, em São Paulo; mestre e doutora em Ciências pela Escola de Sociologia e Política de São Paulo; doutora em Filosofia (Metodologia Científica) pela Escola de Sociologia e Política de São Paulo; Livre-Docente em Sociologia, pela Escola de Sociologia e Política de São Paulo; e Pós-graduada em Ciências Sociais pela Escola de Sociologia e Política de São Paulo.

c) Silva (2010): é professor da Fundação Visconde de Cairu e Professor Adjunto da Universidade Federal da Bahia, sendo graduado em Ciências Contábeis pela Fundação Visconde de Cairu; em Pedagogia pela Faculdade de Educação da Bahia; licenciado em Técnicas Comerciais pela Universidade do Estado da Bahia; Especialista em Psicopedagogia Escolar e Clínica (FEBA), em Contabilidade Gerencial (UFBA); e em Educação à Distância (UCB); mestrado em Contabilidade pela Fundação Visconde de Cairu; e doutorado em Desenvolvimento Curricular pela Universidade do Minho/Portugal. Silva atua na área de metodologia de pesquisa e este livro apresenta uma metodologia voltada, especificamente, para a disciplina da área das Ciências Humanas Aplicadas.

## **APÊNDICE C – Conjunto de categorias iniciais**

Categorias iniciais do Modelo Categorical para trabalhos acadêmicos (teses e dissertações).

---

### **CONCEITO (ANÁLISE CONCEITUAL)**

---

C1. TEMA (o que trata a pesquisa)

C2. DIMENSÃO (lugar, tempo, objeto empírico ou ambiente)

C3. AÇÃO (os objetivos do estudo)

C4. CAUSA (razão ou motivo) E EFEITO (produto de uma causa)

C5. FUNDAMENTO TEÓRICO

C6. MÉTODOS

C7. COLETA DE DADOS

C8. PONTO DE VISTA DO AUTOR;  
PERSPECTIVA

C9. PRINCIPAIS RESULTADOS  
ALCANÇADOS

---

## APÊNDICE D – Conjunto de categorias intermediárias

---

### CONCEITO (ANÁLISE CONCEITUAL)

---

C1. TEMA (o que trata a pesquisa)

C2. AMBIENTAÇÃO (lugar, tempo, objeto empírico ou ambiente)

C3. ESCOPO (os objetivos do estudo)

C4. CAUSA (razão ou motivo) E EFEITO (produto de uma causa)

C5. FUNDAMENTO TEÓRICO

C6. MÉTODOS

C7. COLETA DE DADOS

C8. PONTO DE VISTA DO AUTOR;  
PERSPECTIVA

C9. PRINCIPAIS RESULTADOS  
ALCANÇADOS

---

## APÊNDICE E – Definições para o conjunto de categorias finais

**Conjunto de categorias finais:** C1.Tema; C2.Objeto empírico; C3.Escopo; C4.Ambientação; C5.Tipo de pesquisa; C6.Coleta de dados; C7.Métodos; C8.Fundamento teórico; C9.Fundamento histórico/contextual; C10.Resultados.

### C1.Tema

“Assunto, ideia ou proposição que se desenvolve em um discurso, em uma obra didática ou literária” (CUNHA; CAVALCANTI, 2008, p.358).

“É o assunto escolhido sobre o qual versará o Trabalho; é um assunto que se deseja estudar profundamente, pois não interessa somente saber o tema de modo vago, indefinido, ele deve ser conciso, com limites bem definidos, facilitando o encaminhamento da pesquisa” (SILVA, 2010, p.43).

“O tema de uma pesquisa é o assunto que se deseja provar ou desenvolver [...] é uma proposição até certo ponto abrangente” (MARCONI; LAKATOS, 2010, p.112).

“É o assunto que se deseja estudar e pesquisar [...] significa: selecionar um assunto de acordo com as inclinações, as possibilidades, as aptidões e as tendências de quem se propõe a elaborar um trabalho científico; encontrar um objeto que mereça ser investigado cientificamente e tenha condições de ser formulado e delimitado em função da pesquisa” (MARCONI; LAKATOS, 2010, p.142).

### C2.Objeto empírico

“Unidade documentária; arquivo, imagem, página web, base de dados ou qualquer item de um sistema informático; em hipertexto, um botão que representa informação; artefato tridimensional ou réplica de uma entidade; conjunto de *bits* que são vistos como entidades próprias (p. ex.> documento de texto completo) ou, como parte de outro objeto digital (como uma imagem que faz parte de um livro), geralmente associados com metadados e, algumas vezes, com condições especiais para acesso” (CUNHA; CAVALCANTI, 2008, p.265).

“Que se distinga [...] o objeto de uma questão” (SILVA, 2010, p.44).

“É uma porção ou parcela, convenientemente selecionada do universo (população); é um subconjunto do universo” (SILVA, 2010, p.73).

“Exploração de documentos fontes do objeto pesquisado e registro das informações retiradas nessas fontes e que serão utilizadas no desenvolvimento do trabalho; [...] fonte dirável de informação sobre os fenômenos pesquisados” (SEVERINO, 2007, p.124).

“Conjunto de informações que lhe permitirá a escolha da amostra, que deve ser representativa ou significativa” (MARCONI; LAKATOS, 2010, p.146).

### C3.Escopo

“Intuito, intenção” (FERREIRA, 1988, p.263).

“Determinar a profundidade, abrangência e extensão do assunto; é o que se quer saber ou o que se quer fazer a respeito do sujeito [tema]” (SILVA, 2010, p.44).

“São os fins teóricos e práticos que se propõe alcançar com a pesquisa” (SILVA, 2010, p.50).

“Toda pesquisa deve ter um objetivo determinado para saber o que vai procurar e que se pretende alcançar. [...] o objetivo torna explícito o problema, aumentando os conhecimentos sobre determinado assunto. Para Ackoff (1975, p.27) ‘o objetivo da ciência não é somente aumentar o conhecimento, mas o de aumentar as nossas possibilidades de continuar aumentando o conhecimento’ [...] os objetivos podem definir ‘a natureza do trabalho, o tipo de problema a ser selecionado, o material a coletar (Cervo, 1978, p.49). Podem ser intrínsecos ou extrínsecos, teóricos ou práticos, gerais ou específicos, a curto ou a longo prazo” (MARCONI; LAKATOS, 2010, p.141).

#### **C4.Ambientação**

Ambientar: [ambientação é o ato de ambientar-se]: adaptar-se a um ambiente (FERREIRA, 1988, p.36). Ambiente: [...] lugar, sítio, espaço, recinto (FERREIRA, 1988, p.36).

“Tanto nos trabalhos teóricos quanto nos que se voltam para atividades práticas é importante que o pesquisador estabeleça limites no tempo e no espaço. Isso porque se torna impossível conhecer e analisar dados referentes a um período muito longo ou área muito extensa. O espaço físico deve precisa ser decididamente considerado” (MARCONI; LAKATOS, 2010, p.230).

#### **C5.Tipo de pesquisa**

“Atividades de pesquisa e desenvolvimento, isto é, de investigação científica e tecnológica; construção e desenvolvimento de protótipos; execução de projetos científicos e tecnológicos; combinação das atividades de investigação e desenvolvimento de novos produtos e processos” (CUNHA; CAVALCANTI, 2008, p.281).

“várias são as modalidades de pesquisa que se podem praticar, o que implica coerência epistemológica, metodológica e técnica, para seu adequado desenvolvimento. Pesquisa quantitativa, pesquisa qualitativa [...]; etnográfica [...]; participante [...]; pesquisa-ação [...]; estudo de caso [...]; análise de conteúdo [...]; bibliográfica, pesquisa documental, pesquisa experimental, pesquisa de campo [...]; exploratória, pesquisa explicativa” (SEVERINO, 2007, p.104-124).

#### **C6.Coleta de dados**

“Dados que serão processados; dados que serão submetidos a qualquer das funções básicas de processamento, como sejam: codificação, registro, cálculo, resumo e relatório, gravação e comunicação” (CUNHA; CAVALCANTI, 2008, p.113).

“Dentre os instrumentos utilizados de coleta de dados, podemos citar questionário, entrevista, formulário e painel” (SILVA, 2010, p.60).

“Técnica de coleta de informação: entrevista [...]; história de vida [...]; observação [...]; questionário” (SEVERINO, 2007, p.124-125).

“Etapa da pesquisa em que se inicia a aplicação dos instrumentos elaborados e das técnicas selecionadas, a fim de se efetuar a coleta de dados previstos; [...] são vários os procedimentos para a realização da coleta de dados, que variam de acordo com as circunstâncias ou com o tipo de investigação. Em linhas gerais, as técnicas de pesquisa são: coleta documental; observação; entrevista; questionário; formulário; medidas de opinião ou de atitudes; técnicas mercadológicas; testes; sociometria; análise de conteúdo; história de vida; [...] tabela [...] quadro [...] gráficos” (MARCONI; LAKATOS, 2010, p.149).

#### **C7.Métodos**

“Ordem que se deve impor aos diferentes passos necessários para chegar a um fim determinado. No conhecimento, o caminho que o caminho que se deve seguir para chegar à verdade nas ciências (Descartes). Conjunto ordenado de procedimentos que servem para descobrir o que se ignora ou para provar o que já se conhece. Processo sistemático de aquisição de conhecimento que segue uma série de passos interdependentes” (CUNHA; CAVALCANTI, 2008, p.248).

“Fazem-se necessários os seguintes questionamentos: como, com o que ou com quem, onde?; usa-se mais de um método e mais de uma técnica na realização da pesquisa” (SILVA, 2010, p.53).

“[Podem ser do tipo:] pesquisa bibliográfica [...]; documental [...]; experimental ou de laboratório [...]; *ex-post-facto* [...]; pesquisa-levantamento [...]; de campo [...]; estudo de caso [...]; pesquisa-ação [...]; participante [...]; exploratória [...]; descritiva [...]; explicativa [...]; técnica” (SILVA, 2010, p.53-60).

“É o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo – conhecimentos válidos e verdadeiros – traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista” (MARCONI; LAKATOS, 2010, p.65).

“” Métodos de procedimento seriam etapas mais concretas da investigação, com finalidade mais restrita em termos de explicação geral dos fenômenos e menos abstratas [...] método histórico [...]; comparativo [...]; monográfico [...]; estatístico [...]; tipológico [...]; funcionalista [...]; estruturalista [...]; etnográfico [...]; clínico [...]; quadro de referência” (MARCONI; LAKATOS, 2010, p.88-95).

### **C8.Fundamento teórico**

“Conjunto de regras e leis sistematicamente organizadas, que servem de base a uma ciência e que explicam os fatos” (CUNHA; CAVALCANTI, 2008, p.359).

“Síntese geral que se propõe a explicar um conjunto de fatos cujos subconjuntos foram explicados pelas leis” (SEVERINO, 2007, p.104).

“Expor os referenciais teórico-metodológicos, ou seja, os instrumentos lógico-categoriais nos quais se apoia para conduzir o trabalho investigativo e o raciocínio” (SEVERINO, 2007, p.131).

“A teoria serve como orientação para restringir a amplitude dos fatos a serem estudados – a quantidade de dados que podem ser estudados em determinada área da realidade é finita; [...] serve como sistema de conceptualização e de classificação dos fatos; [...] resumir sinteticamente o que já se sabe sobre o objeto de estudo, através das generalizações empíricas e das inter-relações entre afirmações comprovadas [...] o cientista está anunciando uma lei ao propor as regularidades que se apresentam uniformemente com as manifestações de uma classe de fenômenos; portanto, o “universo” de uma lei é limitado, abrangendo apenas determinada classe de fenômenos” (MARCONI; LAKATOS, 2010, p.99-107).

### **C9.Fundamento histórico/contextual**

“Contexto onde acontece o ciclo informacional e que influencia no seu fluxo e nas suas características, gerando diferenças nos tipos de informação: informação tecnológica, informação para negócios, informação científica, etc.” (CUNHA; CAVALCANTI, 2008, p.105).

“Documento que descreve fatos, ou a evolução de um fato e seus detalhes, em ordem cronológica” (CUNHA; CAVALCANTI, 2008, p.187).

“O método histórico consiste em investigar acontecimentos, processos e instituições do passado para verificar a sua influência na sociedade” (MARCONI; LAKATOS, 2010, p.89).

### **C10.Resultados**

“Consequência, efeito, seguimento”(FERREIRA, 1988, p. ).

“Evidências para a comprovação ou a refutação das hipóteses [...] segundo a relevância dos dados [...] evidenciar as conquistas alcançadas com o estudo; indicar as limitações e as reconsiderações” (MARCONI; LAKATOS, 2010, p.215).

“O desenvolvimento de produtos pode ser visto como um processo, o que significa realizar diversas atividades desde a geração do conceito até a descontinuidade do produto” (JUGEND; SILVA, 2005).

Rozenfeld *et al.* (2006): é um conjunto de atividades que são realizadas em uma seqüência lógica com o objetivo de produzir um bem ou serviço que tenha valor para um grupo específico de clientes.

## APÊNDICE F – Dados estatísticos do processo de extração de conceitos

	ITEM	C1	C2	C3	C4	C5	C6	C7	C8 C9	C10
1	d037-2003	X	X	X	X	X	X	X	X	
2	t013-2005	X	X	X	X			X		
3	d025-2005	X	X	X	X	X				X
4	d026-2005	X	X	X	X	X	X	X		
5	t010-2005	X	X	X	X			X		
6	d029-2005	X	X	X	X					X
7	d022-2006	X	X	X	X			X	X	
8	d021-2006	X	X	X	X	X	X	X	X	X
9	d023-2006	X	X	X	X	X	X	X	X	X
10	d020-2006	X	X		X					
11	t011-2005	X	X	X	X			X	X	X
12	d031-2005	X	X	X	X			X	X	
13	t014-2004	X	X	X	X			X		
14	d018-2007	X	X	X	X		X	X	X	
15	d038-2002	X	X	X	X		X			
16	d036-2004	X	X	X	X	X	X			
17	d030-2005	X	X	X	X		X			
18	d028-2005	X	X	X	X	X		X	X	
19	t015-2004	X	X	X	X			X	X	X
20	t012-2005	X	X	X	X			X	X	
21	d015-2007	X	X	X	X	X			X	
22	d017-2007	X	X	X	X	X		X	X	
23	t009-2007	X	X	X	X			X		
24	d016-2007	X	X	X	X	X	X	X	X	
25	d010-2008	X	X	X	X	X	X	X	X	
26	t006-2008	X	X	X	X		X			X
27	d024-2006	X	X	X	X			X	X	
28	d009-2008	X	X	X	X			X		
29	t001-2009	X	X	X	X	X	X	X	X	
30	d019-2006	X	X	X	X	X			X	
31	d007-2009	X	X	X	X			X	X	X
32	d008-2009	X	X	X	X	X	X	X		
33	t003-2008	X	X	X	X			X		X
34	d004-2009	X	X	X	X	X		X	X	
35	d014-2007	X	X	X	X	X	X			
36	d011-2008	X	X	X	X				X	
37	d012-2008	X	X	X	X	X		X	X	X
38	d027-2005	X	X			X	X			
39	d013-2008	X	X	X	X			X	X	
40	t002-2009	X	X	X	X				X	
41	t004-2008	X	X	X	X	X	X	X	X	
	<b>TOTAIS</b>	41	41	39	40	19	16	28	24	10
	<b>%</b>	100%	100%	95%	98%	46%	39%	68%	59%	25%

## APÊNDICE G – Termos de indexação dos resumos

1	AACR	45	Documentação científica
2	AIDS	46	Documentos arquivísticos
3	Análise de assunto	47	<i>Dspace</i>
4	Análise de citações	48	Educação física
5	Análise de clusters	49	Empresas de informática
6	Análise documentária	50	Entrevistas
7	Análise facetada	51	Estatística, mensuração
8	Análise fatorial	52	Estrutura profunda DEPAM
9	Arquitetura da informação	53	Estudo de caso
10	Arquivos	54	Estudos bibliométricos, cienciométricos e infométricos
11	Atendimento ao cliente	55	Estudos da produção e da produtividade científica
12	Atendimento em <i>call centers</i>	56	Estudos de autoria
13	Automação de unidades de informação	57	Estudos de canais, veículos, ciclos e modelos de comunicação
14	Avaliação de bases de dados	58	Estudos de citação
15	Avaliação de periódicos	59	Estudos de usuário
16	Avaliação de serviços e de unidades de informação	60	Estudos sobre fontes de informação
17	Avaliação de sites	61	Estudos sobre outros serviços e unidades de informação
18	Avaliação e desenvolvimento de coleções	62	Ferramenta de análise de citações
19	Balcão de informações ( <i>Help-Desk</i> )	63	Ferramenta de avaliação
20	Base de dados Qualis	64	Fundamentação epistemológica
21	Bases de dados	65	Gerência de recursos informacionais (GRI)
22	Bibliometria, ciencimetria, infometria	66	Gerência organizacional
23	Biblioteca (digital, virtual e eletrônica)	67	Gestão documental
24	Bibliotecas tradicionais	68	Gráficos e planilhas
25	Cabeçalhos de assunto	69	Hipertexto e hipermídia
26	Campo científico	70	História da arquivologia, da biblioteconomia, da documentação e da ciência da informação
27	Caracterização e comportamento do usuário	71	Incidente crítico
28	Catálogo/catalogação cooperativa	72	Indexação (automática e manual)
29	Catálogos <i>on-line</i>	73	Informação científica e tecnológica
30	Centros populares de documentação e comunicação	74	Informação textual (impresso e digital)
31	Ciência da Computação	75	Informática
32	Ciências físicas	76	Instituição bancária
33	Ciências sociais	77	Instituições de saúde
34	Classificação	78	Interação homem-computador
35	Classificação facetada	79	Interdisciplinaridade
36	Compatibilidade (linguagens, vocabulários, conteúdos)	80	Interfaces hipertextuais
37	Competência Informacional	81	Leis bibliométricas
38	Construção civil	82	Linguagens documentárias
39	Descrição Bibliográfica (FRBR)	83	Literatura (literatura infantil)
40	Desenvolvimento de coleções	84	Mapa Conceitual
41	Divulgação científica	85	Mapa Hipertextual MHTX
42	Divulgação científica - Bordieu	86	Mapas de tópicos
43	Divulgação científica - Knorr-Cetina	87	Mecanismos de busca ( <i>search engines</i> )
44	Divulgação científica - Kuhn		

88	Metacognição	129	Produção audiovisual
89	Metadados	130	Prontuário eletrônico
90	Metadados (Modelo metadados - novo)	131	Prontuário médico
91	Metadados Padrão Brasileiro de Metadados para Teses e Dissertações - MTD-BR	132	Protocolo verbal
92	Metadados Padrões Dublin Core	133	Publicações oficiais
93	Metadados Padrões MARC	134	Questionário
94	Método analítico-sintético	135	Recuperação da informação
95	Método de análise de conteúdo	136	Redes eletrônicas de informação
96	Método indiciário	137	Relações semânticas
97	Método Pejtersen (indexação)	138	Representação temática
98	Metodologia de modelagem de dados	139	Requisitos da especificação MoReq
99	Metodologia de pesquisa	140	<i>Scripts</i> de atendimento
100	Metodologia OGMA	141	Serviços de extensão bibliotecária
101	Metodologia WEKA	142	Serviços de saúde
102	Microfilmes	143	Setor privado
103	Modelagem conceitual	144	Setor público
104	Modelagem de dados	145	Setor turístico
105	Modelo entidade-relacionamento (MER)	146	Sintagmas nominais
106	Multidisciplinaridade	147	Sistema de Bibliotecas da UFMG – Pergamum
107	Norma ISO 2788	148	Sistema de informação
108	Norma ISO 704	149	Sistema de informação gerencial
109	Norma técnica (vigente e proposta)	150	Sistema de informação hospitalar
110	Ontologia	151	Sistemas de gerenciamento eletrônico de documentos (GED)
111	Organização do conhecimento	152	Sistemas de recuperação da informação
112	Outras questões teóricas	153	Sistemas para automação de unidades de informação
113	Outros sistemas e tecnologias de comunicação e informação	154	Sites
114	Pesquisa bibliográfica	155	Sociologia do conhecimento e da ciência
115	Pesquisa científica	156	<i>Software livre (OpenLibris)</i>
116	Pesquisa comparativa	157	Tecnologias da informação
117	Pesquisa descritiva	158	Tecnologias da informação arquivística
118	Pesquisa documental	159	Telecomunicação
119	Pesquisa empírica	160	Teoria da comunicação
120	Pesquisa experimental	161	Teoria do conceito
121	Pesquisa exploratória	162	Teorias e conceitos de Informação
122	Pesquisa qualitativa	163	Terminologia
123	Pesquisa teórica	164	Tesouro
124	Planejamento, organização e gerência de serviços e de unidades de informação.	165	Universidade (pública)
125	Portal de Periódicos Capes	166	Usabilidade
126	Prefeituras	167	Uso da informação e de unidade de informação
127	Prestadora de serviços de saúde	168	Uso e impacto das novas tecnologias de comunicação e informação
128	Prestadoras de serviço		

## APÊNDICE H – Metodologias de análise categorial temática

Apresentação de propostas de métodos de Análise de Conteúdo.

MORAES (1999)	OLIVEIRA (2008)
<p><b>Etapa 1:</b> Preparação das informações: identificação das amostras de informação para a análise compatíveis com o objetivo pretendido, e estabelecer um código para identificar os diferentes documentos analisados;</p> <p><b>Etapa 2:</b> Unitarização ou transformação do conteúdo em unidades: definição das “unidades de análise” para cada conteúdo significativo para atender ao objetivo proposto, que deve ser representado por um “termo-isolado”, e codificação sequencial, a partir do código de origem, isto é, dos diferentes documentos analisados;</p> <p><b>Etapa 3:</b> Categorização ou classificação das unidades em categorias: agrupar os dados, segundo o objetivo proposto, a partir de critérios semânticos (temáticos), sintáticos (verbos, adjetivos etc.), léxicos ou expressivos (problemas de linguagem). A categorização pode ser feita <i>a priori</i> ou a partir dos dados, devendo o conjunto obedecer às características de “válidas, exaustivas e homogêneas”, “mutuamente exclusivas” e “consistente”;</p> <p><b>Etapa 4:</b> Descrição: para pesquisa quantitativa, a descrição será feita através de tabelas e quadros, com frequências e percentuais. Para abordagem qualitativa, produzir um texto síntese para indicar o conjunto de significados, preferencialmente a partir de “citações diretas” dos dados originais;</p> <p><b>Etapa 5:</b> Interpretação: inferir (pesquisa quantitativa) se os resultados permitem generalizações a partir da amostra analisada, e interpretar (pesquisa qualitativa) por meio da compreensão dos resultados, através da fundamentação teórica a priori ou da criação da teoria a partir dos documentos analisados.</p>	<p><b>Etapa 1:</b> Pré-análise: escolha dos documentos ou definição do corpus de análise, com a formulação das hipóteses e dos objetivos da análise, além da elaboração dos indicadores que fundamentam a interpretação final;</p> <p><b>Etapa 2:</b> Exploração do material ou codificação: os dados são transformados sistematicamente e agrupados em unidades que permitem uma descrição exata das características pertinentes ao conteúdo dos documentos analisados;</p> <p><b>Etapa 3:</b> Tratamento dos resultados, inferência e interpretação: explicitar as informações obtidas na análise, através de quantificação simples (frequência) ou mais complexas, tal como a análise fatorial, de forma a apresentar os dados em diagramas, figuras, modelos etc. Outra forma, é uma descrição cursiva de cada categoria, seguidas de exemplificações de unidades de registros significativas, em forma de tabelas e gráficos, quadros com descrições cursivas ou outros.</p>

Fonte: Elaborado pela autora.

## APÊNDICE I – Outros resultados da análise do *corpus* da pesquisa

Para uma demonstração mais detalhada, a seguir são apresentados os gráficos da tabulação realizada em algumas categorias: C1. Tema, C2. Objeto, C7. Métodos, C5. Tipo de pesquisa, e C6. Coleta de dados, nessa ordem.

### Categoria C1: Tema

A categoria C1 representa 100% dos temas que são abordados nos estudos na linha de pesquisa OUI, que fazem parte do *corpus* desta pesquisa. Para a análise do GRÁFICO 3, a seguir, é preciso considerar que há duas vertentes na linha de pesquisa denominada Organização e Uso da Informação (OUI): uma vertente “organização da informação” e outra vertente “estudos de uso”.

GRÁFICO 3 – Relação dos temas abordados no corpus



Fonte: Elaborado pela autora.

Primeiramente, apresenta-se que em “outros temas” estão incluídos 11 temas, com uma ocorrência cada, que podem ser assim classificados: análise de assunto (organização), arquitetura da informação (organização), arquivos (organização), compatibilidade de linguagens (organização), desenvolvimento de coleções (uso), interdisciplinaridade (organização), linguagens documentárias (organização), sistemas de recuperação da informação (organização), tecnologias da informação (organização), outros sistemas de informação (organização) e metadados (organização).

Na vertente “organização” ainda estão incluídos os temas: representação da informação, indexação, bases de dados, recuperação da informação, organização do conhecimento e organização da informação. Na vertente de “uso” estão incluídos ainda os temas: “estudos de usuário” e “estatística e mensuração”.

QUADRO 14 – Síntese estatística dos temas

VERTENTE “ORGANIZAÇÃO”	%	VERTENTE DE “USO”	%
Análise de assunto, Arquitetura da informação, Arquivos, Compatibilidade de linguagens, Interdisciplinaridade, Linguagens documentárias, Sistemas de recuperação da informação, Tecnologias da informação, outros sistemas de informação e metadados	25		
Representação da informação	14		
Indexação	7	Estudos de usuário	27
Bases de dados	5	Estatística e mensuração	5
Recuperação da informação	5	Desenvolvimento de coleções	2
Organização do conhecimento	5		
Organização da informação	5		
TOTAL	66%	TOTAL	34%

Fonte: Elaborado pela autora.

Assim, a vertente “organização” totaliza 66% dos temas das pesquisas e a vertente “uso” totaliza 34% dos temas. Isso sinaliza que, na linha OUI, os temas da vertente “organização” são mais numerosos, o que pode ser compreendido pelo fato de que somente um, entre os nove docentes desta linha de pesquisa, se dedica, especialmente, aos estudos de uso da informação. Mas é preciso esclarecer que na linha de pesquisa “Informação, Cultura e Sociedade” (ICS) há investigações cujos temas abordam o “uso” de informação, principalmente em estudos de comportamento do usuário. Entretanto, o escopo desta pesquisa não abarcou resultados das outras linhas de pesquisa, ficando somente essa explanação como uma pressuposição, para, talvez, um estudo futuro.

### **Categoria C2: Objeto empírico**

A categoria C2 representa 100% dos objetos empíricos utilizados para investigação nos estudos na linha de pesquisa OUI que fazem parte do *corpus* desta pesquisa, apresentados no GRÁF. 4.

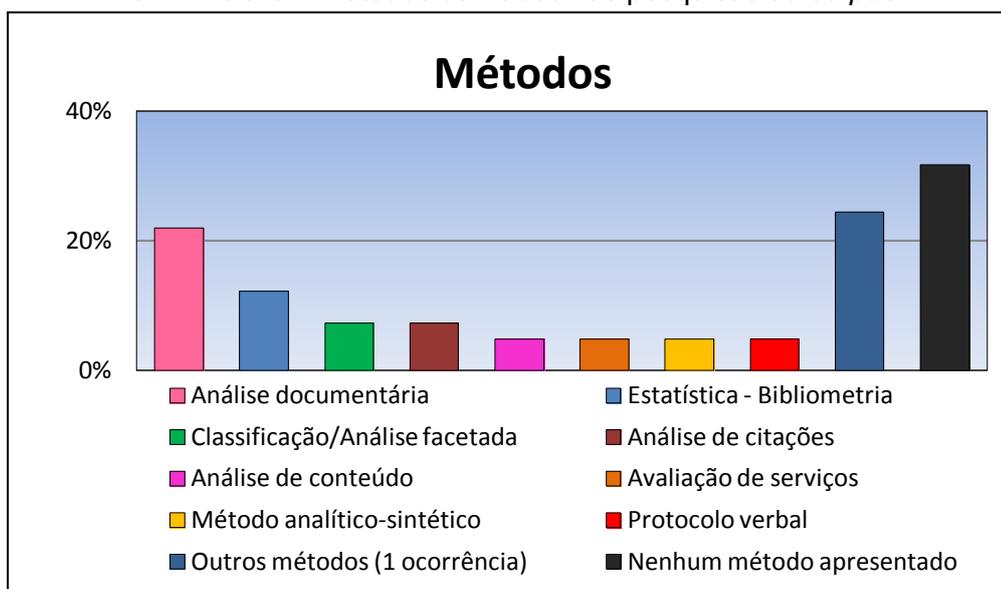
GRÁFICO 4 – Relação dos objetos empíricos utilizados no *corpus*

Fonte: Elaborado pela autora.

Observa-se bastante heterogeneidade de objetos empíricos, ainda que “documentação científica” e “informação textual (impresso e digital)” tenham alcançado o percentual de 27% do total. Entretanto, nesses dois itens estão incluídos as definições retiradas da literatura, referências bibliográficas, informações estratégicas, literatura especializada, termos de indexação de artigos científicos. Um ponto interessante a ser destacado é o fato de três trabalhos utilizarem os prontuários médicos para os estudos na linha de pesquisa OUI. Parece haver, atualmente, uma preocupação especial com o fluxo de informações médicas, principalmente após os hospitais e clínicas iniciarem a implantação de prontuários eletrônicos.

### **Categoria C7: Métodos**

A categoria C7 teve preenchimento para 28 documentos do *corpus*, o que equivale a 68% do total. É importante salientar que um mesmo documento muitas vezes indicou a utilização de mais de um método de pesquisa, conforme o GRÁFICO 5, a seguir.

GRÁFICO 5 – Métodos utilizados nas pesquisas do *corpus*

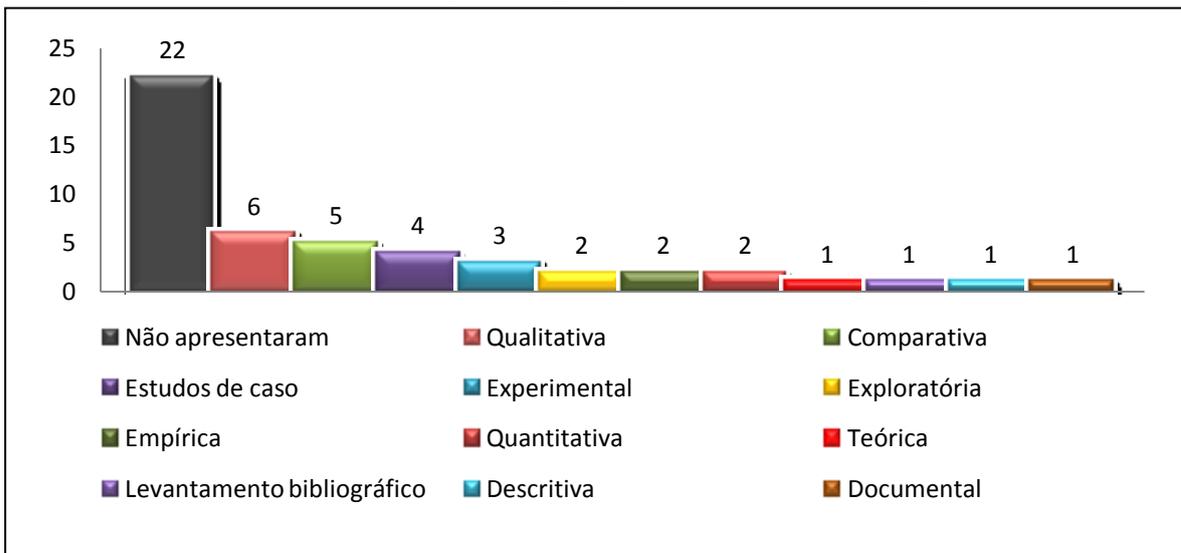
Fonte: Elaborado pela autora.

No item “outros métodos” há dez diferentes métodos, com uma ocorrência cada. Os métodos são: indiciário (citações), estrutura profunda DEPAm (análise de conteúdo), gestão documental (para arquivos), incidente crítico (uso/usuários), modelagem de dados (organização), modelagem conceitual (organização), OGMA (análise de texto), WEKA (mineração de dados), entidade-relacionamento (modelagem de dados) e requisitos MoReq (gestão de arquivos eletrônicos). O método da análise documentária e o método da classificação/análise facetada, que aparecem em primeiro e terceiro lugar de uso, respectivamente, são geralmente aplicados quando a pesquisa aborda a organização de informações. O segundo método mais utilizado é o de estatística/bibliometria, que geralmente é utilizado nas pesquisas de estudo de usuário, numa abordagem quantitativa, mas geralmente houve uma análise qualitativa a partir dos números encontrados. Foi interessante notar, que o método estatística/bibliometria foi utilizado em cinco trabalhos, mas nenhum dos resumos caracterizou a pesquisa como quantitativa. Parece haver um movimento no sentido de entender que pesquisas qualitativas trazem resultados mais interessantes do que os estudos quantitativos. Entretanto, acredita-se que a combinação de abordagens enriquece bastante os resultados de uma pesquisa, a quantitativa para validação dos dados e a qualitativa para a interpretação dos mesmos.

### **Categoria C5 e C6: Tipo de pesquisa e Coleta de dados**

Apesar de o preenchimento das categorias C5 e C6 ter sido possível em menos de 50% dos documentos, considerou-se interessante ter uma visão parcial dos tipos e dos instrumentos utilizados nas pesquisas. Por essa razão, esses dois gráficos estão apresentados por último e serão considerados como resultados parciais da análise.

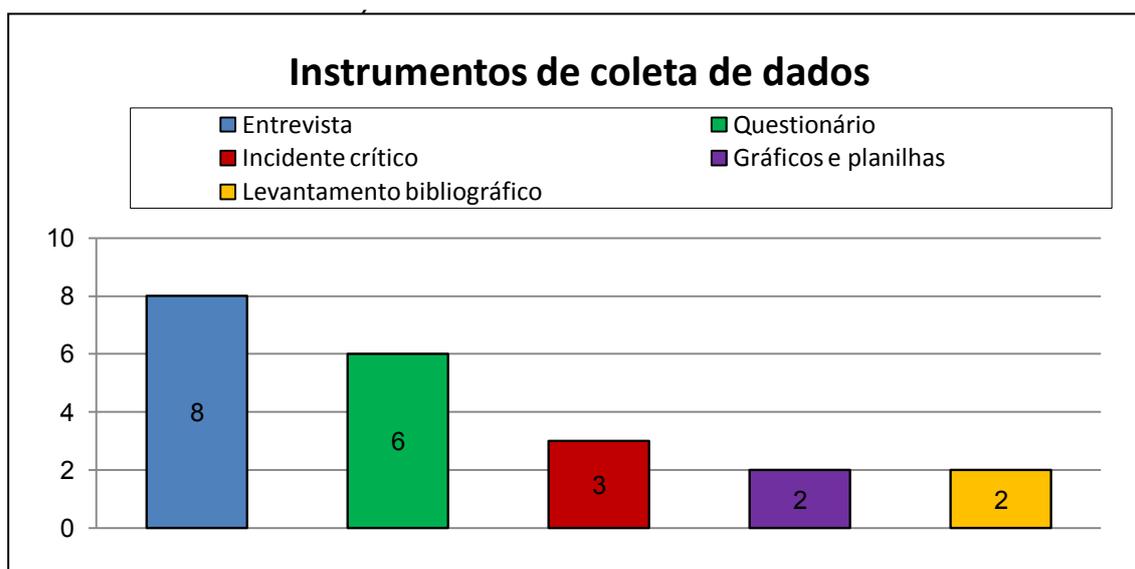
O GRÁFICO 6 mostra o resultado referente ao preenchimento do campo C5 de 19 documentos do *corpus*, que são 46% do total, indicando, na primeira linha, o número de resumos que não incluíram o tipo de pesquisa.

GRÁFICO 6 – Tipos das pesquisas referentes ao *corpus*

Fonte: Elaborado pela autora.

Esclarece-se que uma mesma pesquisa às vezes indica mais de um tipo de pesquisa, o que é bastante comum em estudos na área da CI. De acordo com o gráfico, prevalecem mesmo as pesquisas de cunho qualitativo de análise de dados, sobretudo quando trata de estudos de usuário. Outro dado interessante refere-se às pesquisas comparativas, pois nota-se que elas são utilizadas para indicar melhores formas de organização da informação, pois geralmente se recorre às comparações entre dois tipos de organizações diferentes, em um mesmo tipo de ambientação, para provar um ponto/fato qualquer. A alta incidência de “estudos de caso” parece indicar que as pesquisas de campo têm sido utilizadas para conhecer as conjunturas da sociedade. Entretanto, parece haver certa resistência da comunidade acadêmica a esse tipo de pesquisa, sobretudo porque geralmente são análises minuciosas sobre um único caso, prejudicando as generalizações tão desejadas pela maior parte dos pesquisadores.

E, finalmente, o GRÁF. 7 mostra o resultado parcial referente ao preenchimento do campo C6 de 16 documentos do *corpus*, que representam 39% do total.

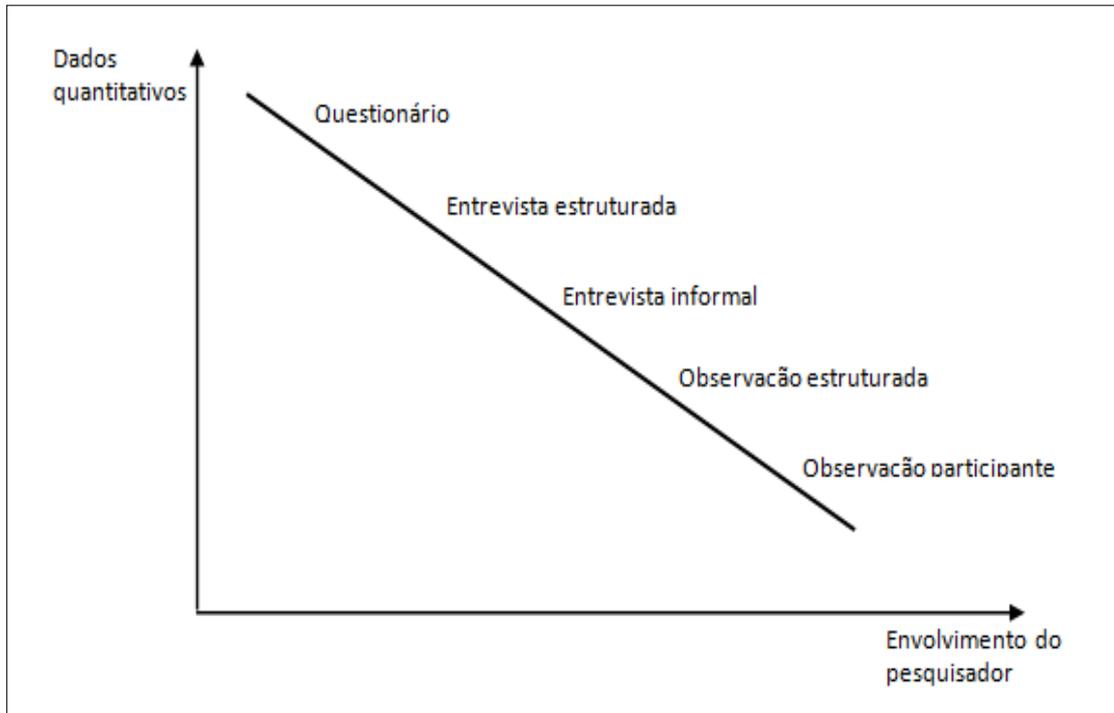


Fonte: Elaborado pela autora.

Por este resultado (parcial), nota-se que as entrevistas e os questionários são os instrumentos de coleta de dados mais utilizados. Além disso, é preciso esclarecer que em uma mesma pesquisa houve o uso concomitante de distintos instrumentos de coleta de dados.

Isso demonstra que o grau de envolvimento dos pesquisadores com os respondentes, conforme FIG. 19.

FIGURA 19 – Envolvimento do pesquisador com respondentes



Fonte: Cação (2008)<sup>81</sup>

Partindo da FIG. 19, observa-se que o pesquisador, geralmente, tem um grau de envolvimento com seus respondentes que fica entre médio a baixo, em uma abordagem mais positivista, ainda que haja uma análise qualitativa dos dados coletados.

Esse resultado, ainda que considerado parcial, indica que conhecer as particularidades das pesquisas pode ser muito importante para orientar futuras pesquisas na área da CI. Permite minimizar os retrabalhos, por exemplo, o que acontece na elaboração de questionários. Se for possível localizar facilmente os questionários utilizados nas pesquisas, isso pode evitar que se comece a elaboração de um questionário do zero.

<sup>81</sup> CAÇÃO, Rosário. Análise qualitativa de dados com o NVIVO. 31 de out. 2008. Disponível em: <<http://slidesha.re/11XyZq>>. Acesso em: 12 mar. 2011.

## ANEXOS

ANEXO A – Modelo de busca facetada: protótipo *TerveSuomi.fi*

terveysuomi.fi  
prototyyppi

Kansalaisen portti luotettavaan terveystietoon

suomeksi | på svenska | in English

HOME INDEX A-Z

**HealthFinland is a channel to a wide selection of reliable health information produced by Finnish research and expert organizations, government institutions and health organizations.**  
[Läs mera...](#)

**Topic areas**

<b>Food and nutrition</b> diet and eating habits, foodstuffs, nutrients...	<b>Intoxicants</b> alcohol, intoxicant problems, smoking...	<b>Conditions and symptoms</b> conditions, symptoms, treatment methods...
<b>Exercise</b> (hälsomotion), fitness and physical activity, sports...	<b>Catastrophes</b> chemical accidents, disasters, terrorism...	<b>Accidents and first aid</b> accident prevention, accidental falls, wounds and injuries...
<b>Weight control</b> body mass index, fatness, overweight...	<b>Epidemics and infectious diseases</b> bacteria, infectious diseases, vaccines...	<b>Finding help</b> care institutions, health centres, public health service...
<b>Sexuality</b> sexual health, sexual intercourse, sexually transmitted diseases...	<b>Mental health and sleep</b> depression, mental disorders, sleep and sleep disorders...	<b>Environment</b> air, residential environment, water...
<b>Family and children</b> breastfeeding, delivery, pregnancy...	<b>Violence and crises</b> marginalisation, violence...	<b>Occupational health</b> koulutus, toimintakyky, työ...

**Largest topics**  
(hälsomotion) adult type diabetes adults babies child  
diabetes dietetics health health promotion infectious diseases  
instructions juvenile type diabetes men physical  
training prophylaxis research risk factors smokers

**Index**  
0-9 A B C D E F G  
H I J K L M N O P  
Q R S T U V W X  
Y Z Å Ä Ö

**News**

- Cross-language searches improved**  
20.10.2008  
The portal now defaults to searching over materials in any language. Most of the material is in Finnish. Searches can still be restricted to a specific language using the language selection facility above the search results.
- Portaalin aineistoja päivitetty**  
18.9.2008  
Portaalin aineistoihin kuuluvat Savonia-ammattikorkeakoulun raskaus- ja vauva-aiheiset artikkelit on nyt päivitetty.  
  
Aineistot ovat vielä keskeneräisiä ja niitä kehitetään parhaillaan portaalin tuotantoversioon.  
  
Portaalin tuotantoversio julkaistaan huhtikuussa 2009 osoitteessa [terveysuomi.fi](#).
- TerveSuomi-portaalin prototyyppi julkistettu**  
12.9.2008  
Kansanterveyslaitoksen johdolla kehitteillä olevan kansalaisen terveystietoportaaln TerveSuomi.fi prototyyppi on valmistunut osana Tekesin rahoittamaa kansallista FinnONTO-kehityshanketta 2003-2010.  
  
[Julkistamislaisuuden ohjelma ja lisätietoja](#)

Disponível em: <<http://demo.seco.tkk.fi/terveysuomi/home>>.

## ANEXO B – Modelo de busca facetada: *Relation Browser Demo*

### Relation Browser Demo – Board Games

Prepared by Earl Bailey, INLS 818, Fall 2008; revised by Rob Capra  
 Data and links source: <http://www.boardgamegeek.com>

type search terms here to start a new search

Current Query

Facet List

Rank (344)	Players (2092)	Play Time (907)
<input type="button" value="Top 10 (10)"/>	<input type="button" value="1 (224)"/>	<input type="button" value="10 minutes or less (115)"/>
<input type="button" value="Top 100 (99)"/>	<input type="button" value="2 (2092)"/>	<input type="button" value="30 minutes or less (907)"/>
<input type="button" value="Top 200 (198)"/>	<input type="button" value="3 (1833)"/>	<input type="button" value="31-45 minutes (331)"/>
<input type="button" value="Top 500 (344)"/>	<input type="button" value="4 (1858)"/>	<input type="button" value="46-60 minutes (452)"/>
<input type="button" value="Bottom 100 (101)"/>	<input type="button" value="5 (1193)"/>	<input type="button" value="61-90 minutes (248)"/>
<input type="button" value="Bottom 10 (10)"/>	<input type="button" value="6 (884)"/>	<input type="button" value="91-120 minutes (239)"/>
	<input type="button" value="7 (309)"/>	<input type="button" value="More than 120 minutes (354)"/>
	<input type="button" value="8+ (274)"/>	

Results List

2531 results. 2531 searched. Showing 1-50.

[Puerto Rico](#)  
 The players are plantation owners in Puerto Rico in the days when ships had sail  
<http://www.boardgamegeek.com/boardgame/3076>

[Power Grid](#)  
 Power Grid is the updated release of the Friedemann Friese crayon game Funkensch  
<http://www.boardgamegeek.com/boardgame/2651>

[Twilight Struggle](#)

Fonte: Disponível em: <<http://id189.ils.unc.edu/rb07/eabailey818/>>.