

RENE FAUSTINO GABRIEL JUNIOR

SOCIALIZAÇÃO DOS SABERES: METODOLOGIA PARA DESENVOLVIMENTO DE
UM REPERTÓRIO TEMÁTICO EM CIÊNCIA DA INFORMAÇÃO

CURITIBA

2011

RENE FAUSTINO GABRIEL JUNIOR

SOCIALIZAÇÃO DOS SABERES: METODOLOGIA PARA DESENVOLVIMENTO DE
UM REPERTÓRIO TEMÁTICO EM CIÊNCIA DA INFORMAÇÃO

Dissertação apresentada ao Curso de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, Linha de Pesquisa Informação, Conhecimento e Estratégia, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientadora: Profa. Dra. Leilah Santiago Bufrem

CURITIBA

2011

Gabriel Junior, Rene Faustino

Socialização dos saberes: metodologia para desenvolvimento de um repertório temático em ciência da informação / Rene Faustino Gabriel Junior. – Curitiba, 2010. 180 f.: graf. (color), 29 cm.

Orientadora: Leilah Santiago Bufrem.

Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) – Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná.

1. Tratamento temático de informação. 2. Sistemas de recuperação de informação I. Título

CDD 025.066

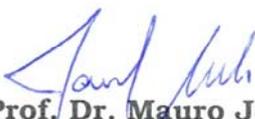
TERMO DE APROVAÇÃO

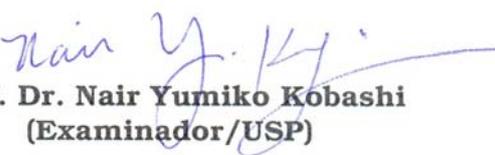
Rene Faustino Gabriel Jr.

**“SOCIALIZAÇÃO DOS SABERES: METODOLOGIA PARA
DESENVOLVIMENTO DE UM REPERTÓRIO TEMÁTICO EM CIÊNCIA DA
INFORMAÇÃO”**

**DISSERTAÇÃO APROVADA COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO GRAU DE MESTRE NO PROGRAMA DE PÓS-
GRADUAÇÃO EM CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO
DA UNIVERSIDADE FEDERAL DO PARANÁ, PELA SEGUINTE BANCA
EXAMINADORA:**


Prof.^a Dr.^a Leilah Santiago Bufrem
(Orientadora/UFPR)


Prof. Dr. Mauro José Belli
(Examinador/UFPR)


Prof. Dr. Nair Yumiko Kobashi
(Examinador/USP)

16 de fevereiro de 2011

AGRADECIMENTOS

À professora Leilah Santiago Bufrem por sua orientação, seu apoio e pelo incentivo na entrada para a vida acadêmica, agradecendo principalmente pela liberdade no desenvolvimento e correções de rumos tão essenciais nesta fase de formação docente, condição básica e necessária para transmissão do saber.

À professora Nair Yumiko Kobashi e ao professores José Augusto Chaves Guimarães e Mauro José Belli pelas contribuições dadas no momento da qualificação.

Agradecimento novamente ao professor Mauro José Belli, como chefe do departamento e responsável pela manutenção e funcionamento do servidor da Brapci que tanto foi requisitado em diversas etapas do desenvolvimento da pesquisa.

À Viviane Gonçalves, colega de pós-graduação e de trabalho que foi parceira em diversos trabalhos acadêmicos, por sua amizade e trocas de conhecimentos em toda nossa trajetória.

Aos colegas mestrandos Juliana Lazzarotto Freitas, Leonardo Cossio, pela ajuda e contribuições feitas quanto ao conteúdo e forma do trabalho.

Aos colegas do grupo de estudo E3PI da graduação que tanto trabalharam para concretização da Brapci, Marcel Ferradas Quadros, Camila Gabrieli Araujo, Caroline Munhoz Corrêa, Daniel Cerqueira Costa, Francisco Daniel de Oliveira Costa, Kaoane Stival Paula, Lauro César Mendes de Oliveira e a Tidra Viana Sorribas.

Agradecimento especial a minha esposa Viviane de Fátima Tulio por sua compreensão em tempos de acúmulos de trabalhos, cuidando de nossos filhos André Vitor Gabriel e Eduardo Augusto Gabriel. E é com eles que reparto a alegria e a satisfação do trabalho concluído. Muito obrigado.

RESUMO

Proposta de metodologia multidisciplinar para construção de bases de dados temáticas com dispositivos plurifuncionais, fundamentada no arcabouço teórico das áreas de Ciência da Informação, Ciência da Computação e Gestão da Informação. Constrói a proposta a partir da necessidade de organização da informação demandada pela ampliação significativa de informações provenientes da evolução das tecnologias de informação e políticas de acesso aberto. Fundamenta a organização do conhecimento nos princípios da Ciência da Informação, analisando teorias que justificam a necessidade de regras e padrões com a utilização de vocabulário controlado para a construção de ferramentas de apoio à organização e ao tratamento temático da informação, como terminologias, tesauros e ontologias. Gera uma metodologia, efetivada pela construção de um protótipo, consolidando um modelo empírico para socialização dos saberes, validado em uma realidade concreta com o estudo de caso na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). O resultado é favorável à aplicação da metodologia por possibilitar o acesso à comunidade acadêmica a um rol de informações nacionais reunidas em um único local, bem como subsidiar dados para análises do comportamento e das tendências da área temática.

Palavras-chave: Recuperação de informação. Tratamento temático de informação. Organização de informação. Base de dados.

ABSTRACT

GABRIEL JUNIOR, Rene Faustino. **Socialization of domain knowledge: a proposition for developing a methodology founded on reality.** 180 f. Dissertation (Master of Science, Management and Information Technology) – Universidade Federal do Paraná, Curitiba, 2011.

A proposition to a multidisciplinary methodology to the development of thematic databases with multifunctioning devices, founded on the theoretical framework from the fields of Information Science, Computer Science, and Information Management. It develops the proposition from the need for an organization of information which is demanded by the significant enlargement of the amount of informations deriving from the evolution of information technologies and open access policies. It finds the organization of knowledge on the principles of Information Science, analyzing theories that justify the necessity of rules and standards with the use of a controlled vocabulary for the development of instruments that give support to the organization and thematic treatment of information, such as terminologies, thesaurus, and ontologies. It generates a methodology, that is made effective by the construction of a prototype, structuring an empirical model for the socialization of knowledge, which is validated by the concrete reality according to a case study from the Brazilian Database for Periodicals on the Domain of Information Science (BRAPCI). The result is favorable to the use of this methodology, since it allows the academic community a range of national informations gathered in a single site, as well as it provides data for the analysis of the behaviors and tendencies of this thematic field.

Keywords: Information retrieval. Thematic treatment of information. Organization of information. Database.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - DIAGRAMA DA ESTRUTURA DO ESTUDO	20
FIGURA 2 - DIAGRAMA DA ESTRUTURA DA FUNDAMENTAÇÃO TEÓRICA	22
FIGURA 3 - NECESSIDADE, INTERAÇÃO, USO DA INFORMAÇÃO E USABILIDADE	28
FIGURA 4 - ÁRVORE DE PORFÍRIO	30
FIGURA 5 - ÁRVORE BANIANA DE RANGANATHAN	32
FIGURA 6 - ESQUEMA DE RELAÇÕES HIERÁRQUICAS	34
FIGURA 7 - ESQUEMA DE CLASSIFICAÇÃO GENÉRICA.....	35
FIGURA 8 - ESQUEMA DE RELAÇÃO PARTITIVA	35
FIGURA 9 - RELAÇÃO SEMÂNTICA E OS INSTRUMENTOS DE CADA CATEGORIA	38
FIGURA 10 - MODELO DE COMUNICAÇÃO CIENTÍFICA DE GARVEY E GRIFFITH.....	61
FIGURA 11 - MODELO DE COMUNICAÇÃO CIENTÍFICA DE HURD	62
FIGURA 12 - ETAPAS DO DESENVOLVIMENTO DO PRODUTO	83
FIGURA 13 - MODELO DE RELACIONAMENTO DE DADOS	86
FIGURA 14 - REPRESENTAÇÃO GRÁFICA DA METODOLOGIA BRAPCI.....	90
FIGURA 15 - METODOLOGIA ITERATIVA DE DESENVOLVIMENTO DE SISTEMA	91
FIGURA 16 - WIREFRAME DO MÓDULO DE MANUTENÇÃO	93
FIGURA 17 - WIREFRAME DO MÓDULO PÚBLICO	93
FIGURA 18 - FLUXO DA LINGUAGEM NATURAL PARA CONTROLADA – INDEXAÇÃO PÓS-COORDENADA PROPOSTA PELA METODOLOGIA	104
FIGURA 19 - PROCESSO PROPOSTO PARA INDEXAÇÃO E MANUTENÇÃO DA BRAPCI.....	108
FIGURA 20 - MODELO DE COLETA DE DADOS OAI-PMH.....	109
FIGURA 21 - PROCESSO DE CONFERÊNCIA DA METODOLOGIA BRAPCI.....	118
FIGURA 22 - <i>LOGIN</i> DO MÓDULO MANUTENÇÃO	131
FIGURA 23 - TELA INICIAL DO MÓDULO DE MANUTENÇÃO.....	132
FIGURA 24 - TELA DE CONTROLE DE EDIÇÃO DA REVISTA.....	133
FIGURA 25 - TELA DE CONTROLE DE EDIÇÃO DA REVISTA COM APRESENTAÇÃO DOS TRABALHOS	133
FIGURA 26 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE I	134
FIGURA 27 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE II	134
FIGURA 28 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE III	135
FIGURA 29 - TELA DE CADASTRO DE AUTOR	136
FIGURA 30 - TELA DE CADASTRO DE PALAVRAS-CHAVE.....	137
FIGURA 31 - RESUMO DO PROCESSO DE INDEXAÇÃO	137
FIGURA 32 - FERRAMENTAS DO SISTEMA.....	138
FIGURA 33 - FERRAMENTAS DO SISTEMA.....	139
FIGURA 34 - RESULTADO DA BUSCA	140

FIGURA 35 - RELAÇÃO DE CONCEITOS DA EPISTEMOLOGIA EM CIÊNCIA DA INFORMAÇÃO	141
FIGURA 36 - INSTRUMENTO DE RELACIONAMENTO DE CONCEITOS NA BRAPCI.....	143
FIGURA 37 - TESAURO BRAPCI, COM ONTOLOGIA NAVEGACIONAL	144
GRÁFICO 1 - CRESCIMENTO DA BASE EM NÚMERO DE TRABALHOS E ARTIGOS POR MÊS.....	145
GRÁFICO 2 - PRODUÇÃO DE TRABALHOS POR ANO NA BRAPCI.....	146
GRÁFICO 3 - TÍTULOS DE REVISTAS EM ATIVIDADE POR ANO	149
GRÁFICO 4 - QUANTIDADE ARTIGOS PUBLICADOS POR ANO EM COAUTORIA.....	150
GRÁFICO 5 - INCIDÊNCIA DA TEMÁTICA “REDES SOCIAIS” NA LITERATURA PERIÓDICA EM CI	151
GRÁFICO 6 - INCIDÊNCIA DA TEMÁTICA “REDES SOCIAIS” NA LITERATURA PERIÓDICA EM CI POR PUBLICAÇÃO.....	152
GRÁFICO 7 - REDE SOCIAL POR AFINIDADE TEMÁTICA	153
GRÁFICO 8 - QUANTIDADE DE ACESSOS A BASE POR NÚMERO DE USUÁRIOS	154
FIGURA 38 - LOCALIZAÇÃO GEOGRÁFICA DOS USUÁRIOS DA BRAPCI	155

LISTA DE QUADROS

QUADRO 1 - RELACIONAMENTO DOS TESAUROS.....	44
QUADRO 2 - LISTA DE VERBOS OAI-PMH	74
QUADRO 3 - SEÇÕES DAS PUBLICAÇÕES DISPONÍVEIS NA METODOLOGIA BRAPCI	99
QUADRO 4 - REMISSIVAS DE NOMES DE AUTOR	106
QUADRO 5 - PERFIS DOS USUÁRIOS DA BASE.....	107
QUADRO 6 - ATIVIDADES PERMITIDAS A CADA PERFIL DE USUÁRIO	108
QUADRO 7 - LINK PARA ACESSO À REVISTA CIÊNCIA DA INFORMAÇÃO PARA COLETA DE DADOS COM O PROTOCOLO OAI-PMH.....	110
QUADRO 8 - RESPOSTA DA CONSULTA DA REVISTA PELO PROTOCOLO OAI-PMH.....	111
QUADRO 9 - DESCRITORES DA CIÊNCIA DA INFORMAÇÃO	112
QUADRO 10 - REPRESENTAÇÃO GRÁFICA DOS STATUS DOS FASCÍCULOS	117
QUADRO 11 - QUADRO DE ATRIBUIÇÃO DE PONTOS PARA RESULTADO DA PESQUISA.....	121
QUADRO 12 - LISTA DE PUBLICAÇÕES DISPONÍVEIS NA BRAPCI	148

LISTA DE TABELAS

TABELA 1 - TOTAL DE TRABALHOS / ARTIGOS NA BRAPCI	145
TABELA 2 - PRODUÇÃO DE ARTIGOS POR ANO NA BRAPCI	146
TABELA 3 - NÚMERO DE REVISTAS E FASCÍCULOS PUBLICADOS POR ANO	149
TABELA 4 - TIPOS DE AUTORIAS NA BRAPCI	150
TABELA 5 - ACESSO À BRAPCI POR CIDADE.....	155

LISTA DE SIGLAS

AA	- Arquivos Abertos
ABNT	- Associação Brasileira de Normas Técnicas
ANSI	- <i>American National Standards Institute</i>
ARS	- Análise de Redes Sociais
BRAPCI	- Base de Dados Referências de Artigos de Periódicos de Ciência da Informação
BSCSA	- Biblioteca do Setor de Ciências Sociais Aplicadas
Capes	- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	- Ciência da Computação
CDD	- Código Decimal de Dewey
CDU	- Código Decimal Universal
CI	- Ciência da Informação
CNPq	- Conselho Nacional de Desenvolvimento Científico e Tecnológico
DC	- Dublin Core®
DCMI	- Dublin Core® Metadata Initiative
DER	- Diagrama entidade de relacionamento
E3PI	- Grupo de Estudo Educação, Pesquisa do Profissional da Informação
IBICT	- Instituto Brasileiro de Informação em Ciência e Tecnologia
IR	- <i>Information Retrieval</i>
LC	- Linguagem Controlada
LD	- Linguagem Documentária
LISA	- <i>Library and Information Science Abstracts</i>
LN	- Linguagem Natural
MB	- Metodologia Brapci
ME	- Módulo Pesquisador
MIT	- Massachusetts Institute of Technology
MM	- Módulo de Manutenção
MP	- Módulo Público
OA	- <i>Open Archives</i>
OAI	- <i>Open Archives Initiative</i>
OAI-PMH	- <i>Open Archives Initiative - Protocol for Metadata Harvesting</i>

OCR	- <i>Optical character recognition</i>
OO	- Orientados a objetos
OWL	- <i>Ontology Language Overview-</i>
PDF	- <i>Portable Document Format</i>
RBU	- Repertório Bibliográfico Universal
RDF	- <i>Resource Description Framework</i>
RI	- Recuperação da Informação
SEER	- Sistema Eletrônico de Editoração de Revistas
SGBD	- Sistema Gerenciador de Banco de Dados
SQL	- <i>Structured Query Language</i>
SRI	- Sistema de Recuperação de Informação
TCT	- Teoria Comunicativa da Terminologia
TGT	- Teoria Geral de Terminologia
TIC	- Tecnologia da Informação e Comunicação
TTI	- Tratamento Temático da Informação
UML	- <i>Unified Modeling Language</i>
VC	- Vocabulário Controlado
W3C	- <i>World Wide Web Consortium</i>
WWW	- <i>World Wide Web</i>
XML	- <i>Extensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO	14
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 RECUPERAÇÃO DA INFORMAÇÃO	23
2.2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO	29
2.2.1 Relacionamento hierárquico ou de subordinação	34
2.2.2 Relações não hierárquicas ou horizontais	36
2.2.3 Semântica	37
2.2.3.1 Glossário	39
2.2.3.2 Taxonomia	40
2.2.3.3 Tesouros	42
2.2.3.4 Ontologia	44
2.3 TRATAMENTO TEMÁTICO DA INFORMAÇÃO	46
2.3.1 Linguagem natural e vocabulários controlados	49
2.3.1.1 Linguagem natural (LN) ou livre	50
2.3.1.2 Vocabulário controlado (VC)	51
2.3.2 Terminologia	53
2.3.2.1 Teoria do Conceito Dahlberg	55
2.3.2.2 Teoria Geral da Terminologia (TGT)	56
2.3.2.3 Teoria Comunicativa da Terminologia (TCT)	57
2.4 COMUNICAÇÃO CIENTÍFICA	58
2.4.1 Periódico científico	64
2.4.2 Base de dados	65
2.4.4 Tecnologias	70
2.4.4.1 <i>Unified Modeling Language</i> (UML)	71
2.4.4.2 Iniciativa de Arquivos Abertos	72
2.4.4.3 Interoperabilidade	75
2.4.4.4 <i>eXtensible Markup Language</i> (XML)	76
3 PROPOSTA METODOLÓGICA	80
3.1 PLANEJAMENTO	82
3.2 POLÍTICA DE DESENVOLVIMENTO DE COLEÇÕES	94
3.2.1 Política de indexação	95
3.2.2 Política de desenvolvimento de acervo	98
3.4 PERFIS DOS USUÁRIOS DA BASE	107
3.5 COLETA DE FASCÍCULOS E ARTIGOS CIENTÍFICOS	108
3.6 ESTRUTURA E INFRAESTRUTURA PARA O SERVIDOR WEB	112
3.7 SISTEMA GERENCIADOR DE BANCO DE DADOS (SGBD)	113
3.8 GESTÃO DE PROCESSOS DA METODOLOGIA BRAPCI	116
3.9 NUVEM DE TAGS	118
3.10 CONTROLE E REGISTRO DE ACESSO	120
3.11 METODOLOGIA DE RELEVÂNCIA DE RECUPERAÇÃO POR PONTOS	121
3.12 POSSIBILIDADES DE ANÁLISE	122
3.12.1 Análise de citação	123
3.12.2 Análise de colaboração da produção de trabalhos	125
3.12.3 Análise da produção de autores	126
3.12.4 Análise de redes sociais	126
3.12.5 Análise da metodologia da pesquisa	127

4 APLICAÇÃO EM UMA REALIDADE – ESTUDO DO CASO BRAPCI	128
4.1 IMPLEMENTAÇÃO DO PROTÓTIPO	130
4.1.1 Módulo de manutenção	131
4.1.2 Módulo Público	138
4.1.3 Módulo Pesquisador.....	141
4.2 VOCABULÁRIO CONTROLADO NA BRAPCI	142
4.3 DADOS BIBLIOMÉTRICOS E CIENTOMÉTRICOS.....	144
4.3.1 Análise de cooperação científica.....	149
4.3.2 Análise de conteúdo	151
4.3.3 Análise de redes sociais.....	152
4.4 AVALIAÇÃO E RESULTADOS	153
4.4.1 Monitoramento de acesso	153
4.4.2 Usabilidade do sistema	156
4.5 MANUAL DE POLÍTICA DE COLEÇÃO E INDEXAÇÃO	157
4.6 DIGITALIZAÇÃO DE DOCUMENTOS	157
5 CONSIDERAÇÕES FINAIS	159
REFERÊNCIAS	162
APÊNDICES.....	172

1 INTRODUÇÃO

A pesquisa científica, para alcançar sua efetividade social e cumprir seu papel de socialização de conhecimento precisa, segundo Le Coadic (2004) e Meadows (1999), ser comunicada, e essa comunicação deve ser realizada utilizando processos formais para que o conhecimento científico seja validado pelos pares ou pela sociedade.

Neste processo especial de comunicação os periódicos científicos sempre foram importantes para o desenvolvimento da ciência, desde seu surgimento em meados do século XVII até os tempos atuais. Durante o final do século XX, os periódicos científicos ganharam um novo elemento disseminador com as inovações das áreas de Tecnologia da Informação e Comunicação (TIC). Essas inovações proporcionaram mecanismos de apoio aos cientistas em sua produção, disseminação, compartilhamento e arquivamento do conhecimento científico (MÁRDERO ARELLANO; FERREIRA; CAREGNATO, 2005).

Desde o final da Segunda Guerra Mundial, a produção científica teve um acréscimo no número de artigos publicados, proporcionando o surgimento de novos títulos de periódicos. A super produção de trabalhos científicos trouxe uma preocupação para os cientistas da época, no que diz respeito à organização, análise e definição de critérios de qualidade da produção. Por outro lado, a ampliação da qualidade dessa produção pode nortear as bibliotecas na definição do planejamento de assinaturas de periódicos, orientado pelo seu uso e representatividade para a área. A Ciência da Informação (CI) emerge então como uma área para fundamentar e criar teorias sobre a produção, disseminação e uso do conhecimento explicitado nas publicações periódicas, livros e literatura cinzenta.

As bibliotecas, perante o volume crescente no número de títulos de periódicos científicos, não mais dispunham de recursos para catalogar e indexar todos os títulos dentro de seu acervo, uma vez que todo o material intelectual produzido pelos cientistas careceu de uma forma mais eficiente de armazenamento, compartilhamento e recuperação dessas informações para, assim, facilitar seu uso. As bases de dados de catálogos de coleções surgem como nova forma de organização e acesso à informação, agregando valor ao trabalho do usuário (pesquisador) e possibilitando maior autonomia no processo de busca. O sistema de

busca em bases de dados permite pesquisar em diversos títulos de periódicos com uma única estratégia, possibilitando a interação do pesquisador diretamente com a fonte de informação, sem a necessidade de intermediários.

Entretanto, ocorre eventualmente um interesse comercial na organização e gestão dessas bases de dados por empresas que não visam à socialização do conhecimento, mas uma forma de explorar comercialmente a produção intelectual no que tange à obtenção de lucros. Grandes bases de dados temáticas comerciais foram criadas e revendidas a um alto preço por meio de assinaturas para as próprias universidades produtoras e consumidoras de conhecimentos. Desse modo, a instituição ou o pesquisador paga para gerar a pesquisa, paga para publicar e depois também paga para acessar a informação publicada.

Em 1999, um movimento internacional começou a se mobilizar contra a mercantilização e a acumulação capitalista da informação (BOLAÑO, 2000). Este movimento internacional é iniciado pelos produtores de conhecimento em prol dos *Open Archives (OA)* ou Arquivos Abertos (AA). Procurando incitar a proposição de uma política voltada para a disseminação social do conhecimento, o movimento tem como proposta a disponibilização de forma integral e gratuita da publicação científica produzida nas universidades. No âmbito dessa proposta, foi criada a *Open Archives Initiative (OAI)*, organização que estabeleceu padrões, protocolos e ideais para a construção dos repositórios abertos, possibilitando a troca de informações entre provedores de dados e serviço.

Esse movimento ganhou força Internacional, e o Brasil, por intermédio do Instituto Brasileiro de Ciência, Informação e Tecnologia (IBICT), foi convidado a aderir formalmente a essa iniciativa que, de acordo com Kuramoto (2006), por problemas políticos e administrativos não pôde participar. Porém, somente em 2005, com uma nova diretoria no IBICT, o Brasil subscreveu a “Declaração de Berlim”, que tem como ideal disseminar o conhecimento, cuja missão

[...] estará incompleta se a informação não for tornada rapidamente acessível e em larga escala à sociedade. Novas possibilidades de difusão do conhecimento, não apenas através do método clássico, mas também, e cada vez mais, através do paradigma do acesso livre via Internet devem ser apoiadas. Nós definimos o acesso livre como uma fonte universal do conhecimento humano e do patrimônio cultural que foi aprovada pela comunidade científica. Para concretizar esta visão de uma representação global e acessível do conhecimento, a Web do futuro tem de ser sustentável, interativa e transparente. Conteúdos e ferramentas de software devem ser livremente acessíveis e compatíveis (KURAMOTO, 2010).

Com a subscrição da “Declaração de Berlim” passaram a ser desenvolvidas no Brasil políticas governamentais de incentivo à produção de periódicos científicos dentro dessas diretrizes, proporcionando financiamento de servidores, capacitação e suporte à criação de títulos que tivessem como fundamento a socialização do conhecimento e o livre acesso.

Houve um incremento significativo na disseminação da produção dos pesquisadores em periódicos em arquivos abertos, porém a maioria das iniciativas de disponibilização de informação concentrou-se em fontes primárias (provedor de dados), com a criação e migração do papel para dualidade (papel e eletrônica) de revistas científicas e bancos de teses e dissertações. Foi inexpressiva a política desenvolvida para produção de fontes terciárias (provedor de serviço), com exceção da adoção de “pacotes prontos” de desenvolvedores internacionais. Esses “pacotes” pouco agregaram ao desenvolvimento da pesquisa para os cientistas, por suas opções restritas a grandes repositórios de dados com interfaces pouco intuitivas, cujo fundamento para o seu desenvolvimento foi a Ciência da Computação (CC), ignorando-se, dessa forma, muitas demandas inerentes da CI.

Em estudo preliminar, utilizando-se o software *ProCite5*, foi possível a formação de um *corpus* com a concentração de registros do conhecimento em uma única base de dados. Porém, esta construção acarretou sérias restrições de acesso e manutenção, principalmente por seu conteúdo ser restrito a um único computador (monousuário), o que impossibilitava a indexação e catalogação por mais de um usuário simultaneamente. No final de 2007, um projeto de pesquisa aprovado e fomentado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) possibilitou a aquisição de um servidor de hospedagem exclusivo para compartilhar os dados do *ProCite5* para a comunidade. Esse projeto intitulado de Base de Dados Referenciais de Artigos de Periódicos de Ciência da Informação (Brapci) foi a primeira versão da plataforma *online*, que proporcionou a conversão dos dados do *ProCite5* para um Sistema Gerenciador de Banco de Dados (SGBD) com busca em campos específicos. O uso de duas plataformas acarretava problemas como a exportação dos dados, que era realizada manualmente e precisava ser repetida periodicamente por um técnico especializado em informática.

Com a publicação da primeira versão, observou-se ser necessário ampliar consulta de textos em ambiente digital. Seria imperativa uma pesquisa

fundamentada na definição de conceitos, métodos, critérios e demarcações, assim como para o desenvolvimento de uma metodologia capaz de sustentar o crescimento do *corpus* da Brapci. A percepção de que bases de dados dessa natureza não são concretizações estáticas, passíveis de uma análise sincrônica, mas se constituem como organismos em constante desenvolvimento, com a incorporação de novos fascículos de periódicos, fundamenta-se na permanente supervisão e controle das formas de entrada de autores e descritores, de modo a garantir a qualidade necessária aos dados disponibilizados para a sociedade.

Este estudo visa oferecer contribuição ao campo da CI, tanto pelo objeto de investigação em si, como pelo caráter metodológico que ela adquire, pautada pela interdisciplinaridade com a CC e a Gestão da Informação (GI). Também é possível encontrar relevância nessa temática sob a ótica de objetivos formativos mais imediatos e/ou pragmáticos: pela contribuição ao desenvolvimento de repositórios temáticos específicos, facilitando o acesso à pesquisa, e pelo interesse na discussão de elementos oriundos do relacionamento entre a CI e a CC, importantes para ambas as áreas, mas principalmente para a socialização de saberes.

No cenário atual, as informações estão dispersas, o usuário tem acesso a milhares de informações em seus mais variados suportes, esta enorme quantidade de informações inviabiliza a organização de todo o conhecimento delas decorrentes utilizando uma única metodologia. Somente com o auxílio da tecnologia é possível uma gestão eficiente dos processos de seleção, processamento e recuperação de informação para atender as necessidades dos usuários. Rowley (2002) ressalta que um serviço eficaz voltado para os clientes depende não apenas do atendimento prestado, mas do sistema em que se apóia a organização desses serviços.

A construção de repositórios temáticos é relevante porque proporciona a concentração de áreas de afinidade em um único local, e esta reunião de dados pertinentes possibilita, com as ferramentas adequadas, a análise de informações por meio de uma ampla gama de possibilidades ao pesquisador. Além disso, ressalta-se a facilidade para o usuário na recuperação de informações em uma única fonte.

Alvarenga (2006, p. 90) contrapõe-se aos argumentos contrários ao uso de métodos da Biblioteconomia para o tratamento de documentos na Internet, principalmente aventados na comunidade da informática e da eletrônica por considerarem que a Biblioteconomia e a CI têm suas teorias voltadas para a organização física de acervos. A autora argumenta que os defensores dessas ideias

desconhecem o trabalho de pesquisa subjacente à construção e atualização dos sistemas de classificação tradicional e que existem teorias que podem auxiliar a classificação e indexação do mundo virtual, sendo uma delas a teoria da classificação facetada de Ranganathan (1967). Esta classificação é parte integrante das teorias da CI, mas pouco utilizada para organização de informação no Ocidente, embora possibilite estruturação semântica dos conceitos constantes em documentos, o que viria ao encontro das necessidades para organização de documentos digitais.

A CC, por meio de algoritmos e concepções de Inteligência Artificial, defendia a ideia de que grandes massas de dados e informações podem ser processadas, organizadas e disponibilizadas aos usuários conforme sua necessidade informacional. Mas o que se observou é que esse tipo de ferramenta tornou-se limitada, principalmente por tentar gerar conhecimento pelo processamento de caracteres computacionais (dados binários), ignorando o contexto dos assuntos, o texto como uma unidade de sentido e as necessidades informacionais do usuário, tornando-se um instrumento de busca por caracteres computacionais, redundante e de baixa relevância.

O que Alvarenga (2006, p. 91) quer demonstrar é que a CI e a Biblioteconomia têm condições de desenvolver teorias e práticas para a organização do conhecimento na Internet, mas para que isso aconteça é necessário que os profissionais da informação explorem as reais possibilidades de sua área do conhecimento e o estabelecimento de parcerias com os profissionais da CC em busca de uma transdisciplinaridade, também aplicável à GI, utilizando-se as tecnologias como meio para obtenção dos fins. Além disso, o processamento de dados com a utilização das tecnologias possibilita, segundo Rados, Valerim e Blattman (1999, p. 11), agregar valor por meio de operações de implementação no tratamento e disseminação da informação, além do monitoramento e o seu uso.

Este estudo propõe-se a preencher uma lacuna nas áreas de CI e CC com uma pesquisa metodológica cujo objetivo é o desenvolvimento de uma metodologia multidisciplinar para socialização de saberes validada em caso concreto, com foco no usuário da informação. Com essa proposta, procura integrar conhecimentos teóricos e práticos da CC, da CI e da GI. Por contar com dispositivos plurifuncionais, a metodologia pauta-se no armazenamento, recuperação e estudos da informação.

Para desenvolvimento satisfatório do objetivo geral, este foi desmembrado em três objetivos específicos, a saber:

- a) consolida um quadro teórico para esboçar a proposta metodológica;
- b) construir a metodologia;
- c) validar a proposta metodológica.

A construção de uma metodologia de base de dados temática só se justifica com o uso das teorias de organização de conhecimento da CI aliada à GI e às TICs. Nesta condição é possível desenvolver uma metodologia multidisciplinar que, embora ultrapasse os limites de cada área do conhecimento, preserve as especificidades inerentes a cada uma delas.

O desenvolvimento da pesquisa gerou uma metodologia, validada pela construção de um protótipo, consolidando um modelo empírico para socialização dos saberes. A área do conhecimento para aplicação da metodologia foi a Ciência da Informação. A escolha deveu-se à motivação pessoal do pesquisador, principalmente pela sua formação em Biblioteconomia, assim como pela oportunidade de consolidar saberes do Grupo de Pesquisa Educação, Pesquisa e Perfil Profissional da Informação (E3PI) da Universidade Federal do Paraná (UFPR) que já realizava, desde 2002, uma base estruturada em *ProCite5*, com a produção acadêmica e profissional em periódicos nacionais correntes e históricos em CI, de 1972 até 2008.

A busca da pluralidade de fundamentos foi contemplada pela revisão de literatura, a qual permitiu a construção de uma metodologia de construção de base de dados que pode ser replicada para outras áreas científicas.

Nesta organização o capítulo introdutório visa posicionar o contexto do estudo, a justificativa, bem como os objetivos e a metodologia utilizada para seu desenvolvimento. A estruturação deste trabalho apresenta-se na Figura 1.



FIGURA 1 - DIAGRAMA DA ESTRUTURA DO ESTUDO

Fonte: O autor (2010)

O segundo capítulo busca fundamentar a organização do conhecimento nos princípios da CI, analisando teorias que justifiquem a necessidade de regras e padrões com a utilização de vocabulário controlado para a construção de ferramentas de apoio à organização e ao tratamento temático da informação, como terminologias, tesouros e ontologias. Estuda também a comunicação científica e o papel das revistas científicas na construção e propagação do conhecimento e analisa como o desenvolvimento de novas tecnologias possibilitou a diminuição de tempo entre a produção e o consumo da informação. Explora a temática de arquivos abertos, bem como a política de disseminação de conhecimento e seus instrumentos tecnológicos de interoperabilidade com bases de dados, principalmente no que tange à coleta de informações, com o objetivo de agregar valor por meio de bases temáticas (provedores de serviço). Fundamenta-se nas teorias de recuperação de informação para traçar um ponto teórico com intuito de definir modelos de sistemas de recuperação de informação.

O capítulo três descreve a trajetória para construção da metodologia Brapci. Partindo de um estudo exploratório em publicações científicas, busca teorias que contribuem para o entendimento e fundamentação teórica dos três pilares orientadores do estudo, a Ciência, a Gestão e a Tecnologia da Informação. Partindo das necessidades de informações dos usuários, o estudo realiza um planejamento e considerações sobre a efetividade de uma política de indexação, com o emprego de vocabulários controlados com uma abordagem mista da linguagem natural e a

controlada, propondo-se um método pós-coordenado e criação de instrumentos como glossários e tesouros oriundos dos descritores da linguagem natural para controlada. Neste capítulo exploram-se as metodologias da CC para documentação e modelagem de dados, empregando uma metodologia dinâmica, mais voltada aos resultados pragmáticos do que à documentação detalhada de sistema. A GI contribui com instrumentos de controle e monitoramento da informação, com fundamentação de relatórios e análises que garantam a manutenção da qualidade dos processos de organização da informação. Propostas de análise de conteúdo também são descritas neste capítulo.

O capítulo quatro apresenta a efetivação da metodologia aplicada a uma realidade, com os procedimentos de construção e avaliação dos módulos “manutenção”, “público” e “pesquisador”. Detalha a empregabilidade do vocabulário controlado, controle de autoridade e políticas da base de dados adequados à realidade da Brapci. A avaliação da usabilidade do sistema e a estruturação dos processos de monitoramento, coleta e organização dos dados são abordadas nesse capítulo, assim como são analisados e estudados pelo grupo de pesquisa E3PI os procedimentos implementados, concretizando-se em estudos publicados em revistas e eventos científicos.

O capítulo final resgata os objetivos do trabalho e sintetiza a aplicação da metodologia em uma realidade, considerando a proposta interdisciplinar e as contribuições advindas de sua aplicação. Pondera também que a metodologia pode contribuir para a geração de indicadores sobre a produção científica brasileira como índice de citação, índice de colaboração e fator de impacto.

2 FUNDAMENTAÇÃO TEÓRICA

Como se trabalhou com uma proposta de metodologia para organização da informação e sua recuperação em ambiente digital, foi necessário o entendimento das bases do novo modelo de comunicação científica, das teorias de organização do conhecimento e de sua ligação com as tecnologias de informação. A definição clara dos conceitos empregados visa um melhor entendimento da proposta da metodologia aqui defendida. Para tanto, recorreu-se a autores das áreas de Organização e Representação do Conhecimento, Produção e Comunicação da Informação em CT&I e Informação e Tecnologia.

Na temática de organização do conhecimento verificou-se a existência de vertentes defendidas por diversos autores, sendo que nesta pesquisa não se buscou uma fundamentação exaustiva, mas suficiente para embasar a construção da metodologia.

A estruturação do referencial teórico é apresentada na Figura 2, iniciando com a recuperação da informação.

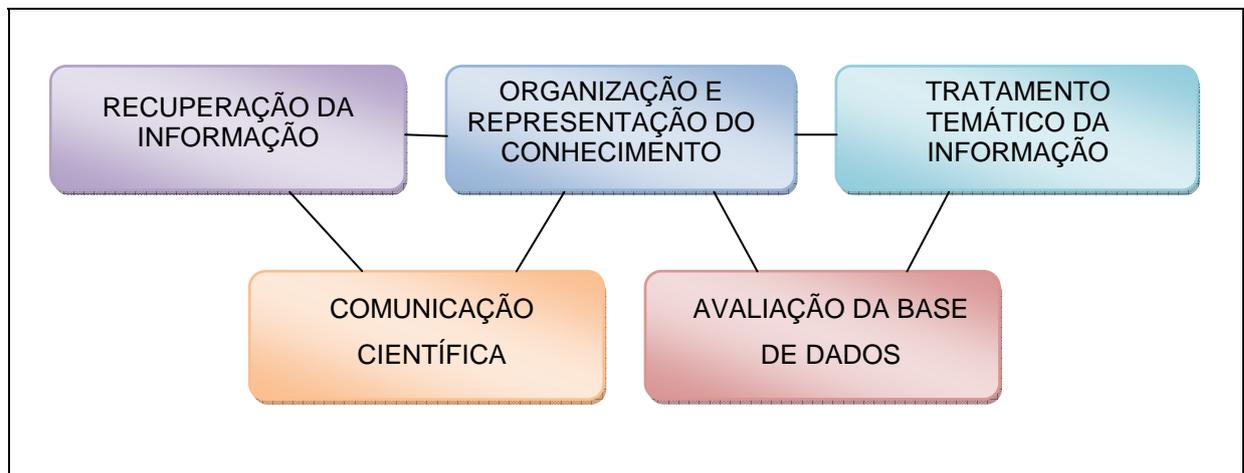


FIGURA 2 - DIAGRAMA DA ESTRUTURA DA FUNDAMENTAÇÃO TEÓRICA
Fonte: O autor (2010)

Parte-se do estudo da fundamentação sobre recuperação da informação por considerá-la um dos objetivos principais da organização da informação.

2.1 RECUPERAÇÃO DA INFORMAÇÃO

Em uma abordagem mais ampla, define-se que a CI emprega o conceito de documento, e que este é o que representa o suporte de informação que será estruturado e organizado, independente de seu suporte, podendo ser em papel, eletrônicos, virtual, magnético ou qualquer outra mídia que contenha informação.

Por sua visão de documentos, com o objetivo na organização e recuperação, Paul Otlet é considerado por muitos autores o precursor e o pai da CI, com sua preocupação na recuperação de documentos, mesmo que em sua época ainda não se tinha a nítida orientação do que viria a ser a recuperação da informação.

A consolidação da recuperação da informação ocorreu durante a Segunda Guerra Mundial, quando um cientista do *Massachusetts Institute of Technology* (MIT), Vannevar Bush identificou um problema que basicamente era “a tarefa massiva de tornar mais acessível, um acervo crescente de conhecimento”. Para enfrentar o problema da explosão informacional e o irreprimível crescimento exponencial da informação e de seus registros, particularmente em ciência e tecnologia, a solução proposta por Bush era a de usar as incipientes tecnologias de informação para combater o problema (SARACEVIC, 1996).

No início e meados da década de 1950, uma massa crítica de cientistas, engenheiros, bibliotecários e empreendedores começaram a trabalhar com entusiasmo sobre o problema e a solução definidos por Bush. Até 1960, a busca por uma solução gerou um grande esforço e foi relativamente bem financiada por agências de fomento, tornando-se uma atividade organizada. Muitas vezes, houve discussões acaloradas e controvérsias sobre a melhor solução “técnica”, ou sistema que pudesse resolver os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos no contexto social, institucional ou individual do uso e das necessidades de informação (SARACEVIC, 1996).

Calvin Mooers, um dos pioneiros da CI, cunhou o termo recuperação da informação em 1951. O termo consolidou-se e hoje é parte integrante da área da informação (CC e CI). Nesta época, Mooers (1951) identificou os problemas a serem abordados e definiu que a recuperação da informação engloba aspectos intelectuais da descrição da informação e sua especificação para busca, e também quaisquer sistemas, técnicas ou máquinas empregadas para realizar a operação.

Com a disponibilidade de acesso à informação, por meio das tecnologias de informática e telecomunicação nas bibliotecas acadêmicas, houve significativa mudança no perfil de seus usuários. Estes passaram a formular sua estratégia de busca e obter a informação desejada, diminuindo seu contato com o bibliotecário para esse fim, mas demandando novos serviços às bibliotecas (CUENCA; NORONHA; ALVAREZ, 2008).

As TICs determinaram profundas transformações no modo pelo qual as informações científicas transitam no meio acadêmico, veiculadas pelas revistas científicas, repositórios, eventos e comunicação eletrônica entre pares. Com o desenvolvimento destas tecnologias, os veículos utilizados pelos pesquisadores para a realização de suas investigações passaram a ser bastante diversificados. A grande quantidade de informações disponíveis trouxe dificuldades em sua recuperação, tanto para os pesquisadores, na organização de seus arquivos pessoais, quanto para os bibliotecários, suscitando muitas dúvidas (GRIEBLER; MATTOS, 2007).

As TICs alteraram o comportamento do usuário, que passou a ter autonomia no acesso à informação, independente de onde esteja, na universidade, em sua casa, no laboratório de informática ou em seu gabinete. A biblioteca funciona de certo modo, em tempo integral, e sua localização e horário de atendimento já não são mais tão fundamentais para a obtenção de determinado tipo de publicação. Entretanto, tornou-se necessária a melhoria da interface de acesso do usuário às coleções de bases de dados e de periódicos de que a biblioteca dispõe, sejam estas impressas ou eletrônicas, pagas ou gratuitas, ajudando a aumentar o êxito na busca da informação desejada (GRIEBLER; MATTOS, 2007).

O conceito de “auto-atendimento” é entendido por Fourie (2001 *apud* GRIEBLER; MATTOS, 2007) como desintermediação da informação e se apresenta como uma tendência na sociedade, proporcionando autonomia e independência para os usuários da informação.

Os sistemas de recuperação da informação tiveram sua origem nas grandes bases de dados desenvolvidas pelas instituições americanas: *National Library of Medicine* (NLM), *Department of Defense* (DOD) e *National Aeronautics and Space Administration* (NASA) (LANCASTER, 1993). No atual contexto, esses sistemas se difundem como uma consequência da explosão da informação. Assim, a necessidade da avaliação tornou-se ainda mais aguda, e alguns autores vêm se

dedicando a este tema e desenvolvendo instrumentos que reúnem critérios para avaliação (MESQUITA *et al.*, 2006).

Para Souza e Alves (2009),

[...] as tecnologias de informação e comunicação tornaram-se indispensáveis, sobretudo, ao provimento e à manutenção de sistemas que propiciam o ambiente para o tratamento descritivo e analítico, bem como a recuperação da informação com qualidade, facilidade e rapidez.

A implantação da *World Wide Web* (WWW), a partir de 1994, e o estabelecimento e expansão da Internet têm proporcionado oportunidades ímpares de acesso e recuperação da informação, viabilizadas sobretudo, pelas bibliotecas digitais surgidas nesse contexto. As bibliotecas digitais têm incorporado de forma satisfatória os conceitos e as práticas inerentes à organização e ao tratamento da informação, herdadas das bibliotecas tradicionais. E para isso, vêm combinando a estruturação e coleta de informação ao uso de tecnologias de informação e comunicação. Essa transformação ocorrida com a incorporação das tecnologias web alterou profundamente a natureza das bibliotecas, de tal modo que o potencial informacional de uma biblioteca digital dificilmente será suplantado, se comparado ao de uma biblioteca tradicional (CUNHA, 2007).

A qualidade da recuperação da informação é um item que deve ser observado atentamente quando se pretende avaliar uma base de dados, pois sua eficiência está intrinsecamente relacionada com o êxito de o usuário localizar a informação desejada. É uma etapa de grande importância e depende das etapas precedentes: indexação e armazenamento da informação (RODRIGUES; SILVA, 2007).

Para Laan (2002), no momento da elaboração da busca, o usuário utiliza uma linguagem de especialidade própria (linguagem natural), muitas vezes corriqueira, e que nem sempre irá coincidir com a linguagem do Sistema de Recuperação da Informação (SRI) (vocabulário controlado) de uma base de dados. A dificuldade de comunicação entre esses dois interlocutores (usuário x SRI), muitas vezes, está relacionada ao fato de que só deve haver um termo para cada conceito, chamado de termo preferido ou descritor. Os outros termos que designem este conceito são chamados de não-preferidos ou não-descritores, como, por exemplo, os sinônimos, os quase-sinônimos e as variantes ortográficas. Quem vai determinar

os descritores e os não-descritores são os vocabulários controlados, os tesouros, os cabeçalhos de assunto entre outros instrumentos de representação (RODRIGUES; SILVA, 2007).

Como nos vocabulários controlados, em geral, só os termos autorizados (descritores) são registrados, estes se tornam a única possibilidade de acesso às informações. Se o usuário utilizar outro termo, ele não conseguirá recuperá-las, a menos que esteja disponível como um termo equivalente no sistema, com a utilização de remissivas. Segundo Laan (2002), essa problemática não se caracteriza pelo desconhecimento do significado do léxico por parte dos interlocutores, mas pelo uso de códigos diferentes para o mesmo conceito. Desse modo, é importante que o profissional da informação e o usuário utilizem códigos coincidentes para que a recuperação da informação tenha êxito (RODRIGUES; SILVA, 2007).

Como solução para esse problema decorrente da dissonância de linguagens, Campos (2001) afirma que, “no ambiente de produção do conhecimento, percebem-se dois espaços: o comunicacional, no qual as descobertas e os avanços do conhecimento viram registros, e o informacional, no qual existe um requerente de informação e um sistema possuidor de informação”. Entre esses dois espaços, encontra-se o que a autora denominou de “Sistema T” – Sistema Terminológico, um ambiente controlado pelas leis da linguagem. A terminologia, por trabalhar com níveis de especialidade de uma área, deverá ser criada para possibilitar a comunicação entre os espaços comunicacional e informacional. Campos (2001) defende que no espaço informacional é necessária a criação de instrumentos que possibilitem a comunicação entre os usuários e o SRI (RODRIGUES; SILVA, 2007).

Percebe-se que as bases de dados dependem de uma linguagem documentária consistente para que a busca tenha maior proveito. Não adianta as fontes informacionais de seu acervo serem relevantes, se não houver a possibilidade clara de o usuário encontrar as informações que deseja. A partir dessa dependência, torna-se necessário elaborar instrumentos de recuperação da informação eficientes, focados em áreas especializadas e que atendam a problemas específicos.

Segundo Belkin (1980), citado por Sayão (1996), quando um usuário inicia uma busca, ele não sabe com exatidão o que pretende procurar, criando uma espécie de angústia, amenizada somente no momento em que navega pela base. Esse estado transitório o autor chama de “estado anômalo de conhecimento”, ou

seja, o alto grau de indefinição do usuário em relação ao assunto que procura, sendo o seu desejo de informação altamente nebuloso antes de consolidar a busca. Essa condição psíquica é inerente à natureza humana, no entanto o profissional da informação pode dirimir esse problema apresentando melhores condições de busca ao usuário, para que o espaço de tempo da sua angústia seja menor ou até inexistente.

Para Ahmed, McKnight, e Oppenheim (2004), os profissionais da informação mantiveram, durante anos, o domínio do manuseio dos sistemas de recuperação da informação devido a razões como a complexidade nos comandos das estratégias de pesquisa e o custo dos sistemas. Estes profissionais eram mediadores da pesquisa, situados entre o usuário e o sistema de informação. Com o crescimento da Internet e principalmente com a *World Wide Web*, verificou-se uma mudança radical no comportamento do usuário, que passa a ser o objeto de estudo na recuperação da informação. A literatura já classifica o usuário em uma escala que vai de novato ao experiente. Segundo Mesquita e colaboradores (2006), estudos realizados com usuários novatos comprovam que no processo de recuperação de informação eles podem ter dificuldades, mas conseguem efetivar buscas satisfatórias e relevantes nas bases de dados. Entretanto, se a mesma busca for realizada por um profissional da informação, a quantidade e a pertinência de registros recuperados tendem a serem maiores. Ainda segundo o autor, o usuário novato é mais lento e comete mais erros durante a pesquisa (MESQUITA *et al.*, 2006).

Parte deste ponto a justificativa de utilizar na metodologia interfaces de consultas mais amigáveis, com graus de busca, das mais simples às avançadas. Na recuperação de informação a utilização de lógicas de busca deve ser empregada para um refinamento dos resultados. A utilização de operadores booleanos como recursos lógicos da busca, com a incorporação tanto pela busca simples como avançada deve ser considerada para que bons resultados possam ser extraídos da base. O usuário quando consulta uma base de dados cria estratégias de busca utilizando diferentes ferramentas e lógicas e deve ter flexibilidade para utilizar os operadores booleanos (*AND*, *NOT* e *OR*) no momento de sua busca, sendo o sistema responsável em interpretar e traduzir a consulta de forma legível e que seja lida como uma consulta nas tabelas do banco de dados (ZANIELLI; CATARINO, 2004).

Por muito tempo, os sistemas de informação foram projetados e desenvolvidos na tentativa de categorizar o usuário, estruturando e organizando as informações com o objetivo de atender a pergunta “para quem?” (COSTA; RAMALHO, 2010). Porém, as instâncias enunciativas, segundo Bakhtin (1979, *apud* KOBASHI, 2007), são indissociáveis no processo de produção de sentido. Enfatiza-se, assim, a fala, a *parole*, a enunciação, o diálogo e o que dele resulta: o enunciado, ao qual se atribui o sentido de acordo com os referenciais e as experiências do usuário.

A nova pergunta norteadora para atender as necessidades do usuário desses novos sistemas de recuperação de informação seria, desse modo, “informação, para fazer o quê?”. Segundo Costa e Ramalho (2010), esta evolução refere-se ao fato da atual compreensão de que um indivíduo pode estar incluso em várias categorias de usuários, de acordo com os papéis desempenhados nos momentos de sua vida, enquanto um ator social.

Le Coadic (2004) ressalta a existência de erro em “estudos ditos de usos e usuários da informação (*information users*)” que são “na verdade mais frequentemente estudos de usos dos sistemas de informação por meio dos usuários e não estudos dos usuários por meio dos usos dos sistemas de informação”. Desse modo, para o autor (LE COADIC, 2004), usar informação é trabalhar com a matéria informação para obter um efeito que satisfaça a uma necessidade de informação. O objetivo final de um produto ou de um sistema de informação deve ser analisado em termos dos usos da informação e dos efeitos resultantes desses usos nas ações dos usuários, devendo ser “orientados para o usuário”. Para o autor (LE COADIC, 2004) o ciclo de conhecimento do usuário em seu processo de uso da informação deve passar por quatro fases (Figura 3).

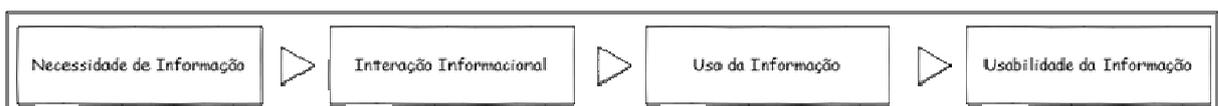


FIGURA 3 - NECESSIDADE, INTERAÇÃO, USO DA INFORMAÇÃO E USABILIDADE
Fonte: LE COADIC (2004)

A busca por informação parte sempre de uma necessidade e conhecê-la é imprescindível para o sistema. Importa compreender por que as pessoas se envolvem num processo de busca da informação. O que levaria uma pessoa a buscar, então, informação? A existência de um problema a resolver, de um objetivo

a atingir e/ou a constatação de um estado anômalo de conhecimento, insuficiente ou inadequado (LE COADIC, 2004, p. 38).

Nesse processo, a interação informacional seria a forma pela qual o usuário utiliza o sistema de informação, os diálogos e a formulação de estratégias, as alterações das perguntas realizadas ao sistema e o seu comportamento. Essas análises, com a formulação da pergunta inicial ao sistema, bem como as alterações realizadas nas estratégias, consistem em bases dinâmicas características dos fenômenos de uso da informação (LE COADIC, 2004, p. 43).

O uso da informação, por sua vez, é uma prática social, que faz parte do conjunto das artes de fazer. Pesquisando-se as atividades do usuário, obtém-se melhor conhecimento de suas práticas, modos de agir, usos, modos de operação e esquemas de atuação. O estudo sobre o uso da informação busca identificar como é realizado o uso dos recursos correntemente oferecidos, a fim de ampliá-los ou redefini-los, assim como avaliar as necessidades dos usuários, para saber até que ponto os serviços oferecidos atendem as suas necessidades de informação (LE COADIC, 2004, p. 48).

Nesse sentido, a usabilidade da informação é o resultado da avaliação sobre até que ponto um produto ou um sistema de informação presta-se ao uso, voltado a sua utilidade, eficácia e confiabilidade (LE COADIC, 2004, p. 49).

2.2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

A ação de classificar é intrínseca ao homem e manifesta-se na preocupação em categorizar e organizar todo e qualquer material cognoscível, o que faz presumir a importância que o homem deu aos registros da história. Para Aquino, Carlan e Brascher (2009), o hábito de classificar é uma atividade social e é realizado muitas vezes de maneira inconsciente.

A história da organização da informação remete à história da classificação do conhecimento, que teve como ponto de partida o modelo estabelecido por Aristóteles, prevalente por aproximadamente dois mil anos (300 a.C. a 1600 d.C.). Aristóteles dividiu a ciência em três partes: teórica, que visa ao conhecimento em si (física, matemática e metafísica); prática, que busca o conhecimento como guia de

conduta, cujo propósito é a ação (ética, política, economia e retórica); e a produtiva, cujo propósito é a criação de um produto (poesias e artes). Essas divisões, segundo Naves (2006, p. 37), não são mutuamente exclusivas, podendo uma disciplina ser identificada em mais de uma divisão.

Na classificação aristotélica, os gêneros supremos e irreduzíveis (ou predicamentos): substância (existente em si); quantidade (determinação atribuída pelas partes ao lado das partes); qualidade (determinação que afeta intrinsecamente o objeto); relação (referência a substância ou um acidente que estabelece com outra substância); tempo (posição em relação ao curso de eventos extrínsecos); lugar (determinação de um ser em referência a outro); posição (determinação de um corpo com referência a um lugar); ação (exercício das faculdades ou de poder sobre a substância de modo a produzir um efeito); paixão (recepção sofrida por um substância, efeito produzido por um agente); posse (determinação que se cria com a adjacência de outro objeto) (CAMPOS, 1978).

Na continuidade histórica da organização da informação, Porfírio, filósofo grego do século IV, aplicou o princípio de oposição de Platão e Aristóteles para apresentar uma classificação dicotômica do conhecimento, representando as subáreas como ramos, ficando conhecida como árvore de Porfírio representada na Figura 4.

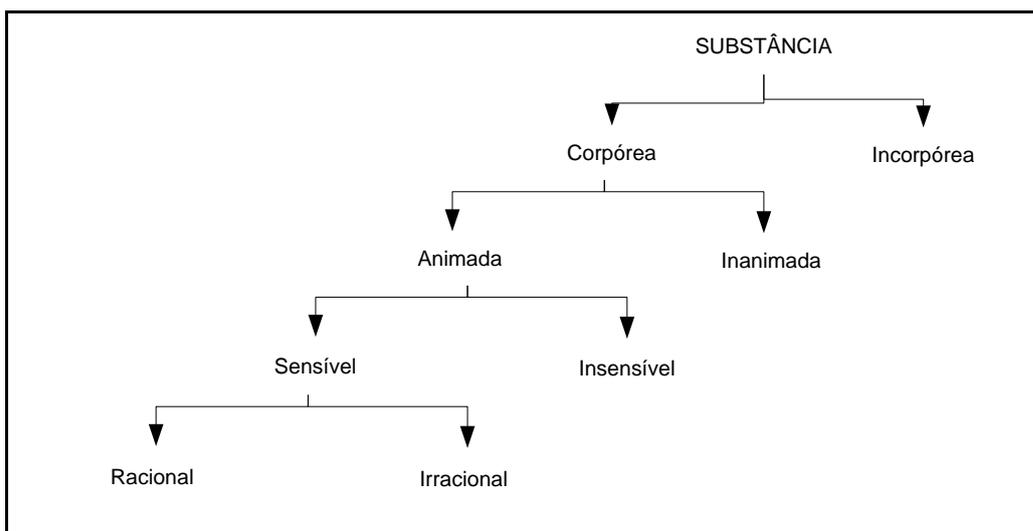


FIGURA 4 - ÁRVORE DE PORFÍRIO

Fonte: NAVES (2006)

Kant foi outro filósofo que procurou demonstrar a elaboração de categorias para explicar que o pensamento humano reveste-se de autoridade por si só, não

constituindo, absolutamente, de posição arbitrária. Ao rever a formulação aristotélica das categorias, Kant reelabora-as com justificativas e apresenta uma tábua na obra *Crítica da Razão Pura* para entender como se relacionam reciprocamente e adquirem sentido (ARANALDE, 2009).

A caracterização de Kant, segundo Aranalde (2009), está relacionada aos juízos, ao modo como se fala dos objetos, sendo que os objetos das categorias matemáticas são oriundos da intuição. Em contrapartida, as categorias dinâmicas dizem respeito à existência dos objetos em suas relações entre si e com o entendimento. Os grupos se subdividem em três categorias, sendo que as duas primeiras formam uma dicotomia enquanto a terceira resulta da síntese das duas, uma combinação entre elas:

- a) quantidade: unidade, pluralidade e totalidade;
- b) qualidade: realidade, negação e limitação;
- c) relação: inerência-subsistência, causalidade, dependência e comunidade;
- d) modalidade: possibilidade \Leftrightarrow impossibilidade, existência \Leftrightarrow não existência e necessidade \Leftrightarrow contingência.

Ranganathan (1967) adotou para a classificação do conhecimento um estilo aforístico, uma aproximação do princípio de “unidade de pensamento atômico”, tendência devida a sua familiaridade com o pensamento indiano e com a literatura védica. O autor (RANGANATHAN, 1967) levantou pela primeira vez a questão de como os assuntos são formados e a resposta que obteve foi que a formação dos assuntos se dá por meio de cinco caminhos: dissecação, laminação, desnudação, reunião/agregação e superposição. Chegou a essa divisão recorrendo à dicotomia difundida por Kant e à árvore de Porfírio e concluiu que existia um crescimento prolífico multidimensional, sugerindo o padrão de mapear o conhecimento com a policotomia ilimitada (NAVES, 2006).

Para Campos (2001), a categorização fornece a visão de conjunto dos agrupamentos que ocorrem na estrutura analisada, possibilitando o entendimento global da área e das categorias fundamentais, que funcionam como o primeiro corte classificatório, sendo este extremamente importante por ser o determinante da estrutura da representação e da classificação.

As categorias possibilitam, de acordo com essa concepção, a definição de certos juízos básicos que emitimos sobre o mundo na tentativa de interpretá-lo e compreendê-lo, permitindo assim descrever o mundo e as coisas que o compõem

como unidades cognitivas que guiam a apreensão da multiplicidade, pois é sob tais unidades que se distinguem e classificam os vários modos em que o ser se apresenta. A classificação, por sua vez, é o processo mental que busca agrupar as coisas conforme seus graus de semelhança e separá-las conforme seus graus de diferenças, captado do ponto de vista do ser humano e como este se coloca no mundo para tentar compreendê-lo (ARANALDE, 2009; AQUINO; CARLAN; BRASCHER, 2009).

Essa classificação possibilita um número ilimitado de divisões das áreas do conhecimento e assim, para superar as limitações da árvore de Porfírio, o autor (RANGANATHAN, 1967) buscou a representação na árvore baniana (Figura 5). Nesta representação, cada tronco representa uma categoria, partindo de cinco categorias fundamentais: tempo, espaço, energia (ação), matéria e personalidade.

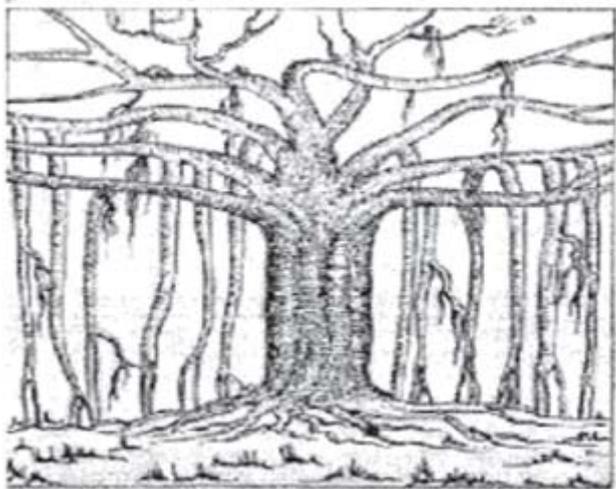


FIGURA 5 - ÁRVORE BANIANA DE RANGANATHAN
Fonte: RANGANATHAN (1967)

As tentativas de empregar as classificações da filosofia para a categorização bibliográfica iniciam em 1791 na França, com a publicação do primeiro código nacional de catalogação. Porém, somente em 1840 os catálogos foram organizados por autor e assunto. Em 1850, foi criado o primeiro catálogo coletivo para classificação bibliográfica utilizado por bibliotecas públicas nos Estados Unidos (ORTEGA, 2004).

Em 1876, o norte-americano Melvin Dewey publicou a primeira classificação temática, a Classificação Decimal de Dewey (CDD) que, em 1904, viria a servir de modelo para a criação da Classificação Decimal Universal (CDU) e posteriormente

do Repertório Bibliográfico Universal (RBU), ambos elaborados por Otlet e La Fontaine.

Para Robredo (2003, p. 41), a CDU “aprimorou a CDD com subdivisões sistemáticas normatizadas e notações ampliadas, razão pela qual ganhou um elevado grau de aceitação nas bibliotecas e centros de documentação especializada”.

Em 1934 foi o ano em que Paul Otlet publicou o “*Traité de documentation*”, fruto da mudança de orientação do campo da Bibliografia para o da Documentação, acreditando que a análise de documentos não poderia se reduzir a uma bibliografia (descrição de documentos), mas chegar ao nível de uma bibliologia, ou seja, uma ciência e uma técnica gerais para estudo do documento (ORTEGA; LARA, 2010). Para o autor, livro e documento constituem “um suporte de certa matéria e dimensão [...] em que se incluem signos representativos de certos dados intelectuais” (OTLET, 1937).

Considerado fundador do que hoje se denomina CI, por ele chamada de Documentação, Paul Otlet ampliou a noção de documento, que passou a abranger o livro, a revista, o jornal, a peça de arquivo, o impresso, a fotografia, a medalha, a música, o filme, o disco e toda a parte documental que precede ou sucede a emissão radiofônica (OTLET, 1937).

Com a noção de categorias nos sistemas de classificação introduzida por Ranganathan, renovaram-se as possibilidades de organização da informação, pois sempre que se realiza uma classificação busca-se um relacionamento entre os conceitos utilizados, e quando da construção de sistemas de organização de conhecimento estas relações precisam ser identificadas. Neste processo pode-se utilizar dois tipos de relacionamento: hierárquico, quando um conceito é subordinado a outro, ou não hierárquico quando a relação pode ser estabelecida horizontalmente.

Parte-se da concepção de que um conceito é uma unidade abstrata e que para sua representação empregam-se os termos, ou suas partes, de uma linguagem dada como unidade de informação (CABRÉ, 1993). A arte de representar os conceitos implica o reconhecimento das relações que se verificam entre eles.

2.2.1 Relacionamento hierárquico ou de subordinação

No relacionamento hierárquico, Cintra e outros (2002, p. 59) pressupõem que um conceito específico é parte da extensão de um conceito mais amplo, sendo que o específico tem todas as características do conceito mais amplo e uma característica a mais que o diferencia. A noção de extensão, por sua vez, vem sempre associada à noção de intenção.

Num sistema de classificação hierárquico há classes que são subdivididas em subclasses, por sua vez divididas em unidades subsequentes, sempre agregando característica ou propriedade particular que a distingue da classe posterior. Essa relação de subordinação entre as classes e suas subclasses forma os conceitos de subordinação, superordenação e coordenação.

A subordinação ocorre das subclasses em relação à classe a qual se encontram relacionadas (Figura 6). Já a relação de superordenação ocorre das classes em relação às suas subclasses e as classes no mesmo nível de divisão designam-se por classes coordenadas (AQUINO; CARLAN; BRASCHER, 2009).

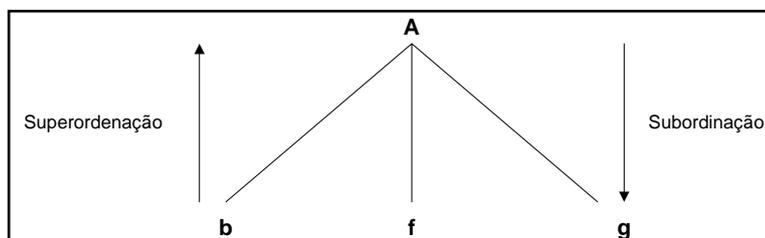


FIGURA 6 - ESQUEMA DE RELAÇÕES HIERÁRQUICAS

Fonte: CINTRA *et al.* (2002, p. 51)

Neste contexto podem existir três tipos de relações hierárquicas:

- a) Relações genéricas (Figura 7), que marcam relações de gênero, portanto globais ou gerais, e podem ser: baseadas na identidade parcial do conjunto de características das noções superordenadas e subordinadas nelas envolvidas; de gênero, entendido como noção superordenada que comporta as mesmas características das noções subordinadas, a partir dela; resultantes da conjunção de características comuns; indicativas de que todo conceito pertencente à categoria do conceito específico (a

espécie) é parte da extensão do conceito amplo (gênero), (ORGANIZAÇÃO INTERNACIONAL DE NORMALIZAÇÃO, 1987).

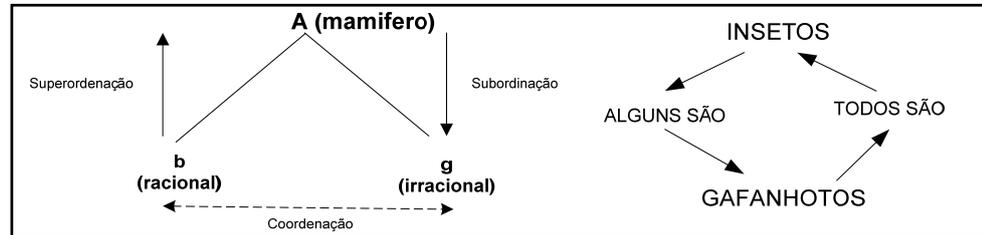


FIGURA 7 - ESQUEMA DE CLASSIFICAÇÃO GENÉRICA

Fonte: Cintra *et al.* (2002, p. 53)

- b) Relações específicas: relações de espécie, logo particulares, e definem-se como relações hierárquicas subordinadas que, além de compartilhar das mesmas características das noções que lhes são superordenadas e, segundo a ISO 704 (1987), devem abranger todas as características do conceito mais amplo e diferenciar-se com pelos menos uma característica distintiva adicional, que serve para diferenciar conceitos específicos no mesmo nível de abstração, estabelecendo uma disjunção, a partir da conjunção dada.
- c) As relações partitivas (Figura 8) constituem-se a partir da integração da parte com o todo, referindo-se à noção superordenada como a um objeto considerado como um todo e as noções subordinadas a objetos considerados como suas partes. A relação da parte depende do conceito do todo e não pode ser definida previamente à definição do conceito do todo.

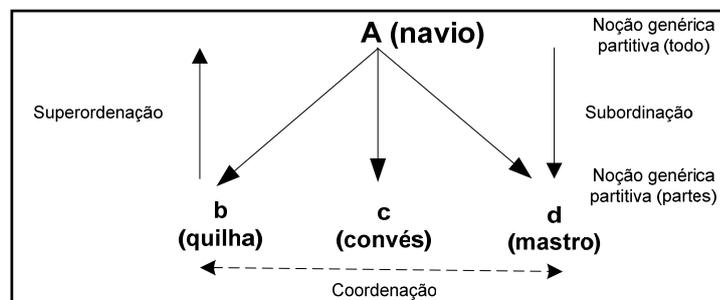


FIGURA 8 - ESQUEMA DE RELAÇÃO PARTITIVA

Fonte: CINTRA *et al.* (2002, p. 54)

Outra forma de relacionamento entre os conceitos é resultante das relações não hierárquicas.

2.2.2 Relações não hierárquicas ou horizontais

As relações não hierárquicas definem-se pela negativa, elas recobrem o conjunto de relações que não são passíveis de serem descritas como hierárquicas. São relações que apresentam uma dependência resultante de uma contiguidade espacial ou temporal do tipo: causa/efeito, antes/depois, esquerda/direita, acima/abaixo, produtor/produto e material/produto.

As associações entre termos pertencentes a categorias diferentes são dadas a partir do universo de referência individual. Tais referências necessitam um relacionamento funcional, “aquele em que um termo, que denote atividade ou operação, se liga conceitualmente, a uma entidade ou propriedade” (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 1984).

A unidade conceitual, para Cabré (1993), pode existir por si mesma, mas não a unidade de informação, que tem que ser captada por um sujeito receptor e, como consequência, transmitida pelo agente emissor. As estruturas horizontais referem-se às classificações facetadas relacionadas às suas palavras-chave, enquanto as relações hierárquicas ou verticais representam as grandes classificações como a CDD e a CDU.

Em aplicações como bases de dados, os esquemas de classificação empregados pelas bibliotecas tradicionais (acervos bibliográficos) são considerados muitos rígidos para atender as necessidades dinâmicas do ambiente digital. Estas classificações objetivam a identificação temática de uma obra, ideal para o armazenamento e recuperação física, porém não exploram as diversas variações de um mesmo tema, deixado para um segundo plano as necessidades dos diversos tipos de usuários.

Neste sentido, Beghtol (1986 *apud* DIAS; NAVES, 2007) afirma que um texto tem uma atinência relativamente permanente, mas um número variado de significados. Há, portanto, uma forte relação entre a atinência do documento e seu(s) significado(s) identificado(s) pelos indivíduos. Esse varia de acordo com o valor de uso que a pessoa atribui ao documento numa certa época, enquanto o mesmo documento pode ter diferentes significados para o mesmo leitor em diferentes circunstâncias, mas o documento, este imutável, adquire uma atinência fundamental.

O processo de análise conceitual demanda forte carga cognitiva, pois é efetuado pela mente humana, por meio da interpretação, da coordenação de signos e da abstração de significados. Voltada ao estudo do significado das palavras e de expressões, a semântica considera as “funções e natureza das palavras no processo de comunicação”.

2.2.3 Semântica

A palavra semântica é originária do grego (*semantikos*, derivada de *sema*, sinal) e refere-se ao estudo do significado opondo-se com frequência à sintaxe. Sendo um ramo da linguística é definida tradicionalmente, segundo Ullmann (1987), como estudo do significado (de partes) de palavras, frases, sentenças e textos.

O signo é uma palavra que, isoladamente, pode ter sentido para determinado indivíduo, mas seu significado está estreitamente ligado à estrutura lexical, isto é, ao contexto no qual o signo (uma palavra ou termo) está inserido. Já o sentido é uma abstração pessoal, ou seja, é como cada indivíduo entende o signo.

Para Rector e Yunes (1980, p. 14), uma explicação de propriedades semânticas requer mais do que a análise do sentido das palavras, isto é, para que se entenda o sentido de uma sentença e suas relações semânticas com outras expressões, é preciso saber não só o significado de suas unidades léxicas, mas também, como estas se relacionam – a dependência da estrutura sintática da sentença.

De acordo com Ullmann (1987), na análise semântica, o campo de estudo é o vínculo do signo com a realidade que o exprime. O objetivo da investigação é buscar dentro de universos possíveis o significado correto dos signos, distinguindo e eliminando os demais que a ele se encontram associados, extraíndo o máximo de toda a imprecisão dos termos oriundos da linguagem natural.

Ao estabelecer uma relação com as tecnologias da informação e suas ferramentas, Obrst (2009) constrói um modelo, atribuindo três níveis de semântica: a semântica baixa ou fraca (com interoperabilidade sintática); a média (com interoperabilidade estrutural) e a alta ou forte (com interoperabilidade semântica). A relação possibilita a comparação das linguagens empregadas pela CI e CC e suas

equivalências, identificando-se em cada categoria seus instrumentos e ferramentas. O modelo de representação de Obrst é mostrado na Figura 9.

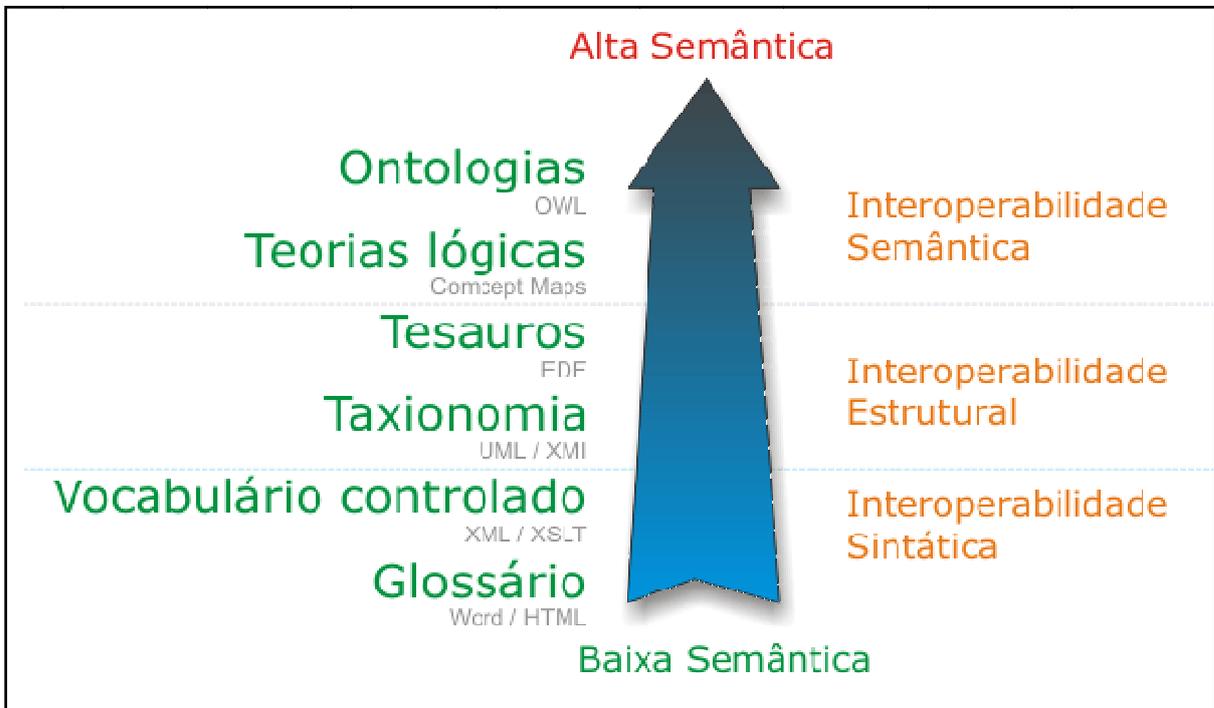


FIGURA 9 - RELAÇÃO SEMÂNTICA E OS INSTRUMENTOS DE CADA CATEGORIA
Fonte: Adaptado de OBRST (2009)

Obrst (2009) ainda ressalta que todos os sistemas de computação precisam utilizar teorias lógicas complexas e que os sistemas podem ser projetados para utilizar todos os tipos de relações. A semântica fraca (*weak semantics*), por exemplo, pode ser utilizada em sistemas de informação mais simples e com pouca complexidade semântica. Este tipo limita-se ao lexical do termo e seu significado, tendo aplicação na construção de dicionários, glossários e controle de vocabulários. Dos instrumentos empregados neste tipo de semântica, Obrst (2009) identifica: Dicionário, Glossário e Vocabulário Controlado, sistemas que podem ser construídos com editores de textos, *hiperlinks* na Web ou metadados com *Extensible Markup Language* (XML).

Porém, existem sistemas nos quais a semântica fraca não é o suficiente para atender as necessidades de organização da informação, principalmente quando se necessitam relações que extrapolem a relação léxica e incorporem esquemas de relacionamento entre os termos, com classificações hierárquicas, de relacionamento de subordinação e relação, tendo como exemplos a taxionomia e os tesouros. Nestes casos podem ser empregados sistemas baseados na média

semântica, existindo uma interoperabilidade estrutural e não somente sintática. Sistemas neste formato podem ser construídos empregando-se modelos *Resource Description Framework* (RDF) e XML. Atualmente os sistemas computacionais têm como base este modelo semântico.

A forte semântica (*strong semantics*), possibilita a estruturação por modelos lógicos, com a modelagem orientada a objetos, centrando-se na classe de entidade, seus atributos, propriedades, relações, associações, restrições e regras. Segundo Obrst (2009) é o tipo recomendado para a organização da informação na Internet, porém ainda o próprio autor admite que este modelo ainda necessite de mais algumas décadas para que os computadores possam processar essas atividades.

Sistemas computacionais precisam se basear em modelos de realidade definidos pelas pessoas que analisam e alimentam os computadores com essa informação, devendo-se partir sempre de uma unidade básica de representação, que vai possibilitar as relações lógicas.

A *American National Standard* (ANSI) desde 1974 desenvolve a *National Information Standards Organization* (NISO) com o objetivo de recomendar o uso de Vocabulários Controlados (VC), apresentando as diretrizes e convenções para construir, testar, exibir, prover manutenção e gerir vocabulários controlados monolíngues. Concentrando-se o foco nos que são empregados para representar conteúdos em sistemas de organização do conhecimento, destacam-se listas de cabeçalho de assunto, sinônimos, taxonomias e tesouros.

Com base no quadro de Obrst (2009), e em relação à carga semântica empregada em modelos computacionais e interoperabilidade sintática, destacam-se como elementos de representação o glossário, a taxonomia, o tesouro e as ontologias.

2.2.3.1 Glossário

O glossário vem a ser a primeira explicação para o delineamento e definição de termos obscuros por meio de outros conhecidos, definidos por especialistas em consenso por determinada área do conhecimento. Definindo os termos técnicos, o

glossário é uma espécie de dicionário, consagrado particularmente à explicação de termos mal conhecidos (arcaicos, peregrinos, dialetais etc.).

Faria e Pericão (2008) definem glossário como dicionário ou elucidário de palavras antigas, raras, novas ou pouco conhecidas numa língua determinada. O glossário pode ser um dicionário de indexação, com termos e definições, dispostos em ordem alfabética e compreendendo remissivas do tipo “*ver*” e “*ver também*”. Como vocabulário controlado, ele pode ser utilizado pelo indexador humano para preparar documentos ou ser pré-registrado e servir para a indexação automática realizada por computadores. Seu emprego atende parcialmente as necessidades de indexação e recuperação da informação pelos usuários, pois não existem relações entre os termos, e nem uma unidade lógica que os oriente. O fortalecimento semântico é possível por outros meios como o uso das taxonomias.

2.2.3.2 Taxonomia

O termo taxonomia ou taxionomia tem sua origem no grego *táxis* (ordem) e *onoma* (nome). As taxonomias surgiram no campo de estudo da Biologia e já vêm sendo utilizadas há bastante tempo para classificar de forma lógica e científica os seres vivos. O primeiro estudo foi fruto do trabalho do médico e botânico sueco Carolus Linnaeus, com sua *Biologia Sistemática* (AQUINO; CARLAN; BRASCHER, 2009).

Segundo Terra e outros (2005), uma taxonomia pode ser definida como: “um vocabulário controlado de uma determinada área do conhecimento, e, acima de tudo, um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema sob uma premissa lógica”. Assim, no contexto da CI, taxonomia seria um sistema para classificar e facilitar o acesso à informação. Para obter êxito, as taxonomias têm como objetivos (AQUINO; CARLAN; BRASCHER, 2009):

- a) estabelecer categorias gerais;
- b) coletar e representar os conceitos por meio de termos;
- c) agilizar a comunicação entre especialistas e outros públicos;
- d) encontrar o consenso;

- e) controlar a diversidade de significação;
- f) construir relacionamento semântico entre os termos, através de relações hierárquicas, de equivalência, e de associação; e
- g) oferecer um mapa da área que servirá como guia em processos de conhecimento.

No âmbito da CI as taxonomias podem ser comparadas a estruturas classificatórias como as Tabelas de Classificação, que têm como objetivo reunir documentos de forma lógica e classificada. Atualmente, as taxonomias reúnem todo tipo de documento digital e permite, diferentemente das estratégias de busca, um acesso imediato à informação. Ao contrário das Tabelas, que oferecem um endereço (notação) com a localização dos documentos nas estantes, a taxonomia prescinde de notação (CAMPOS; GOMES, 2008).

Nos ambientes digitais, seu aparecimento e uso estão relacionados com as formas automatizadas de criação da informação, tornando-se alvo de estudos da CI. As taxonomias têm sido empregadas em portais corporativos de empresas e em bibliotecas digitais, servindo como instrumento para a organização e recuperação de informações (AQUINO; CARLAN; BRASCHER, 2009).

Além dessas aplicações, o seu uso tem sido também difundido no contexto da Web Semântica, pois a taxonomia permite que se estabeleçam padrões de alto nível para a ordenação e a classificação de informação por meio do uso de mecanismos de herança (CAMPOS; GOMES, 2008).

As taxonomias como estruturas classificatórias representam os propósitos de organização intelectual de um dado contexto. Neste sentido, são diferentes dependendo do tipo de organização e de informações que pretendem representar. Segundo Campos, Campos e Gomes (2008) diversos estudos vêm sendo desenvolvidos em seus grupos de pesquisas, e entre as conclusões, apontam para a definição de três tipos de taxonomias: a canônica, a de domínio e a de processos e tarefas gerenciais. As duas últimas são caracterizadas pelos autores como policotômicas, ou seja, a partir de um domínio ou tarefa, várias divisões são possíveis. Por sua complexidade, requerem um primeiro recorte por categorias e no interior destas várias divisões e subdivisões são possíveis em cada passo de divisão (CAMPOS; GOMES, 2008).

Segundo Aquino, Carlan e Brascher (2009), as taxonomias em um ambiente digital podem ser classificadas, de acordo com sua utilidade, em três categorias:

- a) descritivas, com a criação de um vocabulário controlado, baseado nas estruturas dos tesouros, estabelecendo-se termos autorizados e seus significados em um determinado contexto, trabalhando com relações semânticas, possibilitando que tenha significado no ambiente em que foram construídas;
- b) para gerenciamento de dados, composta de pequena lista de termos autorizados sem qualquer estrutura hierárquica;
- c) navegacional, baseada no comportamento do usuário em detrimento do conteúdo, estimulando a navegação e organizando a informação de forma flexível, para que a estrutura dos dados possa ser visualizada quando do acesso por meio da navegação.

As taxonomias também possibilitam a construção de tesouros temáticos, com ligações semânticas mais fortes em busca de um modelo lógico.

2.2.3.3 Tesouros

A palavra tesouro, segundo Vickery (1960 *apud* DODEBEI, 2002, p. 64), tem sua origem na Grécia, dos termos *Treasury* ou *Store hour* que significavam tesouro ou armazenagem/repositório. Em 1936 o termo foi incorporado ao *Oxford English Dictionary* definido como uma expressão inglesa como um dicionário, enciclopédia e similares.

Para Cabré (1993), os tesouros são “recopilações” de termos relacionados semanticamente, que servem como ferramenta para organizar e recuperar informação. Os tesouros diferem das taxonomias por utilizarem, segundo Sales e Café (2009b), vocabulários controlados formados por termos-descritores semanticamente relacionados, atuando como instrumentos de controle terminológico. Conforme as autoras, os tesouros podem estar estruturados hierarquicamente (gênero-espécie e todo-parte) e associativamente (aproximação semântica), e são utilizados para indexar e recuperar informações por meio de seu conteúdo. Caracteriza-se o tesouro como vocabulário controlado por sua especificidade e complexidade no relacionamento entre os termos que comunicam o conhecimento especializado em uma área do conhecimento.

Uma das maiores desvantagens dos tesouros quando utilizados como ferramenta de organização de informação, segundo Dodebei (2002, p. 77), é a velocidade de sua atualização. Por ser construído de forma restrita, sua atualização ocorre de forma lenta, principalmente porque precisa de um consenso de vários profissionais para que o termo possa ser aceito e a ele incorporado.

Com o objetivo de padronizar, normalizar e criar diretrizes para a construção de vocabulários controlados dentre eles os tesouros a ANSI/NISO Z39.19 (2005) definiu conceitos para sua construção, utilizando o padrão norte-americano:

Um tesouro é um vocabulário controlado organizado em uma ordem preestabelecida e estruturado de modo que os relacionamentos de equivalência, de homografia, de hierarquia, e de associação entre termos sejam indicados claramente e identificados por indicadores de relacionamento padronizados empregados reciprocamente. As finalidades primordiais de um tesouro são (a) facilitar a recuperação dos documentos e (b) alcançar a consistência na indexação dos documentos escritos ou registrados de outra forma e outros tipos, principalmente para sistemas de armazenamento e de recuperação de informação pós-coordenados (ANSI/NISO Z39.19, 2005), (tradução nossa).

A norma ANSI/NISO Z39.19 (2005) ressalta que os tesouros não são utilizados somente pelos especialistas da informação no momento da indexação, mas também por usuários da informação quando da busca de documentos. O tesouro é o elo entre a linguagem utilizada pelos indexadores e pelos usuários, enquanto os termos e suas relações fazem dele o instrumento essencial para indexador e usuário buscarem o melhor termo (ou termos) dentro de sistemas de informações.

Na literatura encontramos duas vertentes para construção de tesouros, uma mais voltada ao controle da terminologia empregada nos sistema de recuperação e informação e outra com uma semântica mais forte, voltada não somente na terminologia, mas contextualizando o termo (conceito na ontologia) para que suas relações semânticas fique mais visível ao usuário.

Desta segunda vertente, Dodebei (2002, p. 77) afirma que os vocabulários controlados são construídos para indexação e recuperação da informação e devem ser baseados na estrutura conceitual de um determinado campo temático ou áreas de domínio. O conceito, como um item do conhecimento, reflete este mundo real dos objetos e dos fenômenos, com seus atributos e relacionamentos essenciais e acidentais (Cabré, 1993).

O tesauro proporciona em sua explicitação a visualização dos relacionamentos entre os termos, conforme o quadro 1.

SIGLA	DESCRIÇÃO
TG	Termo Geral
TE	Termo Específico
TR	Termo Relacionado
UP	Usado para
USE	Use (Remissiva)
NT	Nota explicativa

QUADRO 1 - RELACIONAMENTO DOS TESAUROS

Fonte: DODEBEI (2002)

Uma das deficiências da empregabilidade dos tesauros está concentrada na sua necessidade de padronização, ocasionando uma demora na concepção e incorporação de novos conceitos quando se trabalha em um ambiente onde a forja de novos termos é realizada constantemente, principalmente para áreas emergentes do conhecimento. Outra deficiência está relacionada a sua especificidade, pelo foco em um público especializado.

Uma das propostas, quando existe a necessidade da relação semântica mais forte, é o uso das ontologias para categorizar o conhecimento.

2.2.3.4 Ontologia

Originário da Filosofia, o termo “Ontologia”, em sentido estrito, pode ser definido como estudo do ser e de suas propriedades fundamentais. Etimologicamente o termo é originário do grego: *ontos* (ser, ente) e *logos* (saber, doutrina). Ainda na área da Filosofia, a Ontologia é considerada sob dois aspectos: o existencial (saber sobre aquilo que é fundamental ou irreduzível, comum a todos os entes singulares); aspecto essencial (determinar as leis, estruturas ou casas do ser em si). Essas concepções filosóficas remontam aos trabalhos de Aristóteles entre 384 e 322 a.C. (FEITOSA, 2002).

Para Lima-Marques (2006, p. 49), o foco dos estudos filosóficos deslocou-se para uma nova questão, a de desvendar a capacidade do espírito humano para captar a realidade, conduzindo aos estudos que constituíram a denominada Teoria do Conhecimento, com a qual a Ontologia foi correlacionada. Para o autor, esta

nova vertente não trata de explicar o real, mas de investigar o processo cognitivo pelo qual o homem capta, estuda e compreende, e de que modo armazena, classifica e representa o conhecimento, incorporando-o às estruturas do saber já consolidadas, agregando-se a outras disciplinas como a Lógica, a Computação, a Ciência da Informação e a Linguística para levar a resultados mais concretos na área de recuperação de informação.

A informática, desde o início dos anos de 1990, segundo Feitosa (2002), utilizou-se do conceito de ontologia, tornando-o um tópico de pesquisa popular nas áreas de inteligência artificial, engenharia do conhecimento, processamento de linguagem natural e representação do conhecimento. Esse interesse é justificado pela crença de que um entendimento comum e compartilhado sobre um determinado domínio de conhecimento poderá ser comunicado tanto entre pessoas como entre computadores.

Sales e Café (2008) consideram a ontologia como um modelo de representação do conhecimento que, a exemplo do tesouro, é utilizada para representar e recuperar informação por meio de estruturas conceituais que, no caso da ontologia, tem como meio de ação o informático.

O modelo de relações que aparece em vocabulários controlados e em tesouros é um modelo didático que revela apenas as categorias às quais os conceitos pertencem. Nas ontologias, as relações precisam ser explicitadas, pois integram uma proposição. Um importante elemento das ontologias é a representação do conhecimento, que envolve uma análise semântica, em especial para a organização de sistemas de conhecimento que, nas ontologias, visam aperfeiçoar a recuperação, ou seja, prover os computadores de mais inteligência. Trata-se de uma área de pesquisa estreitamente ligada à organização do conhecimento, em especial no que tange à rede semântica (SALES; CAMPOS; GOMES, 2008).

2.3 TRATAMENTO TEMÁTICO DA INFORMAÇÃO

O tratamento da informação, segundo Dias (2001), é definido como a função de descrever os documentos, tanto do ponto de vista físico (características físicas dos documentos) quanto do ponto de vista temático (ou de descrição do conteúdo). Para o autor essa atividade resulta na produção de representações documentais (fichas de catálogo, referências bibliográficas, resumos, termos de indexação etc.), constituindo-se de unidades mais fáceis de manipular num sistema de recuperação da informação do que o documento em sua íntegra, além de representar sínteses, tornando mais fácil a avaliação do usuário quanto à relevância para seu contexto da pesquisa.

Para Guimarães e Sales (2010), citando Barité (1998), o termo Tratamento Temático da Informação (TTI) abrange a organização de documentos, pois se centra nas questões atinentes “à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação”.

Por se abordar neste trabalho a construção em área específica do conhecimento e por se considerar que uma base de dados não pode ser generalista, adota-se o termo “Tratamento Temático da Informação” (TTI) dado seu caráter delimitador, caracterizado por especificidades e precisão inerentes à área.

A TTI no Brasil sofreu influência de três linhas distintas, que segundo estudos de Guimarães e Sales (2010), são historicamente constituídas por suas vertentes teóricas respectivas: a catalogação de assunto (*subject cataloguing*) de matriz norte-americana, a indexação (*indexing*) de matriz inglesa e a análise documental (*analyse documentaire*) de matriz francesa.

Nestas vertentes, os autores observam que as teorias de catalogação de assunto (a partir do final do século XIX) e da indexação (meados do século XX) estão mais voltadas à investigação e ao desenvolvimento de produtos (catálogos e índices) e de instrumentos para padronização da linguagem (tesauros e listas de cabeçalhos de assunto). Foi somente a partir da década de 1960 que houve uma preocupação mais específica com a análise documental, desenvolvendo referenciais teórico-metodológicos para os procedimentos envolvidos (GUIMARÃES; SALES,

2010). Esta vertente francesa no Brasil foi muito estudada pelos pesquisadores do GRUPO TEMMA¹.

Kobashi (1996) pondera sobre a consideração dos aspectos do TTI, caracterizando-os em relação aos seus referenciais teóricos, conceituais e metodológicos mais importantes e destacando a natureza de seu diálogo com as disciplinas que têm o texto como objeto. Fujita (1999) ressalta que o tratamento da informação, do conteúdo é realizado pelas operações de análise, síntese e representação.

Corroborando essa ideia, Moreira González (2002) argumenta que a qualidade das tarefas da análise de conteúdo documental (especificado na indexação e condensação) é essencial para permitir uma recuperação bem sucedida da informação e explicação adequada do conteúdo para os usuários.

Neste trabalho, a ênfase recai sobre o referencial teórico construído sobre análise documentária, por não tratar a Organização da Informação somente como instrumento, caracterizado pela vertente americana, que segundo Tobias (1998) tem a tendência de ignorar os vários tipos de usuários e as suas necessidades, desconsiderando o contexto da informação na recuperação pelo usuário. Essa herança é derivada dos conceitos tradicionais biblioteconômicos, cuja necessidade é focada na organização física de material na estante e não nas necessidades reais e individuais de informação pelo usuário.

Segundo Bufrem (2008, p. 45), para que as qualidades das informações possam contribuir para a satisfação informacional do usuário é necessária a revisão dos critérios aplicados à qualidade dos produtos originados dos processos de análise documentária e de conteúdo. A revisão desses critérios não se limita a uma medição estatística e numérica dos fatores qualitativos, mas se realiza mediante observação e juízos centrados em tarefas de identificação e avaliação de repertórios, da qualificação do analista, da existência de normas e diretrizes para sua organização e do tipo de linguagem de indexação. Segundo a autora (BUFREM, 2008), esses fatores qualitativos se expressam no grau de reutilização dos resumos e da indexação de origem, na forma de traslado da superestrutura, na qualidade técnica e na profundidade de análise. Para essa análise, devem ser considerados

¹ Grupo TEMMA da Universidade de São Paulo: Johanna W. Smit, Nair Kobashi, Anna Maria Marques Cintra, Maria de Fátima Gonçalves Moreira Tálamo e Marilda Lopes Gines de Lara.

indicadores de cobertura e solapamento dos repertórios, coerência, densidade informativa, exaustividade, legibilidade, pertinência e profundidade, entre outros.

Os problemas relativos à exaustividade, especificidade e coerência da indexação tornam-se cruciais para o gerenciamento e recuperação de informações nas bases de dados bibliográficos (NUNES; FREITAS, 1997).

Bufrem (2008, p. 47) critica a construção automática de listas de termos como elementos de classificação, corroborando as ideias de Lancaster (1993, p. 215), que sugere acrescentar ao sistema de construção da base de dados um tipo de “tesauro de busca”.

Além de representar o conhecimento, um sistema inteligente, segundo Kobashi (2007),

deve codificá-lo em uma linguagem que possa ser processada computacionalmente. As representações são expressas sob a forma de proposições lógicas. Segundo Sowa (2000), por meio de linguagens lógicas é possível representar apenas conhecimento declarativo. Contudo, segundo esse mesmo autor, o conhecimento do mundo não pode ser encapsulado em enunciados declarativos. Assim, no atual estágio de desenvolvimento dos recursos computacionais e da Inteligência artificial, não parece ser possível pensar a automatização completa dos processos de representação da informação.

Defendendo a vertente americana na criação de produtos e catálogos, Dias (2001) enfatiza que para ter sucesso nas atividades de tratamento de informação é necessário que outras atividades sejam realizadas, desenvolvidas fora do âmbito dos sistemas de informação e de recuperação da informação, com a criação e manutenção de linguagens e códigos e de linguagens de indexação, tais como listas de cabeçalhos de assuntos, sistemas de classificação, tesouros e códigos de catalogação de assunto.

Integrar as necessidades dos usuários com questões terminológicas em áreas específicas e construir instrumentos que facilitem o processo de indexação e catalogação em busca da redução da dicotomia entre a linguagem natural e o vocabulário controlado, sem perder as relações lógicas e o contexto dos conteúdos das informações, é um dos grandes desafios do meio digital.

2.3.1 Linguagem natural e vocabulários controlados

Os vocabulários controlados, também denominados de linguagens documentárias (LDs), são conhecidos como modelos de representação do conhecimento. Enquanto linguagens artificialmente construídas e constituídas de sistemas simbólicos, visam descrever sinteticamente conteúdos documentais e são utilizados nos sistemas informacionais para indexação, armazenamento e recuperação da informação (SALES; CAFÉ, 2008).

Para Xavier e outros (2004), entende-se que linguagem documentária é um sistema de signos estruturados, cuja finalidade é representar e recuperar as informações registradas nos documentos.

Segundo a vertente inglesa, são denominadas linguagens de indexação os instrumentos que auxiliam no processo de representação da informação que, segundo Moura (2009), visam estabelecer uma equivalência aproximativa entre as linguagens do autor do documento, descrita como natural, e a adotada pelo sistema de informação (controlada). Nesse sentido, busca representar o assunto de maneira consistente permitindo a compatibilidade e o diálogo entre a linguagem do autor, do indexador e do pesquisador.

Sales e Café (2008) enfatizam que um dos fazeres da CI é fomentar esforços teóricos e práticos para a precisão da transferência da informação especializada, tratando o conteúdo informacional de forma a permitir a eficiência de sua recuperação, e para obter sucesso nesta ação é aconselhável a adoção de linguagens documentárias.

As linguagens documentárias, por guardarem uma relação estreita com a língua natural e por contarem com um vocabulário ou léxico documentário, são uma estrutura e sintaxe que representam o conteúdo informativo de um documento e para isso utilizam ferramentas de padronização das operações de análise e síntese (DODEBEI, 2002, p. 49). A língua natural é traduzida para uma linguagem sintética, o que significa economia verbal, de símbolos, de espaço e de significado.

As linguagens documentárias avançaram seus escopos, segundo Sales e Café (2008), passando de sistemas alfanuméricos de classificação, ou até mesmo de sistemas internos de organização a instrumentos de controle terminológico.

2.3.1.1 Linguagem natural (LN) ou livre

Para Lancaster (1993, p. 200), a linguagem natural (LN) pode ser definida como a linguagem do discurso técnico-científico e, no contexto da recuperação da informação, a expressão normalmente se refere às palavras que ocorrem em textos impressos, considerando-se como seu sinônimo a expressão “texto livre”.

A LN se desenvolve na interação entre pessoas, sem controle terminológico, rica em sua atualidade, porém a comunicação acaba se tornando passível de variadas interpretações, gerando ambiguidades e interpretações errôneas. Hughes, citado por Nocetti e Figueiredo (1978 *apud* DODEBEI, 2002, p. 48), define a linguagem natural como um sistema de símbolos vocais arbitrários pelos quais o pensamento é transmitido de um ser para outro.

A linguagem natural tem sido, segundo Alvarenga (2006, p. 90), o meio de organização do universo da Web, pois a autora considera impossível um processamento dos objetos digitais com linguagens documentais estruturadas, usadas no tratamento de informação *a posteriori*, nas quais os termos simbolizam conceitos (e não palavras) e seriam usados como chaves de recuperação. Ainda, segundo a autora, o desafio de organizar a Internet extrapola as fronteiras disciplinares e torna-se a busca de uma solução para que o valor do meio possa efetivamente ser usufruído em toda a sua potencialidade.

Para Lopes (2002), nas bases de dados, os campos de título e resumo registram os termos da LN, enquanto os campos de descritores, termos de indexação ou identificadores registram os termos do vocabulário controlado. Na realidade, isso só ocorre quando a base de dados passa por um preparo antes de catalogar os registros na base. Essa atividade causa uma demora (*delay*) na atualidade da base, o que concorre para que muitas bases de dados utilizem os descritores enviados pelo autor e ou editor como elementos de indexação em prol da velocidade do processo, prevendo-se a busca por ferramentas e instrumentos de recuperação *a posteriori*.

2.3.1.2 Vocabulário controlado (VC)

Para o desenvolvimento de um instrumento que possibilite a linguagem utilizada para padronização dos descritores, faz-se necessário o uso de um vocabulário controlado (VC), um instrumento utilizado pelos indexadores com a finalidade de padronizar a representação do conteúdo temático dos documentos.

Para Lopes (2002) o que caracteriza o VC é seu conjunto limitado de termos autorizados para uso na indexação e busca de documentos, enquanto Laan (2002) utiliza uma definição de Barité (1998), que define vocabulário controlado como:

Conjunto estructurado de conceptos destinado a la representación del contenido de los documentos, y que comprende la organización lógica de dichos conceptos en distribuciones por clases o disciplinas, y las relaciones recíprocas y estables que mantiene entre ellos.

Um dos grandes problemas da utilização do VC é sua atualização, pois diversamente da linguagem natural, flexível e dinâmica, ele precisa de um consenso de especialista sobre a temática para definição do descritor que vai representar o tema. Sua atualização é tarefa constante, pois é necessário identificar e criar novos descritores, estabelecer novas relações, substituir, corrigir ou suprimir descritores e introduzir notas para diferenciar ambiguidades (DODEBEI, 2002, p. 48).

Essa rigidez nas definições dos descritores, principalmente pela definição em consenso de especialistas, faz com que o VC não consiga acompanhar a criação de novos termos produzidos pela ciência, principalmente nos últimos anos com as novas tecnologias. Esse entrave acaba provocando uma desatualização nos instrumentos de controle terminológico.

Com a definição das abordagens da Brapci foi possível delinear os caminhos necessários para o seu planejamento, e por conveniência partiu-se do estudo dos termos e as necessidades de organização.

A utilização de um VC não é o ideal para uma base que coleta dados de múltiplas publicações, e a utilização livre da linguagem natural não garante precisão na recuperação de informação (RI). O modelo conciliatório para o desenvolvimento de sistemas de RI, segundo Dias (2001), é a combinação de uma dupla abordagem

(vocabulário controlado e linguagem natural), cujas vantagens serão absorvidas pela metodologia empregada nesta pesquisa.

Para Moreiro González (2002), a qualidade dos descritores é fundamental para a precisão na recuperação da informação, pois, para se obter “precisão”, é necessário que exista qualidade cognitiva nas tarefas de análise de conteúdo documental. Ainda segundo o autor, os termos já vêm indicados pelos trabalhos, e a reindexação pode ser considerada pouco produtiva em grandes bases de dados. Além do consumo de tempo, Tobias (1998) aponta a tendência do indexador de ignorar os vários tipos de usuários e as suas necessidades, tendência que acaba descontextualizando a informação, o que prejudica a sua recuperação pelo usuário.

De acordo com Svenonius (2000 *apud* MOURA, 2009), existem atualmente inúmeras linguagens de indexação que exibem um variável grau de refinamento. Essas linguagens quando utilizadas nos sistemas de recuperação da informação podem provê-lo de qualidade adicional de modo que por seu intermédio se possa transformar informação em conhecimento. Segundo a autora, quando isso acontece a linguagem de indexação torna-se análoga ao próprio conhecimento.

Dias (2001) ressalta que não se pode ignorar que as TICs contribuíram significativamente para facilitar a recuperação de informações em ambiente Web. Esta facilidade vem das facilidades de armazenar grandes quantidades de dados e pelos mecanismos de buscas (*search engines*). Porém o autor destaca que não se pode descartar a necessidade de linguagens documentárias e, por conseguinte, de sistemas de classificação na organização de sistemas de informação. Afirma que o não uso revela uma ignorância grande das funções desses sistemas, sendo uma de suas finalidades a de permitir que consultas aparentemente bem definidas possam ser refinadas por meio do exame de documentos cuja temática remete para outras definições próximas da inicial.

Irving (1997 *apud* XAVIER *et al.*, 2004) ressalta que a indexação temática deve atribuir os descritores extraídos de um vocabulário controlado, identificando a categoria que mais adequadamente expressa o assunto genérico do texto. Partindo dessa concepção, entende-se que indexação é o processo pelo qual é determinado o tema principal, ou assunto, e os subtemas, ou assuntos secundários, tratados em um documento e, posteriormente traduzidos para uma linguagem documentária. Dos elementos que podem ser representados estão o título, o resumo, as palavras-chave, a autoria, a autoridade e o texto completo.

O controle do vocabulário controlado dá-se pelo controle de sua terminologia.

2.3.2 Terminologia

A terminologia tem como unidade padrão o termo, utilizando a definição da ISO1087 (ORGANIZAÇÃO INTERNACIONAL DE NORMALIZAÇÃO, 1990, p. 5) como “designação, por meio de uma unidade linguística, de um conceito definida em uma língua de especialidade”. O termo, segundo Barros (2004, p. 40) é “uma unidade lexical com um conteúdo específico dentro de um domínio específico”.

Segundo Bessé (1998, p. 255 *apud* CAFÉ; MENDES, 2009a),

[...] para que uma palavra tenha o direito ao título de termo, será necessário que ela possa, como elemento de um conjunto (uma terminologia), ser distinguida das outras. A única maneira para expressar este sistema de distinções recíprocas é a operação dita definição.

Para Tristão e outros (2004) a definição do termo é importante, na medida em que ele fornece características do conceito que vão permitir seu agrupamento, a formação das categorias e indicar as relações que se estabelecem entre conceitos. As características usadas na comparação entre conceitos conduzem a um sistema de classificação de conceitos.

O objeto de estudo da terminologia, segundo Barros (2004, p. 34), é o conjunto de termos de um domínio e dos conceitos (ou noções) por eles designados, fornecendo as bases teóricas e metodológicas para inúmeras pesquisas. A autora ainda ressalta que existe confusão na definição de terminologia e nomenclatura.

No contexto histórico, Eugen Wüster (BARROS, 2004, p. 53) é considerado o precursor da Terminologia, fundador da Escola de Viena, cujo princípio régio é o da univocidade entre o conceito e o termo que o designa (um único termo pode designar um conceito), desse modo não existiria ruído na comunicação entre especialistas de um mesmo campo do saber ou de diferentes campos.

Em contexto histórico, segundo Cabré (1993 *apud* BARROS, 2004) a terminologia passou por quatro fases:

- a) 1930-1960: as origens da terminologia, como disciplina científica dão-se na Alemanha com Wüster e na ex-URSS com Lotte, com ênfase no caráter sistemático das terminologias;
- b) 1960-1975: estruturação, influenciada pelo desenvolvimento da informática, com o surgimento dos primeiros bancos de dados terminológicos monolíngues, bilíngues e multilíngues; quando as técnicas documentais de recuperação da informação evoluem consideravelmente. A terminologia adquire dimensões internacionais e a abordagem normativa das línguas e das terminologias desenvolve-se de modo expressivo;
- c) 1975-1985: eclosão, proliferação de políticas de planejamento linguístico, e pela popularização da informática. A terminologia desempenha papel importante em processos de normalização e harmonização terminológicas, modernização vocabular e transmissão de conhecimentos técnicos e científicos;
- d) 1985-década de 1990: fase caracterizada pela expansão territorial e científica da Terminologia, alcançando a América Latina e países dos continentes asiáticos e africanos. Diversificação dos temas tratados, organização de redes internacionais para intercambio científicos. A Terminologia assume dimensões e articula-se no plano internacional.

Ao complementar esse quadro histórico, Barros (2004, p. 35) acrescenta mais uma fase, a partir do início da década de 1990, que segue até os dias atuais, proporcionando à Terminologia revisões, questionamentos e novas abordagens no mundo todo.

O estudo e o entendimento das diversas vertentes da terminologia são de suma importância quando se pretende construir um sistema baseado em conceitos que possa ser representado por descritores, e este deve ser de forma controlada para garantir sua qualidade.

Cabré (1993, p. 43) define que a “[...] Terminologia nasce da necessidade manifestada pelos especialistas de organizar as denominações de seus sistemas de conceitos, com a finalidade de conseguir uma comunicação profissional mais confiável”. Um dos aspectos postulados pela Terminologia clássica é a normalização dos termos, fixando o uso de um termo e descartando a utilização de outros para o mesmo conceito.

A construção da terminologia deve ser dirigida pela padronização do vocabulário técnico e científico, possibilitando a melhoria da comunicação entre os especialistas da área. Para Xavier e colaboradores (2004), esse tem sido um dos objetivos das Ciências da Informação, considerando-se que a linguagem é o veículo para a representação e recuperação de informações.

A padronização do vocabulário contrapõe-se à dinâmica da língua. As normas são estáticas e a língua é dinâmica. O falante está constantemente criando novos termos, alterando os padrões. Portanto, a padronização terminológica especializada não é garantia de uma relação comunicativa bem sucedida, pois depende de formalização, contrapondo-se à linguagem natural.

Para tentar solucionar o problema, parte-se do princípio de que o conceito é a unidade que pela sua carga semântica cria o sentido, que os termos são utilizados para representar o conceito e que este, por sua vez, pode ser representado por vários termos. Moreiro González (2010) afirma que o emprego de polissemia de termos não é prejudicial à terminologia, mas importante para as múltiplas descrições de um mesmo conceito buscando uma relação entre a linguagem natural e a controlada. Para fundamentar a teoria do conceito, parte-se dos primeiros epistemólogos sobre o assunto.

2.3.2.1 Teoria do Conceito Dahlberg

A teoria do conceito parte da afirmação de que os conceitos são sínteses rotuladas de enunciados verdadeiros sobre objetos do pensamento: esses enunciados - asserções - levam ao reconhecimento ou à separação das características dos conceitos que também podem ser consideradas como os elementos dos conceitos (DAHLBERG, 1978). Para Dahlberg, o conhecimento se fixa pela linguagem, havendo entre eles relação direta e dependente. Desse modo, novos conhecimentos apareceram com novos elementos linguísticos e também por meio destes tornaram-se mais claros e precisos. A autora define que conceito é a unidade de conhecimento que surge pela síntese dos predicados necessários relacionados com determinado objeto e que, por meio de sinais lingüísticos, pode ser comunicado.

Ao focalizar o conceito no contexto de um vocabulário controlado, Dodebei (2002), define-o como um item que reflete o conhecimento “do mundo real dos objetos e dos fenômenos, com seus atributos e relacionamentos semânticos e acidentais”. Nesse contexto, o conceito é uma unidade informacional que deve ser expressa por termos e simbolicamente representada por uma expressão (sintagma ou palavra). Assim os conceitos podem ser representados por apenas uma palavra ou por uma expressão, desde que essa expressão designe um único conceito. Desse modo, os conceitos são representações mentais das relações que ocorrem entre objetos na realidade empírica.

Com base na definição de conceitos, foi desenvolvida a Teoria Geral da Terminologia (TGT) para poder ser empregada na área de documentação.

2.3.2.2 Teoria Geral da Terminologia (TGT)

A TGT, segundo Campos (2001), é uma disciplina científica que possibilita uma base para o trabalho terminológico. A TGT visa à elaboração de definições orgânicas, além de estabelecer princípios para criação de novos termos e possibilitar, ainda segundo a autora, a comunicação mais precisa entre especialistas de diversas áreas do conhecimento nos diversos âmbitos da Ciência e Tecnologia.

A TGT foi desenvolvida por Eugen Wüster (1971 *apud* CAMPOS, 2001), sistematizando princípios terminológicos que deram à atividade um caráter científico próprio, estudando os princípios estabelecidos para a determinação de conceitos e diferenciando-se da atividade lexicográfica. A TGT inicia com a definição de um conceito, tendo uma unidade de denominação que é o termo. Uma das diferenças fundamentais entre a Lexicologia e a pesquisa terminológica, segundo Barros (2004), é que a primeira trabalha no âmbito da língua natural, fruto de uma evolução histórica, enquanto a segunda trabalha no âmbito de linguagem artificial, configurada dentro de um determinado grupo de especialistas, construída de maneira a permitir uma relação unívoca entre o conceito e a denominação. A pesquisa terminológica da TGT limita-se a situações bem definidas e não pode ser generalizada.

Para a TGT, conceito é uma unidade de pensamento, constituído de características que refletem as propriedades significativas atribuídas a um objeto, ou

a uma classe de objetos. Sua finalidade, na concepção de Campos (2001, p. 71), é permitir a ordenação mental e a comunicação por meio de símbolo linguístico que é o termo, compondo-se a tríade linguagem-pensamento-realidade. Na TGT, o termo que representa o conceito pode ser constituído de uma ou mais palavras, uma letra, um símbolo gráfico, uma abreviação, uma notação (CAMPOS, 2001, p. 73).

Barros (2004) ressalta que a TGT busca princípios que visam propiciar uma correspondência exata entre conceitos e termos, com o objetivo de facilitar a comunicação nos vários domínios da Ciência e da Tecnologia. Um desses princípios é o da univocidade.

Porém, a TGT possui algumas limitações, principalmente por não poder empregar unidades polilexicais para representar um conceito, tornando-se insuficiente para trabalhar com o dinamismo das fontes eletrônicas de documentos.

2.3.2.3 Teoria Comunicativa da Terminologia (TCT)

Uma tentativa de aperfeiçoar a TGT vem da Teoria Comunicativa da Terminologia (TCT) criada por Maria Teresa Cabré na *Universitat Pompeu Fabra* (Barcelona) no início da década de 1990. A TCT é uma teoria descritiva de base linguística e perspectiva funcionalista focada no caráter comunicativo do termo. Cabré (1999 *apud* SALES; CAFÉ, 2009b) definiu uma teoria generalizada, levando em consideração que a terminologia é interdisciplinar (integrando aspectos da linguística, das ciências cognitivas e das ciências sociais) e transdisciplinar (atua em todas as disciplinas).

Para Cabré (1993), a TCT não considera os termos como unidades isoladas que constituem seu próprio sistema, mas sim, considera-os como unidades que se incorporam no léxico de um falante, no momento em que este adquire o *know how* (saber) de especialista por meio da aprendizagem do conhecimento especializado.

A TCT visualiza os termos como uma unidade de conhecimento, composta por uma forma e um conteúdo, sendo a forma a unidade lexical que denomina o conceito (conteúdo) (BARROS, 2004). Com base nos fundamentos da TCT são extraídos os elementos de observação para análise: o **termo**, considerado seu caráter de unidade de conhecimento pertencente à linguagem natural e as distintas

funções destes no contexto discursivo; o **conceito** e seus diferentes tipos de relações, formadores da estrutura conceitual e os **objetivos** (teóricos e práticos) atribuídos aos modelos em questão (SALES; CAFÉ, 2009a).

A TCT é a que mais se adéqua à organização de documentos digitais, porém apresenta algumas limitações, principalmente quando atua na área de multidisciplinaridade, podendo existir diversas relações para um mesmo conceito dependendo da temática ou subtemática em sua construção.

O importante é ressaltar que o conceito como unidade, representado por termos, imagens, sons, grafos é a unidade que pode gerar uma carga semântica para criar instrumentos de recuperação de informação, promovendo desde uma relação sintática até relações de ontologias.

Para a formação do *corpus* que será organizado, deve-se partir de fontes de informações com credibilidade científica e com métodos e técnicas definidos. Para a Ciência da Informação, o campo que estuda a produção e disseminação de conteúdo relevante para pesquisadores é a Comunicação Científica.

2.4 COMUNICAÇÃO CIENTÍFICA

A Comunicação Científica é definida por Garvey e Griffith (1979, p. 127-163) como o conjunto de atividades associadas à produção, disseminação e uso da informação. Na fase da pesquisa ocorre a geração da informação e a disseminação se dá pela transferência da informação por meio de canais de comunicação heterogêneos, os quais podem ser formais ou informais, orais ou escritos. A divisão da estrutura da pesquisa em *inputs* (política científica, recursos disponíveis, procedimentos metodológicos) e *outputs* (comunicação, acessibilidade e visibilidade) deixa clara a ideia de que não se pode pensar em pesquisa, sem considerar, ao mesmo tempo, as formas como seus resultados serão divulgados (MEADOWS, 1999).

A comunicação científica é um processo inerente ao fazer científico, e sua relevância sempre foi reconhecida pelos cientistas que, ao longo dos tempos, instituíram diferentes canais de intercâmbio. O periódico científico é considerado atualmente o principal canal da comunicação científica formal (OLIVEIRA, 2005).

Pecegueiro (2002) define a importância da divulgação dos resultados das pesquisas e seu acompanhamento pelo pesquisador ao afirmar que:

[..] o acompanhamento do que está sendo produzido na sua área dará ao pesquisador condição de melhor desenvolver seu trabalho, irá atualizá-lo e dará subsídios para que ele possa avançar cada vez mais e melhor. Por isso, é importante a divulgação do resultado – total ou parcial – dos seus estudos, que, após lido, criticado e aceito por seus pares, concederá ao cientista segurança de estar no caminho certo.

A divulgação do andamento e dos resultados das pesquisas é de vital importância para que o ciclo da comunicação científica se complete (pesquisa – divulgação – leitura – validação e aceitação pelos pares – pesquisa), proporcionando o progresso da ciência com a geração de novos conhecimentos ou utilização de conhecimentos já produzidos (OLIVEIRA, 2005).

Na visão de Meadows (1999, p. vii), a comunicação científica situa-se no próprio coração da ciência. É para ela tão importante quanto a própria pesquisa, pois a esta não cabe reivindicar com legitimidade enquanto não houver sido avaliada e aceita pelos pares. Desta forma, o meio para sua explicitação é a comunicação.

No sistema científico e tecnológico, a publicação é o principal indicador de excelência profissional e um meio para sua avaliação. Segundo Furtado (1981), esse fenômeno se verifica conforme critérios da aristocracia vigente, gerando dissonâncias, pois a avaliação e conquista de posições não dependem tanto da informação publicada, mas dos tipos de produtos e processos gerados, da rentabilidade econômica ou social envolvida e do prestígio conquistado. O próprio tipo de atividade e diferentes campos de atuação de cientistas e tecnólogos ocasionam conflitos conceituais e semânticos sobre a natureza da informação. A forma e o conteúdo da publicação variam nessas duas áreas, pelo fato de serem destinadas a leitores distintos, mas as dificuldades não estão limitadas à linguagem, pois podem se originar da própria estrutura institucional. Apesar das diferenças tipológicas entre cientistas e tecnólogos, um elemento comum é exigido, a competência (BUFREM, 2008, p. 40).

O crescimento do número de títulos de periódicos científicos reflete a expansão da atividade científica. Em 1760, existiam em todo o mundo dez revistas científicas. Entre os anos de 1750 e 1950, o número de periódicos apresentou aumento de dez vezes a cada quinze anos. Em 1962, existiam cerca de 35 mil

periódicos no mundo, que geravam em conjunto um milhão de artigos por ano. Estima-se que atualmente haja cerca de cem mil periódicos no mundo (OLIVEIRA, 2005).

Preocupado com a superprodução científica, o governo americano (UNITED STATES OF AMERICA, 1963) produz um documento que define uma política que deveria ser adotada tanto pelos pesquisadores, como profissionais da informação e do governo. Esta política é explicitada no *Relatório Weinberg* (nome do relator do projeto) de 1963. Neste informe, a comunidade científica e governo são responsabilizados pela transferência de informação e por uma proximidade informacional das pesquisas fundamental, aplicada, tecnológica e industrial. O relatório faz um alerta às principais comunidades científicas mundiais que passam, junto a seus governos, a estabelecer políticas de informação denominadas de caráter científico e tecnológico, em que o poder do controle da informação, responsável pelo desenvolvimento técnico-produtivo fica dividido entre os Estados e a Ciência e Tecnologia ou a comunidade científica.

A partir deste relatório existiu uma intervenção político-administrativa em planejamento, gestão, monitoramento e avaliação das atividades e dos resultados de empreendimentos científicos, adquirindo vulto a leitura “científica”, produzindo parâmetros externos e observáveis da produção científica. É a partir relatório que se estabelecem pelo ISI e JCR os primeiros indicadores da produção científica, das instituições de pesquisa e dos pesquisadores, como reconhecimento à “inteligência científica” (GONZÁLEZ DE GÓMEZ, 2003).

Na década de 1960, o paradigma dominante da comunicação científica é representado por Hurd (2000) e Le Coadic (2004, p. 35) como o modelo que pode garantir qualidade para a produção científica, transferindo a responsabilidade da avaliação e crítica dos trabalhos para seus pares, evidenciada na Figura 10, este modelo ficou conhecido como modelo de Garvey e Griffith. Por meio dele, os resultados da pesquisa antes de concluída são apresentados e recebem as primeiras críticas de seus pares, para que após sua conclusão o estudo seja apresentado em seminários, colóquios e eventos para um “amadurecimento” das ideias, até que estejam prontas para serem submetidas a um periódico. Nesta fase a pesquisa é avaliada por pareceristas, especialistas em suas áreas, emitindo-se um relatório da qualidade da pesquisa, passando pelo crivo para ser publicado em um

periódico e consequentemente indexado nas bases de dados e edições anuais de *abstracts* e devidamente pontuado (JCR).

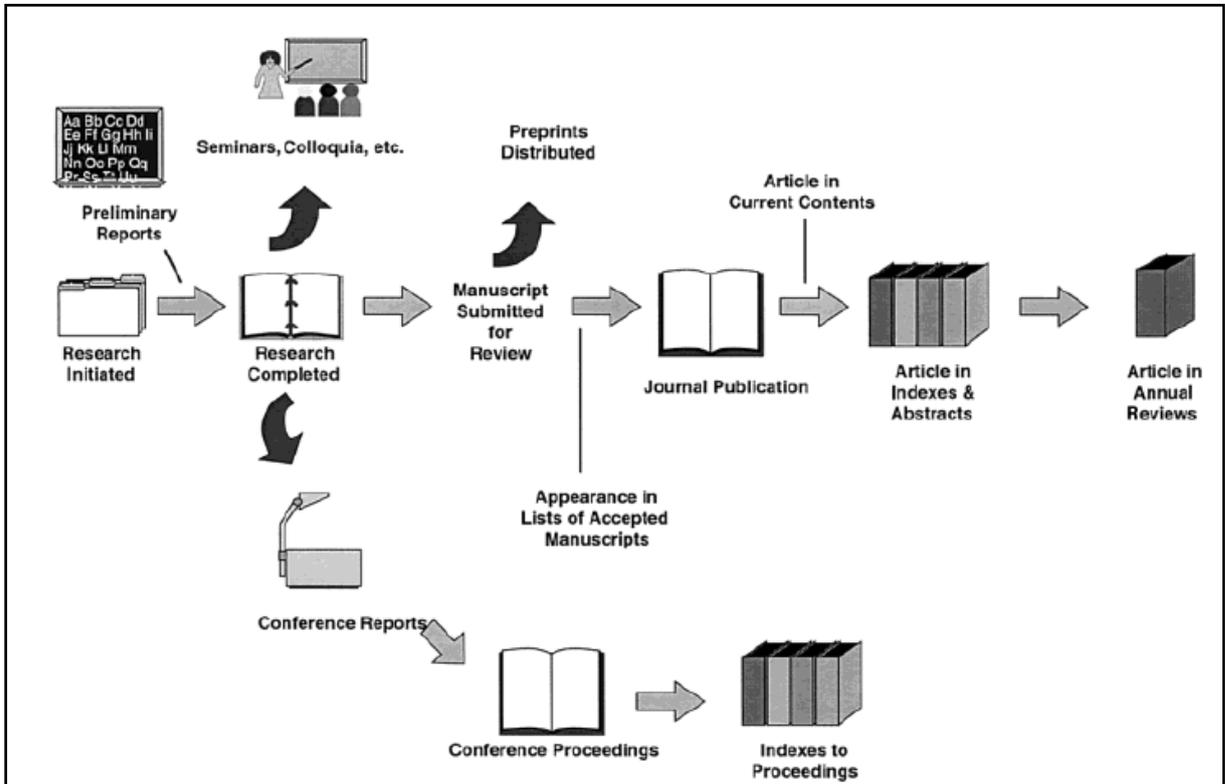


FIGURA 10 - MODELO DE COMUNICAÇÃO CIENTÍFICA DE GARVEY E GRIFFITH
Fonte: HURD (2000)

Este ainda é o modelo que prevalece na comunidade científica nos dias atuais. Com o desenvolvimento das TICs, surgem novas perspectivas sobre a comunicação científica e o modelo de validação por pares começa a ser questionado, novas propostas emergem como substitutivos às formas tradicionais, com novas situações e modos de avaliação.

Em 1977, Lancaster previa que os periódicos científicos iriam alterar o modelo de explicitação e abrangência, indicando que seu suporte iria passar da fase de impresso para a dualidade (impresso e eletrônico), até se firmar definitivamente no suporte eletrônico. Nesta perspectiva de facilidade de comunicação e publicação, Hurd (2000) propõe um novo modelo de comunicação científica para 2020, representado pela Figura 11. A autora (HURD, 2000) sugere a forma como as pesquisas serão avaliadas, caracterizada por um dinamismo maior do que no modelo tradicional. Dessa forma, as pesquisas, antes de avaliadas e publicadas em revistas eletrônicas, são armazenadas em servidores *pre-print* para disponibilização

às comunidades (como forma de registro). Neste modelo proposto por Hurd (2000) novos elementos como *pre-printer*, *e-Journal*, *library digital* e *aggregator site* são formalizados.

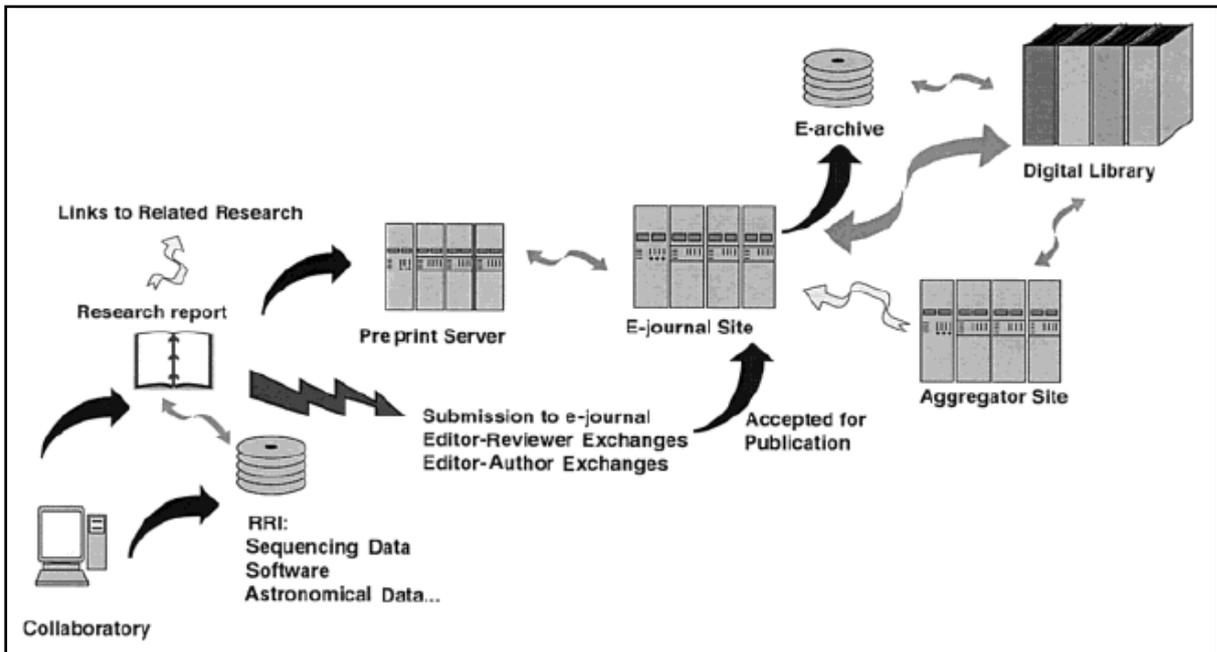


FIGURA 11 - MODELO DE COMUNICAÇÃO CIENTÍFICA DE HURD
Fonte: HURD (2000)

A metodologia proposta neste estudo vem ao encontro das propostas de Hurd (2000) por ser um elemento agregador, “*aggregator site*”, que reitera a necessidade de “provedores de serviço” com o objetivo de concentrar informações temáticas em um único repositório, além de monitorar e gerar indicadores dessa produção, propondo mecanismos mais eficientes e eficazes de recuperação de informação.

Antecipando as possibilidades de comunicação imediata da produção científica, Barreto (1998) ressalta que os meios eletrônicos vieram definitivamente liberar o texto e a informação de uma ideologia envelhecida e autoritária dos gestores da recuperação da informação, defensores de uma pretensa qualidade ameaçada, os fatais intermediários e porta-vozes que vêem seus poderes ameaçados cada vez mais pela facilidade de convivência direta entre os geradores e consumidores de informações. Desse modo, o artigo científico passa não mais a representar a única forma de disseminação de conhecimento, pois novos suportes e tipologias de documentos impõem-se, havendo uma necessidade de definir um

termo que abranja toda essa produção. Esse conceito foi denominado de produção intelectual.

Para efeito desta pesquisa adotou-se a definição de produção intelectual para todos os trabalhos produzidos por autores individuais ou em coautoria, independente de sua tipologia e de conteúdo que estão disponíveis em publicações científicas, relacionando dentre esses: artigos científicos, resumos, resenhas, recensão, opinião, editorial, áudio e vídeo entre outros. Esta definição faz-se necessária para identificar as várias modalidades de trabalhos coletados e armazenados dentro das bases de dados.

O termo Produção Intelectual segue definição encontrada dentro do *site* do Programa de Pós-Graduação em Artes da Unicamp (2010):

Produção Intelectual são realizações de própria autoria ou em parceria, dadas à publicação, que se materializam em forma bibliográfica (livros, capítulos de livros, artigos em periódicos ou anais...); técnica (apresentação de trabalho, organização de eventos, organizar e/ou ministrar cursos de curta duração, relatório de pesquisa...); artística/cultural (apresentação de obras artísticas, obra de arte visual, arranjo musical).

Partindo desta definição, todo trabalho tratado na metodologia deste estudo será identificado como trabalho intelectual ou por suas especificidades quando já caracterizados (artigos, resenhas, entrevistas entre outros).

Por sua vez, toda produção intelectual produzida para ter seu valor científico precisa ser comunicada. Esta atividade de disseminação, em sua fase inicial de comunicação, deve atender tanto ao produto quanto aos autores, promovendo visibilidade social no meio científico. Dois tipos de instrumentos são valorizados nesta fase pela comunidade e são os: congressos e reuniões científicas e o periódico científico.

2.4.1 Periódico científico

A publicação periódica de acordo com a norma NBR 6023:2002 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002) pode ser disponibilizada em qualquer tipo de suporte, deve ser editada em unidades físicas sucessivas com designação numérica e ou cronológicas e destinadas a ser continuada indefinidamente. Já o periódico científico além de atender as normas de publicação periódica, também em seu conteúdo deve conter artigos de autoria simples ou múltiplas que seguem uma metodologia em sua construção e com caráter científico.

O periódico científico desempenha papel fundamental no processo de comunicação científica, registrando publicamente o conhecimento e permitindo que flua tanto do produtor como do consumidor, aprovando novos conhecimentos produzidos pelos integrantes da comunidade científica, viabilizando o intercâmbio e fortalecendo a geração de saberes e constituindo em um canal de comunicação que viabiliza o projeto de desenvolvimento científico da área e do país onde se insere, disseminando o conhecimento científico e compondo a memória científica nacional.

Para Meadows (1999), a concepção do periódico científico resulta da própria evolução conceitual, tendo sua origem do inglês *journal*, que designa coletânea de artigos científicos de diferentes autorias, reunidos em intervalos, impressos, encadernados e distribuídos sob um único título.

As citações dos periódicos em outros textos têm sido objeto de análise e comprovam efetivamente sua importância como elementos estimuladores da produção científica. Segundo Mueller (1999, p. 3), a inclusão dos periódicos em índices e bases internacionais garante aos artigos indexados a visibilidade necessária, facilitando a busca dos autores por literatura recente e aumentando a chance de serem lidos, citados e conseqüentemente cumprindo o ciclo leitura, produção, citação, leitura. Vale ressaltar que entre os artigos analisados sobre o tema, quando da caracterização dos periódicos científicos nacionais, foi decisiva a leitura de Ohira, Sombrio e Prado (2000), sobre as características e disponibilidade desses veículos da produção científica especializada.

Zimba e Mueller (2004) definem visibilidade científica como “o grau de exposição e evidência de um pesquisador ante a comunidade científica”. Essas

autoras concluem ainda que a visibilidade elevada ocorra quando os trabalhos e ideias do pesquisador se tornam acessíveis de maneira fácil, o que aumenta as chances de que possam ser recuperados, lidos e citados. As maiores possibilidades de isso ocorrer acontecem com os trabalhos publicados em periódicos internacionais indexados em bases de dados especializadas e internacionais. No entanto, os periódicos internacionais privilegiam as pesquisas alinhadas ao *mainstream* da área, o que exclui os resultados de pesquisas de interesse regional ou nacional, mesmo que tenham qualidade internacional.

Os periódicos nacionais têm então importante papel na disseminação da informação para a comunidade científica de seu país, principalmente em áreas mais voltadas aos temas de interesse nacional. A indexação dos periódicos em bases de dados para a disseminação da informação e a visibilidade da produção nacional, reside no aumento da possibilidade de que um artigo seja “visto quando cientistas pesquisarem a literatura para novas descobertas em seus campos e decidirem qual trabalho citar em seus próprios artigos” (GIBBS, 1995, p. 76, *apud* OLIVEIRA, 2005).

Muitos periódicos nacionais não conseguem ser indexados em bases de dados internacionais, principalmente por sua restrição temática local, pouco relevante para comunidade científica internacional, mas muitas vezes de suma importância para comunidade local a qual esta inserida. Parte daí a necessidade da criação de bases de dados que não criem critérios rígidos para incorporação de títulos, mas que sejam representativos de uma área temática e que atendam a necessidade de informação da comunidade nela disponibilizada.

2.4.2 Base de dados

Para a concretização de uma base de dados é necessário compreender a necessidade de informação que precisa ser armazenada e modelada em um ambiente computacional. A representação das estruturas de informação do mundo real para uma base de dados foi a proposta de Chen em 1990, conhecida como Modelo de Entidade e Relacionamento (MER) e que sugere a possibilidade de capturar as estruturas de dados empregando três conceitos fundamentais: entidade, atributo e relacionamento. Por definição a entidade é a que representa um determinado grupo de informações, como os títulos das revistas, os fascículos ou os

artigos. O atributo é o que define as características de cada entidade, da forma detalhada e o relacionamento estabelece as relações entre as entidades. No modelo de Chen (1990), a entidade e o atributo são obrigatórios, e o relacionamento opcional.

As estruturas de relacionamento que fazem ligações entre entidades podem ser de três tipos: **um para um**, quando um dado de uma entidade se relaciona somente com um dado de outra entidade; **um para muitos**, quando um dado da entidade pode se relacionar com vários outros dados de outra entidade; e **muitos para muitos**, quando os dados de uma entidade podem se relacionar com vários outros dados de outra entidade.

Os atributos são os que descrevem uma entidade, e seus valores ocupam a maior parte dos dados armazenados na base de dados. As entidades podem ser classificadas em dois tipos: supertipo e subtipo, não mutuamente excludentes. Para exemplificar, tem-se uma entidade PUBLICAÇÃO e uma entidade FASCÍCULO, diz-se que PUBLICAÇÃO é um supertipo de FASCÍCULO e FASCÍCULO é um subtipo de PUBLICAÇÃO.

Na área da CI, Rowley (1994, p. 66) define base de dados (biblioteca digital) como “uma coleção de registros similares entre si e que contém determinadas relações entre esses registros”. Registro é a informação que a base contém e que diz respeito a um documento ou um item. Já, para Cianconi (1987, p. 55), base de dados é “um conjunto de dados inter-relacionados, organizados de forma a permitir a recuperação de informação” (RODRIGUES; SILVA, 2007).

A base de dados pode ser considerada como uma biblioteca digital por disponibilizar conteúdos organizados para a uma comunidade específica. Para Tammaro e Salarelli (2008, p. 145) ao se criar um projeto de biblioteca digital, deve-se em primeiro lugar partir do contexto social predominante no qual os documentos digitais serão criados, compartilhados e usados, incluindo a cultura organizacional das instituições. Uma vez definido o escopo do público, deverá decidir sobre o que terá acesso os serviços, quais coleções serão escolhidas, quais recursos e serviços serão eliminados para permitir obter recursos financeiros destinados a um programa de biblioteca digital. Os autores ressaltam ainda que as tecnologias sozinhas tendem a empobrecer a humanidade, que necessitam serem combinadas com o desenvolvimento cultural e social das pessoas e seu contexto, tendo na comunidade sua representação pertinente e a influência recíproca que o usuário se faz dela.

Para Rodrigues e Silva (2007), as bases de dados têm condições de estabelecer indicadores da produção técnico-científica, informações acerca do registro intelectual da produção científica nacional, dados importantes para gerar indicadores, análises e estudos de domínios, colaboração científica e conteúdos.

O conceito para construção de uma base de dados envolve as questões sobre a sua natureza e a finalidade do seu conteúdo. A composição de seu *corpus* define-se a partir da extensão e profundidade na cobertura temática, segundo os níveis estabelecidos no planejamento. Segundo Bufrem (2008, p. 40), não se justifica a criação de uma base de dados se não for considerado o usuário como elemento orientador para a avaliação crítica dos processos e do produto gerado.

De acordo com Rowley (1994) as bases de dados podem ser classificadas em dois tipos: bases de dados referências e bases de dados de fontes. A primeira encaminha ou orienta o usuário para outra fonte, que pode ser um documento, uma instituição ou um indivíduo, a fim de obter informações adicionais ou conseguir o texto integral de um documento. Segundo Rodrigues e Silva (2008), essas bases referenciais subdividem-se em três tipos: **bibliográficas**, que informam ao usuário sobre o que foi publicado e onde se publicou (ROWLEY, 1994, p. 68); **catalográficas**, que mostram o acervo de determinada biblioteca ou rede de bibliotecas e, **referenciais**, que identificam informações ou dados, como nomes e endereços de instituições, e outros dados típicos de cadastros.

As bases de dados de fontes contêm os dados originais e constituem um tipo de documento eletrônico, ou seja, contêm informações primárias e, portanto não remetem o usuário a outras fontes (LOPES, 1991; ROWLEY, 1994). Podem ser agrupadas segundo seu conteúdo, como numéricas contêm dados estatísticos e de resultados de pesquisas; texto integral: contém notícias de jornal, especificações técnicas e programas de computador; textuais e numéricas: contêm uma mistura de dados textuais e numéricos (RODRIGUES; SILVA, 2007).

Um *corpus* pode ser definido, segundo Bufrem (2008, p. 51), como um conjunto de conteúdos componentes de uma determinada base. Para a autora destaca-se o estabelecimento do recorte, fundado no interesse social e propósito de um instrumento de recuperação da informação.

Considerando-se que, para a definição de um *corpus* ou conjunto de componentes de uma determinada base é necessário também estabelecer o recorte fundado no interesse e propósito dessa base, a tarefa pode implicar a rejeição de

parcela da realidade, constituindo-se em jogo dialético entre a extensão de elementos disponíveis e a seleção dos elementos pertinentes e adequados às necessidades (BUFREM, 2008, p. 41).

Para que se possam analisar os elementos representativos de cada um dos itens do *corpus*, destacam-se indicadores da estrutura de relações linguísticas utilizadas para condensar as informações, de modo a oferecer um conjunto mínimo de dados que permitam uma recuperação do desejado dentro do conjunto elementos disponibilizado com qualidade dos processos de representação informacional (BUFREM, 2008, p. 50).

A definição criteriosa do recorte, da amostra ou da explicitação do universo da base de dados é justificada pela precisão na recuperação da informação em sua pertinência temática, pois um *corpus* muito grande e diversificado gera uma dispersão nos conteúdos, enquanto se centrado em uma única área qualifica os indicadores recuperados.

2.4.3 Qualidade da Informação

O acompanhamento da literatura sobre a produção científica é uma preocupação que vem se ampliando à medida que se registram formalmente as pesquisas na área de CI. Impõem-se reflexões críticas sobre os modos de construção desse referencial, especialmente para aqueles cuja vocação seja o aprimoramento das bases de dados que os representam (BUFREM, 2008, p. 38).

Para Pereira e outros (1999, p. 215) é necessário definir critérios de qualidade da informação. Quando se trata de organizar uma base de dados, ressalta sua importância como insumo básico para a “sociedade do conhecimento”, sendo a qualidade um dos “aspectos mais importantes da indústria da informação na atualidade”.

A questão da qualidade da informação perpassa pela diferença entre o que o usuário deseja (expectativa) e o que efetivamente o serviço oferece (percepção) (PARASURAMAN; BERRY; ZEITHAM, 1985). Para chegar a esse resultado, é necessário criar subsídios para o usuário opinar sobre o serviço que está sendo oferecido.

A análise sobre indicadores de qualidade é parte das atividades de um processo em que se clarificam e distinguem a provisão da base e o seu monitoramento. Entretanto, deve-se considerar que os indicadores não são um fim em si mesmo, mas instrumento prático para a administração de bases de dados e tomada de decisão.

Pereira e outros (2008) ressaltam que, para a avaliação da qualidade da informação em uma base de dados, deve-se partir de uma visão macro e basear-se em duas vertentes de avaliação: a qualidade das estruturas de armazenamento e recuperação e a qualidade do conteúdo da base.

Para Bufrem (2008, p. 48), a qualidade de uma base de dados depende não apenas do conteúdo e da sua adequada representação, mas da atualização periódica dos dados dentro dos padrões estabelecidos para o acompanhamento das publicações, sua padronização, controle de qualidade e segurança.

Na estrutura de armazenamento e recuperação, a participação do usuário é fundamental para que haja qualidade. O usuário é ator fundamental nesse processo, pois é ele que avalia a qualidade dos serviços oferecidos, a partir do grau de sua satisfação. Outro aspecto importante na qualidade das bases de dados é o estudo do seu conteúdo, pois uma base de dados com erros de conteúdo pode trazer grande prejuízo para a sociedade com dados e informações incorretas (RODRIGUES; SILVA, 2008).

Para o processo de avaliação da informação, De Sordi (2008) propõe algumas dimensões que devem ser analisadas quando se necessita mensurar a qualidade da informação, sugerindo que sejam observados alguns aspectos principais para análise:

- a) abrangência e escopo, verificando-se a qual público-alvo interessa a informação, para evitar excessos desnecessários;
- b) integridade da informação;
- c) acurácia e veracidade, ou a fidelidade da informação aos fatos que ela representa;
- d) confidencialidade e privacidade, ou a capacidade da informação de ser acessada somente a quem de direito;
- e) disponibilidade, ou facilidade de acesso por quem de direito;

- f) atualidade, ou seja, a periodicidade com que a informação é gerada e atualizada nos intervalos de tempo considerados adequados pelo público-alvo;
- g) ineditismo e raridade;
- h) contextualização, ou capacidade de atrair o público-alvo;
- i) precisão, quando a informação está suficientemente detalhada e esmiuçada para o pronto uso;
- j) confiabilidade, quando a fonte e o conteúdo da informação têm credibilidade perante o público-alvo;
- k) originalidade, quando a informação com que se trabalha provém diretamente da fonte geradora ou de algum retransmissor;
- l) pertinência e agregação de valor, que representam o quanto a informação é importante ao público-alvo para atender a sua necessidade;
- m) identidade, quando a denominação da informação é representativa, pertinente e fiel no que tange ao seu conteúdo;
- n) audiência, quando a informação está sendo acessada pelo seu público-alvo.

As dimensões apresentadas por De Sordi (2008) são importantes quando se pretende o reconhecimento e referência da comunidade da área. Com o objetivo de obter o máximo das dimensões, as tecnologias da informação alinham-se para o desenvolvimento da qualidade da informação.

2.4.4 Tecnologias

O avanço da ciência, aliado ao expressivo progresso das telecomunicações e das tecnologias de informação impulsiona a produção de documentos pelas instituições, especialmente daquelas que têm objetivos acadêmicos voltados para pesquisa e conseqüentemente, produtoras do conhecimento. O meio eletrônico pelo uso da Internet é o suporte mais rápido de armazenar, transmitir, compartilhar e recuperar documentos e informações.

A Internet tornou a pesquisa mais rápida e a informação acessível a pessoas localizadas em diferentes partes do mundo, aumentando ainda mais a importância dos periódicos disponibilizados em bases de dados conceituadas em suas áreas de atuação. A concentração de vários títulos de periódicos em uma única base de dados facilita a operação e recuperação da informação por parte do usuário, tendo em um único local de busca as informações mais relevantes sobre o assunto pesquisado.

As tecnologias proporcionam esta concentração de informações em um único local, com técnicas de organização que facilitam o trabalho do indexador, poupam espaço físico para armazenagem de grandes quantidades de informações e aumentam a velocidade de sua recuperação, por meio de algoritmos que permitem a melhor eficiência na recuperação.

Diversas técnicas de recuperação de informação já foram propostas para extrair informações das bases de dados, sendo uma das mais empregadas atualmente a utilização de operadores booleanos. Esta técnica é fundamental no momento da definição da estratégia da pesquisa, uma vez que são eles que definem o tipo de relacionamento entre os termos a serem pesquisados. Como a forma de expressão dos operadores booleanos varia de uma base para outra, o usuário pode confundir determinados símbolos no momento de sua pesquisa, os quais podem ser diferentes dependendo da base de dados utilizada (ZANINELLI; CATARINO, 2004).

Outras técnicas podem ser propostas como modelos de recuperação como a ontologia navegacional ou relações semânticas.

As tecnologias além de facilitarem a recuperação de informação, por meio da interoperabilidade, possibilitaram a comunicação entre softwares na extração e monitoramento de conteúdo. A ideia de ampliar as oportunidades de acesso ao potencial dessas informações gerou iniciativas no sentido de melhor aproveitar esse contexto.

2.4.4.1 *Unified Modeling Language (UML)*

A UML é uma linguagem visual universal para modelagem de sistemas orientados a objetos (OO), voltada para especificação, visualização, construção e

documentação de artefatos de sistemas de software (OMG 2003). Pelas definições dos autores da UML, um sistema pode ser descrito por cinco visões interdependentes, onde cada visão enfatiza um aspecto diferente do sistema. Estas visões são:

- a) visão de casos de uso, que descreve o comportamento do sistema conforme é visto pelos usuários finais, analistas e pessoal de teste, ou seja, apresenta o sistema do ponto de vista externo como um conjunto de interações entre o sistema e os agentes externos;
- b) visão de projeto, que abrange as classes, interface e colaborações componentes do vocabulário do problema e de sua solução;
- c) visão de implementação, que descreve os arquivos utilizados para a montagem e fornecimento do sistema físico, bem como o gerenciamento de versões de componentes do sistema, denominados de artefatos;
- d) visão de implantação, que abrange os nós formadores da topologia de hardware em que o sistema é executado. Essa visão direciona principalmente a distribuição, o fornecimento e a instalação das partes físicas do sistema;
- e) visão de processo, que enfatiza as características de concorrência (paralelismo), sincronização e desempenho do sistema.

Cada uma destas visões é modelada e documentada por meio de artefatos gráficos chamados de diagramas da UML. Estes diagramas são: diagrama de casos de uso, diagrama de classes, diagrama de objetos, diagrama de transições de estados, diagrama de atividades, diagrama de sequência, diagrama de colaborações, diagrama de componentes e diagrama de implantação.

As especificações técnicas da UML não são o foco deste trabalho, razão pela qual aqui é apresentado somente o modelo básico de sua estruturação e construção.

2.4.4.2 Iniciativa de Arquivos Abertos

A filosofia dos AA (OA), uma ideia baseada na democratização do acesso à informação, surge no centro desta revolução que busca facilitar o acesso ao conhecimento por toda a comunidade de forma irrestrita e gratuita (CAFÉ; FACHIN, 2007). Encabeçando esta ideologia, a OAI criou estruturas e normas para definir padrões de disseminação de informações.

A justificativa para a criação e adoção da OAI é motivada pelo fato desta iniciativa facilitar a disseminação dos artigos científicos, o que traz um aumento da oferta e uma rápida divulgação dos trabalhos desenvolvidos pela comunidade científica.

A participação na iniciativa OA pode ser de duas maneiras: por meio de um provedor de dados (*Data Provider*) ou um provedor de serviços (*Service Provider*).

Os provedores de dados administram sistemas e informações, geradores de conteúdos originais (repositórios de documentos digitais), que suportam o protocolo para coleta de dados especificados pelo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) como uma forma de expor os metadados de seus documentos.

Os provedores de serviços coletam metadados obtidos dos provedores de dados por meio do protocolo OAI-PMH para oferecer serviços com valor agregado, como uma biblioteca digital descentralizada e que pode, por exemplo: oferecer serviços de busca em vários repositórios ao mesmo tempo. Analogamente, podemos dizer que o google.com é um provedor de serviços em relação à World Wide Web.

O OAI-PMH define seis tipos de requisição. Um *harvester* não é obrigado a usar todos os tipos, mas um repositório necessita interpretar e responder a todos. O funcionamento do protocolo definido pelo OAI é composto por perguntas que o sistema coletor (provedor de serviço) faz ao sistema provedor de dados, e que podem ser realizadas por seis “verbos” definidos na versão 2.0 do protocolo, descritos no quadro 2.

VERBO	FUNÇÃO
Identify	Recupera informações sobre o repositório
ListMetadataFormats	Recupera todos os formatos dos metadados utilizados pelo repositório
ListSets	Recupera todas as edições ou seções disponíveis na publicação
ListIdentifiers	Resumo de todos os registros, mostrando apenas o cabeçalho de cada um
ListRecords	Lista todos os registros armazenados no repositório
GetRecord	Recupera um registro completo, com todos os dados

QUADRO 2 - LISTA DE VERBOS OAI-PMH
Fonte: OAI (2010)

O provedor de serviço (*Service provider*) é quem agrega valor às coleções de metadados dos provedores de dados, com mecanismos de busca, cruzamento de arquivos, autores e palavras-chave. A utilização dos protocolos OAI-PMH para coleta de metadados de outros provedores de dados (Periódicos científicos) possibilita a concentração de diversos títulos em um único local executando uma busca múltipla com uma única estratégia.

Hurd (2000) chama a atenção para um novo papel dentro das bases de dados, com a formulação de parcerias entre organizações e o surgimento de serviços de agregador (*Aggregator Site*). O agregador fornece serviços de busca e *links* como uma interface de acesso (*front-end*) para o conteúdo dos periódicos científicos. Para a autora o modelo é utilizado por grandes bases de dados como a *Blackwell*, a *Dawson* e a *EBSCO*, que normalmente firmam parcerias com as editoras e conectam um usuário a um artigo no servidor do editor após a validação adequada de seu direito de acesso ao jornal (bases pagas).

2.4.4.3 Interoperabilidade

A interoperabilidade é a facilidade que as “máquinas” têm em trocar metadados, sem a necessidade de interferência humana. Essa facilidade garante uma agilidade e confiabilidade nos dados transmitidos a bases indexadoras.

De acordo com Lagoze e Van de Sompel (2001 *apud* CAFÉ; FACHIN, 2007), a interoperabilidade tem inúmeras facetas, incluindo nomes uniformes, formatos de metadados, modelos de documentos e protocolos de acesso (*The Open Archives Initiative: Building a low-barrier interoperability framework Carl Lagoze Digital Library Research Group Cornell University Ithaca*).

Os metadados dão significado semântico a dados aparentemente isolados, provendo um contexto ao processo de busca (VIANA; MÁRDERO ARELLANO; SHINTAKU, 2005). Definem-se metadados como dados codificados e estruturados que descrevem as características de recursos de informação e são importantes principalmente na organização, gestão e recuperação da informação digital. Para tanto, são adotados procedimentos técnicos de catalogação, indexação e categorização dos conteúdos informacionais, o que possibilita a integração de fontes diversificadas e heterogêneas de informação.

Os elementos de metadados têm o propósito primário de descrever, identificar e definir um recurso de informação com o objetivo de modelar e filtrar o acesso. O uso de padrões de metadados possibilita aos sistemas de informação e de gestão do conhecimento a integração e o compartilhamento de recursos e aplicações. Elementos como autor, título, assunto são exemplos de metadados e podem ser usados para descrever tanto um livro em um catálogo de uma biblioteca *online* ou não, quanto para descrever uma *homepage*, uma base de dados ou qualquer outro recurso eletrônico em ambiente *Web* (SOUZA; ALVES, 2009, p. 193).

Inúmeras instituições brasileiras envolvidas na produção, organização, armazenamento e recuperação de informação digital têm gradativamente incorporado o uso de padrões de descrição de recursos de informação eletrônicos para a geração de metadados. A adoção de padrões é indispensável para promover a integração entre sistemas e garantir a interoperabilidade entre aplicações, condições necessárias à disponibilização de informação qualificada na Internet (SOUZA; ALVES, 2009).

O protocolo Dublin Core (DC) é um conjunto de quinze elementos metadados, os quais podem ser definidos como o mais baixo denominador comum para a descrição de recurso de informação, equivalente aos dados de uma ficha catalográfica. Suas características principais são: o entendimento semântico universal dos elementos metadados, o escopo internacional e a extensibilidade, visando permitir adaptações às necessidades adicionais de descrição (DUBLIN CORE, 2010).

O Elemento de Metadados Dublin Core Simple consiste de quinze elementos de metadados que são comunicados pelos provedores de dados aos sistemas de coletas (DUBLIN CORE, 2010, tradução nossa):

- a) *Title*: Título;
- b) *Creator*: Criador;
- c) *Subject*: Assunto;
- d) *Description*: Descrição;
- e) *Publisher*: Publicador;
- f) *Contributor*: Contribuidor;
- g) *Date*: Data;
- h) *Type*: Tipo;
- i) *Format*: Formato;
- j) *Identifier*: Identificador;
- k) *Source*: Origem;
- l) *Language*: Idioma;
- m) *Relation*: Relação;
- n) *Coverage*: Abrangência;
- o) *Rights*: Direitos;

Cada elemento Dublin Core é opcional e pode ser repetido. O DCMI estabeleceu maneiras padronizadas para refinar os elementos e encorajar o uso de esquemas de codificação e vocabulário.

2.4.4.4 *eXtensible Markup Language* (XML)

O XML, ou Extensible Markup Language, é uma metalinguagem (linguagem para definir outras linguagens), desenvolvido e mantido pela *World Wide Web*

Consortium (W3C), um consórcio internacional fundado com o objetivo de desenvolver protocolos e políticas que visem a evolução da web (MÜLLER FERNANDES, 2006).

O XML foi projetado para facilitar a transmissão e interoperabilidade de dados entre vários sistemas, permitindo que dados genéricos sejam enviados, recebidos e processados na forma simples, mantendo a compatibilidade com todos os sistemas, e transmissão confiável.

O XML é uma linguagem que pode adicionar marcas ou etiquetas com o objetivo de descrever o conteúdo do documento para indicar a semântica de sua apresentação, geralmente de forma textual.

2.5 AVALIAÇÃO DE BASE DE DADOS

A avaliação de um produto pelo seu público e sua aceitação é fundamental para o uso e sobrevivência e como uma base de dados pode ser considerada como um produto de “consumo”, ela carece de avaliação periódica sobre sua qualidade. Valls e Vergueiro (2006) ressaltam que a avaliação da qualidade em base de dados ainda é pouco estudada, e que a maior parte das experiências de avaliações de qualidade de produtos e serviços de informação no Brasil está concentrada em bibliotecas universitárias. De modo geral, a avaliação de serviços de informação empresta as teorias de outras áreas do conhecimento, principalmente da administração. Para Rodrigues e Silva (2008), um referencial prático e teórico sobre a avaliação de base de dados é quase inexpressivo.

Porém, na literatura de referência o processo de avaliação deve ser um dos objetivos das bases de dados, sendo mensurado por diversos critérios e indicadores. Segundo Lancaster (1993), uma base de dados bibliográficos deve ser avaliada pela sua utilidade ao responder as necessidades de informação, de acordo com quatro critérios principais (MESQUITA *et al.*, 2006):

- a) cobertura, ou seja, o quanto completo é o conteúdo da base de dados em relação a um assunto;

- b) recuperação, já que, tanto quanto ter uma boa cobertura de um assunto, é preciso que estes itens sejam passíveis de recuperação por meio de uma estratégia de busca não muito complexa;
- c) previsibilidade, para que um documento seja reconhecido como relevante a partir das informações contidas na base de dados;
- d) atualidade, que é uma medida da velocidade com que novas publicações são incluídas na base.

Em sua visão, García Lopez (2007 *apud* BUFREM, 2008, p. 41) amplia os critérios, e considera que devem ser apurados para avaliação de uma base de dados, aspectos como: grau de cobertura ou alcance temático, com indicação da área temática e revistas indexadas; cobertura geográfica e linguística; grau de inclusão, incluindo-se tipos de documentos; estrutura, incluído o número de campos, sua utilização ou não para indexação e tamanho e número de registros, além do nível de crescimento, ou seja, quantidade de itens novos a cada unidade de tempo. Esses indicadores evidentemente devem ser levados em conta a partir de um estudo das necessidades dos utilizadores da base, como orientação para sua criação, implantação e atualização. A outra face da avaliação pode ser realizada a partir das análises de uso e satisfação dos usuários/clientes de uma base.

Das técnicas para se avaliar o desempenho de Sistemas de Recuperação de Informação (SRI) baseadas no usuário final destacam-se a relevância, pertinência ou precisão e funcionalidades do sistema. Outros estudos abordam o monitoramento do uso de bases de dados, utilizando os protocolos de comunicação e análise de *logs* (ERICSSON; SIMON, 1993 *apud* MESQUITA *et al.*, 2006). As diferentes metodologias têm suas aplicações, dependendo dos objetivos de cada caso, porém, todas têm vantagens e desvantagens.

Para Schamber e Bateman (1996 *apud* MESQUITA *et al.*, 2006), a avaliação de bases de dados deve obedecer a um conjunto de requisitos, citados abaixo:

- a) reunir uma diversidade considerável de critérios que possam embasar a avaliação, além de perguntas abertas para obter explicações ou informações adicionais;
- b) coletar dados quantitativos analisáveis, como frequência de uso, rankings e estimativas;
- c) ser fácil e rápida de analisar, de preferência eletrônica e sequencialmente;

- d) ser adequada para aplicação em grupos variados de usuários e em qualquer tipo de ambiente;
- e) ser adaptável a vários objetivos de pesquisa;
- f) ser integrável a outros instrumentos de coleta de dados, como os que objetivam o estudo da relevância ou pertinência, dependendo das necessidades e interesses de cada caso.

A criação de indicadores e critérios de avaliação possibilita a geração de dados quantitativos que, interpretados e analisados de forma qualitativa mesmo em uma pequena amostra, podem identificar necessidades e problemas não antes previstos no planejamento.

Com o arcabouço teórico estabelecido, parte-se para o desenvolvimento da metodologia Brapci de base de dados e a construção de um modelo que possa ser aplicado em uma realidade para sua avaliação e mensuração de seu impacto social.

3 PROPOSTA METODOLÓGICA

Neste capítulo é apresentada a metodologia para a construção da base de dados temática, descrevendo-se todos os procedimentos de planejamento e estudo para sua construção, tendo como foco principal a aplicação das teorias da CI, da CC e da GI em prol do usuário que presta manutenção a base, bem como, o que a utiliza para recuperação de informação.

A metodologia proposta neste estudo, em sua fase de estudo de caso, utiliza conhecimentos da área da CI para a formação de um *corpus* para construção, análise e validação. Esta ênfase para a área de CI não necessariamente implica que a metodologia só poderá ser utilizada nesta área, pois o que se pretende é sua replicação a outras áreas do conhecimento.

No percurso metodológico parte-se de uma busca exploratória na base de dados Brapci (disponível desde 2009), no banco de teses e dissertações do IBICT, no portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e na *Library and Information Science Abstracts* (LISA). A pesquisa utiliza como estratégia de busca a identificação de trabalhos que utilizem como palavras-chave os termos: “Tratamento da Informação”, “Organização da Informação”, “Base de Dados”, “Comunicação Científica”, “Terminologia” e “Recuperação da Informação”. Dos trabalhos selecionados, realiza-se uma leitura e busca o referencial que fundamentou esta pesquisa, montando uma relação de autores da epistemologia para formulação da proposta metodológica. Não foi realizada uma delimitação de ano na recuperação.

Foram recuperados nesta estratégia 22 artigos para “tratamento da informação”, 37 para “organização da informação”, 78 para “base de dados”, 137 para “comunicação científica”, 58 para “terminologia” e 146 para “recuperação da informação”.

Dos documentos identificados, foi realizada uma leitura rápida para a identificação dos autores que fundamentam essas temáticas, buscando-se um referencial representativo. Entre as obras identificadas, livros e capítulos de livros, algumas foram adquiridas com os recursos da bolsa Capes, outras esgotadas foram emprestadas da biblioteca do setor de Ciências Sociais Aplicadas da UFPR ou por outros meios de acesso. Outras fontes foram localizadas por meio do *Google* para

leitura, incluindo-se livros e relatórios publicados no início e meados do século passado.

Para fundamentar as teorias da CC foi realizada outra busca, porém restrita ao banco de teses e dissertações do IBICT. Esta restrição é justificada pelo grande número de técnicas existentes na criação de banco de dados e por terem um ciclo de vida relativamente pequeno, devido a sua dependência em relação à tecnologia dominante da época. Entre as opções selecionadas optou-se pela utilização da *Unified Modeling Language* (UML) iterativa, por se considerar uma metodologia ágil e focada no projeto orientado a objetos e desenvolvimento em partes.

A proposta de uma metodologia para este estudo decorre da conjugação e integração de conhecimentos teóricos e práticos, incluindo técnicas e métodos com o objetivo de efetivar um produto de socialização da informação. Adota-se uma abordagem transdisciplinar das áreas de CI e CC que, segundo Lazarte (2000, p. 48), envolve o estudo de problemas complexos que exigem conhecimentos de várias áreas, procurando superar a fragmentação cognitiva de cada especificidade com a reunião de especialistas com conhecimentos em suas disciplinas. Uma das dificuldades da abordagem multidisciplinar está no diálogo entre os especialistas, com linguagens, visões, propostas e conclusões diferentes, porém esta diversidade de opiniões enriquece o desenvolvimento da metodologia. Mas existe a necessidade de uma convergência de ideias para possibilitar, segundo o autor, transcender os limites de cada disciplina.

Com a fundamentação teórica foi possível desenvolver a proposta metodológica para organização da informação em uma base temática. Todas as decisões sobre as estruturas empregadas na construção da base de dados foram tomadas em conjunto com o Grupo de Pesquisa E3PI, garantindo assim uma visão multidisciplinar das três áreas de conhecimento (CI, CC e GI).

Para a construção da metodologia são propostas seis etapas, abrangendo-se a definição de conceitos até o produto para validação que compreende:

- a) o planejamento;
- b) o levantamento de requisitos;
- c) a documentação do projeto (UML);
- d) o desenvolvimento do protótipo;
- e) o estudo concreto (caso Brapci);
- f) a validação e avaliação.

Antes de construir o protótipo (modelo relacional) para a base de dados foi necessário explorar todos os conceitos empregados em sua construção. Essa exploração possibilitou o planejamento adequado de forma a abranger as necessidades informacionais de curto e longo prazo de pesquisadores e usuários da base. Assim, sob o ponto de vista da definição, as representações dos conteúdos explorados na metodologia da base deverão atingir uma simplificação relativa a cinco aspectos: coleta, organização, acesso, recuperação e avaliação.

3.1 PLANEJAMENTO

Para atingir o objetivo, procurou-se estruturar elementos com vistas a uma metodologia que utilizasse os conceitos das áreas visando o planejamento, desenvolvimento e implantação da base. Para esta construção e com vistas a um modelo conceitual do protótipo funcional utilizou-se a metodologia advinda da teoria de desenvolvimento de produtos de Rozenfeld *et al.* (2006) representada na Figura 11, que prevê três etapas: **pré-desenvolvimento** que envolve a coleta de informações das necessidades sobre a base, um diagnóstico detalhado das necessidades dos usuários, análise da ferramenta atual, *benchmarking* de novas ferramentas e estudos de linguagens e tecnologias existentes; **desenvolvimento** com a definição do conteúdo e classificação das funcionalidades do sistema, definição da arquitetura da informação, modelagem do banco de dados, elaboração de *wireframes* e implantação; **pós-desenvolvimento** com a avaliação do produto, correções, incorporação de módulos adicionais ou complementares.

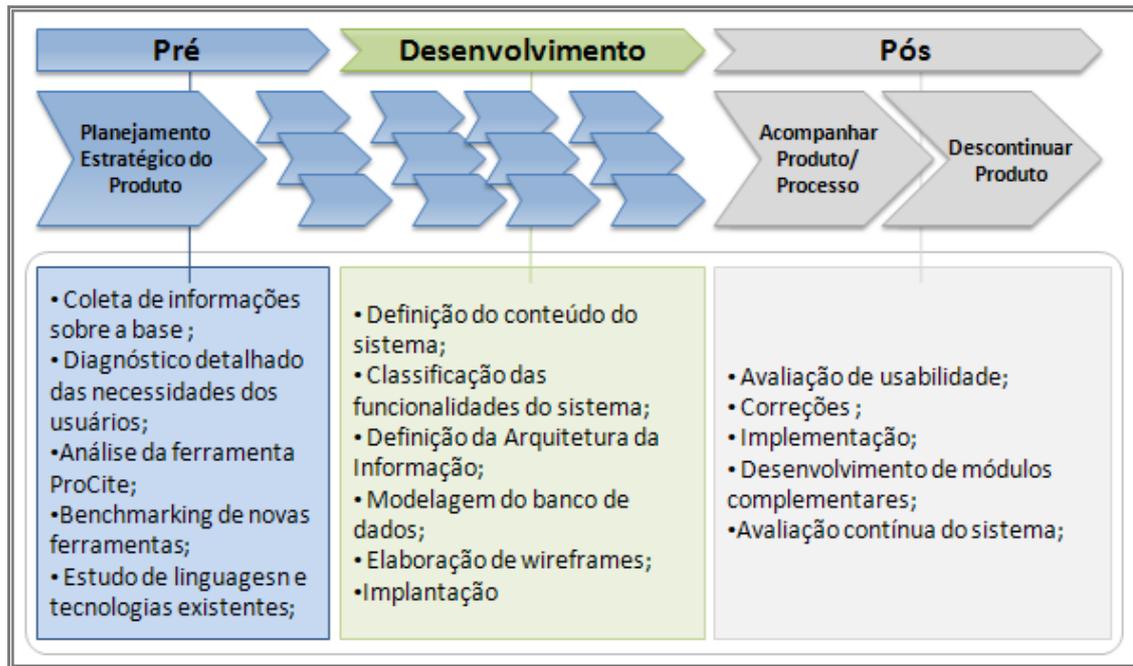


FIGURA 12 - ETAPAS DO DESENVOLVIMENTO DO PRODUTO

Fonte: Adaptado de ROZENFELD *et al.* (2006)

A Figura 12 apresenta as etapas de planejamento para desenvolvimento do produto utilizando a metodologia de Rozenfeld *et al.* (2006). O planejamento da metodologia denominada de “Metodologia Brapci” (MB) visa atender as necessidades de organização e recuperação de informação em uma base de dados, tendo como diretrizes os critérios de qualidade das informações e necessidades do usuário. A metodologia aqui proposta não se encerra com a efetivação do protótipo, mas em constantes revisões e incorporação de novas funcionalidades que possam contribuir para sua melhoria.

Nesta metodologia, o usuário é definido como pessoa física que utiliza a base como instrumento de pesquisa, bem como aquele que a utiliza nos processos de organização e tratamento como o indexador, o revisor e coordenador.

Para a compreensão das necessidades de construção da metodologia é necessário, segundo Rozenfeld e outros (2006), analisar o ambiente atual, ter uma visão sistêmica de todos os processos por meio de análises e diagnósticos das atividades desempenhadas por todos, possibilitando a efetivação da proposta coerente e alinhada aos propósitos do planejamento.

Para Rosenfeld e Morville (2002), o simples desenvolvimento de um site ou Intranet, sem uma arquitetura das informações em seu planejamento, pode resultar em um produto final alienado da finalidade para a qual foi construído. Ainda,

segundo os autores, alguns sites oferecem estruturas lógicas, o que leva a encontrar respostas e concluir satisfatoriamente os objetivos da pesquisa, já outros não apresentam a mínima organização e frustram as necessidades de navegação.

Em reunião realizada com os indexadores, utilizando-se a técnica de foco em grupo, atingiu-se o objetivo específico de identificar dificuldades na execução de atividades inerentes ao processo de cadastro de artigos ambientadas no sistema, até então se utilizando o software *ProCite5*. As dificuldades foram reportadas oralmente para os coordenadores, cujas demandas foram registradas em relatórios escritos. As necessidades coletadas foram sistematizadas pelo grupo E3PI, com uma reunião sobre o tema “dificuldades da base de dados na indexação”. A partir do reconhecimento das demandas, conseguiu-se categorizá-las e classificá-las, identificando:

- a) a dificuldade de compreender as mensagens do sistema (mensagens em inglês);
- b) a necessidade de revisão periódica em busca de novos fascículos nas revistas em papel;
- c) a necessidade de revisão periódica em busca de novos fascículos nas revistas eletrônicas;
- d) os erros na transcrição de informações dos trabalhos para a base, incluindo-se nomes dos autores, títulos e grafia de palavras-chave;
- e) a impossibilidade de copiar e colar as informações pelo *ProCite5*, ou seja, evitar a coleta da formatação original, pois o sistema importa também a fonte, cor e tamanho;
- f) a impossibilidade de dois indexadores trabalharem ao mesmo tempo, ou seja, base monousuário;
- g) a realização de *backups* semanais em CDs ou outros discos;
- h) as múltiplas cópias da base em diversos computadores (locais), o que acarretava perda de tempo do indexador trabalhando na base incorreta;
- i) a perda constante da integridade dos dados, quando na ocorrência de problemas no computador;
- j) a perda parcial da base de dados, quando na impossibilidade de um *backup* atualizado;
- k) a falta de uma visão sistemática da revista, com indicação das edições não indexadas;

- l) a interrupção na incorporação de novos trabalhos, com prejuízos da integridade dos dados.

Adicionada às dificuldades observadas, ainda existia a necessidade de uma exportação dos dados do *ProCite5* para a consulta em um ambiente virtual na Internet. Essa exportação era realizada de forma periódica e carecia da participação de um técnico especialista, não podendo ser realizada de modo automático, comprometendo a atualização dos dados ao público.

Depois de realizada a coleta de todas as demandas e analisadas pelo grupo de estudo, concluiu-se que a utilização da ferramenta existente não supria as necessidades de atendimento às principais demandas dos usuários indexadores.

A proposta visualizada incluía o desenvolvimento de um produto “novo” (nova metodologia) que possibilitasse a manutenção e disponibilização das informações da base de dados em um ambiente integrado, multiusuário, com facilidade de acesso e de forma gratuita. Nestas condições foi decidido que o sistema deveria ser totalmente ambientado na *Web*, e que todos os softwares utilizados e os sistemas de gerenciamento de banco de dados (SGBD) utilizariam licenças gratuitas.

A MB tem como premissas de sua política de construção:

- a) um ambiente multiusuário;
- b) o uso em qualquer lugar (com acesso a Internet);
- c) a disponibilidade de ferramentas tecnológicas como elementos facilitadores;
- d) o emprego de conceitos de usabilidade;
- e) um sistema de coleta automática de publicação periódica;
- f) a possibilidade de aplicação da metodologia a outros tipos de publicações como: eventos, livros, capítulos de livros, teses e dissertações;
- g) a produção de indicadores sobre a produção temática.

Desse modo, procurou-se, com a adoção da MB transcender os limites de uma base de dados, como fonte de pesquisa de modo a propiciar a preservação histórica da produção e ainda municiar o pesquisador de instrumentos que possibilitassem facilitar análises sobre a produção intelectual da área e estudos de uso e usuários da informação.

Para melhor distribuição das atividades de planejamento e definição, a MB desdobrou-se a partir de três funcionalidades gerais definidas aqui como módulos. Esse desdobramento facilitou tanto a definição dos conceitos utilizados como a construção da metodologia, sendo eles definidos como:

- Módulo Manutenção (MM): responsável por toda organização e manutenção dos cadastros, tabelas e informações pertinentes à base, assim como pela coleta, indexação e padronização das informações no sistema;
- Módulo Público (MU): responsável em prover ao usuário visitante informações de pesquisa simples, utilizando título, palavras-chave, resumo e autor como elementos de consulta em busca textual;
- Módulo Pesquisador (MP): contempla todos os conceitos do módulo público (MU) e incorpora ferramentas para análises de conteúdo, análises de redes (*networks*), análise de domínio, seleção de fontes (referências) e outros indicadores e futuras ferramentas que podem ser desenvolvidas.

Definido o modo de atuação, o grupo de estudos E3PI realizou o levantamento dos requisitos básicos necessários para a elaboração do projeto. Nesta etapa, além de apurados os requisitos, foi realizada a construção do “relacionamento de dados”, utilizada para propiciar ao desenvolvedor uma visão sistemática do produto.

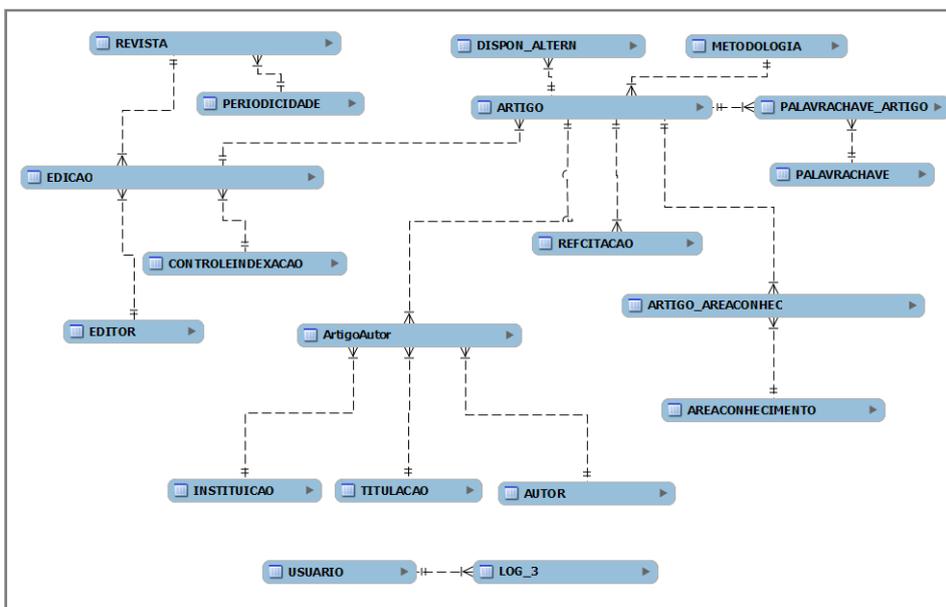


FIGURA 13 - MODELO DE RELACIONAMENTO DE DADOS
Fonte: COSTA (2009)

Para a construção do relacionamento de dados partiu-se do elemento principal e desencadeante do processo de indexação, o título da publicação (título do periódico, título do evento), que se relacionando com suas edições (fascículos) e os trabalhos a ela vinculados constitui-se de uma edição. Na construção, o elemento “ARTIGO” é o que mantém relacionamentos com o cadastro de autores, seções da publicação, palavras-chave, metodologias de pesquisa aplicadas e cruzamento de citações. A estrutura de relacionamento está representada na Figura 13.

Com a definição da estrutura, foi necessária a seleção de uma metodologia para o desenvolvimento e documentação do sistema que sustentasse a concepção tecnológica. Por utilizar uma abordagem dinâmica, e com foco voltado ao desenvolvimento pragmático, foi selecionada a metodologia UML 2.1 iterativa para o desenho do sistema. Esta metodologia relativiza a formalização documental do sistema em prol da possibilidade de desenvolvimento rápido e direto com o usuário.

A primeira etapa da metodologia UML 2.1 iterativa é a definição do escopo do produto e suas formas, o que já foi definido na Figura 13. A segunda etapa é a identificação de todos os Casos de Uso (UC) que representam funcionalidades para o sistema, havendo ou não interações com o usuário.

O grupo E3PI definiu em reunião os casos de uso para o módulo de manutenção referente aos aspectos:

- a) gestão de periódico (publicações);
- b) gestão de edição (fascículo);
- c) gestão de trabalho (artigo, resenha, entrevista, entre outros)
- d) gestão das seções das publicações;
- e) gestão de usuários do módulo manutenção;
- f) gestão de autores;
- g) gestão de instituições afiliadas aos autores;
- h) gestão de metodologias e técnicas;
- i) gestão de palavras-chave e idioma;
- j) ferramentas de conceitos e tesouros;
- k) ferramentas de áreas tema (*thema*) para ontologias;
- l) gestão de revisão de trabalhos na indexação;
- m) gestão de comunicações e avisos do sistema;
- n) gestão de conteúdos do site;
- o) gestão de acesso e níveis de perfis;

- p) gestão de trabalho (registro de acesso);
- q) sistema de coleta de metadados e processamento de metadados (OAI);
- r) sistema gerador de relatório sobre a base e seu uso;
- s) exportação de dados para o módulo público/pesquisador.

Os casos de uso do módulo público definidos foram referentes aos aspectos:

- a) pesquisa simples em todos os campos;
- b) pesquisa direta em campos específicos (autor, título, palavras-chave e resumo);
- c) pesquisa com funções booleanas (*AND*, *OR*, *NOT*);
- d) referência das fontes (ABNT);
- e) navegação por revista e edição.

Os casos de uso do módulo pesquisador foram definidos a partir de funcionalidades referentes aos aspectos:

- a) cadastro de pesquisador, referente a identificação do usuário;
- b) seleção de trabalhos, armazenamento das estratégias de busca e composição de *corpus* para análise isolada do todo;
- c) envio de seleções de pesquisas por e-mail, com suas referências;
- d) resumo analítico da seleção de pesquisa e análises, com indicadores bibliométricos² da seleção do usuário;
- e) busca por conceitos (ontologias) ou ontologia navegacional, propiciando ao usuário consulta pelo vocabulário controlado e visualização de suas relações conceituais.

A concepção da base como fonte de estudos, tanto para os visitantes e pesquisadores (externos), quanto para o grupo de pesquisa, está contemplada no desenvolvimento da metodologia. Todos os acessos são registrados por meio de um *log*, contendo as páginas acessadas, a forma de busca, os termos, a data, o horário e o *Internet Protocol* (IP). Esses dados possibilitam ao grupo uma análise do que está sendo pesquisado, de que forma e quando.

Uma das ferramentas que possibilita o sucesso na utilização da base de dados é o seu instrumento de recuperação de informação. Segundo a concepção de Mooers (1951) a “recuperação da informação engloba aspectos intelectuais da

² Relações entre os autores, índice de citação, referências utilizadas pelos autores, autores mais citados, revistas que mais publicaram sobre o assunto, ano de publicação dos temas.

descrição da informação e sua especificação para busca, e também qualquer sistema, técnicas ou máquinas que são empregados para realizar a operação”. Esta concepção pode ser aplicável hoje aos novos sistemas de recuperação de informação (SRI) utilizados por supercomputadores, que estabelecem uma grande diferença entre os antigos sistemas pelos quais o usuário era assessorado por um profissional da informação que dominava as estratégias de buscas dos SRI (AHMED; MCKNIGHT; OPPENHEIM, 2004) e os novos, que possibilitam ao usuário acesso direto e sem intermediários à informação do acervo. Essa forma de consulta faz com que o usuário utilize uma linguagem de especialidade própria (linguagem natural), muitas vezes corriqueira, e nem sempre coincidente com a linguagem de indexação (LAAN, 2002).

Em um estudo prospectivo de cenários, Zins (2007) propõe um modelo pelo qual o usuário e o acervo têm ligação direta, e a estrutura física e profissional é responsável por preparar a informação para esse acesso, com base em estudos aprimorados sobre suas estratégias de buscas. Cabe ao profissional da informação a responsabilidade de criar novos instrumentos para facilitar esse acesso. Assim, a tarefa de analisar as formas de buscas, o modo como são realizadas e a satisfação dos usuários são pontos importantes para o desenvolvimento da base.

O SRI da MB deve preconizar o uso de instrumentos para recuperação da informação, dispondo de:

- a) busca com base em operadores booleanos (*AND*, *OR*, *NOT*);
- b) sugestões léxicas;
- c) busca semântica (fraca, média e forte);
- d) busca por conceitos (ontologia).

Para um melhor entendimento da MB é representada na Figura 14 uma visão macro por meio de um diagrama das abordagens e suas funcionalidades.

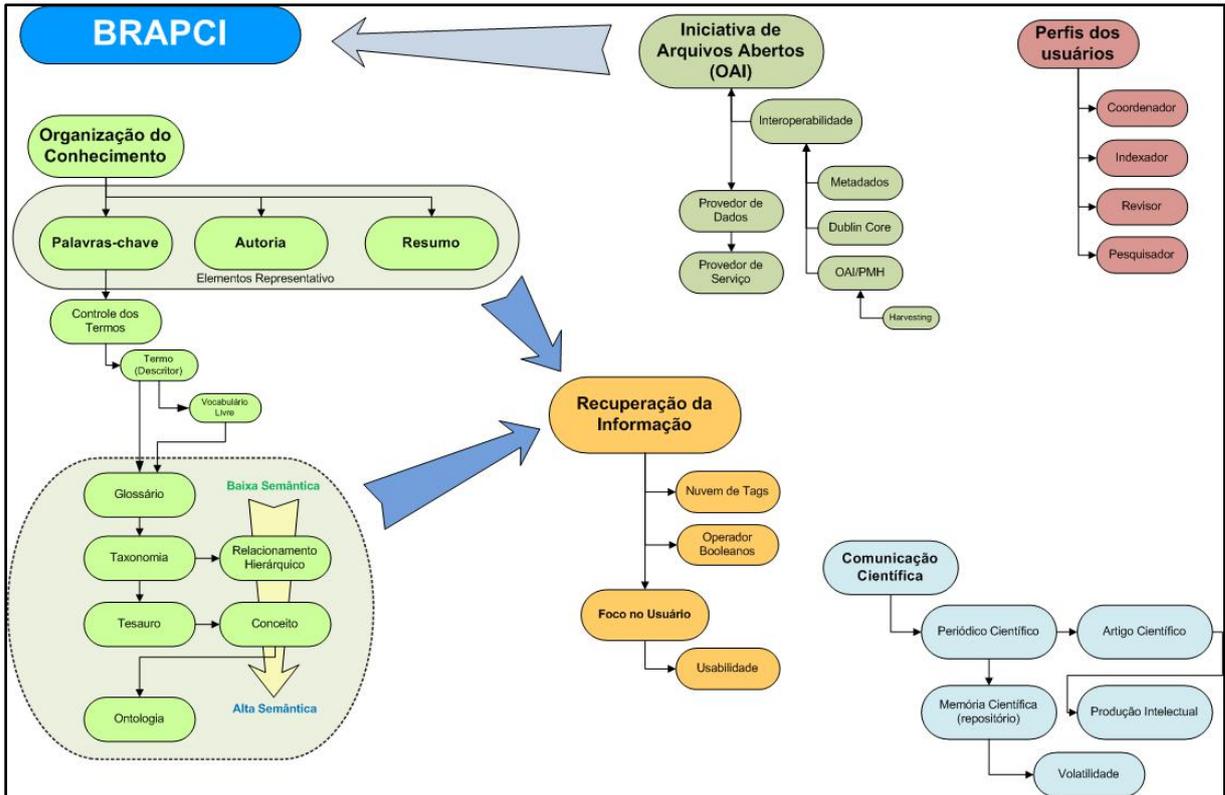


FIGURA 14 - REPRESENTAÇÃO GRÁFICA DA METODOLOGIA BRAPCI
Fonte: O autor (2010)

Para o desenvolvimento da interface da base com o usuário foram respeitados os critérios de usabilidade definidos por Nielsen e Thair (2002). Essa utilização garante uma interação maior entre o usuário e o sistema, principalmente na sua forma de aprender e relembrar os processos desenvolvidos na manutenção da base. Entre os critérios, destaca-se o tempo menor de capacitação de novos usuários, bem como a facilidade de uso da base.

Para o desenvolvimento do protótipo optou-se pelo uso da metodologia UML iterativa, técnica que reflete a ideia da construção de um todo composto de partes, formando ciclos, ou iterações. A metodologia realiza a mesma sequência de passos para cada porção do produto ou nível de elaboração, partindo de conceitos abrangentes com a codificação individual de cada funcionalidade. De acordo com o Larman (2010) cada ciclo da iteração começa com a identificação da:

- concepção, que dura geralmente poucos dias, o suficiente para estabelecer uma visão de alto nível dos objetivos, do escopo e das prioridades. Visa coletar os 10% de requisitos mais relevantes, identificar os principais riscos e elaborar uma estimativa de esforços, realizada em oficinas de requisitos, com histórias e cenários de uso, de prototipação,

de rascunho de interface e da identificação de regras, restrições e requisitos não-funcionais;

- b) elaboração, ou compreensão dos requisitos, que se estende em quantidade e em profundidade. Busca o núcleo da arquitetura e as partes críticas são desenvolvidas e testadas de forma que os principais riscos sejam minimizados e seja possível obter uma estimativa razoável de duração e esforço. Essas metas são alcançadas por meio de iterações curtas, contendo oficinas de requisitos, modelagem, programação e testes;
- c) construção, cuja base é uma estrutura estável criada nas etapas anteriores, quando se realiza a implementação do sistema, razão pela qual esta fase tende a ser a mais longa do processo. Uma versão *alpha* com a documentação é produzida usando iterações curtas com pouca modelagem e muita programação, com testes e avaliações dos usuários;
- d) transição, ou prontidão do sistema para entrar em produção. Uma versão *beta* é produzida e avaliada. Com base no *feedback* obtido, alguma programação e ajuste são feitos para chegar à versão que entrará em produção, acompanhada da documentação completa.

Esta técnica permite a construção de versões básicas ou funcionalidades novas no sistema em prazos curtos e possibilita novos requisitos para implementação futura, no momento em que se tornam necessários.



FIGURA 15 - METODOLOGIA ITERATIVA DE DESENVOLVIMENTO DE SISTEMA
Fonte: Adaptado de BOEHM (1988)

Esta técnica de desenvolvimento é uma evolução da metodologia espiral utilizada na década de 1990, e não está presa ao trâmite burocrático com a definição de documentação exaustiva antes da construção de um produto funcional das metodologias UML anteriores, mas ao cerne da interação entre o sistema, o desenvolvedor e o usuário.

Processos metodológicos de desenvolvimento de sistemas rígidos transferem o foco da produção do sistema para si ao fazerem com que as pessoas e o projeto tenham que se adequar a eles. Tentar tirar dos indivíduos a responsabilidade pela produção do software não faz sentido em um contexto em que eles são fundamentais, uma vez que pessoas sem processos podem produzir software, mas processos sem pessoas não produzem. Métodos ágeis apóiam-se em poucas regras rigidamente definidas e em princípios que oferecem certa flexibilidade para que as práticas do dia-a-dia sejam adaptadas (LARMAN, 2007).

Definida a metodologia de desenvolvimento do trabalho para construção do protótipo parte-se para a concepção do sistema. A fim de facilitar o entendimento do grupo e construção de telas, adota-se a técnica de *wireframes*, utilizada pelos arquitetos da informação para construir um modelo não-funcional (rascunho) de como deverão ser categorizadas as informações e a forma de visualização. As Figuras 16 e 17 representam dois *wireframes*, respectivamente o do módulo de manutenção e o do módulo público.

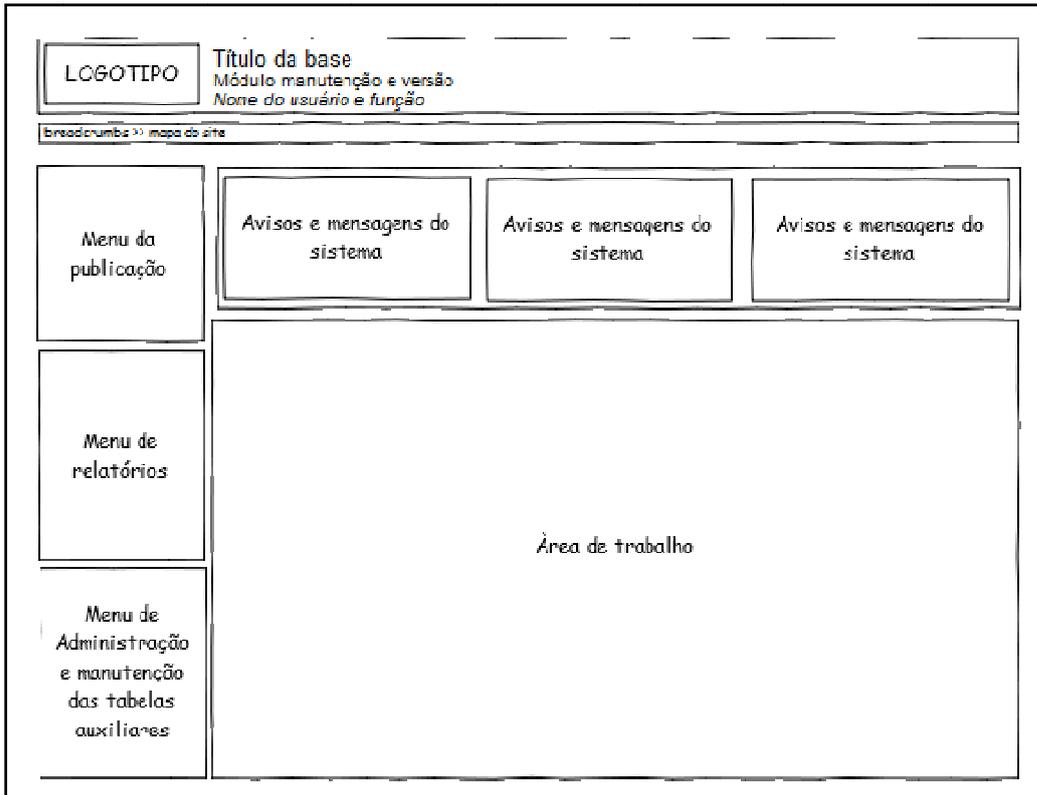


FIGURA 16 - WIREFRAME DO MÓDULO DE MANUTENÇÃO
 Fonte: O autor (2010)



FIGURA 17 - WIREFRAME DO MÓDULO PÚBLICO
 Fonte: O autor (2010)

Nesta fase do planejamento, os processos de levantamento de requisitos não podem ser realizados por uma única classe profissional. Para verificar todas as

possibilidades e funcionalidades, a participação nesta fase deve ser interativa e contar com o conhecimento de diversos especialistas com o objetivo principal de obtenção de um consenso sobre todas as ações e representações do sistema. Para o desenvolvimento da metodologia foi necessário um envolvimento interdisciplinar das áreas de CI e CC, envolvendo: analista de sistema; profissional da informação (bibliotecário, gestor da informação); arquiteto da informação; modelador e gestor de sistema de gerenciamento de bases de dados (SGBD); *webdesign* e programador.

Com a concepção do sistema definida, foi necessária a estruturação da política de indexação, para definição da exaustividade da indexação e análise dos recursos disponíveis.

3.2 POLÍTICA DE DESENVOLVIMENTO DE COLEÇÕES

A criação de um instrumento norteador como a Política de Desenvolvimento de Coleções deve-se ao fato de se considerar importante a existência de um instrumento formal que estabeleça critérios e prioridades com relação à seleção e aquisição do material que irá compor os acervos da base de dados. A formalização de uma política possibilita que a coleção cresça de forma consistente, qualitativa e quantitativamente, e que estabeleça as diretrizes a serem seguidas no processo incorporação de novos títulos, de catalogação e indexação e quais os tipos de acervos serão abrangidos em seu *corpus*.

A política de desenvolvimento de coleções estabelece os padrões que serão empregados pela base, detalha os processos aplicados nas atividades operacionais e ainda consiste de um elemento básico para qualquer tomada de decisão, contemplando os objetivos, visão e missão da base.

Para uma melhor explicitação da política, esta foi dividida em dois aspectos, o primeiro referente à política de indexação e o segundo sobre a política de crescimento do acervo.

3.2.1 Política de indexação

No planejamento de um serviço de indexação, deve-se levar em conta, segundo Carneiro (1985), um grande número de variáveis, pois estas irão afetar consideravelmente o desempenho e a eficiência na recuperação da informação pelo usuário no momento preciso. Definir essas variáveis, estabelecer princípios e critérios que servirão de guia na tomada de decisão, racionalizando processos e garantindo consistência das operações são os objetivos de uma política de indexação.

Deste modo, é necessário que se estabeleçam as políticas a serem empregadas nos processo de organização da informação. Carneiro (1985) considera alguns fatores que são imprescindíveis no planejamento de um sistema de recuperação de informação:

- a) a identificação da organização à qual estará vinculado o sistema de indexação, sendo fundamental para a determinação do tipo de serviço implantado;
- b) a identificação da clientela a que se destina o sistema, pois sua principal finalidade do sistema é fornecer informações na forma e momento exigidos pelos usuários;
- c) os recursos humanos, materiais e financeiros, limitações quanto às despesas com o capital, questões operacionais e as decorrentes da não existência de serviços convenientes de informação.

A combinação desses fatores vai definir pontos importantes na base, como a exaustividade na indexação, pois, quanto mais especificidade, mais recursos serão exigidos, tanto financeiros como humanos na operacionalização da indexação. Por este motivo deve-se ter clareza na definição de uma política de indexação. O conhecimento da política por toda equipe é fundamental para decisões coerentes, alinhadas para construção e crescimento racional da base.

A criação da política de indexação, ainda de acordo com Carneiro (1985), deve abordar os seguintes aspectos:

- a) cobertura dos assuntos: o planejador do sistema deve ter uma ideia dos assuntos a serem cobertos pelo sistema, áreas e temáticas, definindo a profundidade ou superficialidade de como serão tratadas;

- b) seleção e aquisição dos documentos-fontes: determinando como será estabelecido a seleção das fontes que serão incluído-se no sistema, sua extensão de cobertura, a qualidade dos documentos e interesse dos usuários assistidos na área do sistema;
- c) processo de indexação: influência das variáveis que afetam o processo de recuperação, seu nível de exaustividade, nível de especificidade, escolha da linguagem, capacidade de revocação e precisão do sistema;
- d) estratégia de busca: forma de interação do usuário com a base, podendo ser delegada quando é realizada por um especialista da informação ou não delegada quando realizada diretamente pelo usuários, esta definição é importante para o planejando do sistema;
- e) tempo de resposta do sistema: medição entre o recebimento da consulta e o fornecimento de uma resposta satisfatória ao usuário;
- f) forma de saída: disposição de como os resultados são apresentados, quando a existe baixa precisão é necessário munir o usuários com mais informações na reposta, como resumos;
- g) avaliação do sistema: retorno dos usuários com a satisfação do sistema.

A definição dessas características (CARNEIRO, 1985), é importante para a economia de recursos financeiros e retrabalho humano. Desta forma, uma boa definição favorece um crescimento contínuo da base e principalmente a abertura para que o usuário possa contribuir com informações sobre a sua experiência e satisfação no uso do sistema de informação.

Para definição da política de indexação, a MB define que a base deve trabalhar somente com uma área do conhecimento específico, ou áreas que tenham relação muito próxima. Os procedimentos de aquisição e seleção de novas fontes de informação pautam-se por duas formas:

- a) automática, por meio do protocolo OAI-PMH, quando o sistema monitora novas publicações diretamente na fonte e as incorpora para indexação;
- b) manual, para publicações que não possibilitam esta coleta, definindo uma periodicidade para coleta manual de novos trabalhos.

No processo de indexação, a metodologia define o uso da linguagem natural para coleta e processamento, realizando uma indexação pós-coordenada com a criação de instrumentos como glossários, tesouros e modelos de conhecimento conforme áreas temáticas e assuntos específicos.

A MB define que as informações serão disponibilizadas na Internet, utilizando os recursos tecnológicos disponíveis, para prover ao usuário a informação de forma livre e completa, podendo em alguns casos indicar o local de acesso ao conteúdo integral. Partindo dessa premissa, a metodologia propõe duas formas da base de dados:

- a) como diretório, indicando onde se localiza o acesso a conteúdo integral das fontes de informação, mantendo na base somente informações importantes para indexação;
- b) como repositório, indicando todos os dados da publicação, bem como o acesso diretamente de sua base ao conteúdo integral do trabalho.

Essas definições dependem dos recursos e da capacidade de armazenamento disponível em cada estrutura.

Por utilizar recursos tecnológicos, disponibilizando o conteúdo para acesso por meio da Internet, faz-se necessário um sistema de busca orientado ao usuário, não delegado, ou seja, o próprio usuário pode definir como vai realizar a busca. O importante é ressaltar que, na metodologia proposta, todas as consultas realizadas pelos usuários são registradas para posterior análise dos gestores da base.

O tempo de resposta para o usuário, nesta metodologia, depende mais dos recursos de *hardware* do hospedeiro da base de dados do que da forma como o usuário elabora sua estratégia, pois quanto mais veloz o equipamento, menor o tempo de resposta. No modelo, esta resposta não deve ultrapassar três segundos por consulta.

Por se trabalhar com documentos revisados e não revisados (pelos indexadores da base), optou-se na metodologia em fornecer ao usuário a resposta mais completa possível, indicando em uma única consulta o título do trabalho, seção de publicação, fonte de publicação, nome dos autores, resumo e as palavras-chave, bem como a possibilidade de visualizar sua referência na norma ABNT e Vancouver. Com o resultado da busca, o sistema possibilita a marcação e seleção de trabalhos representativos conforme leitura do usuário. Esta seleção forma novo *corpus* que pode ser armazenado, impresso, enviado por *e-mail*, contando ainda com as funcionalidades do sistema para gerar indicadores de análises.

Por não haver nenhuma ligação direta do usuário com os gestores da base de dados, o sistema disponibiliza três formulários para o usuário: a) reportar sua interação com o sistema; b) informar inconsistência da base ou quebra de *link*; c)

entrar em contato com o grupo E3PI. Todos os envios são analisados pela equipe de gestão com providências necessárias para sua correção e posterior retorno ao usuário.

Com a definição do tipo de tratamento de assuntos dentro do sistema, é necessário, nesta metodologia, criar um subsistema que gere um vocabulário controlado para estruturar os termos e posteriormente criar e definir conceitos.

3.2.2 Política de desenvolvimento de acervo

Uma base de dados para ser representativa deve relacionar-se com uma área do conhecimento e manter sua especificidade, principalmente para o controle de sua terminologia e facilidade de recuperação da informação pelo seu público alvo, o que vem ao encontro das teorias do tratamento temático da informação. Desta forma o *corpus* de uma base deve ser representado por todos os trabalhos publicados em repositórios científicos de sua área de concentração.

A definição de uma política de incorporação de novos títulos deve ser clara e explícita, fundamentando o coordenador e/ou editor com subsídios para inserção de novos títulos. Das sugestões para incorporação de novos títulos a uma base de dados, pode-se estabelecer os seguintes critérios para definição da política do acervo:

- a) estar avaliada em algum sistema de avaliação como o sistema Qualis³, no Brasil;
- b) estar no escopo da base de dados;
- c) ter artigos representativos para a área;
- d) ter um ciclo de vida de no mínimo dois anos e/ou mais de três fascículos, para periódicos;
- e) ter importância e representatividade dos autores para a área.

³ Qualis é o conjunto de procedimentos utilizados pela Capes para estratificação da qualidade da produção intelectual dos programas de pós-graduação. Tal processo foi concebido para atender as necessidades específicas do sistema de avaliação e é baseado nas informações fornecidas por meio do aplicativo Coleta de Dados. Como resultado, disponibiliza uma lista com a classificação dos veículos utilizados pelos programas de pós-graduação para a divulgação da sua produção.

Ressalta-se que os critérios acima citados são sugestões ao editor da base, e que, conforme a área, estes devem ser definidos para a montagem da política do acervo.

Na construção da política de incorporação deve-se também analisar quais as seções que serão disponibilizadas para consulta aos usuários, pois as revistas científicas produzem muitas seções que não podem ser consideradas como publicações científicas, muitas vezes publicando documentos, resumos de trabalhos apresentados em eventos, ou coletâneas de teses e dissertações defendidas dentro das instituições mantenedoras. Desta forma, a MB por definição armazena todos os tipos de trabalhos, qualquer que seja sua seção, porém disponibiliza para consulta no módulo público somente as seções definidas na política como visíveis para consulta, conforme os dados apresentados no Quadro 3.

SEÇÃO DE PUBLICAÇÃO	VISÍVEL PARA CONSULTA
Apresentação	Não
Artigo científico	SIM
Artigo Científico Temático	SIM
Artigos completos em anais de congresso	Não
Áudio e Vídeo	Não
Comunicação	SIM
Debate	SIM
Documentos	SIM
Editorial	SIM
Ensaio	SIM
Entrevista	SIM
Expediente	Não
In Memoriam	SIM
Normas de publicação	Não
Opinião	Não
Outros (não processados pela indexação)	Não
Palestra	Não
Ponto de vista	SIM
Prêmio	Não
Recensão	SIM
Relato de caso	SIM
Resenha	SIM
Resumos em anais de congressos	Não
Revisão de literatura	SIM
Sessão de estudos	Não
Teses e Dissertações	Não
Trabalhos acadêmicos	Não
Trabalhos apresentados em eventos	Não
Trabalhos de conclusão de curso	Não

QUADRO 3 - SEÇÕES DAS PUBLICAÇÕES DISPONÍVEIS NA METODOLOGIA BRAPCI
Fonte: O autor (2010)

Esta política de disponibilização foi definida pelo Grupo E3PI, entretanto, a qualquer momento ela pode ser revista, possibilitando a inclusão de novas seções aqui não previstas.

3.3 VOCABULÁRIO CONTROLADO NA METODOLOGIA BRAPCI

A MB parte do princípio de que a informação disponível para ser analisada precisa estar organizada, com qualidade e ser precisa. Esta qualidade só pode ser almejada com o controle e padronização dos dados nela inseridos, principalmente com relacionamento de tabelas de dados e padronização nas entradas dos registros. Porém, esta rigidez na padronização não pode ser um entrave na disponibilização de informações, e estas devem estar em constante atualização com a incorporação de novos trabalhos, proporcionando ao usuário pesquisador uma fonte de pesquisa atualizada e fidedigna com a fonte original de informação.

Partindo desta concepção, a MB propõe uma abordagem mista, sem o rigor na coleta de dados com padronização de entradas, entretanto utiliza instrumentos para uma indexação pós-coordenada das informações quando analisadas pelos indexadores no momento de sua revisão. Essa abordagem gera uma perda da qualidade em um primeiro momento, com poucos critérios de padronização na entrada dos dados, mas com a vantagem de manter a base atualizada. O tratamento e padronização das informações são realizados em uma segunda etapa quando ocorre a revisão por um indexador. Cada nova regra criada pelos gestores da base é incorporada, como forma de “ensinar” o sistema a processar informações que coincidam com estas, facilitando e reduzindo trabalho repetitivo e economizando tempo na organização. Um ponto forte da metodologia é de que todos os trabalhos inseridos na base, em algum momento, devem passar por uma revisão e complementação do indexador.

Sem um controle de vocabulário é impossível garantir a qualidade da recuperação da informação, pois a busca textual não garante precisão, sendo dependente das estratégias e variações sintáticas utilizadas pelo usuário. A metodologia propõe a utilização da linguagem natural, empregados pelos autores e editores de trabalhos como a do discurso técnico-científico (LANCASTER, 1993).

Na proposta da metodologia os descritores oriundos na LN são comparados com outros descritores já utilizados e normalizados para criação de um vocabulário controlado, construindo-se, desta forma, instrumentos para organização e representação do conhecimento. Em um primeiro momento com a criação e atualização de listas de assuntos e formação de glossários (FARIA; PERICÃO, 2008), para posteriormente evoluir para outros instrumentos e formas de apresentação.

Os termos que não apresentam relação com o vocabulário são armazenados automaticamente e passam para análise dos indexadores, que optarão pela sua incorporação ao glossário, com inserção de um novo termo, ou pelo uso de remissivas (USE) para outro termo.

Para avaliar a qualidade de uma base de dados em artes, Strehl (1998) realizou uma análise dos descritores de artigos científicos, identificando diversos problemas em sua representação, desde a utilização de sinônimos, uso de singular e plural, relação de assuntos com subcategorias e descritores redundantes. Foram observados termos utilizados para descrever nomes geográficos, nome de autores e ou instituições e tempos cronológicos da história. Partindo-se dos problemas relatados pela autora em relação aos descritores e cotejando-os com os termos utilizados pelos autores para representar seus conteúdos, optou-se por criar seis categorias:

- a) descritores unilexicais e polilexicais, para definir um assunto;
- b) descritores cronológicos, para definir um período histórico;
- c) descritores geográficos, para definir nomes de cidades, estados, regiões geográficas;
- d) descritores onomásticos, para definir nomes e instituições;
- e) outros, caso o descritor não se categorize em uma das opções acima;
- f) não analisado, para avaliação do indexador quando incorporado um novo descritor.

Cada novo termo inserido na base, como regra, é classificado como não analisado para que o indexador possa identificá-lo e classificá-lo.

A estruturação de instrumentos para a construção de vocabulários controlados não pode ser “engessada” como nos instrumentos tradicionais do vocabulário controlado (glossários, tesouros), mas se basear em ferramentas dinâmicas que permitam a organização da informação fundamentada em conceitos e

relacionamentos semânticos, utilizando sua versão sintática somente como forma de explicitação dos conceitos, concepção apoiada na Teoria Comunicativa da Terminologia de Cabré (1993).

O que a MB propõe é que a identificação e a classificação de novos termos representem apenas a inserção no glossário da base, como demonstrado na relação semântica de Orbst (2008), com uma fraca semântica. Os glossários são utilizados na padronização da entrada de dados dos descritores, mas uma base de dados que pretende ter relevância e precisão em suas consultas por informações pelos usuários, carece de outros instrumentos que possam aumentar a carga semântica dos termos, convertendo ou incorporando a outros instrumentos como os tesauros e modelos conceituais.

O conceito, como ponto de partida para determinar a forma representativa mais adequada, é entendido, conforme a teoria analítica de Dahlberg (1978), como unidade de conhecimento e a definição, como elemento essencial para sua fixação. Essa concepção corrobora a proposta da construção de instrumento de apoio à organização do conhecimento, tendo no conceito o foco e não na sua representação por signos. A grande vantagem pragmática dessa abordagem está na integração de diversas formas de representação de um mesmo conceito, podendo utilizar a forma escrita em seus diferentes idiomas, bem como possibilitar o emprego de outros meios de representação como sons, imagens ou movimentos.

Na MB propõe-se o uso da Teoria Comunicativa da Terminologia de Cabré (1993), com a adoção de conceitos para representar os assuntos. Porém a TCT é limitada à definição de um único termo para representar o conceito, o que na metodologia aqui proposta não será empregado. Considerando os avanços tecnológicos e a velocidade no processamento de informação por computadores, a proposta é de ampliação do número de descritores para representar um conceito, definindo um como preferencial e outros como alternativos. A metodologia também propõe a existência de mais de um termo preferencial na descrição de um conceito, mas somente em idiomas diferentes, podendo haver três termos preferenciais em um conceito.

Desta forma, a metodologia se alinha às mais modernas concepções de organização de conhecimento que, segundo Moreiro González (2010), não podem ser restritas a uma visão única da abordagem do conceito, mas múltiplas visões para

os diferentes olhares para o mesmo conceito, sem perder o controle da linguagem empregada.

Outra proposta da metodologia é a criação de tesouros e ontologias para áreas temáticas e subtemáticas dentro da própria base, porém geradas pela análise cognitiva de pesquisadores ou especialistas em determinadas subtemáticas. Essa construção é efetivada com o relacionando de subordinação e superordenação (CINTRA *et al.*, 2002) entre os conceitos e aplicação em tesouros temáticos. O que deve se ressaltar na MB é que o tesouro aqui empregado não tem como elemento de organização os termos, mas sim os conceitos.

A construção de categorias foi a forma adotada para representar classes de conceitos pela metodologia e teve como respaldo a necessidade sentida desde os estudos aristotélicos e que até hoje não foi suprida apenas com as possibilidades tecnológicas, nem mesmo com as mais recentes inovações do mundo da informação. Desta forma a geração de conceitos na metodologia será sempre cognitiva, ou seja, realizada pelo indexador humano.

A definição da organização por conceitos possibilita a criação de diversas ontologias temáticas em uma mesma área do conhecimento de acordo com a diversidade do usuário da informação. O relacionamento e hierarquias dos conceitos podem ser realizados com taxionomias específicas em uma única temática ou em diversas subtemáticas, o que vem ao encontro da proposta de classificação de Ranganathan (1967) em uma estrutura multifacetada, demonstrada pela árvore baniana, permitindo a visualização de uma interligação entre diversos conceitos, não exclusivamente de forma hierárquica, mas também de forma relacional.

Na efetivação da proposta da metodologia com a criação do protótipo são disponibilizadas ferramentas de criação de ontologias (relacionamento de conceitos), propondo-se manter a compatibilidade com outros instrumentos disponíveis para este fim, como o *Protégé*⁴, que possibilita a construção de ontologias (OWL), ou emprego de *folksonomia*, quando o usuário contribui para o sistema com sua análise classificatória e a disponibiliza para outros usuários.

O uso das ontologias parte principalmente da facilidade de um pesquisador menos experiente de identificar as relações lógicas com outros conceitos,

⁴ O *Protégé* é uma ferramenta que permite construir ontologias de domínio, personalizar formulários de entrada de dados, inserir e editar dados, possibilitando então, a criação de bases de conhecimento guiadas por uma ontologia.

principalmente quando aplicadas a ontologia navegacional, que segundo Campos (2001) vão fornecer a visão de conjunto dos agrupamentos que ocorrem na estrutura, possibilitando o entendimento global da área (ou subárea).

A concepção da metodologia proposta parte da construção de uma ontologia para que se evidenciem as relações entre conceitos, por meio da construção de um vocabulário controlado, para garantir a convenção, a coesão e a consistência da entrada e manutenção dos dados. Esta necessidade é justificada principalmente porque as ciências estão em constante evolução, novos termos são criados, atualizados ou substituídos por outros de forma dinâmica. Nesta metodologia, como expresso na Figura 18, o pré-controle ocorre diretamente da linguagem natural, garantindo-se a velocidade de atualização e controle da terminologia, provendo a qualidade aos descritores da base e a possibilidade de definição de novos conceitos.

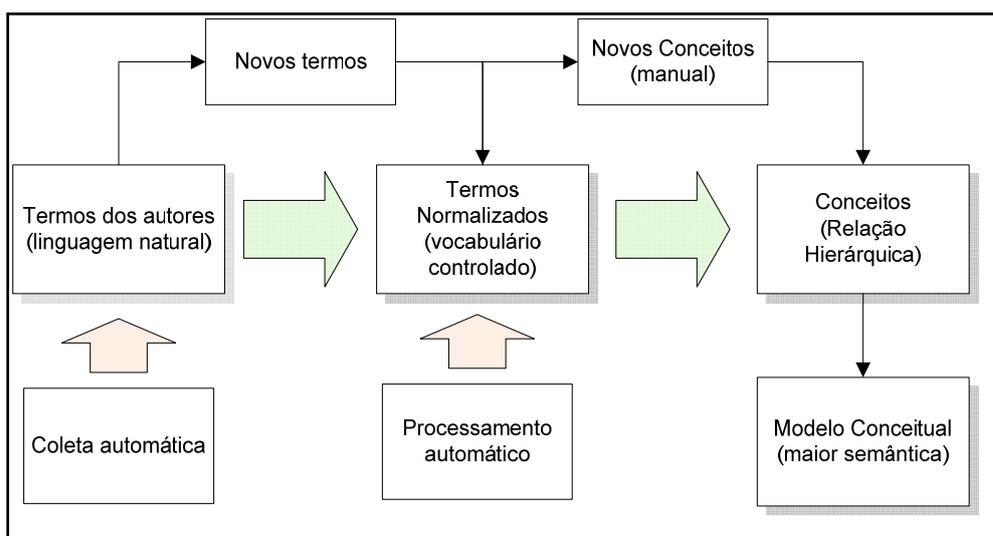


FIGURA 18 - FLUXO DA LINGUAGEM NATURAL PARA CONTROLADA – INDEXAÇÃO PÓS-COORDENADA PROPOSTA PELA METODOLOGIA

Fonte: O autor (2010)

A MB recomenda a coleta automática dos descritores (palavras-chave) de todos os trabalhos intelectuais disponíveis de forma automática por meio do protocolo OAI-PMH ou inserida manualmente pelos indexadores. Com a coleta os descritores são comparados aos já existentes na tabela de termos (não controlada), e em caso de não existirem, são incorporados à tabela. A entrada de novos termos é monitorada e analisada, identificando qual sua tipologia e associando-se à descrição de um conceito já existente como remissiva (USE), ou à criação de um novo conceito e indicação como termo preferencial.

Vale ressaltar que na MB um processo não precisa ser realizado imediatamente quando na incorporação de novos dados, pelo contrário, pode ser feito a qualquer momento. Porém, a informação, sem passar por análise ou tratamento, quando disponibilizada ao usuário pode conter inconformidade (erros). Entretanto, há o benefício de sua atualização e à medida que forem analisadas e indexadas, serão automaticamente corrigidas.

A padronização da forma de entrada garante à base de dados melhor qualidade, maior precisão e menor redundância para o pesquisador. A normalização, no entanto, não se restringe aos descritores, mas deve ser pensada tanto para o nome dos autores como para a construção dos resumos.

A proposta da MB é criar modelos ou mapas de representação de conteúdos, e para essa construção são necessários conhecimentos das formas de categorização apresentadas na fundamentação teórica, principalmente com a subordinação, a superordenação, além das outras formas de relações. Deve ficar claro também, que por questões estruturais e conjunturais, nem todos os instrumentos aqui propostos serão concretizados, no entanto a base conceitual para sua efetivação já foi definida.

Como ressaltado anteriormente, em uma base de dados que pretende realizar estudos de autoria e coautoria, deve disponibilizar instrumentos de padronização de entrada de nomes de autores, ação fundamental para garantir a precisão na recuperação da informação. Do mesmo modo que problemas foram encontrados nas palavras-chave, na representação dos nomes inconsistências foram identificadas. Por falta de padronização do próprio autor na escrita ou abreviação do seu nome, em uma observação da tabela de autores da Brapci, podem ser visualizadas as seguintes formas armazenadas:

- a) sobrenome e nome completo do autor;
- b) sobrenome e nome abreviado do autor;
- c) sobrenome e iniciais do nome do autor;
- d) nome e sobrenome;
- e) iniciais do nome e sobrenome.

Além dessas formas, alguns registros apresentaram inconsistência em sua grafia, identificando-se, por exemplo, o mesmo autor, com diferentes modos de entrada. Esses problemas são provenientes de eventos como:

- a) sobrenome com erros de escrita;

- b) sobrenome diferente para uma mesma autora, conforme seu estado civil;
- c) sobrenome homônimo, quando somente são representadas as iniciais do nome;
- d) sobrenomes espanhóis e hispanoamericanos, cuja entrada deve ser pelo penúltimo sobrenome.

Seguindo ainda o objetivo da análise bibliométrica, o controle não pode se restringir aos autores, mas deve possibilitar a identificação e categorização de sua instituição de vínculo, titulação e forma de contato (*e-mail*). Para essa efetivação é necessário criar instrumentos auxiliares que possibilitem a padronização das instituições e formação acadêmica de cada autor, porém este controle não pode ser vinculado ao autor diretamente, mas ao trabalho intelectual que ele produziu, ou seja, o nome da instituição vinculada e sua formação dependem do ano em que um artigo foi produzido, podendo o autor mudar de instituição ou aperfeiçoar sua formação acadêmica com o decorrer do tempo, e essas informações devem ser armazenadas para quando da necessidade de uma análise diacrônica de uma temática ou autor.

Como no controle dos termos, a metodologia define que os autores sejam identificados e inseridos na base conforme seu nome registrado no trabalho científico e que seja realizada uma pesquisa *a posteriori*, pela qual todos os autores e todas as formas de grafia sejam agrupados. Em caso de um mesmo autor ser representado de formas diferentes é recomendado o uso das remissivas (USE) como demonstrado no Quadro 4.

NOME CADASTRADO	NOME PADRONIZADO
CAFÉ, Lígia	CAFÉ, Lígia*
CAFÉ, Lígia Maria Arruda	use CAFÉ, Lígia
CAFÉ, Lígia Arruda	use CAFÉ, Lígia
Lígia Arruda Café	use CAFÉ, Lígia
Lígia Maria Arruda Café	use CAFÉ, Lígia
MIRANDA, Ligia Maria Café de	use CAFÉ, Lígia

QUADRO 4 - REMISSIVAS DE NOMES DE AUTOR

Fonte: O autor (2009)

NOTA:* nome definido pela Plataforma Lattes como forma de citação.

Desta forma, toda produção de um autor é concentrada em um único nome padronizado, evitando a dispersão e erros de autoria quando em uma análise deste tipo. A proposta é não alterar o nome pelo qual o trabalho foi representado, mas

disponibilizar a informação como foi cadastrada, criando *links* para o agrupamento do autor ensejando a recuperação precisa e análise de autoria.

3.4 PERFIS DOS USUÁRIOS DA BASE

Na MB existe uma designação de atividade e responsabilidade que cada membro do módulo de manutenção deve exercer. Essa designação é atribuída pelo perfil de cada usuário, estabelecendo-se atividades e relatórios aos quais este poderá ter acesso. Desta forma, as atividades e os processos serão executados por grupos específicos de usuários, com a categorização das respectivas funcionalidades, bem como dos perfis autorizados para cada tarefa. A definição dos perfis dos usuários da base, conforme consenso do Grupo E3PI, é representada no Quadro 5 e tem sido considerada fundamental para definir as atividades ou funcionalidades possíveis de se concretizarem na base.

Perfil dos usuários da base
Coordenador
Editor
Revisor
Indexador
Pesquisador
Visitante

QUADRO 5 - PERFIS DOS USUÁRIOS DA BASE
Fonte: O autor (2010)

Com a definição dos perfis de usuários na base é possível delinear o mapeamento das funções do sistema. Categorizadas as permissões de uso, elaborou-se um processo para indexação e manutenção da Brapci que contempla a possibilidade de ocorrerem atividades simultâneas e online diretamente no sistema (Figura 19).

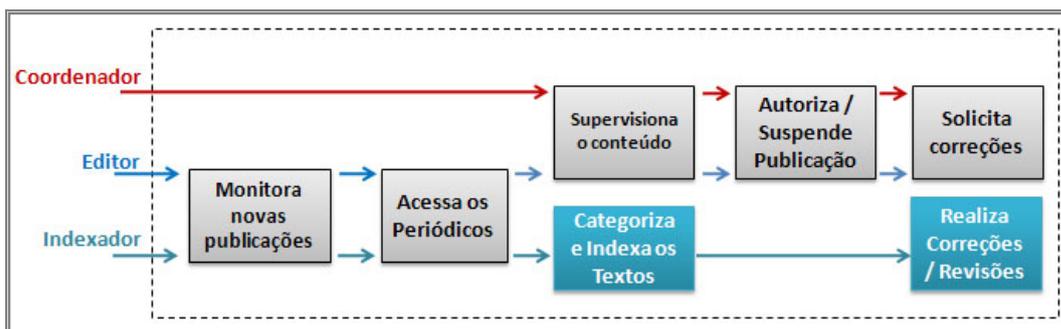


FIGURA 19 - PROCESSO PROPOSTO PARA INDEXAÇÃO E MANUTENÇÃO DA BRAPCI
Fonte: COSTA (2009)

Cada perfil definido pelo Grupo E3PI é autorizado a realizar determinadas atividades dentro da base, conforme descritas no Quadro 6, que apresenta todos os perfis de acesso.

Atividade	Perfil autorizado a realizar
Indexar novo Periódico	Coordenador
Indexar novas Edições	Editor / Indexador
Indexar novos Artigos	Todos
Realizar modificações nos Periódicos	Coordenador / Revisor
Realizar modificações nas Edições	Editor / Revisor / Indexador
Realizar modificações nos Artigos	Editor / Indexador
Excluir Periódicos	Nenhum
Excluir Edições	Nenhum
Excluir Artigos	Editor / Indexador
Autorizar/Suspender Publicação da Edição	Coordenador / Editor
Gerar Relatórios	Todos
Criar Usuários	Coordenador / Editor
Atribuir perfil ao Usuário	Coordenador / Editor
Coletar registros (OAI-PMH)	Todos
Consultar a base (módulo público)	Todos
Selecionar trabalhos e análises	Pesquisador

QUADRO 6 - ATIVIDADES PERMITIDAS A CADA PERFIL DE USUÁRIO

Fonte: Adaptado de COSTA (2009)

3.5 COLETA DE FASCÍCULOS E ARTIGOS CIENTÍFICOS

Para garantir uma atualização constante da base, esta deve utilizar-se de recursos tecnológicos que possibilitem a coleta automática, sempre que possível, de trabalhos científicos nos repositórios. A utilização de arquivos abertos pela quase totalidade das revistas ativas na área de Ciência da Informação no Brasil possibilita a otimização do monitoramento e coleta de trabalhos pelos computadores. Esta

ferramenta é acomodada pela interoperabilidade do computador da base com os das revistas que utilizam softwares baseados no *Open Journal System* (OJS).

A metodologia possibilita ainda a integração, por meio do protocolo OAI-PMH utilizado pelo Sistema Eletrônico de Editoração de Revistas (SEER), com revistas e periódicos nacionais, otimizando o processo de indexação e permitindo o aumento da abrangência do *corpus* da base. Desse modo, haverá possibilidade de transferência de conhecimentos para outros domínios, respeitadas as especificações e peculiaridades que permeiam cada área.

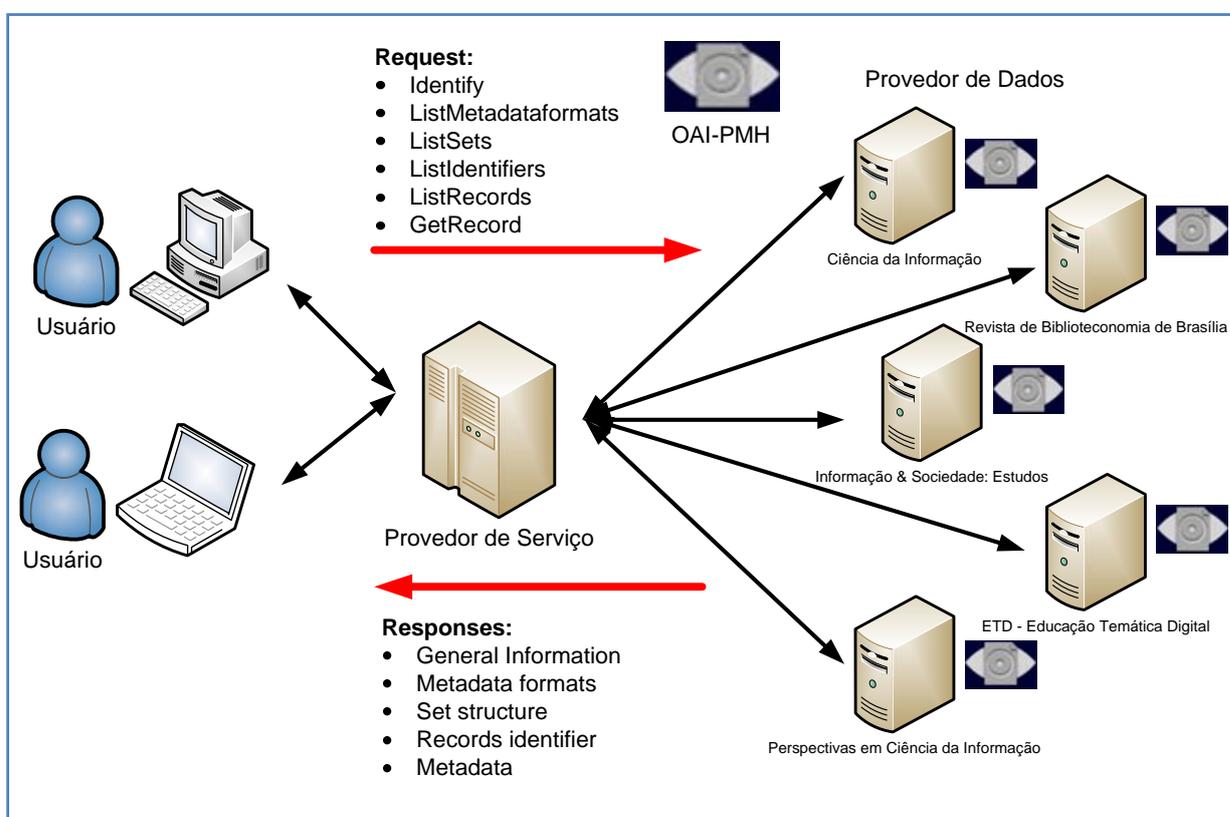


FIGURA 20 - MODELO DE COLETA DE DADOS OAI-PMH
Fonte: Adaptado de OPEN ARCHIVES INITIATIVE (2010)

A base estrutural do protocolo OAI-PMH é composta de consultas por meios de “verbos” que um servidor faz a outro, trocando desta forma metadados em linguagem XML, o que possibilita a identificação de novos fascículos e novos trabalhos disponíveis na publicação sem a necessidade de entrar no site do provedor para verificar se houve atualizações. As revistas que publicam trabalhos intelectuais são denominadas de provedor de dados (*data provider*) e os sistemas agregadores (propostos pelo modelo de Hurd, 2000) são denominados de provedor de serviço (*service provider*). A troca de metadados entre os computadores pode

ocorrer sem perdas, no que tange à posição dos campos. No exemplo do Quadro 7 é demonstrada a forma pela qual é recuperado um artigo científico da revista Ciência da Informação quando consultado pelo protocolo.

http://revista.ibict.br/ciinf/index.php/ciinf/oai/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:org.js.ibict.ciinf:article/529

QUADRO 7 - LINK PARA ACESSO À REVISTA CIÊNCIA DA INFORMAÇÃO PARA COLETA DE DADOS COM O PROTOCOLO OAI-PMH

Fonte: O autor (2010)

Quando na solicitação de uma consulta pelo link, a resposta dada pela revista está representada no Quadro 8.

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2010-06-25T12:46:50Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
    identifier="oai:org.js.ibict.ciinf:article/529">http://revista.ibict.br/ciinf/index.php/ciinf/oai/</request>
  <GetRecord>
  <record>
  <header>
    <identifier>oai:org.js.ibict.ciinf:article/529</identifier>
    <timestamp>1995-04-30T17:12:25Z</timestamp>
    <setSpec>ciinf:ART</setSpec>
  </header>
  <metadata>
  <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title xml:lang="pt-BR">A ciência da informação no contexto da pós-graduação do IBICT</dc:title>
  <dc:creator>Heloisa Tardin Christovão</dc:creator>

  <dc:subject xml:lang="pt-BR">Ciência da Informação</dc:subject>
  <dc:subject xml:lang="pt-BR">Ciência da informação; Pós-Graduação; IBICT; Currículo; Mestrado</dc:subject>
  <dc:subject xml:lang="pt-BR"></dc:subject>
  <dc:description xml:lang="pt-BR">Relata o surgimento do Curso de Pós-Graduação (Mestrado) em Ciência da Informação do IBICT e do Curso de Documentação Científica/Curso de Especialização em Informação e Documentação (CDC). Analisa a origem das disciplinas que constituem o currículo. Apresenta uma visão internacional e local (IBICT/DEP) das mudanças da &quot;ciência da informação&quot;. </dc:description>
  <dc:publisher xml:lang="pt-BR">IBICT</dc:publisher>
```

```

<dc:contributor xml:lang="pt-BR"></dc:contributor>

<dc:date>1995-04-30</dc:date>
<dc:type xml:lang="pt-BR"></dc:type>
<dc:format>application/pdf</dc:format>
<dc:identifier>http://revista.ibict.br/ciinf/index.php/ciinf/article/view/529</dc:identifier>
<dc:source xml:lang="pt-BR">Ciência da Informação; Vol. 24, No 1 (1995)</dc:source>
<dc:language>pt</dc:language>

<dc:coverage xml:lang="pt-BR"></dc:coverage>
<dc:rights>A revista se reserva o direito de efetuar, nos originais, alterações de ordem normativa, ortográfica e gramatical, com vistas a manter o padrão culto da língua, respeitando, porém, o estilo dos autores. As provas finais não serão enviadas aos autores.&lt;br /&gt;Os trabalhos publicados passam a ser propriedade da revista Ciência da Informação, ficando sua reimpressão total ou parcial, sujeita à autorização expressa da direção do IBICT. Deve ser consignada a fonte de publicação original. Os originais não serão devolvidos aos autores.&lt;br /&gt;As opiniões emitidas pelos autores dos artigos são de sua exclusiva responsabilidade.&lt;br /&gt;Cada autor receberá dois exemplares da revista.&lt;br /&gt;</dc:rights>
</oai_dc:dc>
                </metadata>
            </record>
        </GetRecord>
</OAI-PMH>

```

QUADRO 8 - RESPOSTA DA CONSULTA DA REVISTA PELO PROTOCOLO OAI-PMH
 Fonte: O autor (2010)

A recuperação das informações realizadas entre computadores contém campos estruturados garantindo a modelagem dos dados, mas para leitura pelo ser humano observam-se problemas na transmissão dos conteúdos como caracteres não legíveis e agrupamento de informações em um único campo. Nos dados recuperados acima, existe a frase “das mudanças da "ciência da informação"”, incluindo sinais gráficos que os computadores podem processar e reconverter para a leitura humana.

Na observação do campo estruturado referente às palavras-chave, percebe-se que não existe uma padronização na transmissão, mesmo tendo os campos específicos para cada tipo de informação com o idioma do descritor, estes vêm dentro de um único campo. Numa análise preliminar com a coleta desse campo observa-se que cada uma das revistas tem uma forma. Algumas utilizam como separador de palavra-chave o “ponto e vírgula”, outras o “ponto” e outras ainda utilizam a *Tag* do Dublin Core o “*dc:subject*” para um dos descritores, o que seria o correto para interpretação entre computadores, conforme representado no Quadro 9.

<dc:subject xml:lang="pt-BR">Ciência da informação; Pós-Graduação; IBICT; Currículo; Mestrado</dc:subject>
--

QUADRO 9 - DESCRITORES DA CIÊNCIA DA INFORMAÇÃO

Fonte: O autor (2010)

Ressalta-se que a coleta automática e incorporação à base não exime os indexadores e revisores de análises dos conteúdos que serão disponibilizados para consulta, indicando as novas incorporações com o *status* para “indexar”.

3.6 ESTRUTURA E INFRAESTRUTURA PARA O SERVIDOR WEB

A qualidade de uma base de dados não pode se prender somente a sua estrutura lógica, mas deve se preocupar com a qualidade física de suas instalações. A infraestrutura deve prover a base de dados um funcionamento contínuo, com o mínimo de falhas, reduzindo ao máximo o período de inacessibilidade por motivos técnicos. Na metodologia da Brapci, que é ambientada na *Web*, existe a necessidade de seu sistema estar conectado a uma linha de Internet todo o tempo (*fulltime*) e um servidor que dê suporte à quantidade de consultas previstas para o sistema.

Outro fator importante na definição da política da MB é de que ela utilize somente *softwares* que tenham sua licença livre. Esta definição se alinha às orientações que o Grupo E3PI em conjunto com o Centro de Computação Eletrônica (CCE) da UFPR estipularam para o servidor. Desta forma, adotou-se o sistema operacional *Debian*, com aplicação servidora *Apache* e o módulo *PHP5* e o *MySQL* como Sistema Gerenciador de Banco de Dados (SGBD).

O servidor de hospedagem é um IBM (X3200) com 4 Giga de memória RAM e disco rígido com capacidade de 250Gbytes. Este equipamento está alocado em uma sala de servidores do Setor de Ciências Sociais Aplicadas conectado à Rede UFPR em conexão de 1 Gigabyte. Existe um sistema contra queda de luz suprido por um *no-break* dedicado NHS Senoidal com capacidade de 4200VA/2940W. Este equipamento mantém o servidor em funcionamento sem energia externa por aproximadamente duas horas e meia.

Para garantir a permanência dos dados, é realizado diariamente um *backup* a frio (sem a necessidade de reinicializar o servidor), com armazenamento histórico por sessenta dias. Um segundo *backup* é realizado, porém apenas copiando as informações de um servidor para outro, mantendo uma cópia duplicada em cada um dos dois servidores.

Por motivo de segurança a sala de servidores do Setor de Ciências Sociais Aplicadas conta com restrição de acesso físico, com a entrada somente de pessoas autorizadas. A temperatura é constante, em torno de 15°C, com filtragem do ar para evitar resíduos sólidos no ar (poeira).

3.7 SISTEMA GERENCIADOR DE BANCO DE DADOS (SGBD)

SGBDs são coleções de softwares que permitem por meio de uma linguagem apropriada estruturar e criar modelos de dados com regras e restrições para armazenamento e manipulação do banco de dados.

O SGBD possibilita o armazenamento de informações de forma organizada e estruturada, proporcionando integridade e consistência na entrada de dados, segurança quanto ao acesso e controle à interação entre usuários simultaneamente, verificando as ocorrências de falhas e proporcionando procedimentos de *backup* e restauração (SILBERSCHATZ; KORTH; SUDARSAHN, 1999).

Segundo Kern (1994), o tratamento da informação a partir de um banco de dados, proporciona ao sistema algumas vantagens como:

- a) gerenciamento e controle dos usuários pelo administrador do DB, proporcionando privilégios de consulta, alteração ou exclusão de registros mesmo antes das aplicações serem desenvolvidas, favorecendo a segurança dos dados;
- b) estruturação das tabelas do banco, reduzindo ou eliminando a redundância de dados, favorecendo a manutenção e integridade dos dados e melhorando sua consistência;
- c) padronização, os tipos de dados mais importantes são atribuídos com nomes e tamanhos definidos no planejamento do banco;

d) concentração de dados em um ambiente, possibilitando o acesso por qualquer aplicativo com privilégios para consulta.

A arquitetura de um SGBD é estabelecida a partir de um modelo de dados, que é formada pela representação de uma abstração, que pode ser interpretada tanto por um profissional em TI, bem como por um leigo não familiarizado com esta linguagem. O projeto de um banco de dados envolve o desenvolvimento de um modelo formal relativo à estruturação dos dados (KERN, 1994).

Uma das etapas do desenvolvimento de uma base de dados é a modelagem, técnica esta da CC, que proporciona a estruturação do modelo conceitual em tabelas relacionais dentro de um SGBD, criando regras de validação e restrições que são executadas no desenvolvimento da aplicação. A execução desta atividade é realizada com o uso das definições conceituais da MB apresentadas na Figura 13.

A modelagem é a transformação ou interpretação das necessidades de informação do sistema construída por meio de relacionamento de campos e regras de validação da informação que serão inseridas dentro das tabelas. Esse trabalho demanda precisão em sua modelagem podendo causar problemas sérios e de alto impacto em todo o sistema quando mal estruturado. Uma má estruturação na modelagem de dados pode impactar na velocidade (*performance*) do sistema.

Por ser uma metodologia que poderá ser aplicada em diversas áreas e estruturas, teve-se como diretriz básica a não utilização de recursos específicos de uma arquitetura de SGBD, mas a compatibilização com os principais sistemas de banco de dados disponíveis. Desta forma, optou-se em construir o sistema com o uso da *Structured Query Language* (SQL), em sua modalidade básica, isso é, sem a utilização de *functions* (funções) e *procedures* específicos, empregando-se somente chaves primárias e índices na modelagem de dados. Este uso também se justifica, quando na migração de dados de uma plataforma para outra, ocasionando o mínimo ou nenhum impacto na migração ou restauração de *backups*.

Na estruturação dos bancos de dados para a MB, em reunião com o Grupo E3PI, optou-se por trabalhar com dois bancos de dados dentro do SGBD, um acoplado ao módulo de manutenção (*brapci_base*) e outro ao módulo público e pesquisador (*brapci_publico*).

O primeiro banco de dados é composto de tabelas relacionadas à coleta, seleção e preparo das informações, com dados sobre as revistas, edições

(fascículos), artigos e autorias, além de inúmeras tabelas auxiliares de cadastro com o objetivo de manter a padronização do sistema. O segundo banco é responsável por disponibilizar ao visitante e pesquisador informações sucintas, com separação dos campos de busca e armazenando de todos os acessos dos visitantes e suas estratégias de busca (*logs*).

O software para gestão da SGBD escolhido foi o *phpMyAdmin*, compatível com *MySQL*, e é utilizado pelo próprio ambiente Web, facilitando o acesso e manutenção dos campos do banco de dados. Entende-se por banco de dados, conforme o conceito de Boghi (2002), todas as tabelas e suas relações sem instrumentos de manipulação e consulta, sendo um elemento da base de dados.

A modelagem de dados foi realizada em um software específico para esse fim, baseada no modelo de relacionamento de dados de Costa (2009), apresentado na Figura 11, pelo qual foram criadas as tabelas de acordo com as estruturas necessárias para subsidiar a operação da Brapci.

No banco de dados **brapci_base** destacam-se as tabelas:

- a) *journals* (publicações) - informações sobre a publicação, site, última coleta, vigência, periodicidade etc.
- b) edições (fascículos) - informações sobre os fascículos publicados, data, ano, número e volume relacionando com a tabela *journals*;
- c) trabalhos - informações sobre o trabalho intelectual do autor ou autores, com o título, resumo, palavras-chave, páginas, seção relacionando com a tabela edições;
- d) autores - cadastro dos autores e suas remissivas;
- e) palavras-chave - tabela dos termos utilizados em linguagem natural pelos autores;
- f) conceitos - tabelas com codificação dos conceitos relacionando com sua explicitação pelos termos;
- g) relacionamento de conceitos - tabela que relaciona os conceitos pela hierarquia, temática, subtemática.

No banco de dados **brapci_publico** destacam-se as tabelas:

- a) trabalhos para consulta - em uma única tabela todas as informações referentes a publicação;
- b) descritores - relação dos descritores identificados no banco de dados Brapci_base;

- c) consulta - registro das consultas realizadas pelos usuários;
- d) acesso - registro dos acessos dos usuários.

Todas as tabelas e definições de seus campos estão detalhadas no apêndice A - Diagrama Entidade de Relacionamento (DER).

A modelagem de dados possibilitou o início do desenvolvimento da Brapci em um produto real, ainda como protótipo, mas com possibilidade de operacionalização. O processo inicial para o desenvolvimento foi a criação das telas do sistema que possibilitassem o cadastro dos usuários e seus perfis, telas para cadastro das revistas, telas para cadastro das edições e telas para cadastro dos artigos.

Como a Brapci já continha uma base de dados estruturada no *ProCite5*, foi necessária a criação de uma rotina que possibilitasse a importação dessas informações para a nova base de dados. Essa importação gerou inconsistências, pois os critérios de padronização da nova metodologia conflitaram com a flexibilidade da base anterior. Essas inconsistências foram observadas na escrita de nomes, títulos de artigos, artigos sem resumo ou palavras-chave, além da quantidade de trabalhos que não correspondiam ao total real do fascículo.

As dificuldades não previstas pelo Grupo E3PI foram discutidas e assim surgiu a proposta de uma revisão geral total dos conteúdos registrados que foram importados do *ProCite5*. Esta revisão iniciaria com a alteração de todos os *status* dos trabalhos como se eles tivessem sido inseridos como novos, ou seja, para serem reindexados.

3.8 GESTÃO DE PROCESSOS DA METODOLOGIA BRAPCI

A MB por ser multiusuário, estando disponível para manutenção em qualquer ponto de acesso a Internet, carece de instrumentos que monitorem e indiquem aos operadores do sistema quais tarefas foram realizadas e quais ainda precisam ser realizadas.

Desta forma, é possível identificar o total de trabalhos que não foram indexados, quantos não foram revisados e principalmente o volume de trabalhos novos incorporados a base.

Nesta metodologia o elemento de controle das revistas é realizado pelo *status* do fascículo, possibilitando ao indexador uma visualização rápida da atual situação de cada edição representada no Quadro 10.

Ícone	Situação do fascículo de cada revista
	Trabalhos coletados e não processados (indexação)
	Trabalhos indexados e prontos para primeira revisão
	Trabalhos com a primeira revisão pronta, encaminhado para segunda revisão
	Trabalhos conferidos e finalizados

QUADRO 10 - REPRESENTAÇÃO GRÁFICA DOS STATUS DOS FASCÍCULOS

Fonte: O autor (2010)

Desta forma o indexador ao adentrar na publicação específica para trabalhar, será informado de quais fascículos estão finalizados e quais precisam ser indexados. A troca do *status* do fascículo se dá de forma automática quando em sua conclusão.

Na indexação de um trabalho intelectual o perfil responsável por esta atividade deve rever e corrigir as inconsistências dos campos, para o indexador analisar e interferir na:

- a) identificação dos autores;
- b) nome da instituição afiliada;
- c) e-mail de contato, outras formas de contato;
- d) titulação do autor (Dr., Msc, estudante, profissional, *stricto sensu*);
- e) título original e título no segundo idioma;
- f) resumo no idioma original;
- g) palavras-chave no idioma original;
- h) resumo (*abstract*) no segundo idioma;
- i) palavras-chave (*keywords*) no segundo idioma;
- j) observações (pendências) do processamento;
- k) número das páginas inicial e final;
- l) metodologia da pesquisa;
- m) referências.

Na finalização do processo de indexação, o operador do sistema aciona a funcionalidade “concluído, enviar para revisão”, este procedimento encaminha a atividade para o revisor, transferindo de sua alçada o trabalho específico.

Com a alteração da rotulagem do *status* do trabalho, este passa para a fase de conferência, que conceitualmente deveria ser realizada por dois revisores diferentes. Entretanto, na operacionalização percebeu-se que uma segunda revisão seria necessária somente em caso de dúvida, sendo que a grande maioria dos trabalhos poderia ser analisada somente por um revisor.

O revisor tem por obrigação efetuar a revisão do trabalho realizado pelos indexadores. Na impossibilidade de resolução de problemas referentes à indexação existe um campo no sistema que possibilita ao indexador a inserir comentários dos motivos pelos quais não foi possível finalizar a atividade. Assim, em um campo denominado “pendências de indexação”, que é visitado rotineiramente, o revisor tem acesso aos problemas encontrados, e com maior experiência busca suas resoluções.

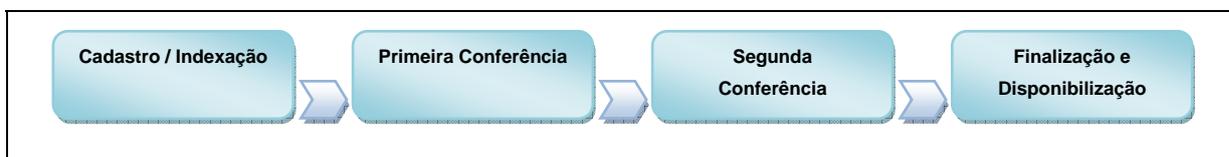


FIGURA 21 - PROCESSO DE CONFERÊNCIA DA METODOLOGIA BRAPCI
Fonte: O autor (2010)

Como relatado, o processamento de um trabalho científico coletado ou cadastrado percorre quatro processos até sua revisão total (Figura 21).

Os procedimentos de indexação e consultas geram listas de termos que são organizadas e empregadas na construção de listas que orientam o usuário quais os termos mais empregados na base, essas listas são as nuvens de *tags*.

3.9 NUVEM DE TAGS

A nuvem de *tags* é um elemento de interface do usuário, geralmente associado com os conjuntos de dados denominado de *folksonomia*. Uma nuvem de *tags* é uma representação visual dos termos de conteúdo usados em um site. Os

termos mais significativos são visualizados em uma fonte maior, conforme sua incidência ou relevância para o site. Sua apresentação geralmente é em ordem alfabética. Assim, a procura de uma *tag* por ordem alfabética e de popularidade torna-se possível. A seleção de uma única marca dentro de uma nuvem de *tags* conduzirá geralmente a uma coleção de itens que estão associados com seu termo (SINCLAIR; CARDEW-HALL, 2008).

Em essência, utiliza-se a nuvem de *tags* para representar o vocabulário emergente de uma *folksonomia* em uma ferramenta de navegação social (DIELBERGER *et al.*, 2000). Na visão de Dielberger e colaboradores (2000), apesar de as nuvens de *tags* serem extremamente comuns entre os serviços de marcação, a literatura acadêmica parece optar em grande parte pelo silêncio sobre o assunto. As nuvens de *tags* foram empregadas inicialmente em sites de *blog* para classificar o assunto.

O acoplamento dos termos empregados pelos pesquisadores em suas estratégias de busca é possível porque o sistema armazena todas as formas de consulta do usuário. Este armazenamento possibilita a geração de relatório com a incidência dos termos em ordem decrescente, com limitação de 50 termos. Nesta lista gerada pelas consultas dos pesquisadores (*folksonomia*) são identificadas a maior (*f_{max}*) e a menor (*f_{min}*) frequência dos termos. Estes valores são aplicados na fórmula 1 para calcular o tamanho da fonte da letra será grafado o termo.

$$TagSize = 1 + C \frac{\log(f_i - f_{min} + 1)}{\log(f_{max} - f_{min} + 1)}$$

onde:

TagSize: resultado do tamanho da fonte
C: escala máxima do tamanho da fonte
f_i: frequência incidência do termo
f_{min}: menor valor de todos os descritores
f_{max}: pico de incidência do descritor.

Fonte: SINCLAIR; CARDEW-HALL (2008)

Para a base de dados, a nuvem de *tags* não só representa os termos e a temática mais consultada pelos pesquisadores, mas uma forma de disponibilizar conteúdo para ser indexado pelos robôs de busca (*Google*, *Yahoo*, *Bing*). A indexação por esses mecanismos aumentou significativamente a visibilidade e o

número de acesso da Brapci, devido à múltipla indexação de conteúdo derivada dos termos das *tags*. Este fato não pode ser desprezado conforme demonstrado nas pesquisas de Quadros (2010), pelas quais mais de 75% dos acessos à Brapci são oriundos dos mecanismos de busca e somente 25% de sites referenciais e acesso direto.

3.10 CONTROLE E REGISTRO DE ACESSO

Com o objetivo de subsidiar a investigação científica, a MB tem como diretriz o controle e monitoramento das consultas realizadas na Base. Esses dados são armazenados de forma contínua e analisados quando solicitados.

Uma das aplicações que utiliza os resultados das consultas é a criação das *Nuvens de Tags* da página inicial do módulo público da Brapci. Nesta geração o sistema ordena os cinquenta termos mais consultados e conforme sua incidência aumenta o tamanho de sua letra.

O monitoramento ainda possibilita a visualização das estratégias de busca que os usuários estão empregando na recuperação da informação, bem como a identificação dos temas mais consultados em determinado período.

No registro da estratégia de busca, a MB propõe o registro e dados adicionais do usuário como a data, hora e local (IP) de utilização do sistema, controlando a produção dos indexadores.

Ferramentas de monitoração de visitas do site também podem ser empregadas com controle externo. É o caso do *Google Analytics* que cria outros indicadores como origem geográfica de acesso, tempo de permanência no site e índice de rejeição. A incorporação deste instrumento é realizada pela inserção de um código na própria página de consulta do módulo pesquisador.

3.11 METODOLOGIA DE RELEVÂNCIA DE RECUPERAÇÃO POR PONTOS

Definir critérios que forneçam ao usuário orientação da relevância de trabalhos pesquisados alinhados as estratégias pragmáticas na busca deva vir ao encontro das necessidades de criar mecanismos filtrantes e informativos da representatividade do assunto. Como proposta da metodologia para recuperação da informação e com o intuito de orientar o pesquisador da relevância do assunto pesquisado com o conteúdo do trabalho apresentado, foi desenvolvida uma metodologia de pontos que calcula por meio de um índice a relevância do trabalho com o descritor utilizado na busca. A metodologia identifica em três campos (título, resumo e palavras-chave) a presença do termo de consulta. Para cada localização é atribuída uma pontuação específica, e a somatória desta define uma representação gráfica com estrelas visualizada pelo usuário.

A metodologia de pontos visa orientar o usuário da relevância do conteúdo do trabalho com o termo pesquisado. Quanto maior a quantidade de estrelas (somatória de pontos), maior a identificação do conteúdo com o tema. A atribuição de valores para cada elemento esta descrita no Quadro 11, atribuindo:

PONTOS	CAMPO
1	Incidência no Resumo
2	Incidência no Título
4	Incidência nas Palavras-chave

QUADRO 11 - QUADRO DE ATRIBUIÇÃO DE PONTOS PARA RESULTADO DA PESQUISA
Fonte: FREITAS; BUFREM; GABRIEL JR (2010)

Cabe ressaltar que a metodologia de pontos depende da estratégia de busca do usuário e da forma como o autor representa sua obra (título, resumo e palavras-chave), e que os resultados podem ser distorcidos quando o autor emprega na representação descritiva de sua obra figuras de linguagens ou analogias, fazendo com que um assunto relevante para a temática não seja recuperado devido à imprecisão do autor na descrição de seu trabalho.

3.12 POSSIBILIDADES DE ANÁLISE

A MB por definição deve transcender um “repositório”, criando um ambiente que propicie ao pesquisador diversos tipos de análises de conteúdo, comportamento e tendências da literatura.

A criação e o desenvolvimento de ferramentas que possibilitem a análise de conteúdo estão abarcados na proposta da MB, tendo em sua fase inicial a disponibilização de alguns instrumentos que são avaliados pelo Grupo E3PI e posteriormente disponibilizados no módulo pesquisador.

O constante desenvolvimento e avaliação possibilitam à MB a incorporação de novos instrumentos e tipos de análises. Um ponto importante a ser ressaltado é que a metodologia pretende ser livre (código aberto), podendo ser replicada e aplicada por qualquer pessoa ou instituição que demonstre interesse em seu uso. A política de acesso ao código fonte possibilita que o sistema receba críticas e contribuições para idealização e construção contínua.

Na primeira versão do sistema, a metodologia propõe o uso dos tipos de análises, tanto para o foco qualitativo, como quantitativo, dependendo apenas das finalidades que o pesquisador provê aos indicadores produzidos.

As possibilidades de análise desdobram-se proporcionalmente ao desenvolvimento de metodologias de caráter quantitativo, cujas realizações vão sendo definidas conforme os objetos e objetivos das pesquisas. Nesse contexto, a bibliometria, como uma prática multidisciplinar, começou a ser usada para identificar comportamentos da literatura e sua evolução.

Definida pela primeira vez por Otlet, em 1934, como parte da Bibliografia “que se ocupa da medida ou da quantidade aplicada ao livro” (1986, p. 20), a Bibliometria procura um perfil dos registros do conhecimento, servindo-se de um método quantificável, voltando-se para o estudo de alguns problemas sociológicos e tentando atingir uma realidade concreta (BUFREM; PRATES, 2005).

As possibilidades apresentadas por esse tipo de estudo são defendidas por Santos e Kobashi (2009) ao afirmarem que os estudos métricos se aproximam, cada vez mais, das ciências ditas “moles” (CHS), pois estas últimas oferecem teorias e modelos que permitem interpretar os dados em contextos distintos. São também passíveis de realização as análises com enfoques e métodos diferenciados.

3.12.1 Análise de citação

O estudo das mais remotas práticas de utilização das técnicas bibliométricas permite constatar que elas serviam para mensurar fisicamente o livro, pode-se observar, sob um enfoque diacrônico, que em suas aplicações históricas, elas foram expandindo um domínio de aplicação científica que hoje se ramifica em sub-áreas do conhecimento e de aplicações técnicas. Uma das possibilidades de análise bibliométrica é a análise de citação. Integrante da bibliometria, busca investigar as relações entre os documentos citantes e os documentos citados, considerados como unidade, no todo, ou em suas diversas partes: autor, título, origem geográfica, ano e idioma de publicação (SILVEIRA, 1993, p. 18).

Vanz e Caregnato (2003) ao citarem Morel (1977), entendem que a bibliometria tem como premissa o fato que os cientistas elaboram seus estudos a partir de publicações anteriores e as reconhecem mencionando-as em seus textos em uma lista de referências. O comportamento dos cientistas fica evidenciado a partir das citações, representando as tendências de concentração em campos, em instituições, em países e no uso de determinados periódicos.

As possibilidades dos estudos bibliométricos estão em conhecer o comportamento de uma propriedade em relação a outra já conhecida, identificando, por exemplo, os autores que trabalham ou que trabalharam em determinado assunto; os periódicos mais citados, os mais representativos da área e vida média da literatura sobre a área, entre outros aspectos relacionados à literatura científica. Uma das formas de criar indicadores e identificar os autores fundantes e representativos de uma temática é a análise de citações e das referências empregadas na fundamentação teórica dos trabalhos. Porém, este tipo de análise requer cuidados especiais quando no tratamento e organização da informação contida dentro dos trabalhos publicados nas revistas científicas dentro da base de dados.

A primeira dificuldade está no número de artigos publicados anualmente pelas revistas. Para ilustrar essa dificuldade, tomando como base a área de CI, os dados da Brapci revelam que a média anual entre os anos de 2006 e 2009 chega a aproximadamente 490 artigos por ano. Cada artigo publicado, conta com no mínimo dez referências, as quais devem ser processadas, assim como se faz necessária uma linguagem de marcação precisa para análise de citações. São mais de cinco mil

referências anualmente que precisam ser marcadas, identificadas, categorizadas e em alguns casos, corrigidas para que possam ser reconhecidas pelo sistema.

A linguagem de marcação é fundamental para organização e geração de indicadores de pesquisa. A metodologia propõe e define métricas para o uso da linguagem de marcação de referências, mas considera que seu uso em 100% dos artigos é uma tarefa exaustiva e requer um grande número de indexadores somente para este fim. Neste processamento, também precisam ser consideradas as divergências entre o autor citado e a obra propriamente, encontrando-se em alguns casos, a imprecisão na escrita do título da obra, ou na forma como foi referenciada.

Deste modo, o processo de marcação, mesmo que realizado de forma semi-automática, precisa ser revisado para a realização de possíveis correções. Esta necessidade existe quando se pretende criar indicadores de citação fidedignos. Das considerações sobre a linguagem de marcação, ressalta-se que ela é retirada das referências das citações do autor, cabendo identificar: nome dos autores, título da obra, local de publicação, volume, número, ano, página, título da publicação, idioma, cidade da publicação e o tipo de obra (revista, livro, capítulo de livro entre outros) para que possa ser realizada a marcação.

Em uma base de marcação estruturada, existe a possibilidade de extrair indicadores como:

- a) distribuição das citações ou referências dos artigos pelo tipo de fonte (livro, documento eletrônico, artigo, capítulo de livro);
- b) distribuição dos autores por sua origem geográfica e idioma de publicação;
- c) vida média da temática selecionada, representando o ano em que a produção atingiu a metade do total de artigos publicados sobre o assunto, indicador conhecido como *half life*;
- d) grupo de autores mais referenciados sobre a temática;
- e) aplicação do método de Lokta⁵;
- f) aplicação da Lei de Bradford (idade média)⁶;
- g) aplicação dos conceitos de Zipf⁷;
- h) obras mais citadas.

⁵ Método de medição da produtividade de cientistas.

⁶ Lei de dispersão do conhecimento científico.

⁷ Modelo de distribuição e frequência de palavras num texto.

Cabe ressaltar que, neste tipo de estudo, existe a necessidade de todas as referências estarem cadastradas na base, o que demanda tempo suficiente para sua organização e revisão.

Para validação de cada referência citada pelo autor, a metodologia propõe os seguintes procedimentos:

- a) identificação do símbolo “___” e substituição pelo nome do autor;
- b) importação e vinculação de todas as referências com o trabalho que os citou;
- c) primeira conferência da referência com a identificação dos autores e tipo da obra;
- d) segunda conferência com a identificação do ano da obra;
- e) terceira conferência com a identificação do tipo de obra;
- f) quarta conferência com a identificação da instituição publicadora.

Para cada fase concluída o trabalho recebe um *status*, caso o registro passe por intervenção manual, este retorna para o primeiro nível de revisão.

3.12.2 Análise de colaboração da produção de trabalhos

A identificação da colaboração dos autores pode ser realizada analisando-se a quantidade de publicações que foram realizadas individualmente ou coletivamente. A análise da cooperação dos autores possibilita a exploração da formação e configuração das relações sociais na área de conhecimento da base. Esta representação possibilita uma análise diacrônica ou sincrônica das relações entre os autores e ou instituições.

A metodologia proporciona ao pesquisador, no ME e logado⁸, a geração de gráficos com relações e indicadores que mostram a relação de autoria simples e múltipla em até cinco autores.

A geração desses relatórios é possível porque na construção da metodologia pensou-se na possibilidade desse tipo de análise, identificando e categorizando

⁸ O pesquisador precisa estar identificado na base, e para isso precisa ter um login e uma senha.

cada autor individualmente. Neste tipo de gerador de informações, podem-se aplicar delimitadores temporais por ano.

3.12.3 Análise da produção de autores

Em uma busca por determinado tema, ou período, a identificação dos autores mais produtores pode ser importante para o conhecimento da temática, ou até para buscas de outros trabalhos destes autores.

Alinhada a essa ideia, a MB proporciona ao pesquisador (logado), um indicador com a quantidade de trabalhos e a relação dos autores mais produtores de acordo com os critérios de seleção identificados pelo pesquisador.

O resultado do número de autores limita-se a apresentar os cinquenta mais produtores. Para trabalhos realizados em colaboração, cada um dos autores é somado isoladamente, resultando em uma totalização individual de cada autor. Uma das falhas decorrentes desta análise é a despadronização da escrita do nome dos autores, possibilitando uma somatória em alguns casos imprecisa, quando um mesmo autor está grafado diferente, e este ainda não passou pela revisão dos indexadores da base para o acoplamento de seus vários nomes em apenas um que o represente.

3.12.4 Análise de redes sociais

A análise de redes sociais (ARS) é uma abordagem oriunda da Sociologia, da Psicologia Social e da Antropologia (FREEMAN, 1996), possibilitando o estudo das ligações relacionais entre atores sociais. Os atores cujas ligações são analisadas pela ARS podem ser tanto pessoais quanto organizacionais, identificados como unidades individuais ou sociais coletivas. A ARS interessa a pesquisadores de vários campos do conhecimento que, na tentativa de compreenderem o seu impacto sobre a vida social, deram origem às diversas metodologias de análise que têm como base as relações entre os indivíduos, em uma estrutura em forma de redes.

A metodologia proporciona ao usuário pesquisador os resultados de dados para serem exportados para outros aplicativos com a geração de gráficos e representações das relações dos autores conforme a temática selecionada. O software compatível com a metodologia é o *Pajek*⁹, selecionado por atender as necessidades gráficas e ser disponibilizado em livre acesso e gratuito.

3.12.5 Análise da metodologia da pesquisa

A MB vem ao encontro de trabalhos que buscam compreender as metodologias, métodos e técnicas de pesquisa empregadas no desenvolvimento dos estudos registrados na base. A categorização proposta nesta pesquisa é parte integrante do projeto da professora doutora Leilah Santiago Bufrem, que busca identificar e categorizar as metodologias empregadas na área de CI, com o objetivo de sua compreensão e consolidação.

Como produto desta categorização, a MB fornece suporte ao cadastro de metodologias e geração de relatórios para sua análise, bem como informa ao usuário pesquisador a metodologia empregada em cada trabalho. Como resultado desta categorização, o sistema disponibiliza online um “resumo” de todas as metodologias, métodos e técnicas categorizadas e com as definições e os autores que as fundamentam.

Para categorização da pesquisa, por definição do Grupo E3PI, optou-se em dividi-la quanto a seus: fins, meios, enfoque, técnica de pesquisa e tipo de análise. Ressalta-se que a metodologia possibilita esse tipo de análise e categorização, porém é necessário um empenho dos indexadores na efetivação da leitura de todos os trabalhos e sua categorização, conforme proposta.

⁹ O Pajek é um programa em código aberto para *Windows*, desenvolvido para análise e visualização de grandes redes que possuam milhares ou até milhões de nós.

4 APLICAÇÃO EM UMA REALIDADE – ESTUDO DO CASO BRAPCI

Da concepção inicial de reunir a literatura pertinente da área de CI em um único local que facilitasse a busca e recuperação da informação para pesquisadores, acadêmicos e comunidade em geral até o desenvolvimento do produto final que pudesse ser acessado por todos em um ambiente virtual e coletivo, foram demandados muitos recursos, como tempo e empenho do Grupo E3PI. A concretização dessa ideia foi possível com a aprovação de dois projetos pelo CNPq que garantiu aporte financeiro para aquisição de equipamentos e desenvolvimento da base.

A realização de uma base de dados com escopo temático para área de Ciência da Informação é justificada por estar relacionada com a área de estudo do pesquisador, integrante do Grupo E3PI, da UFPR, composto por docentes e discentes do Mestrado em Ciência, Gestão e Tecnologia da Informação e graduandos do curso de GI. Além disso, contou com uma base de dados pré-estruturada em um banco de dados “antigo” (ambientado no *ProCite5*), no qual a estrutura preexistente possibilitou uma análise, mesmo que de forma parcial, de quais informações seriam necessárias para subsidiar pesquisas futuras.

A concretização da metodologia resultou em um produto final, a Brapci, sendo possível consolidar as etapas de coleta, organização e disponibilização de trabalhos científicos na área de CI publicados em revistas científicas com avaliação Qualis e conseqüentemente sua validação por seus usuários.

Resgatando a história da Brapci, relata-se que a indexação de artigos científicos completou em 2010 onze anos de trabalho com a coleta, seleção e organização de artigos representativos para a área, publicados em revistas científicas. O projeto foi idealizado por Bufrem (2008) e o pelo Grupo E3PI, com intuito de reunir em um único banco todos os trabalhos publicados em revistas relevantes para área da CI, organizados de modo a oferecer possibilidade de se extraírem não somente dados quantitativos, mas qualitativos, para análise e desenvolvimento de pesquisas para reconhecimento da área e exercício da crítica da área de CI no Brasil.

Nos três primeiros anos da implantação do projeto (2000-2003), segundo Costa (2009), foram levantados os títulos dos periódicos da área (treze títulos) e,

paralelamente, realizada busca no acervo da Biblioteca do Setor de Ciências Sociais Aplicadas (BSCSA) da UFPR, onde foram catalogados os periódicos do acervo físico relacionados à área de CI. Os fascículos não encontrados foram solicitados aos editores ou, após localização no Catálogo Coletivo Nacional de Publicações Seriadas (CCN) do IBICT, a outras bibliotecas, para que fossem preenchidos os dados de acordo com os campos definidos para cadastro na base do *ProCite5*. Nos anos seguintes, compreendidos entre 2004 e 2008, foram realizadas algumas revisões nos critérios da política de títulos ao acervo, possibilitando a incorporação de novos títulos à base. Essa revisão ampliou significativamente a quantidade de títulos selecionados, dos treze iniciais para 27 revistas em 2008, totalizando 4.637 artigos (BUFREM, 2008).

Em meados de 2009 o protótipo foi disponibilizado para ser testado, permanecendo os dois sistemas, o *ProCite5* e a base *online*, por um período de três meses. Com a efetivação do protótipo foram identificados e incorporados à base mais três títulos de publicações, implementando-se o mecanismo de coleta automática de registros (OAI-PMH), que ampliou para 1.984 o número de registros adicionais. Este número significativo de novos registros foi ocasionado pela coleta automática e inclusão de todos os tipos de seções disponibilizados pelo periódico, não se restringindo aos artigos científicos. Enquanto anteriormente no suporte papel eram indexados somente artigos científicos. Os trabalhos coletados dentro da política da base para indexação estão: artigos científicos; comunicações; ensaio; ponto de vista; relato de caso, editorial e revisão de literatura.

A MB tem como diretriz o não descarte ou exclusão dos documentos coletados automaticamente, seja qual for a seção, ficando estes em uma posição estacionária na base para futura utilização ou quando na alteração da política de disponibilização de conteúdo da Brapci.

Em dezembro do ano de 2010, a Brapci reunia 23 publicações periódicas vigentes e dez históricas produzidas no Brasil, concentrando 6.632 artigos em CI. O último periódico incorporado a base foi a Revista InCID da Universidade de São Paulo em outubro de 2010.

No protótipo, o primeiro módulo a ser concretizado foi o de manutenção (MM), o que possibilitou as primeiras visualizações sistemáticas e primeiras avaliações dos processos dentro da base. Este módulo foi o que demandou mais

trabalho, por ser o que gerar a estruturação do sistema. Na sequência foram produzidos os módulos público e pesquisador.

A explicitação das informações indexadas é realizada pelo Módulo Público, por questões operacionais e de atualização. Como contraponto aos procedimentos burocráticos e operacionais, o Grupo E3PI optou por disponibilizar artigos “novos” coletados para acesso imediato ao público, mesmo sem passar pelo processo de revisão. Esta opção proporciona à Brapci uma perda relativa da qualidade na indexação, o que é compensado pela sua rápida atualização, chegando a casos de uma publicação estar disponível na base no mesmo dia de sua edição *online*. O tempo máximo de incorporação à Brapci é de sete dias para revistas que utilizam o protocolo OAI-PMH.

Com o protótipo em uso, o processo de validação foi necessário para sua consolidação e efetivação. Essa validação para a área de CC é denominada de “Homologação” e nessa fase os usuários realizam o uso do sistema com dados muito próximos do real, efetuando operações que deveriam ser utilizadas no trabalho rotineiro em diversos processos.

4.1 IMPLEMENTAÇÃO DO PROTÓTIPO

Com a proposta da metodologia para criação de uma base de dados temática, partiu-se para a concretização do produto final que viria a ser implementado e testado pelos usuários. Em sua primeira fase, os usuários restringiram-se aos integrantes do grupo de pesquisa e aos indexadores, com a incorporação e revisão dos trabalhos nela armazenados.

Esta primeira fase deu-se pelo módulo de manutenção, ambientado já na Internet, possibilitando múltiplos acessos simultâneos e disponíveis no endereço eletrônico (<http://www.brapci.ufpr.br/mm>).

4.1.1 Módulo de manutenção

O módulo de manutenção (MM) é o ambiente restrito do sistema, no qual os gestores e indexadores executam as atividades de organização das informações, revisão dos trabalhos e disponibilização final para os usuários.

Por ser de acesso restrito, uma das funcionalidades desenvolvidas pelo módulo de manutenção é a identificação do usuário por uma tela de *login* visualizada na Figura 22, a qual possibilita somente a pessoas autorizadas utilizarem o Módulo de Manutenção.



FIGURA 22 - LOGIN DO MÓDULO MANUTENÇÃO
Fonte: O autor (2010)

Devidamente identificado, o usuário assume o perfil que a ele corresponde, definido pelos gestores do sistema, apresentando em sua primeira tela após a identificação, um resumo das atividades (Figura 27) que o perfil deve realizar, bem como um quadro de informações com novidades sobre a base. Usuários novos podem ser cadastrados pelos gestores, ressaltando-se que este cadastro do usuário pode ser realizado de forma completa, com informações úteis a diversos processos dentro da vida acadêmica, como data de nascimento, e-mail, telefone, CPF, RG, código de matrícula, entre outros. Estes dados não são obrigatórios para efetuar o cadastro, mas são recomendados, principalmente para que possam ser compartilhadas informações com o Grupo E3PI, e caso haja necessidade de

localização ou preenchimento de fichas acadêmicas, encontram-se disponíveis para serem consultados.

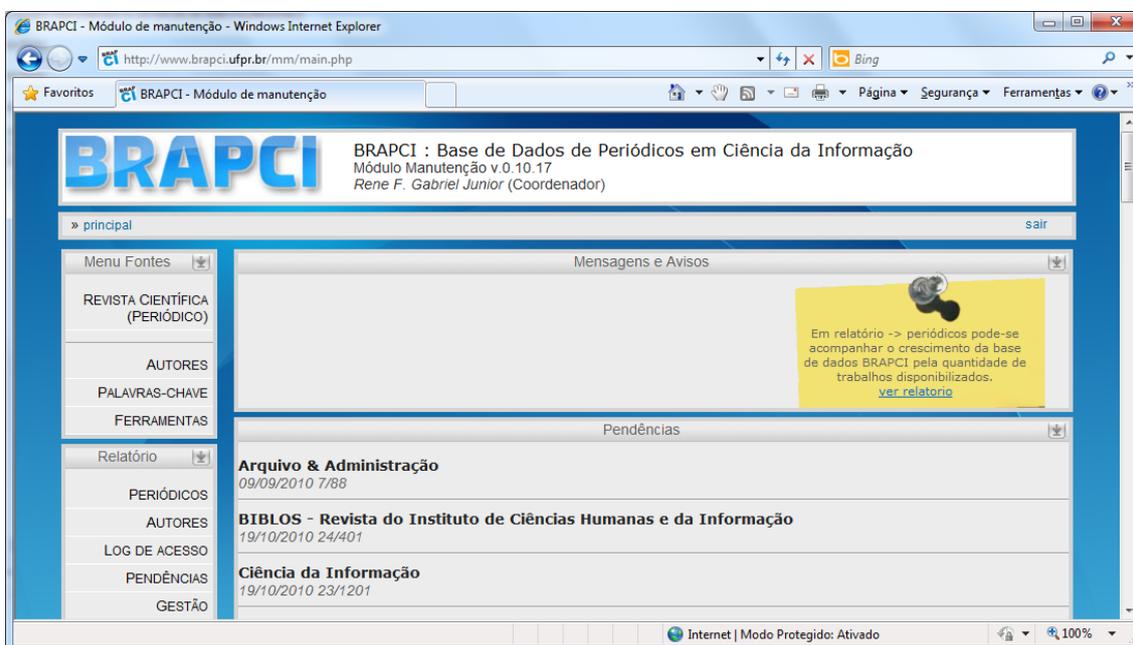


FIGURA 23 - TELA INICIAL DO MÓDULO DE MANUTENÇÃO
Fonte: O autor (2010)

O desenvolvimento do protótipo pelo programador *Web* seguiu as definições estipuladas em seu planejamento, como demonstrado na Figura 23, a tela real com incorporação de cores e fundo que seguiu o modelo de seu *wire frame*.

Por se tratar de uma metodologia extensível a diversos tipos de informação, no *menu* principal abriu-se uma opção para cadastro dos diferentes tipos de fontes de informações que serão armazenadas na base de dados. As fontes cadastradas são apresentadas no “Menu Fontes”. No caso da Brapci, está disponível somente a categoria “Revistas Científicas” por delimitação do grupo, mas podem ser incorporados outros tipos de fontes como teses e dissertações, livros, capítulos de livros, eventos ou qualquer outra fonte que seja necessário indexar.

A Figura 24 apresenta ao indexador o resumo (visual) das atividades de trabalho a serem realizadas para preparar as edições. A ferramenta representa o fascículo a ser processado, e os *checks*¹⁰ em verde, indicam que já foram revisados e concluídos.

¹⁰ Representação gráfica realizada por ícones que representam ações finalizadas.



FIGURA 24 - TELA DE CONTROLE DE EDIÇÃO DA REVISTA
Fonte: O autor (2010)

Para cada nova edição coletada pelo sistema, uma nova indicação de trabalho é reportada aos usuários indexadores e gestores. A identificação visual facilita a localização dos trabalhos a serem indexados. Quando selecionada uma das edições, clicando-se em seu número, o sistema apresenta todos os trabalhos intelectuais que a compõem, como representado na Figura 25.

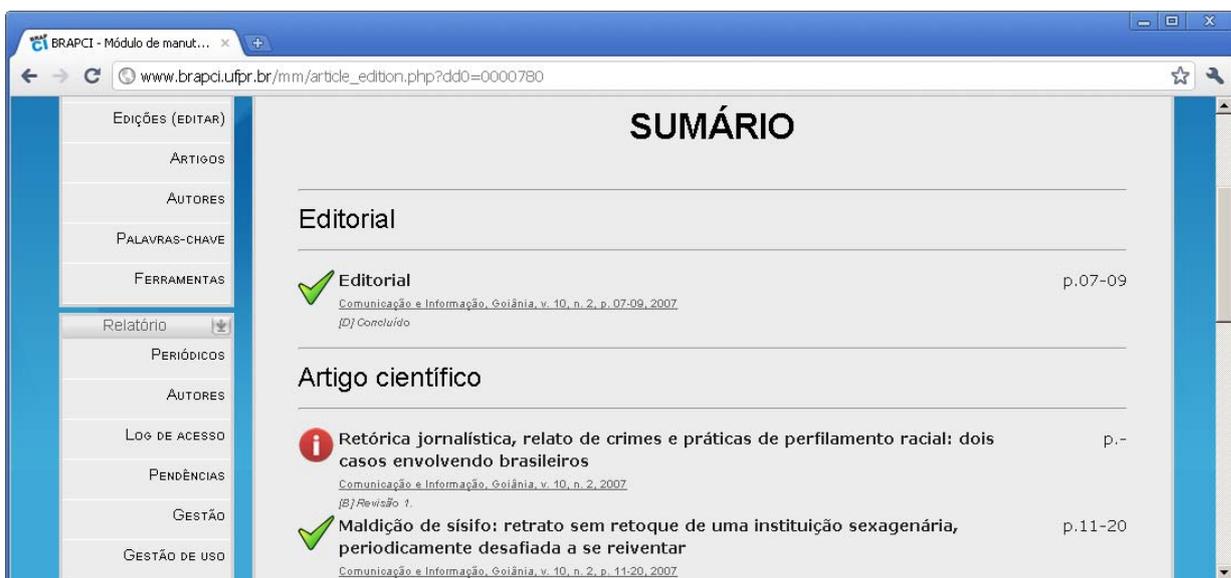


FIGURA 25 - TELA DE CONTROLE DE EDIÇÃO DA REVISTA COM APRESENTAÇÃO DOS TRABALHOS
Fonte: O autor (2010)

A indicação gráfica que mostra ao indexador quais trabalhos precisam ser indexados e ou revisados pode ser acessada clicando-se no título de um dos trabalhos. Esta ação detalha todos os dados do trabalho a ser indexado/revisado.

As Figuras 26, 27 e 28 apresentam o detalhamento, bem como as informações que devem ser tratadas pelo indexador.

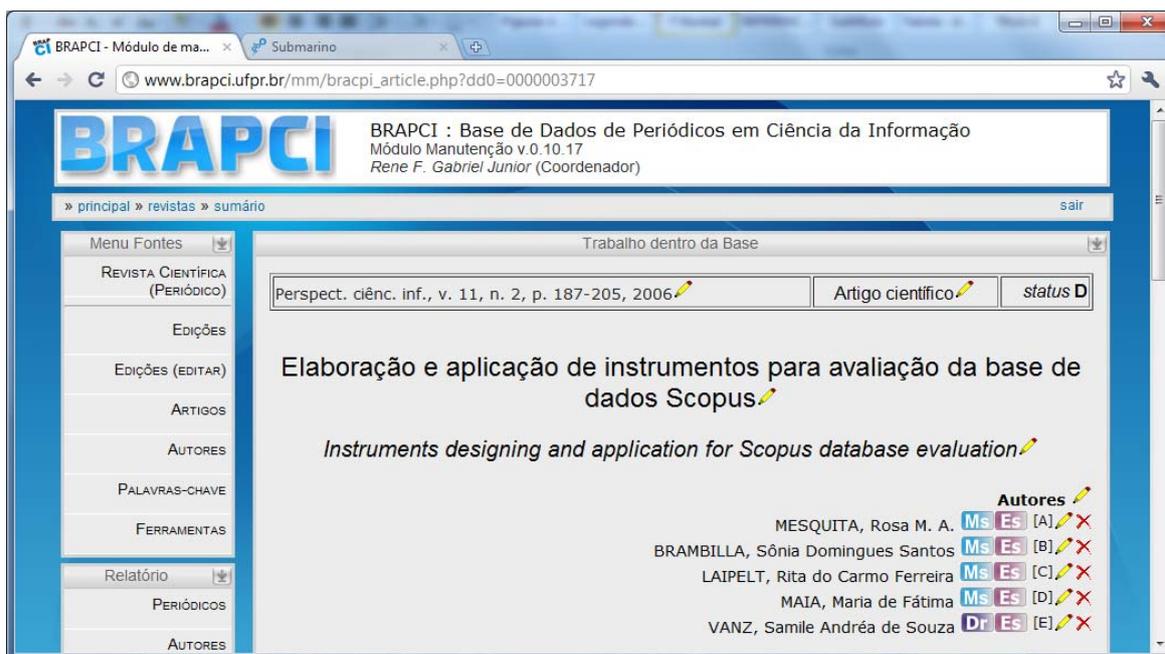


FIGURA 26 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE I
Fonte: O autor (2010)

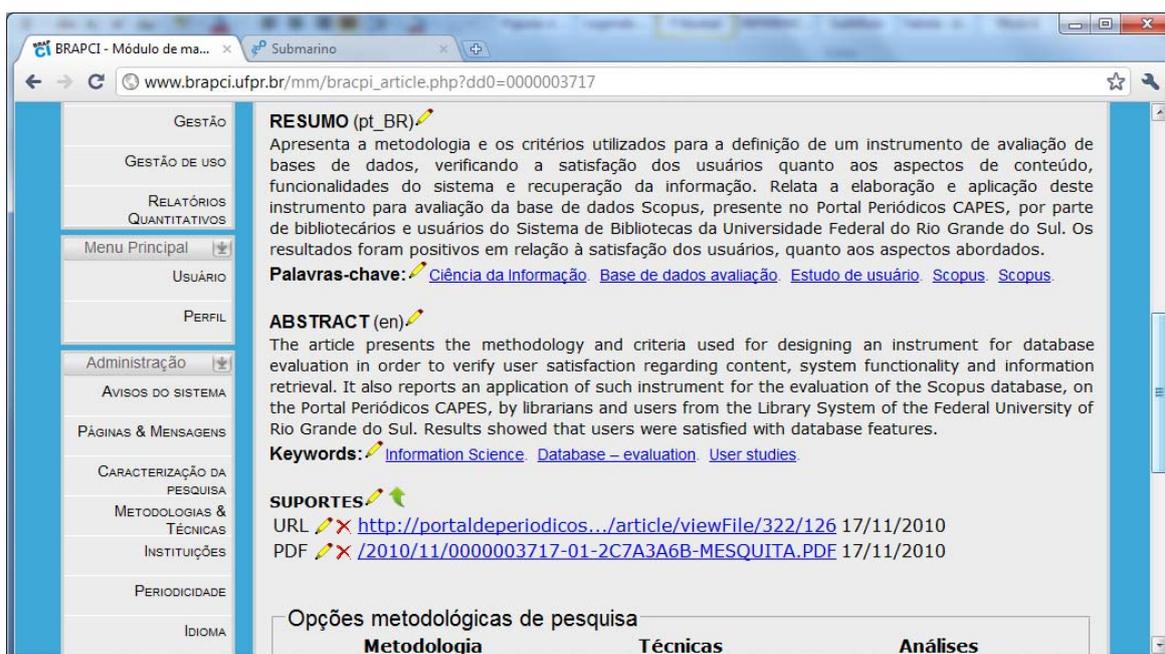


FIGURA 27 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE II
Fonte: O autor (2010)

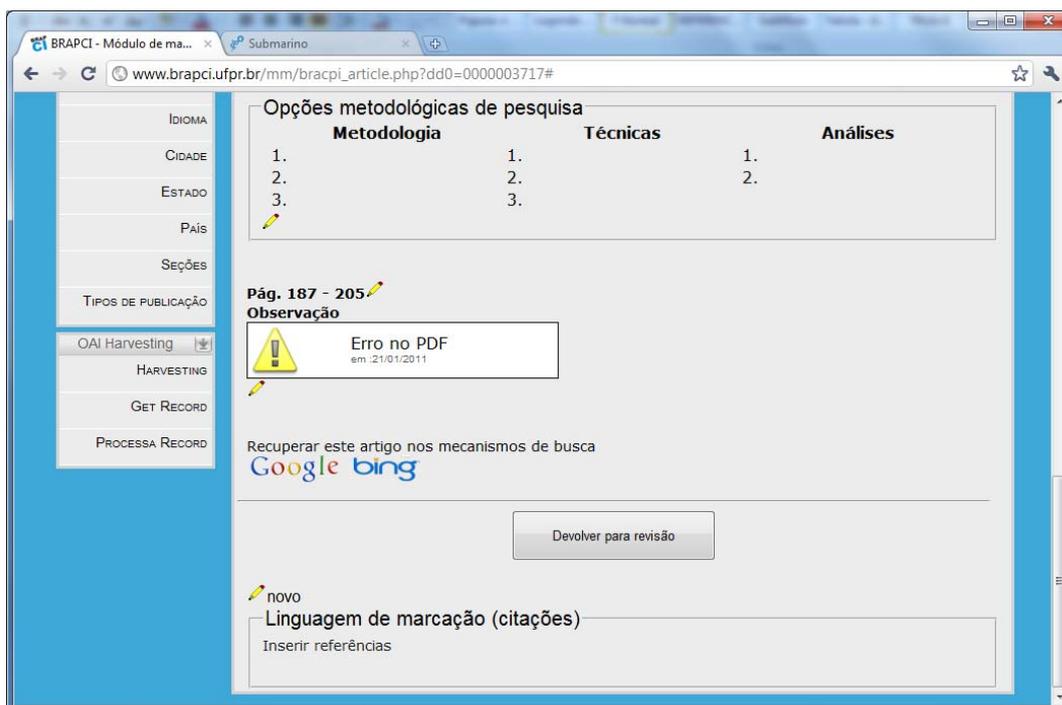


FIGURA 28 - TELA DE DETALHAMENTO DE UM TRABALHO PARA INDEXAÇÃO – PARTE III
Fonte: O autor (2010)

Entre as informações importantes para o indexador sobre o artigo estão:

- a) a edição a qual está vinculado;
- b) a página inicial e final (em edições eletrônicas é considerada a página zero);
- c) o título original e título alternativo do trabalho;
- d) o nome dos autores, titulação, vínculo institucional e forma de contato com os autores;
- e) o resumo e o *abstract*;
- f) as palavras-chave e *keywords*;
- g) o suporte ou *link* do trabalho completo (optando-se pelo PDF a ser incorporado à base);
- h) uma tabela classificatória para identificar a metodologia, abordagem e técnica aplicada à pesquisa;
- i) campo de observação, para preenchimento quando o trabalhos apresentam algum problema que impeça sua indexação;
- j) os botões de ações para prosseguir com o processamento dos textos.

No processo de indexação, existem ainda outras telas que auxiliam o processo de cadastro dos registros. As duas telas são apresentadas na Figura 29 como cadastro de autor e na Figura 30 como cadastro de palavras-chave.

www.brapci.ufpr.br/mm/brapci_autores_ed.php?dd1=000008992&dd2=&dd3=&dd10= - Google Chrome

www.brapci.ufpr.br/mm/brapci_autores_ed.php?dd1=000008992&dd2=&dd3=&dd10=

Nome do autor (completo)

Instituição Universidade Federal do Paraná(UFPR)

Qualificação

e-mail

posição 1

Professor (Docente)

Estudante (Discente)

Doutor (Doutorando, marcar também opção estudante)

Mestre (Mestrando, marcar também opção estudante)

Programa Strict Sensus

Profissional

Gravar >>

Dados complementares

Telefone

Endereço

FIGURA 29 - TELA DE CADASTRO DE AUTOR

Fonte: O autor (2010)

No cadastro de autor, buscou-se sistematizar os dados de forma a solicitar seu vínculo institucional, sua atuação (professor, estudante), sua titulação (Doutor, Mestre ou Profissional) e se está ligado a programas *Stricto-Sensu*. Outros dados complementares podem ser fornecidos, mas de forma não obrigatória. A identificação detalhada dos autores possibilita uma análise minuciosa da colaboração entre autores e entre instituições, provendo subsídios para estudos de rede e colaboração científica.

A inserção das palavras-chave do artigo é realizada nesta entrada, representada pela Figura 30. Uma funcionalidade adicionada à metodologia foi recomendada pelos indexadores que perceberam a necessidade de inserir todos os termos de uma única vez. Esta funcionalidade foi adicionada ao sistema, possibilitando que um algoritmo realizasse a separação dos termos (quando separados por ponto “.”, vírgula “,” ou ponto e vírgula “;”), comparando-os com os já existentes na tabela de termos, inserindo os não localizados e incorporando este ao trabalho que está sendo indexado.

Os novos termos incorporados posteriormente são analisados e relacionados com outros quando semelhantes, categorizados conforme seu tipo (termo, data, onomástico ou autoridade) para construção dos vocabulários controlados.

FIGURA 30 - TELA DE CADASTRO DE PALAVRAS-CHAVE
Fonte: O autor (2010)

O módulo de manutenção possibilita a gestão de informação e dos fluxos das atividades a serem realizadas e executadas pelos operadores do sistema. A tela inicial do sistema apresenta um resumo analítico da situação atual do processo de indexação (Figura 31).

Titulo	indexação	192ª revisão	revisado	total	atualizado		
<i>Arquivistica.net</i>	0	0	70	70	09/09/2010		
Arquivo & Administração	51	58.0%	7	8.0%	30	88	09/09/2010
<i>Biblionline</i>	0	0	108	108	21/10/2010		
BIBLOS - Revista do Instituto de Ciências Humanas e da Informação	230	57.4%	8	2.0%	163	401	19/10/2010
<i>Brazilian Journal of Information Science</i>	0	0	32	32	30/10/2010		
Cadernos de Biblioteconomia	84	100.0%	0	0	84	19/10/2010	
<i>Ciência da Informação</i>	0	0	1201	1201	19/10/2010		
Comunicação & Informação	89	43.4%	43	21.0%	73	205	19/10/2010
DataGramaZero	0	11	3.4%	313	324	19/10/2010	
Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS	19	8.8%	8	3.7%	189	216	20/12/2010
Encontro Nacional de Pesquisa em Ciência da Informação	0	4	100.0%	0	4	19/10/2010	
Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação	17	4.8%	13	3.6%	327	357	19/10/2010
Estudos Avançados em Biblioteconomia e Ciência da Informação	25	92.6%	2	7.4%	0	27	19/10/2010

FIGURA 31 - RESUMO DO PROCESSO DE INDEXAÇÃO
Fonte: O autor (2010)

O módulo de manutenção é o que apresenta a maior complexidade e envolve diversos relatórios e ferramentas para gestão. Este trabalho não visa descrever todas as atividades e possibilidades do sistema, que estão descritas em seu manual de operacionalização, mas apenas relatar as possibilidades empíricas do sistema.

Dentre algumas ferramentas que se encontram disponíveis no sistema estão: coleta de dados por meio do OAI-PMH; exportação de dados para o módulo público; geração das nuvens de *tags*; criação de ontologias e tesauros e processamento da linguagem de marcação empregada na análise de citação, funcionalidades estas representadas na Figura 32.



FIGURA 32 - FERRAMENTAS DO SISTEMA
Fonte: O autor (2010)

4.1.2 Módulo Público

O módulo público é a interface que permite ao usuário (visitante) a consulta à base. A primeira disponibilização em abril de 2008 de um módulo público com um mecanismo de busca simplificado permitiu à comunidade acadêmica a consulta à Brapci. Nesta fase o processo de publicação de novas versões de conteúdo era realizado manualmente, convertendo os dados de uma plataforma (*ProCite5*) para outra (*MySQL*), problema que foi solucionado no novo sistema.

A versão definida nesta metodologia foi disponibilizada ao público em meados do mês de março de 2010 (Figura 29), apresentando-se mais intuitiva e

possibilitando a recuperação de informação com a apresentação direta do título do trabalho, autores, publicação, palavras-chave e resumo. O uso dos operadores booleanos já está contemplado nesta versão. A interface com o usuário é representada na Figura 33.



FIGURA 33 - FERRAMENTAS DO SISTEMA

Fonte: O autor (2010)

Nesta mesma tela, o pesquisador pode delimitar o período de abrangência da pesquisa por um recorte específico, de acordo com suas necessidades. Ainda na tela inicial é apresentada uma nuvem de *tags* com os termos mais consultados pelos usuários desde a disponibilização do módulo público.

Como forma de mensuração de uso, todos os acessos e consultas realizados na Brapci são armazenados para posterior análise dos gestores do sistema, possibilitando um resgate histórico de seu uso e ainda os registros das temáticas mais pesquisadas em determinado período.

O resultado apresentado ao usuário pesquisador em uma busca, por definição do Grupo E3PI, deveria ser o mais abrangente. Desta forma foi definido que cada registro recuperado da base deveria conter o título do trabalho, o responsável pela publicação, o volume, a edição e número, o nome dos autores, o resumo e as palavras-chave (Figura 34).

[Sistema hipermídia ajudando a construir a pesquisa escolar](#) . *Transinformação* v. 15, n. 2, maio/ago. 2003
 (ARTIG)
 MACHADO, Ana Maria Nogueira; VIDOTTI, Silvana Aparecida Borsetti Gregório

Este trabalho surgiu da constatação das condições em que alunos de 1a. a 4a. séries do 1o. grau, hoje ensino fundamental, vinham à biblioteca para elaborar a chamada pesquisa escolar. O objetivo deste artigo é discutir o desenvolvimento do sistema hipermídia 'Ajudando a construir a pesquisa escolar', com enfoque na elaboração de referências, de acordo com a NBR 6023/2000 da Associação Brasileira de Normas Técnicas. O sistema foi construído no software Personal Brain TM que possui características estruturais em forma de rede. Estas possibilitam a criação de um ambiente hipermídia informacional, pessoal ou coletivo, numa relação natural entre as informações armazenadas em documentos e os temas abordados pelos usuários. Acredita-se que o sistema aqui discutido, pode ser utilizado como um instrumento motivador da pesquisa escolar.

Palavras-chave: pesquisa escolar. sistemas hipermídias educacionais. nbr 6023 - referências. motivação. personal brain tm

[link externo](#) | [ABNT](#) | [VANCOUVER](#)
 Método de

FIGURA 34 - RESULTADO DA BUSCA

Fonte: O autor (2010)

O uso da base para pesquisas acadêmicas e profissionais pode ser percebido nas solicitações dos usuários por e-mail, principalmente na solicitação de trabalhos que não estavam disponíveis online em texto completo para *download*. O uso acadêmico e os resultados de diversas fontes demandaram a necessidade de copiar a referência do trabalho diretamente no site. Em uma primeira fase, optou-se pela formatação das normas ABNT e Vancouver por serem as mais utilizadas nas ciências humanas e sociais aplicadas.

Pela quantidade de artigos disponibilizados em determinada temática, houve a necessidade da criação de uma metodologia que demonstrasse ao pesquisador a relevância de sua consulta aos artigos recuperados. Desta forma, um estudo proposto por Freitas, Bufrem e Gabriel Jr (2010) resultou na metodologia de atribuição de pontos, conforme a incidência dos termos correspondentes às buscas encontrados nos campos de acesso ao acervo indexado.

Para validar esta metodologia na recuperação de informação, Freitas (2010) realizou uma pesquisa empírica, utilizando-se de uma análise de conteúdo. A autora selecionou a temática da Epistemologia da Ciência da Informação e realizou diversas consultas na base de dados empregando os termos de busca "Epistemologia" e "Gnosiologia", com a delimitação temática na CI.

Para a análise foram selecionados os artigos que obtiveram entre seis e sete pontos. A autora identificou que aproximadamente 80% desses artigos recuperados tinham relação direta com o assunto. O desvio de 20% foi ocasionado pela influência de trabalhos da área de Educação que provocaram ruído na temática específica. Para validar a metodologia, Freitas (2010) ainda procurou analisar os outros

trabalhos que foram recuperados, porém com pontuação inferior à do recorte. Nesta análise, a autora observou que existiam alguns trabalhos importantes para a temática com pontuação baixa e, realizando uma análise mais criteriosa, observou que em seu título ou palavras-chave recorriam a figuras de linguagem, utilizando termos imprecisos para representação do trabalho.

Com a recuperação dos artigos de Freitas, Bufrem e Gabriel Jr (2010) montou-se um quadro que representa os diversos descritores identificados na base.

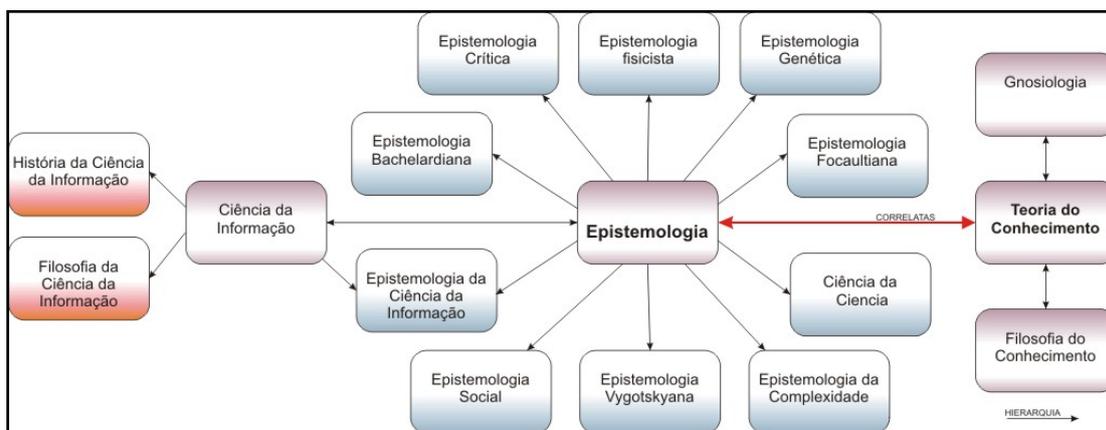


FIGURA 35 - RELAÇÃO DE CONCEITOS DA EPISTEMOLOGIA EM CIÊNCIA DA INFORMAÇÃO
 Fonte: Freitas; Bufrem e Gabriel Jr (2010)

A metodologia de atribuição de pontos demonstrou-se importante para identificar trabalhos sobre a temática, principalmente para pesquisadores que ainda estão em estado anômalo do conhecimento.

4.1.3 Módulo Pesquisador

O módulo pesquisador é o que apresenta informações compiladas aos usuários, conforme sua seleção de trabalhos. Esta funcionalidade do sistema, até o presente momento está sendo testada pelo Grupo E3PI, apresentando os primeiros resultados em análises de citação e bibliometria.

4.2 VOCABULÁRIO CONTROLADO NA BRAPCI

Como proposta da MB, o controle do vocabulário parte da linguagem natural empregada pelos autores na identificação das palavras-chave de seus trabalhos. Estas palavras são armazenadas e avaliadas pelos gestores do sistema, e sendo caracterizada conforme a tipologia estabelecida (termo indexador, data, autoridade ou não usado).

Em estudo preliminar dos termos utilizados pelas revistas científicas para representar os assuntos por meio de suas palavras-chave, percebeu-se que eles não representam somente os assuntos, mas descrevem outras categorias. Os termos utilizados pelos autores não apresentam uma normalização, mas têm sido indexados, pois decorrem do que o autor acredita ser representativo de seu trabalho.

Neste estudo preliminar, utilizando-se recursos computacionais e organizando-se uma planilha de dados, foi possível identificar 16.256 termos para descrever os assuntos de 6.689 trabalhos. Numa classificação por idiomas foram identificados 9.965 (61,3%) termos em português, 6.012 (37,0%) em inglês e 279 (1,7%) em espanhol. Os problemas mais comuns dos termos referem-se a erros de padronização e à não utilização de instrumentos de padronização em suas fontes originais (revistas) como glossários ou tesouros.

Partindo-se dos termos representativos coletados dos trabalhos dos autores, pode-se criar instrumentos que possibilitem o controle de vocabulário para indexação. Entre esses instrumentos, destacam-se o glossário e a geração de índices de palavras-chave. Outro instrumento aplicado usa as teorias de Cabré (1993) com a concepção de conceitos, os quais são representados por termos, que a qualquer momento podem ser atualizados ou substituídos, permanecendo, entretanto, as relações entre os conceitos.

Em um estudo piloto vem sendo desenvolvida a construção de um tesouro baseado em ontologias e protocolo OWL para compartilhamento deste instrumento. Neste estudo já foram identificados 350 conceitos e geradas mais de 150 remissivas. O que é interessante observar é que este instrumento pode ter diversos termos, em diversos idiomas para representar um conceito, e que existe em cada idioma um termo preferencial, mas no processo de busca ou indexação automática, pode empregar qualquer um deles, e todos remetem ao mesmo conceito.

The screenshot displays the BRAPCI web application interface. On the left is a vertical navigation menu with categories like 'Menu Fontes', 'Relatório', 'PERIÓDICOS', 'AUTORES', 'LOG DE ACESSO', 'PENDÊNCIAS', 'GESTÃO', 'GESTÃO DE USO', 'RELATÓRIOS QUANTITATIVOS', 'Menu Principal', 'USUÁRIO', 'PERFIL', 'Administração', 'AVISOS DO SISTEMA', 'PÁGINAS & MENSAGENS', 'CARACTERIZAÇÃO DA PESQUISA', 'METODOLÓGICAS & TÉCNICAS', and 'INSTITUIÇÕES'. The main content area is titled 'Relatórios Quantitativos da Base BRAPCI' and contains a table with columns 'Tipo', 'Idioma', and 'Valor'. Below the table, there are sections for 'CONCEITOS GERAIS' and 'CONCEITOS ESPECÍFICOS'. A diagram illustrates the relationship between 'Biblioteca' and 'Unidades de informação', with 'Biblioteca' acting as a central node connected to 'Unidades de informação' and a list of specific library types: 'Biblioteca escolar', 'Biblioteca Especial', 'Biblioteca Especializada', 'Biblioteca Nacional', 'Biblioteca universitária', and 'Biblioteca virtual'. The diagram also includes a 'Termos alternativos' node.

FIGURA 36 - INSTRUMENTO DE RELACIONAMENTO DE CONCEITOS NA BRAPCI
Fonte: O autor (2010)

Outro ponto importante a ressaltar é que o modelo de relacionamento entre os conceitos é construído pela representação mental de quem o está elaborando e pelo seu conhecimento da área. Desta forma, a MB optou em construir a relação dos conceitos em temas específicos (*thema*). Desta forma os conceitos permanecem, alterando-se apenas as relações que estes têm conforme a temática (*thema*) que se está trabalhando, ou seja, dois conceitos podem ter relações diferentes para a área de CI, por exemplo, quando na visão da administração e na visão da organização da informação. Esta possibilidade cria duas ontologias diferentes em uma mesma área.



FIGURA 37 - TESAURO BRAPCI, COM ONTOLOGIA NAVEGACIONAL
Fonte: O autor (2010)

4.3 DADOS BIBLIOMÉTRICOS E CIENTOMÉTRICOS

A reunião de dados é um desafio que o pesquisador por vezes enfrenta para localizar itens pertinentes a fim de subsidiar suas necessidades de informações para resolução de problemas. Esse desafio vem sendo compensado, segundo Santos e Kobashi (2010), com a criação de métodos e técnicas de tratamento, análise e visualização de informação, baseado principalmente em princípios estatísticos e/ou lingüísticos.

Dentre esses princípios, a MB possibilita a geração de gráficos para facilitar a análise do pesquisador; Os Gráficos 1 e 2 demonstram o crescimento da base. Observa-se que com a implementação da coleta automática, em dois meses, o acervo cresceu de 4.139 para 5.716 trabalhos, representando um acréscimo de 43,4% no número total de artigos.

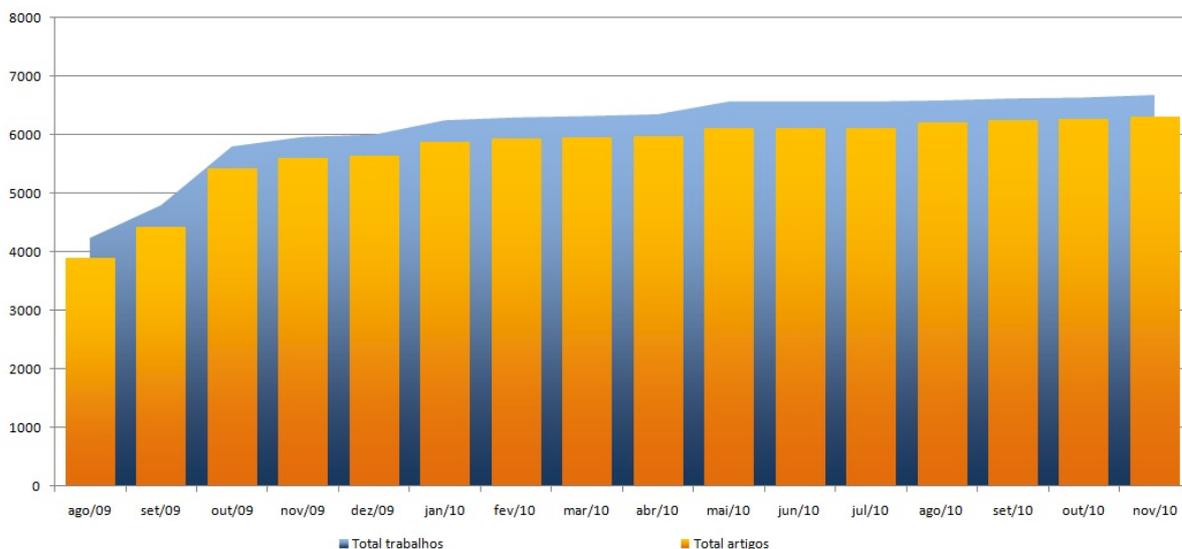


GRÁFICO 1 - CRESCIMENTO DA BASE EM NÚMERO DE TRABALHOS E ARTIGOS POR MÊS
Fonte: O autor (2010)

A Tabela 1 apresenta de forma quantitativa e diferenciada os trabalhos e os artigos indexados na base de dados após a efetivação do módulo de manutenção, observando-se que a diferença dos trabalhos para os artigos, refere-se à publicação de trabalhos que não são artigos, como apresentações, expedientes, teses e dissertações, trabalhos de conclusão de cursos, entrevistas, editoriais entre outros.

TABELA 1 - TOTAL DE TRABALHOS / ARTIGOS NA BRAPCI

MÊS	TOTAL TRABALHOS	TOTAL ARTIGOS
Agosto / 2009	4297	4139
Setembro / 2009	5244	4658
Outubro / 2009	5716	5355
Novembro / 2009	5753	5392
Dezembro / 2009	5986	5622
Janeiro / 2010	6057	5693
Fevereiro / 2010	6064	5705
Março / 2010	6088	5724
Abril / 2010	6296	5932
Maio / 2010	6360	5996
Junho / 2010	6369	6005
Julho / 2010	6395	6031
Agosto / 2010	6439	6075
Setembro / 2010	6484	6120
Outubro / 2010	6634	6270
Novembro / 2010	6677	6313
TOTAL ATUAL	6677	6313

Fonte: O autor (2010)

Em análise do que foi apresentado, percebe-se que na área da CI o pico de produção de artigos científicos no ano de 2008 foi de 526 trabalhos publicados, visualizados no Gráfico 3. Este gráfico representa o crescimento da Ciência da

Informação do Brasil desde 1972. Em 1995, observa-se um novo ciclo de produção científica na área, seguido por uma estabilização e novos picos de produção entre 1998 e 2000. Em 2006, as Revistas em CI tiveram um novo salto e se mantiveram estáveis. A queda no gráfico referente a 2009 é explicada, principalmente por alguns periódicos científicos não terem sua periodicidade efetivada, com fascículos atrasados neste ano. A tabela 2 apresenta o número de artigos do acervo da base desde 1972 até 2010.

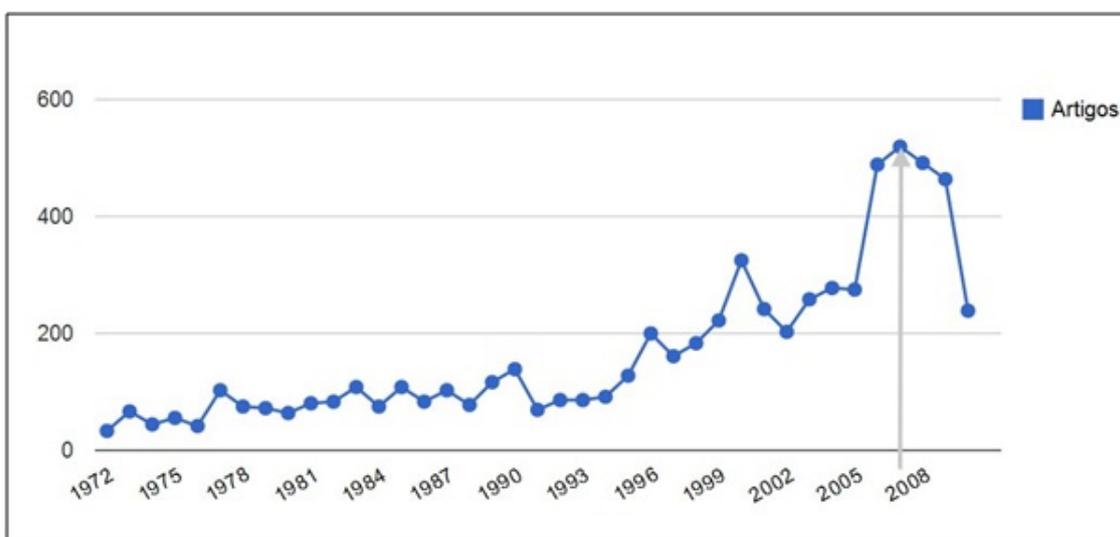


GRÁFICO 2 - PRODUÇÃO DE TRABALHOS POR ANO NA BRAPCI
Fonte: O autor (2010)

TABELA 2 - PRODUÇÃO DE ARTIGOS POR ANO NA BRAPCI

ANO	ART	ANO	ART	ANO	ART	ANO	ART
1972	34	1982	82	1992	85	2002	203
1973	66	1983	107	1993	86	2003	258
1974	44	1984	76	1994	91	2004	277
1975	55	1985	109	1995	128	2005	275
1976	43	1986	84	1996	199	2006	490
1977	104	1987	103	1997	162	2007	520
1978	76	1988	78	1998	184	2008	493
1979	72	1989	117	1999	221	2009	463
1980	63	1990	139	2000	324	2010	240
1981	81	1991	69	2001	241		
SUB-TOTAL	638		964		1721		3219
TOTAL	6542*						

Fonte: Autor, com base nos dados levantados na Brapci (2010)
NOTA: *total sem revisão

Desde sua primeira versão em ambiente monousuário, com treze periódicos, a base em 2010, incorporou até o mês de novembro de 2010 um total de 33 publicações. A relação das revistas disponíveis encontra-se no Quadro 12.

A base conta com títulos de revistas descontinuadas, algumas disponíveis somente no suporte papel (encontradas em bibliotecas brasileiras), as quais são identificadas no campo publicação do Quadro 12 como “histórica”. As publicações ativas estão identificadas como “corrente”, sendo possível identificar ainda, quais estão disponíveis no formato eletrônicos e impresso, ressaltando-se que algumas dessas revistas são disponibilizadas eletronicamente somente a partir de determinado ano de publicação.

(Continua)

Revista	ISSN	Periodicidade	Publicação	Disponib.
Arquivística.net	1808-4826	Semestral	Corrente	Eletrônico
Arquivo & Administração	0100-2244	Quadrimestral	Corrente	Impresso
Biblionline	1809-4775	Anual	Corrente	Impresso e Eletrônico
BIBLOS - Revista do Instituto de Ciências Humanas e da Informação	0102-4388	Semestral	Corrente	Impresso e Eletrônico
<i>Brazilian Journal of Information Science</i>	1981-1640	Semestral	Corrente	Eletrônico
Cadernos de Biblioteconomia	0102-6607	Indefinida	Histórica	Impresso
Ciência da Informação	0100-1965 1518-8353*	Quadrimestral	Corrente	Impresso e Eletrônico
Comunicação & Informação	1415-5842	Semestral	Corrente	Impresso
DataGramaZero	1517-3801	Bimestral	Corrente	Eletrônico
Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS	1807-8893 1808-5245*	Semestral	Corrente	Impresso e Eletrônico
Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação	1518-2924	Semestral	Corrente	Eletrônico
Estudos Avançados em Biblioteconomia e Ciência da Informação	0100-9869	Anual	Histórica	Impresso
ETD - Educação Temática Digital	1517-2539 1676-2592 a partir dez. 2005	Semestral	Corrente	Eletrônico
InCID: Revista de Ciência da Informação e Documentação	2178-2075	Semestral	Corrente	Eletrônico
Inclusão Social	1808-8678	Semestral	Corrente	Eletrônico
Infociência	1415-0018	Anual	Corrente	Impresso
Informação & Informação	1414-2139	Semestral	Corrente	Impresso e Eletrônico
Informação & Sociedade: Estudos	0104-0146 1809-4783*	Quadrimestral	Corrente	Eletrônico
Informare: Cadernos do Programa de Pós-Graduação em Ciência da Informação	0104-9461	Semestral	Histórica	
Liinc em revista	1808-3536	Semestral	Corrente	Eletrônico
Perspectivas em Ciência da Informação	1413-9936	Quadrimestral	Corrente	Impresso

(Conclusão)

Revista	ISSN	Periodicidade	Publicação	Disponib.
Ponto de Acesso	1981-6766	Quadrimestral	Corrente	Impresso e Eletrônico
Revista ACB: Biblioteconomia em Santa Catarina	1414-0594	Anual	Corrente	Impresso e Eletrônico
Revista Brasileira de Biblioteconomia e Documentação	0100-0691 1980-6949*	Semestral	Corrente	Impresso e Eletrônico
Revista da Escola de Biblioteconomia da UFMG	0100-0829	Semestral	Histórica. Mudou nome	Impresso
Revista de Biblioteconomia & Comunicação	0103-0361	Indefinida	Histórica. Mudou nome	Impresso
Revista de Biblioteconomia de Brasília	0100-7157	Semestral	Corrente	Impresso e Eletrônico
Revista Digital de Biblioteconomia & Ciência da Informação	1678-765X	Semestral	Corrente	Eletrônico
Revista do Departamento de Biblioteconomia e Ciência da Informação	0101-045X	Semestral	Histórica. Mudou nome	Impresso
Revista Eletrônica Informação e Cognição	1807-8281	Anual	Corrente	Eletrônico
Revista Online da Biblioteca Prof. Joel Martins	1517-3992	Quadrimestral	Histórica. Mudou nome	Eletrônico
Tendências da Pesquisa Brasileira em Ciência da Informação	1983-5116	Quadrimestral	Corrente	Eletrônico
Transinformação	0103-3786	Quadrimestral	Corrente	Impresso e Eletrônico

QUADRO 12 - LISTA DE PUBLICAÇÕES DISPONÍVEIS NA BRAPCI

Fonte: O autor, com base nos dados levantados na Brapci (2010)

Um dos indicadores da produção científica pode ser construído pelo monitoramento da quantidade de títulos de periódicos que se mantêm vigentes na área. Analisando o Gráfico 3, percebe-se que entre 1972 e 1993 houve uma estagnação no número de títulos de revistas, mantendo-se na média de seis, e somente a partir de 1994 houve um crescimento que ocorreu até 2000, chegando-se à cifra de 17 revistas. Porém, em 2001 e 2002 houve uma redução de 35% do total em títulos, caindo para 11 revistas. Somente em 2002 ocorreu um novo ciclo de crescimento, atingindo seu pico em 2007, com 22 revistas.

O grande número de revistas criadas a partir de 2005 se explica por dois fatores, sendo o primeiro a criação de novos programas em CI no Brasil e o outro a facilidade de criação das revistas eletrônicas, principalmente com incentivos do IBICT e da Capes na adoção da política de arquivos abertos.

O acompanhamento geral da quantidade de “títulos de revistas em atividade por ano” pode ser realizado observando-se o Gráfico 3.

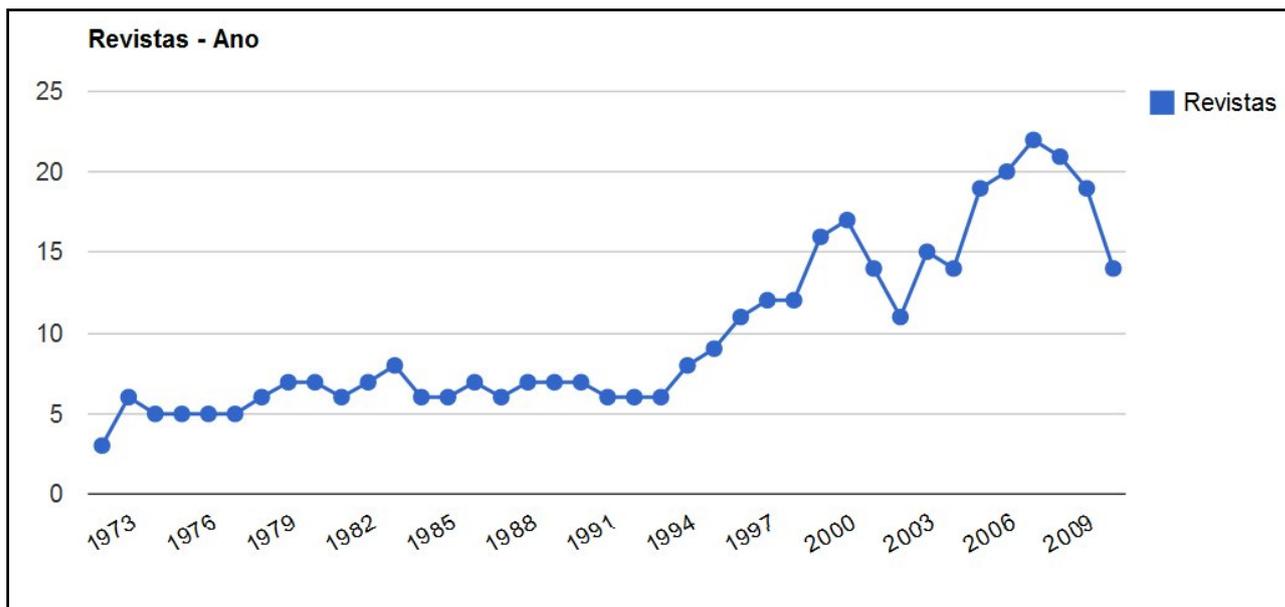


GRÁFICO 3 - TÍTULOS DE REVISTAS EM ATIVIDADE POR ANO

Fonte: Autor, com base nos dados levantados na Brapci (2010)

A Tabela 3 apresenta em números a produção da área, indicando a quantidade de revistas e de fascículos publicados por ano, entre 1972 e 2010

TABELA 3 - NÚMERO DE REVISTAS E FASCÍCULOS PUBLICADOS POR ANO

ANO	REV./ED.	ANO	REV./ED.	ANO	REV./ED.	ANO	REV./ED.
1972	3 / 5	1982	7 / 12	1992	6 / 10	2002	11 / 25
1973	6 / 13	1983	8 / 13	1993	6 / 9	2003	15 / 34
1974	5 / 13	1984	6 / 10	1994	8 / 12	2004	14 / 32
1975	5 / 13	1985	6 / 10	1995	9 / 14	2005	19 / 40
1976	5 / 11	1986	7 / 10	1996	11 / 21	2006	20 / 51
1977	5 / 13	1987	6 / 9	1997	12 / 20	2007	22 / 57
1978	6 / 12	1988	7 / 11	1998	12 / 22	2008	21 / 52
1979	7 / 14	1989	7 / 13	1999	16 / 27	2009	19 / 49
1980	7 / 12	1990	7 / 13	2000	17 / 38	2010	14 / 26
1981	6 / 11	1991	6 / 8	2001	14 / 30		
Total	388 Revistas com 795 fascículos, média de 2,05 edições por ano						

Fonte: Autor, com base nos dados levantados na Brapci (2010)

4.3.1 Análise de cooperação científica

Com resultado na formação do *corpus* da Brapci, observa-se que a área de CI tem uma prevalência da publicação de um único autor (63,9%) e que a publicação

em cooperação representa menos de 37% do total de artigos. Mas essa análise é pontual, representando a somatória de todos os artigos publicados desde 1972 (Tabela 4).

TABELA 4 - TIPOS DE AUTORIAS NA BRAPCI

AUTORES	QUANTIDADE	PERCENTUAL
Um	4178	63,9%
Dois	1452	22,2%
Três	546	8,3%
Quatro	204	3,1%
Cinco	81	1,2%
Seis	34	0,5%
Sete	20	0,3%
Oito	10	0,2
Nove ou mais	15	0,3
Total	6540	100%

Fonte: O autor (2010)

Analisando-se as informações da Tabela 4, observa-se o predomínio da publicação em autoria única mas, quando realizada esta mesma análise de forma diacrônica (Gráfico 4), observa-se que existe uma alteração na configuração da publicação autoral, percebendo um crescimento do número de cooperação a partir de 1990, e que vem numa curva ascendente, representando que em 2008 o número de artigos publicados em coautoria superou os de autoria única.

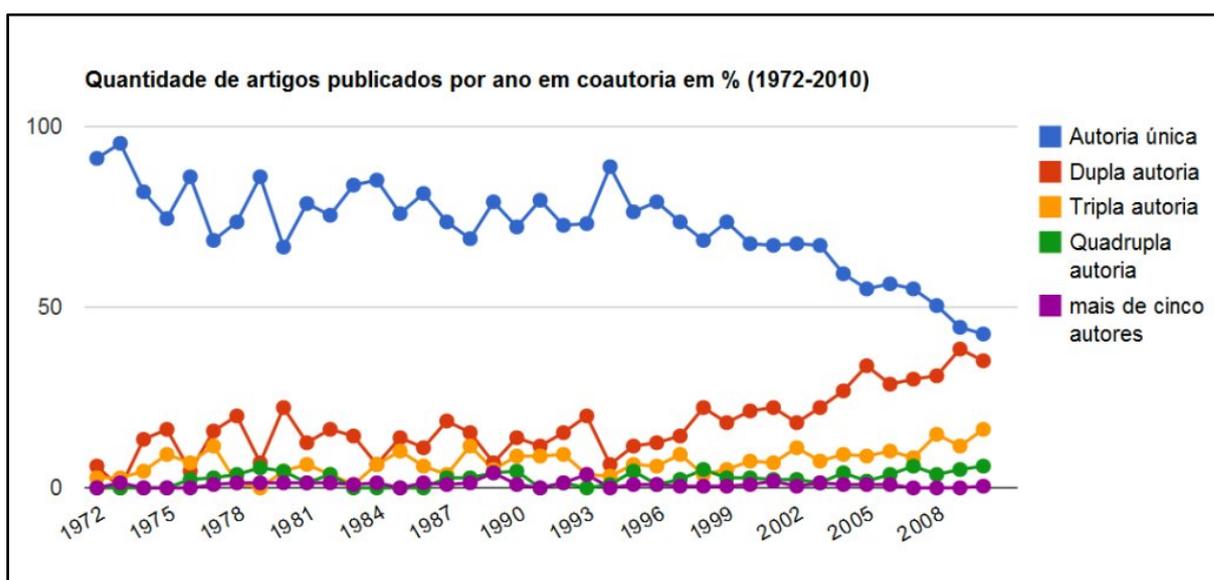


GRÁFICO 4 - QUANTIDADE ARTIGOS PUBLICADOS POR ANO EM COAUTORIA

Fonte: Autor, com base nos dados levantados na Brapci (2010)

4.3.2 Análise de conteúdo

O pesquisador no ME conta com instrumentos que lhe possibilitam análises de conteúdo. Em estudo realizado por Bufrem (2010), foi analisada a produção científica brasileira com a temática “redes sociais” entre 1997 e 2010. A autora localizou 54 trabalhos em que um dos descritores tinha o termo “redes sociais”. Com o *corpus* selecionado, foi possível identificar o período em que este assunto foi mais publicado (entre 2007 e 2008), e quando começou a surgir a temática (1997), representado no Gráfico 5.

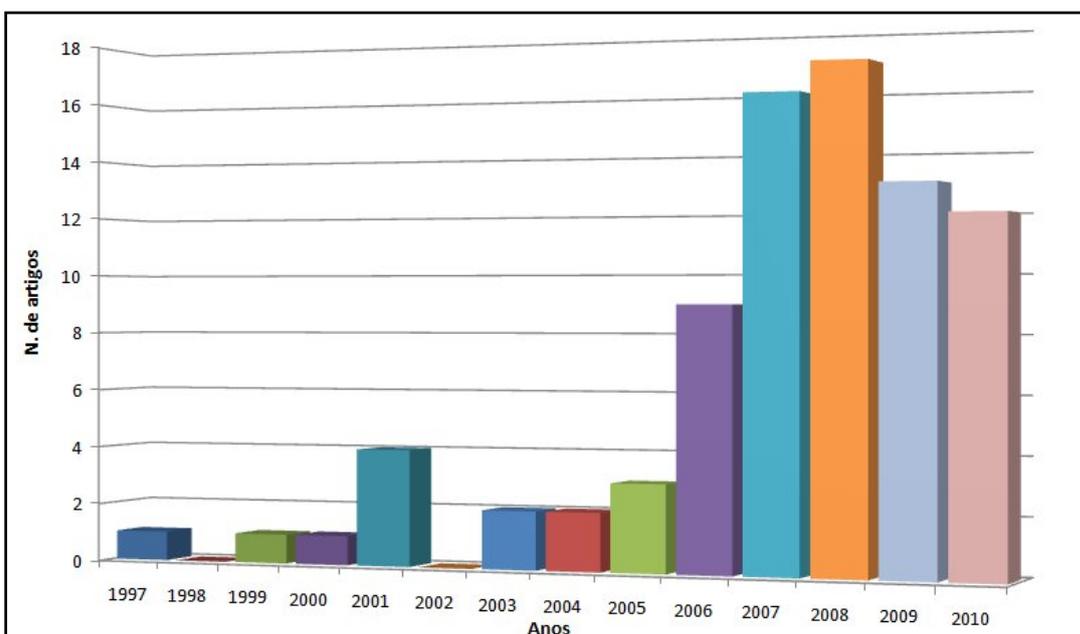


GRÁFICO 5 - INCIDÊNCIA DA TEMÁTICA “REDES SOCIAIS” NA LITERATURA PERIÓDICA EM CI
Fonte: BUFREM (2010)

Outra possibilidade de análise é a relação da temática com as revistas que publicaram sobre o assunto, demonstrado no Gráfico 6.

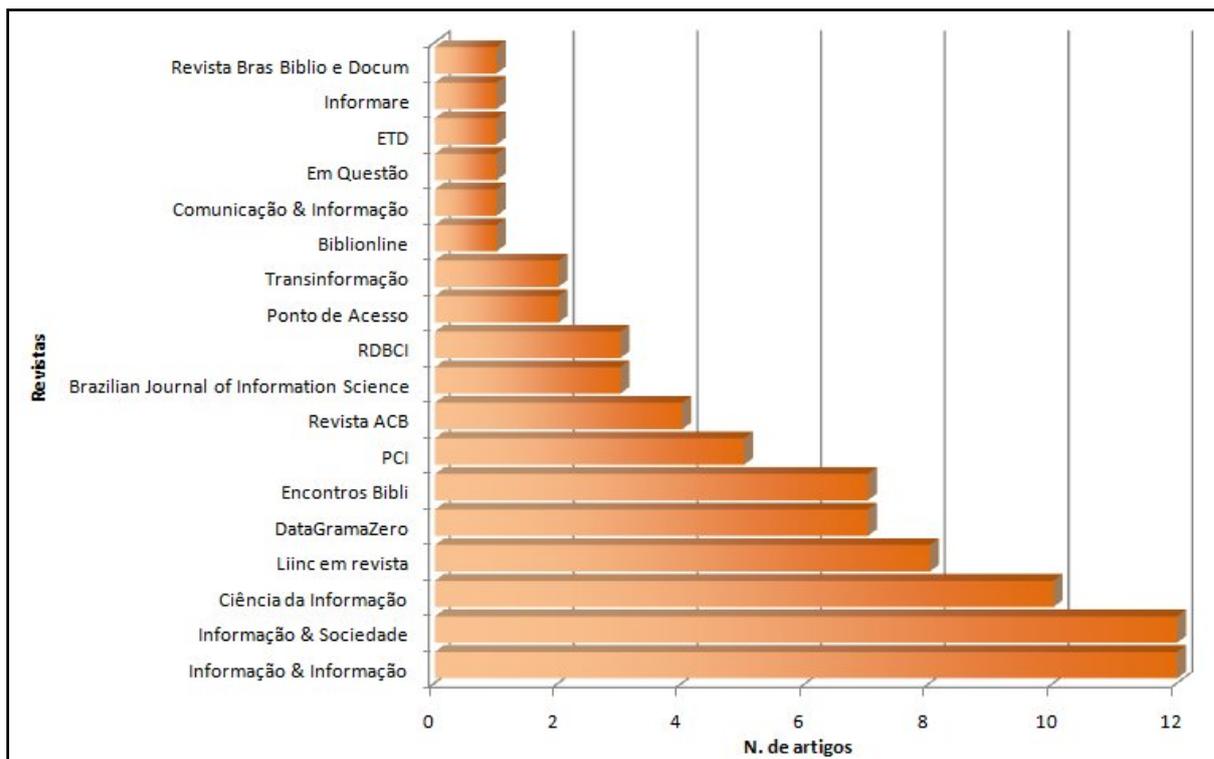


GRÁFICO 6 - INCIDÊNCIA DA TEMÁTICA “REDES SOCIAIS” NA LITERATURA PERIÓDICA EM CI POR PUBLICAÇÃO

Fonte: BUFREM (2010)

Este estudo não teve como objetivo a análise aprofundada dos resultados representados nos gráficos, mas ilustrar com exemplos algumas potencialidades da MB.

4.3.3 Análise de redes sociais

A metodologia propicia dentro de um *corpus* a identificação de autores com afinidade temática. Conforme expressa o gráfico 7, foi explorado o tema “redes sociais”, tendo sido possível a identificação de autores que publicaram sobre o tema, evidenciando-se as relações de cooperação.

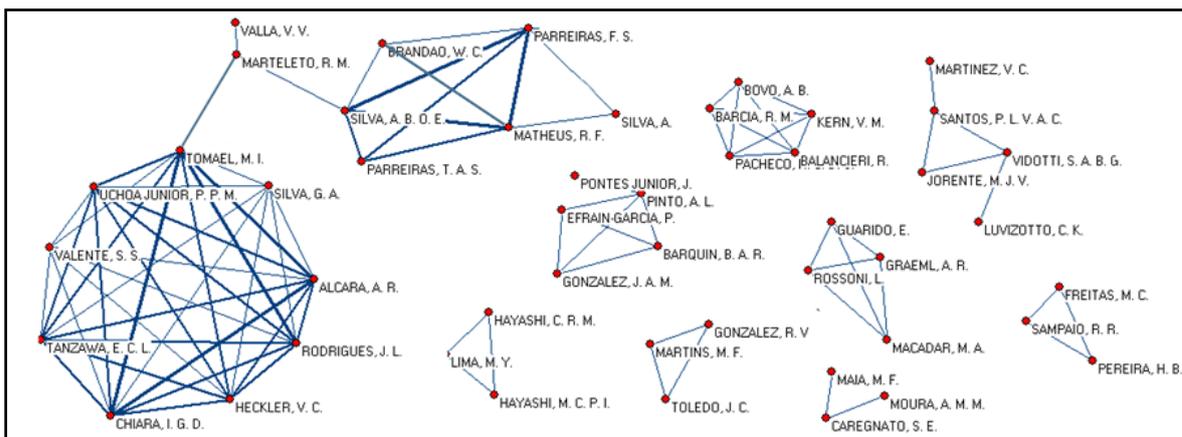


GRÁFICO 7 - REDE SOCIAL POR AFINIDADE TEMÁTICA

Fonte: BUFREM (2010)

Na representação do gráfico 7¹¹, de um estudo realizado na Brapci, pode-se observar que Marteleto foi a precursora de dois grupos de pesquisa sobre o tema, um deles formado predominantemente por pesquisadores da UFMG, destacando-se Silva, Parreiras e Matheus e o grupo da Universidade Estadual de Londrina (UEL), mais denso, liderado por Tomaél. Outros pequenos grupos são identificados, mas a maior concentração de trabalhos está relacionada a esses dois grupos. Esta análise tem como objetivo demonstrar a potencialidade das ferramentas de ARS disponíveis na Brapci.

4.4 AVALIAÇÃO E RESULTADOS

A avaliação da metodologia empregada no desenvolvimento da Brapci é constante, sendo realizada pelo grupo de estudo E3PI da UFPR, levantando problemas e com realização de estudos buscando identificar suas causas com fundamentos teóricos e empíricos.

4.4.1 Monitoramento de acesso

¹¹ Gráfico construído com o software Pajek, disponível em: <<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>>. Acesso em: 22 fev. 2011.

A avaliação e o monitoramento do uso da informação são essenciais para formulação de estratégias, promovendo um planejamento com visão a médio e longo prazo. Um dos instrumentos é o monitoramento de acesso, que na Brapci é realizado de duas formas diferentes. A primeira por meio da ferramenta *Google Analytics* que possibilita relatórios de acesso, tempo médio de permanência, índice de rejeição e localização geográfica dos visitantes.



GRÁFICO 8 - QUANTIDADE DE ACESSOS A BASE POR NÚMERO DE USUÁRIOS
Fonte: *GOOGLE ANALYTICS* (2010)

Os dados obtidos entre 1º de março de 2010 e 30 de novembro de 2010 permitem observar que houve 81.916 visitantes oriundos em sua maioria do Brasil. Em média, dois mil usuários por semana visitaram a Brapci, e destes 39% fazem visitas constantes, demonstrando a aprovação do sistema de busca. Observa-se também uma sazonalidade de consultas, com redução significativa no mês de férias (julho). (Gráfico 8).

Analisando-se as informações também fornecidas pelo *Google Analytics* (Tabela 5), observa-se uma grande concentração de acessos na região sudeste e sul, com as cidades de São Paulo, Rio de Janeiro e Belo Horizonte com mais acessos que a cidade da mantenedora da base (Curitiba). O que pode ser observado é que as cidades que registram mais acessos são as que contam com maior quantidade de programas de graduação e pós-graduação, enquanto as regiões Norte e Nordeste, embora ofertando cursos na área, fazem pouco uso da base. Esta percepção carece de novos estudos para identificar o motivo deste baixo acesso, e se é ocasionado pela dificuldade de acesso a computadores ou por desconhecimento da base.

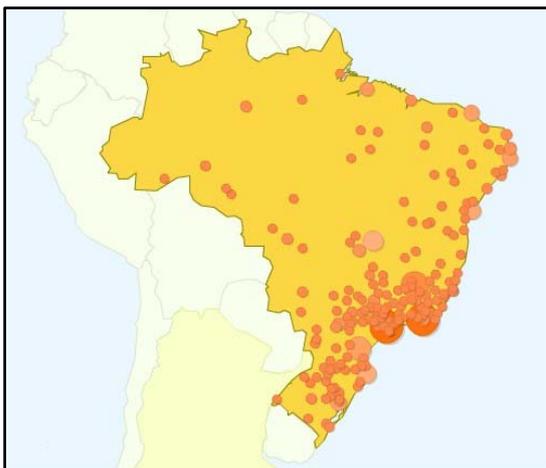


FIGURA 38 - LOCALIZAÇÃO GEOGRÁFICA DOS USUÁRIOS DA BRAPCI
Fonte: GOOGLE ANALYTICS (2010).

TABELA 5 - ACESSO À BRAPCI POR CIDADE

CIDADE	VISITANTES
São Paulo	8.640
Rio de Janeiro	8.378
Belo Horizonte	6.586
Curitiba	5.358
Brasília	4.033
Florianópolis	3.700
Salvador	3.190
Porto Alegre	2.593
Fortaleza	2.415
Recife	2.353

Fonte: GOOGLE ANALYTICS (2010)

A outra forma de monitoramento parte das consultas dos termos e das formas que os pesquisadores utilizam em sua estratégia de busca. Registrando todas as consultas desde 2008. A partir de novembro de 2010 o sistema registra também a quantidade de *download* de PDF realizados na Brapci.

Entre os termos mais consultados pelos usuários estão (por ordem alfabética): ação cultural, aprendizagem, arquivologia, avaliação, biblioteca escolar, biblioteca pública, biblioteconomia, ciência da informação, catalogação, ensino, ética, hábito de leitura, indexação, informação, linguagem natural, literatura infantil, pesquisa e pesquisa escolar. O resultado demonstra ainda um interesse acentuado nas áreas de leitura, biblioteconomia e escola, influenciado pelo grande número de cursos de biblioteconomia no âmbito de graduação no Brasil.

Por meio de uma enquete realizada no final de 2009 pelo grupo de pesquisa, traçou-se o perfil dos usuários da Brapci, correspondendo a 80% de alunos de graduação, 10% de alunos de pós-graduação e outros 10% que correspondem à categoria de professores, pesquisadores e os que não se identificaram em nenhuma das categorias.

O crescimento do número de usuários da Brapci, comparado com o mesmo período de 2009, foi surpreendente, atingindo a marca de 1.200%, considerando-se que a metodologia de acesso pelo módulo público só foi instalada em meados do mês de março de 2010.

4.4.2 Usabilidade do sistema

A primeira validação do Módulo Manutenção foi realizada por usuários indexadores da base, sendo a avaliação baseada nos elementos de usabilidade, comparando-se com usuários que utilizaram somente a versão Procite5 da base. Essa comparação possibilitou a identificação de problemas. O estudo e aplicação da validação foi realizado pela aluna de Iniciação Científica, Kaoane Stival (2009), membro do Grupo E3PI e apresentado em evento científico da UFPR.

Stival (2009) identificou em sua pesquisa que a facilidade com que as pessoas empregam uma ferramenta ou objeto com o fim de realizar uma tarefa específica é definida como usabilidade. A utilização de instrumentos que possam avaliar pontos fortes e necessidades de melhoria do sistema deve constantemente ser monitorada para o aprimoramento do sistema. Nielsen e Thair (2002) identificou 101 critérios de usabilidade para análise de sites da *Web*. Selecionando-se os principais, esses critérios foram agrupados em seis grupos:

- a) facilidade de aprender;
- b) facilidade de relembrar;
- c) eficiência;
- d) eficácia;
- e) satisfação;
- f) controle de erros.

Para a avaliação do protótipo foi aplicado um questionário de trinta questões, contemplando-se os seis grupos de usabilidade a uma amostra de dez usuários indexadores da base, essa amostra era dividida em dois grupos (braços), cinco que conheciam apenas o sistema *ProCite5* e outros cinco que utilizaram a nova metodologia. Dos enviados, obtiveram-se oito questionários respondidos, correspondendo a quatro de cada grupo, qualificando-se com 80% de retorno.

As respostas das questões foram agrupadas conforme os elementos de avaliação e tabuladas em uma planilha do *Microsoft Excel* para visualização dos gráficos e análise. Por meio dos resultados obtidos, foi possível uma comparação entre as usabilidades das interfaces.

Pelas conclusões de Stival (2009), foi possível verificar, por meio da comparação entre as performances dos dois grupos que o protótipo, sob todos os

elementos avaliados, obteve ganho considerável de qualidade em relação ao software anteriormente utilizado (*ProCite5*). Dos elementos analisados no protótipo, o controle de erro foi o que apresentou o menor grau de satisfação, carecendo de uma reavaliação neste quesito e melhoria nas mensagens de informação ao usuário.

4.5 MANUAL DE POLÍTICA DE COLEÇÃO E INDEXAÇÃO

Para orientação da equipe de indexação e os gestores da base, foi constituído um manual da política de coleção e indexação, incorporando-se ainda a este um breve histórico do sistema e da equipe participante desde sua primeira versão. O Apêndice B traz a íntegra do documento, também disponível no site da Brapci (www.brapci.ufpr.br/brapci_politica.pdf).

4.6 DIGITALIZAÇÃO DE DOCUMENTOS

A política de coleção da Brapci definiu que seu acervo seria composto não somente das referências das obras, mas do seu conteúdo integral, contribuindo para o resgate e preservação da memória da CI no Brasil. Por existirem muitas publicações históricas, que foram somente disponibilizadas em papel, tem sido necessária a conversão de seu suporte físico para o eletrônico por meio da digitalização. Para a execução desta tarefa foi adquirido um scanner HP Modelo G4050 para a conversão das páginas das revistas históricas, disponíveis somente neste suporte para serem incorporadas a base.

Softwares específicos para a digitalização e conversão de imagens em texto conhecidos como *Optical Character Recognition* (OCR) podem ser encontrados na Internet, porém dos avaliados pelo Grupo E3PI, nenhum deles atendia as necessidades de conversão, as digitalizações provocavam erros crassos comprometendo toda a digitalização.

Após alguns testes sistematizados, o Grupo E3PI concluiu que a melhor maneira de converter para o formato digital seria digitalizar as revistas por partes,

criando um arquivo em formato PDF para cada trabalho da publicação (ferramenta disponível no software do *scanner*). Optou-se em digitalizar todas as seções, como editorial, apresentação e outras em uma única vez.

Com os arquivos digitalizados em formato PDF, era possível manter a compatibilidade com diversos sistemas de arquivos e ainda garantir a impressão sem perda da qualidade, quando comparado com o original. Utilizando o programa de leitura de PDF, o *Acrobat 8.0 Professional*, foi possível converter a imagem do PDF em um PDF pesquisável (gerando os metadados). A conversão por este software diminui a margem de erro do reconhecimento de texto e ainda possibilita ao usuário pesquisador a cópia de parte do texto digitalizado.

Importa salientar que, em muitas das digitalizações realizadas, a fonte original continha marcações realizadas a lápis, canetas ou marca textos e que estas foram fielmente capturadas na digitalização. Porém na conversão para texto (OCR) essas marcações provocam um reconhecimento incorreto dos caracteres.

Todos os PDF digitalizados são incorporados aos seus respectivos trabalhos, proporcionando ao usuário o acesso direto ao conteúdo integral sem a necessidade de que seja redirecionado a outra fonte.

5 CONSIDERAÇÕES FINAIS

No início do estudo, não se tinha a exata dimensão da complexidade para o desenvolvimento de uma metodologia que envolvesse três áreas do conhecimento, que muitas relações têm entre si, mas que pouca representatividade conseguem na efetivação de propostas concretas.

Foi desafiante o processo de identificar e analisar teorias da CI que pudessem fundamentar este tipo de estudo. Com a fundamentação nas teorias oriundas da filosofia e resgatando-se o histórico das classificações bibliográficas e documentais do final do século XIX, foi possível selecionar e relacionar princípios de classificação que pudessem atender as necessidades da metodologia.

Na busca, consolidou-se a formação de conceitos para identificação de um corpo teórico fundamentado em Wüster (1971), Dahlberg (1978) e Cabré (1993), para a construção de uma metodologia com base empírica. Define-se, com este referencial, o termo como unidade léxica temporal para representação de um conceito, este caracterizado pela sua permanência. Estabelecendo esta relação, partiu-se para os estudos desses conceitos, na tentativa de reconhecer a carga semântica a eles correspondente, passando pelas definições de glossários, tesouros e ontologias.

A leitura de estudos recentes levou à percepção de que o Brasil foi influenciado por três vertentes no tratamento de informação, cada uma com terminologias e com objetivos algumas vezes diferentes.

Uma curiosidade encontrada na literatura, e na empregabilidade dos tesouros foi a tênue linha que separa sua estruturação sintática (pelo termo) de sua estruturação semântica (ontologia), o que permite que se antevejam as duas abordagens no emprego de tesouros, o mesmo ocorrendo no tratamento de ontologias (às vezes restritas ao processamento de termos).

Na formalização de instrumentos auxiliares à construção de glossários e tesouros foi essencial o reconhecimento das relações possíveis entre os conceitos, sobressaindo-se a superação das relações dicotômicas ou hierárquicas, representada por construções como as teorias facetadas de Ranganathan.

O estudo sobre os modelos de comunicação científica resultou no conhecimento de novas propostas como a de Hurd (2000), que foram fundamentais

para sustentar a necessidade de criação de uma metodologia que agregasse mais ao usuário do que um repositório de dados, mas que também possibilitasse a ele, concentrar seus esforços de recuperação em um único local, permitindo a visualização de indicadores para uma análise qualitativa com uma amostra maior de informações, retirando parcialmente do pesquisador o trabalho operacional de coletar e sistematizar dados selecionados.

Para efetivação da proposta de uma metodologia multidisciplinar para socialização de saberes houve uma longa trajetória, buscando-se um quadro teórico que pudesse sustentá-la. A proposta viria ao encontro do que planejou o grupo de pesquisa E3PI e para que pudesse ser efetivada foram realizados outros esforços paralelos conjugados para a construção da MB. Com a proposta definida, acrescentou-se o desafio de sua aplicação em uma realidade, com um estudo de caso na construção da Brapci, e principalmente na avaliação e reestruturação de funcionalidades que pudessem manter os fundamentos teóricos com conceitos de usabilidade para os usuários.

A avaliação da metodologia empregada para construção da Brapci foi realizada, levantando-se em consideração os problemas e as soluções encontrados para sua manutenção e aperfeiçoamento. Foram realizados estudos buscando-se identificar e analisar a produção intelectual veiculada pela Brapci, com a metodologia por ela disponível e a fundamentação em elementos teóricos e empíricos.

Esses estudos possibilitaram a concretização de um corpo de conhecimento constituído pela produção intelectual do grupo E3PI, assim como de outros pesquisadores que têm se reportado à Brapci como fonte de suas investigações.

Pode-se considerar que o objetivo deste estudo foi atingido, e que a proposta metodológica é viável como comprovado no estudo de caso da Brapci. Ressalta-se ainda que a Brapci garante a socialização do conhecimento nela registrado e por ela organizado e comunicado, concentrando em um único local uma literatura especializada, poupando tempo do pesquisador e diminuindo seu trabalho operacional, tanto nas estratégias de busca, quanto na seleção e análise de dados.

Por utilizar uma preferência na abordagem cognitiva, a metodologia demonstrou-se lenta, quando da necessidade de organizar grandes volumes de informação, por seu grau de detalhamento, muitas vezes consumindo muito tempo do indexador. Mas essa lentidão nos processos garante à Brapci a sua qualidade,

estruturando os dados de forma que facilite a análise de dados e possa ser exportada para outras bases de dados.

O que deve ficar explícito é que a proposta desta metodologia não deve se restringir ou encerrar com a conclusão deste trabalho. Pelo contrário, a proposta é o ponto inicial para que novas bases sejam construídas em outras áreas temáticas, adaptando-se ou se modificando a metodologia para que possa atender o mais amplamente possível os requisitos dos usuários em sua busca e organização de informação.

Outras vertentes podem surgir deste trabalho, principalmente no uso de novas técnicas e tipos de análises para geração de indicadores que ajudem no crescimento das pesquisas brasileiras, consolidando-se como um observatório da produção científica nacional.

Por mais que as tecnologias evoluam e exista a criação de protocolos de interoperabilidade, os processos automáticos de indexação não conseguem atender a todas as variações da explicitação de informações de seus autores. As ferramentas computacionais podem, sem dúvida, agilizar o monitoramento, coleta armazenamento de novas informações em banco de dados, contribuindo com a redução do trabalho manual e repetitivo do indexador. Porém a revisão e a ratificação dos dados devem ser realizadas por um indexador humano, que consiga interpretar e buscar padrões para que possa ser “programado” e facilitar novos processos que se enquadrem nestas regras.

Desse modo, a concepção da MB é mais abrangente do que o conceito de base de dados, transcendendo os limites de um recurso documental, como fonte de pesquisa de modo a propiciar a preservação histórica da produção e ainda municiar o pesquisador de instrumentos que possibilitem o estudo sobre indicadores e modos de produção, de uso e comportamento dos usuários no processo de recuperação de informação. Caracteriza-se, portanto, em sua área de concentração como um observatório da produção intelectual de uma área, podendo adquirir dimensões que extrapolam os limites de um repositório.

REFERÊNCIAS

AHMED, S. M. Z.; McKNIGHT, C.; OPPENHEIM, C. A Study of users' performance and satisfaction with the Web of Science IR Interface. **Journal of Information Science**, Cambridge, v. 30, n. 5, p. 459-468, 2004.

ALVARENGA, L. Organização da informação nas bibliotecas digitais. In: NAVES, M. M.; KURAMOTO, H. **Organização da informação**: princípios e tendências. Brasília: Briquet de Lemos, 2006.

ANSI/NISO Z39.19:2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO, 2005. 184 p.

AQUINO, I. J.; CARLAN, E.; BRASCHER, M. B. Princípios classificatórios para a construção de taxonomias. **Revista Ponto de Acesso**, Salvador, v. 3, n. 3, jul./dez. 2009.

ARANALDE, M. M. Reflexões sobre os sistemas categoriais de Aristóteles, Kant e Ranganathan. **Ciência da Informação**, Brasília, v. 38, n. 1, jan./abr. 2009.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências – elaboração. Rio de Janeiro, 2002.

BARITÉ, M. **Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica**. Porto Alegre: ABEED, 1998. 7 p.

BARRETO, A. A. As novas tecnologias de informação e geração do conhecimento. **Comunicação & Informação**, v. 1, n. 1, jan./jun. 1998.

BARROS, L. A. **Curso básico de terminologia**. São Paulo: EDUSP, 2004.

BELKIN, N. J. Anomalous state of knowledges as a basis for information retrieval. **Canadian Journal of Information Science**, v. 5, p. 133-140, 1980.

BESSÉ, B. La définition terminologique. In: Colloque la Définition, 1988, Paris. **[Anais...]** Canadá: Librairie Larrousse, 1990, p. 252-260.

BOEHM, B. W. A spiral model of software development and enhancement. **IEEE**, May 1988. Disponível em: <<http://www.cs.umd.edu/class/spring2003/cmsc838p/Process/spiral.pdf>>. Acesso em: 26 mar. 2010.

BOGHI, C. **Sistema de informação**: um enfoque dinâmico. São Paulo: Ed. Érica, 2002.

BOLAÑO, C. **Indústria cultural, informação e capitalismo**. São Paulo: Hucitec, 2000.

BUFREM, L. S. Práticas de organização e divulgação da produção intelectual em ciência da informação no Brasil. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2008.

_____. Práticas de organização e divulgação da produção intelectual em ciência da informação no Brasil. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 2. , 2010, São Carlos, SP. **Anais...** São Carlos: UFSCar, 2010.

BUFREM, L. S.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.

CABRÉ, M. T. **La terminología**: teoría, metodología, aplicaciones. Barcelona: Antártida; Empúries, 1993.

CABRÉ, M. Teresa. **La terminología**: representación y comunicación. Barcelona: Institut Universitari de Lingüística Aplicada, 1999. 369 p.

CAFÉ, L.; FACHIN, G. R. B. Provedores de dados, provedores de serviços e periódicos em ciência da informação, biblioteconomia e áreas afins. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2007.

CAFÉ, L.; MENDES, F. Estudo sobre estrutura definitiva para desenvolvimento de ontologias. **Informação e Sociedade: Estudos**, João Pessoa, v. 19, n. 2, p. 71-80, 2009.

CAMPOS, A. T. O processo classificatório como fundamento das linguagens de indexação. **Revista de Biblioteconomia de Brasília**, Brasília, v. 6, n. 1, p. 1-8, 1978.

CAMPOS, M. L. A. **Linguagem documentária**: teorias que fundamentam sua elaboração. Niterói: EDUFF, 2001. 133 p.

CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e Classificação: o princípio de categorização. **DataGramaZero**, Rio de Janeiro, v. 9, n. 4, 2008.

CARNEIRO, M. V. Diretrizes para uma política de Indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CHEN, P. **Modelagem de dados**: a abordagem entidade relacionamento para projeto lógico. São Paulo: McGraw-Hill, 1990

CIANCONI, R. Banco de Dados de acesso público. **Ciência da Informação**, Brasília, v. 16, n. 1, p. 53-59, jan./ jun. 1987.

CINTRA, A. M. M. *et al.* **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.

COSTA, F. D. O. Opções metodológicas em pesquisa: modelização do novo sistema gerenciador da base brapci. In: EVINCI, 17., 2009, Curitiba. **Anais...** Curitiba: Universidade Federal do Paraná.

COSTA, L. F.; RAMALHO, F. A. A usabilidade nos estudos de uso de informação: em cena usuários e sistemas interativos de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 1, p. 92-117, jan./abr. 2010.

CUENCA, A. M. B.; NORONHA, D. P.; ALVAREZ, M. C. A. Avaliação da capacitação de usuários para a recuperação da informação: o caso de uma biblioteca acadêmica. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 4, n. 1, jan./jun. 2008.

CUNHA, M. B. Das bibliotecas tradicionais às digitais – o que há de novo? In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO - CBBD, 22., 2007, Brasília, DF. **Igualdade e diversidade no acesso à informação**: da biblioteca tradicional à biblioteca digital: anais. Brasília, DF: FEBAB, 2007.

DAHLBERG, I. Fundamentos teórico-conceituais da classificação. **Revista de Biblioteconomia de Brasília**, Brasília, v. 6, n. 1, jan./jun. 1978.

DE SORDI, J. O. **Administração da informação**: fundamentos e práticas para uma nova gestão do conhecimento. São Paulo: Saraiva, 2008.

DIAS, E. W. Contexto Digital e Tratamento da Informação. **DataGramZero**, Rio de Janeiro, v. 2 n. 5, out. 2001.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

DIEBERGER, A. *et al.* Social navigation: techniques for building more usable systems. **Interactions**, v. 7, n. 6, p. 36-45, 2000.

DODEBEI, V. L. D. **Tesouro**: linguagem de representação da memória documentaria. Rio de Janeiro: Interciência, 2002. 120 p.

DUBLIN CORE. **Dublin core metadata initiative**. 2010. Disponível em: <<http://dublincore.org/>>. Acesso em 12 out. 2010.

FARIA, M. I. R.; PERICÃO, M. G. **Dicionário do livro**: da escrita ao livro eletrônico. São Paulo: EDUSP, 2008. 768 p.

FEITOSA, A. **Organização da informação na web**: das tags à web semântica. Brasília: Thesaurus, 2002.

FREEMAN, L. C. *Some antecedents of social network analysis*. **Connections**, v. 19, n. 1, p. 39-42, 1996.

- FREITAS, J. L. **Proposta de metodologia para a recuperação da produção científica em Ciência da Informação em uma base de dados**. 2010. Trabalho de Conclusão de Curso (Graduação em Ciência e Gestão da Informação) - Universidade Federal do Paraná, 2010.
- FREITAS, J. L.; BUFREM, L. S.; GABRIEL JUNIOR, R. F. 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, jul./dez. 2010.
- FUJITA, M. S. L. Análise e síntese documentárias para compreensão de leitura de textos didáticos: uma proposta de aplicação do sistema de indexação Precis. **Informare**, Rio de Janeiro, v. 5, n. 2, 1999.
- FURTADO, J. Conceitos e tipologia em ciência e tecnologia e sua influência na publicação de informações. **Ciência da Informação**, Brasília, v. 10, n. 1, p. 13-18, 1981.
- GARVEY, W. D.; GRIFFITH, B. C. Communication, the essence of science, Apêndice A, B. In: GARVEY, W.D. **Communication: the essence of science**. Oxford: Pergamon Press, 1979. p. 299. Disponível em: <<http://global-reach.biz/globstats/evol.html>>. Acesso em: 16 maio 2010.
- GONZÁLEZ DE GÓMEZ, M. N. As relações entre ciência, Estado e sociedade: um domínio de visibilidade para as questões da informação. **Ciência da Informação**, Brasília, v. 32, n. 1, p. 60-76, jan./abr. 2003.
- GRIEBLER, A. C. F.; MATTOS, A. M. Novas tecnologias, novas mídias, velhas dificuldades: aprimorando a interface com o usuário para a escolha de base de dados ou periódicos. **Encontros Bibli**, Florianópolis, v. 12, n. 23, 1º sem. 2007.
- GUIMARÃES, J. A. C.; SALES, R. Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 11, n. 1, fev. 2010.
- HURD, J. M. The transformation of scientific communication: a model for 2020. **Journal of the American Society for Information Science**, v. 51, n. 14, p. 1279–1283, 2000.
- KERN, M. V. **Banco de dados relacionais: teoria e prática de projetos**. São Paulo: Érica, 1994.
- KOBASHI, N. Y. Análise documentária e representação da informação. **Informare**, Rio de Janeiro, v.2, n. 2, p. 5-27, 1996.
- _____. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. **DataGramZero**, Rio de Janeiro, v. 8, n. 6, dez. 2007.
- KURAMOTO, H. **Blog do Kuramoto**. 2010. Disponível em: <<http://kuramoto.blog.br/2006/03/page/2/>>. Acesso em 14 ago. 2010.

_____. Sintagmas nominais: uma abordagem no processo de indexação. In: NAVES, M. M. L.; KURAMOTO, H. **Organização da informação: princípios e tendências**. Brasília: Brique de Lemos, 2006.

LAAN, R. H. V. D. **Tesauro e terminologia: uma inter-relação lógica**. 2002. 182 f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul. Porto Alegre, 2002.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília: Brique de Lemos, 1993.

LARMAN, G. **Utilizando UML e padrões: uma introdução à análise e ao projeto orientados**. 3. ed. São Paulo: Bookman, 2007.

LAZARTE, L. Ecologia cognitiva na sociedade da informação. **Ciência da Informação**, Brasília, v. 29, n. 2, p. 43-51, maio/ago. 2000.

LE COADIC, Y. F. **A Ciência da Informação**. 2. ed. Brasília: Brique de Lemos, 2004.

LIMA-MARQUES, M. **Ontologias: da filosofia à representação do conhecimento**. Brasília: Thesaurus, 2006. 72 p.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, jan./abr. 2002.

LOPES, R. R. V. Acesso a bases de dados em linha e em CD-ROM: algumas considerações sobre a sua implantação em países em desenvolvimento. **Ciência da Informação**, Brasília, v. 20, n. 2, jul./dez. 1991.

MÁRDERO ARELLANO, M. Á. M.; FERREIRA, S. M. S. P.; CAREGNATO, S. E. Editoração eletrônica de revistas científicas com suporte do protocolo OAI. In: FERREIRA, S. M. P.; TARGINO, M. G. **Preparação de revistas científicas: teoria e prática**. São Paulo: Reichmann e Autores, 2005.

MEADOWS, A. J. **A comunicação científica**. Brasília: Brique de Lemos, 1999.

MESQUITA, R. M. A. *et al.* Elaboração e aplicação de instrumentos para avaliação da base de dados Scopus. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, 2006.

MOOERS, C. N. Zatoncoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 3, p. 20-32, 1951.

MOREIRO GONZÁLEZ, J. A. **Análise e representação do conteúdo das imagens: 20-24 de setembro de 2010**. Notas de aula. Apresentação.

_____. Criterios e indicadores para evaluar la calidad del análisis documental de contenido. **Ciência da Informação**, Brasília, v. 31, n. 1, p. 53-60, 2002.

MOURA, M. A. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 1, p. 59-73, 2009.

FERNANDES, R. P. M. XML y registros electrónicos: principales estándares en la descripción archivística. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 45-53, set./dez. 2006.

MUELLER, S. P. M. O círculo vicioso que prende os periódicos nacionais. **DataGramZero**, Brasília, v. 0, n. 0, dez. 1999.

NAVES, M. M. A importância de Ranganathan para a organização do conhecimento. In: NAVES, M. M.; KURAMOTO, H. **Organização da informação**: princípios e tendências. Brasília: Briquet de Lemos, 2006.

NAVES, M. M.; KURAMOTO, H. **Organização da informação**: princípios e tendências. Brasília: Briquet de Lemos, 2006.

NIELSEN, J.; THAIR, M. **Homepage usabilidade**: 50 websites desconstruídos. Rio de Janeiro: Campus, 2002.

NUNES, C. O. I.; FREITAS, M. R. M. Exaustividade, Especificidade e Coerência da Indexação da Base de Dados Bibliográficos do NID/FURG (SAB-II). **BIBLOS: Revista do Departamento de Biblioteconomia e História**, Rio Grande, v. 9, n. 1, p. 155-198, 1997.

OBRST, L. **The Semantic Spectrum & Semantic Models**. 2009. Disponível em: <http://ontolog.cim3.net/file/resource/presentation/LeoObrst_20060112/OntologySpectrumSemanticModels--LeoObrst_20060112.ppt>. Acesso em: 20 nov. 2009.

OHIRA, M. . B.; SOMBRIO, M. L. L. N.; PRADO, N. S.. Periódicos brasileiros especializados em biblioteconomia e ciência da informação: evolução. **Encontros Bibli**: Florianópolis, v. 5, n. 10, p. 26-40, out. 2000.

OLIVEIRA, É. B. P. M. Produção científica nacional na área de geociências: análise de critérios de editoração, difusão e indexação em bases de dados. **Ciência da Informação**, Brasília, v. 34, n. 2, maio/ago. 2005.

OBJECT MANAGEMENT GROUP. **OMG Unified Modeling Language Specification Version 2.0**. Disponível em: < http://www.omg.org/gettingstarted/what_is_uml.htm>. Acesso em: 18 de fev. 2011.

OPEN ARCHIVES INITIATIVE. **OAI para principiantes**: tutorial en línea del Open Archives Forum. Disponível em <<http://travesia.mcu.es/portaln/jspui/html/10421/1823/intro.htm>>. Acesso em 13 mar. 2010.

ORGANIZAÇÃO INTERNACIONAL DE NORMALIZAÇÃO (ISO). **Principles and methods of terminology**. ISO Recommendation 704. SL, 1987. 16p.

ORGANIZAÇÃO INTERNACIONAL DE NORMALIZAÇÃO (ISO). **Terminology work -- Vocabulary** -- Part 1: Theory and application. ISO Recommendation 1087. SL, 1997

ORTEGA, Cristina Dotta. Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 5, n. 5, out. 2004.

ORTEGA, C. D.; LARA, M. L. G. A noção de documento: de Otlet aos dias de hoje. **DataGramZero**, Rio de Janeiro, v. 11, n. 2, abr. 2010.

OTLET, P. **El Tratado de Documentación**: el libro sobre el libro: teoría y práctica. Trad. por Maria Dolores Ayuso García. Murcia: Universidad de Murcia, 1996. Tradução de: *Traité de Documentation: le livre sur Le livre: théorie et pratique*. Bruxelles: Mundaneum, 1934. 431 p. Versão original disponível em: <http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf>. Acesso em: 30 out. 2010.

PARASURAMAN, A.; ZEITHAML, V. A.; BERRY, L. L. A conceptual model of service quality and its implications for future research. **Journal of Marketing**, vol. 49 (fall), p. 41-50, 1985.

PECEGUEIRO, C. M. P. A. Temática dos artigos de periódicos brasileiros na área da Ciência da Informação na década de 90. **Transinformação**, Campinas, v. 14, n. 2, p. 117-131, jul./dez. 2002.

PEREIRA, M. N. F. *et al.* Bases de dados na economia do conhecimento: a questão da qualidade. **Ciência da Informação**, Brasília, v. 28, n. 2, maio/ago. 1999.

PROGRAMA DE PÓS-GRADUAÇÃO EM ARTES DA UNICAMP. SIPEX – Produção intelectual. Disponível em: <<http://www.iar.unicamp.br/pg/cpg.producao.intelectual.php?programa=2>>. Acesso em 02 fev. 2010.

RADOS, G. J. V.; VALERIM, P.; BLATTMANN, Ú. Valor agregado a serviços e produtos de informação. Informativo CRB 14. **Revista ACB**, Florianópolis, v. 9, n. 1, p. 11-12, jan./mar. 1999.

RANGANATHAN, S. R. **Prolegomena to Library Classification**. 3rd. ed. Bombay: Asia publ. house. 640 p., 1967.

RECTOR, M.; YUNES, E. **Manual de semântica**. Rio de Janeiro: Ao Livro Técnico, 1980. 171 p.

RODRIGUES, A.; SILVA, A. R. Base de dados de informações jornalísticas sobre a Amazônia: ciência, tecnologia e meio ambiente – BDIJAm: em busca da qualidade. **Ciência da Informação**, Brasília, v. 36, n. 1, 2007.

ROSENFELD, L.; MORVILLE, P. **Information Architecture for the World Wide Web**. 2. ed. Sebastopol, CA: O'Reilly, 2002.

ROWLEY, J. **A biblioteca eletrônica**. Brasília: Brique de Lemos/Livros, 2002.

_____. **Informática para bibliotecas**. Brasília: Briquet de Lemos/Livros, 1994.

ROZENFELD, H. *et al.* **Gestão de desenvolvimento de produtos**: uma referência para a melhoria do processo. São Paulo: Saraiva, 2006.

SALES, L. F.; CAMPOS, M. L. A. C.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 13, n. 2, p. 62-76, 2008.

SALES, R.; CAFÉ, L. Diferenças entre Tesouros e Ontologias. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 14, n. 1, jan./abr. 2009a.

_____. Semelhanças e diferenças entre tesouros e ontologias. **DataGramaZero**, Rio de Janeiro, v. 9, n. 4, 2008.

_____. Tesouros e ontologias sob o olhar da teoria comunicativa da terminologia. In: ISKO-ESPANHA, 9, 2009, Valencia. **Anales...** Valencia, 2009b.

SANTOS, R. N. M.; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 2, n. 1, p. 155-172, 2009.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, 1996.

SAYÃO, L. F. Bases de dados: a metáfora científica. **Ciência da Informação**, Brasília, v. 25, n. 3, set./dez. 1996.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSAHN, S. **Sistema de banco de dados**. São Paulo: Makron Books, 1999.

SILVEIRA, A. **Administração de unidades de informação**: estudo de análise de citação da literatura de marketing da informação como instrumento para a tomada de decisão no desenvolvimento de colocação na biblioteca setorial do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, Florianópolis, BDC/CED/UFSC, 1993. 119 p. Trabalho apresentado a Comissão Examinadora do Concurso Público para provimento do grupo Magistério Superior, na Classe de TITULAR, do quadro de pessoal da Universidade Federal de Santa Catarina.

SINCLAIR, J.; CARDEW-HALL, M. *The folksonomy tag cloud: when is it useful?* **Journal of Information Science**, v. 34, n.1, p. 15–29, 2008.

SOUZA, M. I. F.; ALVES, M. D. R. Representação descritiva e temática de recursos de informação no sistema agência Embrapa: uso do padrão Dublin core. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 7, n. 1, 2009.

STIVAL, K. Opções metodológicas em pesquisa: validação do novo sistema gerenciador da base Brapci. In: EVINCI, 17., 2009, Curitiba. **Anais...** Curitiba: Ed. da UFPR, 2009.

STREHL, L. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. **Ciência da Informação**, Brasília, v. 27, n. 3, set./dez. 1998.

TAMMARO, A. M.; SALARELLI, A. **A Biblioteca Digital**. Brasília: Briquet de Lemos, 2008.

TERRA, J. C. C. *et al.* **Taxonomia**: elemento fundamental para a gestão do conhecimento, 2005. Disponível em: <<http://www.terraforum.com.br>>. Acesso em: 04 set. 2008.

TOBIAS, J. Seeking the subject. **Library Trends**, v. 47, n. 2, p. 209-217, Fall 1998.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistemas de classificação facetados e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, maio/ago. 2004.

ULLMANN, S. **Semântica**: uma introdução a ciência do significado. Lisboa: Fundação Calouste Gulbenkian, 5. ed, 1987.

UNITED STATES OF AMERICA. President's Science and Advisory Committee. **Science, Government and Information**: the responsibilities of the technical community and the government in the transfer of information. Washington, D.C. Government Printing Office, 1963. 52 p. (The Weinberg Report).

VALLS, V. M.; VERGUEIRO, W. C. S. A Gestão da Qualidade em Serviços de Informação no Brasil: uma revisão de literatura, de 1997 a 2006. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, jan./abr. 2006.

VANZ, S.A.S.; CAREGNATO, S.E. Estudos de citação: uma ferramenta para entender a comunicação científica. **Em Questão**, Porto Alegre v. 9, n.2, p. 295-307, 2003.

VIANA, C. L. M.; MÁRDERO ARELLANO, M. A.; SHINTAKU, M. Repositórios institucionais em Ciência e Tecnologia: uma experiência de customização do DSpace. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. **Anais...** São Paulo, 2005. Disponível em: <<http://eprints.rclis.org/archive/00005563/01/viana358.pdf>>. Acesso em: 10 set. 2006.

WÜSTER, E. **Les classifications de notions et de thèmes**: différences essentielles et applications. [S.l.], 1971. (Série INFOTERM 2-71 F).

XAVIER, A. G. *et al.* Base de dados como elemento facilitador para o tratamento dos dados terminológicos. **Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS**, Belo Horizonte, v. 10, n. 2, jul./dez. 2004.

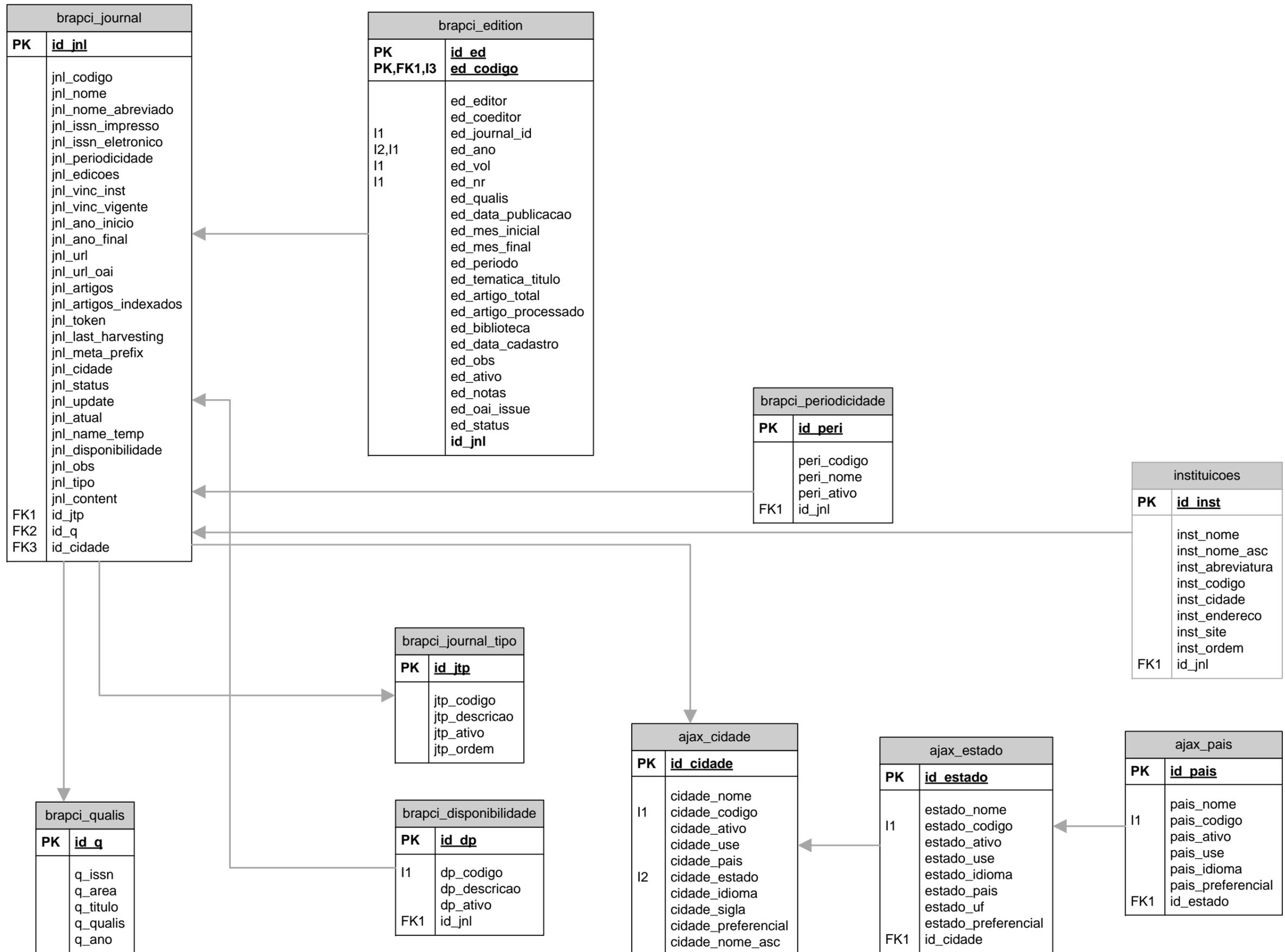
ZANINELLI, T. B.; CATARINO, M. E. Recuperação de informação no banco de dados da Embrapa soja: estudo do AINFO. **Informação & Informação**, Londrina, v. 9, n. 0, 2004.

ZIMBA, H. F.; MUELLER, S. P. M. Colaboração internacional e visibilidade científica de países em desenvolvimento: o caso da pesquisa na área de medicina veterinária em Moçambique. **Informação & Sociedade: Estudos**, João Pessoa, v. 14, n. 1, 2004.

ZINS, C. Knowledge map of information science. **Journal of the American Society for Information Science and Technology**, v. 58, n. 4, p.526-535, Feb. 2007.

APÉNDICES

APÊNDICE A - DIAGRAMA ENTIDADE DE RELACIONAMENTO



brapci_journal1	
PK	<u>id_jnl</u>
	jnl_codigo
	jnl_nome
	jnl_nome_abreviado
	jnl_issn_impresso
	jnl_issn_eletronico
	jnl_periodicidade
	jnl_edicoes
	jnl_vinc_inst
	jnl_vinc_vigente
	jnl_ano_inicio
	jnl_ano_final
	jnl_url
	jnl_url_oai
	jnl_artigos
	jnl_artigos_indexados
	jnl_token
	jnl_last_harvesting
	jnl_meta_prefix
	jnl_cidade
	jnl_status
	jnl_update
	jnl_atual
	jnl_name_temp
	jnl_disponibilidade
	jnl_obs
	jnl_tipo
	jnl_content
	id_jtp
	id_q
	id_cidade

brapci_edition	
	id_ed
	ed_codigo
	ed_editor
	ed_coeditor
	ed_journal_id
	ed_ano
	ed_vol
	ed_nr
	ed_qualis
	ed_data_publicacao
	ed_mes_inicial
	ed_mes_final
	ed_periodo
	ed_tematica_titulo
	ed_artigo_total
	ed_artigo_processado
	ed_biblioteca
	ed_data_cadastro
	ed_obs
	ed_ativo
	ed_notas
	ed_oai_issue
	ed_status
	id_jnl

brapci_edition_article	
PK	<u>id_ea</u>
	ea_journal_id
	ea_article
	ea_position
	ea_section
	ea_edition

brapci_metodologias_tp	
PK	<u>id_bmtf</u>
	bmtf_codigo
	bmtf_descricao
	bmtf_tipo

brapci_metodologias	
PK	<u>id_bmt</u>
	bmt_descricao
	bmt_content
	bmt_tipo
	bmt_ativo
	bmt_ordem
	bmt_codigo
	id_bmtf

brapci_article	
PK,I1	<u>id_ar</u>
I3	ar_journal_id
I2	ar_codigo
	ar_titulo_1
	ar_titulo_1_asc
	ar_titulo_2
	ar_titulo_3
	ar_idioma_1
	ar_idioma_2
	ar_idioma_3
	ar_obs
	ar_status
	ar_disponibilidade
	ar_pg_inicial
	ar_pg_final
	ar_data_envio
	ar_data_aceite
	ar_data_aprovado
	ar_doi
	ar_oai_id
	ar_tipo
	ar_area_conhecimento
	ar_brapci_id
	ar_resumo_1
	ar_resumo_2
	ar_resumo_3
	ar_position
	ar_section
	ar_edition
	ar_notas
	at_metodo_1
	at_metodo_2
	at_metodo_3
	at_tecnica_1
	at_tecnica_2
	at_tecnica_3
	at_analise_1
	at_analise_2
	at_metodologia_log
	at_metodologia_data
	at_metodologia_status
	at_obs
	at_obs_data
	at_obs_log
	at_mt
	ar_dt_inclusao
	id_ea
	id_bmt
	id_se

brapci_article_keyword	
PK	<u>id_ak</u>
	kw_article
	kw_keyword
	kw_ord
FK1	id_ar

brapci_keyword	
PK	<u>id_kw</u>
	kw_word
	kw_word_asc
	kw_codigo
	kw_use
	kw_idioma
	kw_tipo
	kw_hidden
FK1	id_ak

brapci_article_author	
PK	<u>id_ae</u>
	ae_journal_id
	ae_article
	ae_position
	ae_author
	ae_instituicao
	ae_aluno
	ae_professor
	ae_ss
	ae_pos
	ae_contact
	ae_mestrado
	ae_doutorado
	ae_profissional
	ae_bio
	ae_telefone
	ae_endereco
FK1	id_ar

brapci_autor	
PK	<u>id_autor</u>
	autor_codigo
	autor_nome
	autor_nome_asc
	autor_nome_abrev
	autor_nome_citacao
	autor_nasc
	autor_lattes
	autor_alias
	autor_fale
	autor_tipo
FK1	id_ae

brapci_article_suporte	
	id_bs
	bs_article
	bs_type
	bs_adress
	bs_status
	bs_journal_id
	bs_update

brapci_section	
PK	<u>id_se</u>
	se_codigo
	se_descricao
	se_ativo
	se_processar
	se_tipo
	se_ordem
	se_cod

(1:n)

(1:n)

(1:1)

(1:1)

(1:n)

(1:1)

(1:1)

(1:n)

(1:n)

(1:1)

I3

FK2

FK3

FK4

APÊNDICE B
POLÍTICA DE DESENVOLVIMENTO DE COLEÇÃO - BRAPCI

BRAPCI

POLÍTICAS DE COLEÇÕES DA BASE DE DADOS REFERENCIAL DE ARTIGOS EM PERIÓDICOS EM CIÊNCIA DA INFORMAÇÃO



Sumário

APRESENTAÇÃO.....	2
DISPOSIÇÕES GERAIS	3
POLÍTICAS DE CONTROLE E MONITORAMENTO DAS FONTES DE INFORMAÇÃO DA BASE BRAPCI	6
COLETA DE TRABALHOS EM REPOSITÓRIOS OAI-PMH.....	15

APRESENTAÇÃO

A proposta da criação de uma Política de Desenvolvimento de Coleções para a Base de Dados Referencial de Artigos em Periódicos em Ciência da Informação (Brapci) deve-se ao fato de se considerar importante a existência de um instrumento formal que estabeleça diretrizes para a tomada de decisão pela equipe que mantém a base.

A criação de um instrumento norteador como a Política de Desenvolvimento de Coleções deve-se ao fato de se considerar importante a existência de um instrumento formal que estabeleça critérios e prioridades com relação à seleção e aquisição do material que irá compor os acervos da base de dados. A formalização de uma política possibilita que a coleção cresça de forma consistente, qualitativa e quantitativamente, e que estabeleça as diretrizes a serem seguidas no processo incorporação de novos títulos, de catalogação e indexação e quais os tipos de acervos serão abrangidos em seu corpus.

A política de desenvolvimento de coleções estabelece os padrões que serão empregados pela base, detalha os processos aplicados nas atividades operacionais e ainda consiste de um elemento básico para qualquer tomada de decisão, contemplando os objetivos, visão e missão da base.

A Política de Desenvolvimento de Brapci define critérios para a composição do *corpus* em sua base de dados. É o conjunto de atividades caracterizadas por um processo decisório que determina a conveniência de se adquirir, manter ou descartar materiais bibliográficos e ou especiais, tendo como base critérios previamente definidos que por sua vez, são definidos por meio das diretrizes estabelecidas para formação ideal do acervo, tornando-se um instrumento para planejamento e avaliação.

A temática da coleção deve ser selecionada e desenvolvida para atender os interesses e necessidades de seus usuários, facilitando sobremaneira o acesso, a recuperação e a disseminação da informação. Portanto, o êxito da coleção está diretamente ligado a uma política de seleção.

É importante ressaltar que a Política de Desenvolvimento de Coleções numa base consiste num elemento básico para qualquer tomada de decisão. Ela contempla aspectos relativos à função e objetivos da base, usuários e necessidades, abrangências e níveis das coleções, tipos de materiais, critérios e responsabilidade pela seleção, modalidade de aquisição, critérios para alocação de recursos financeiros, de descarte e outros.

Entretanto, faz-se necessário o estabelecimento dos agentes envolvidos no processo decisório. Para tanto, a criação de uma Comissão Permanente de Seleção de Coleções possibilita que seus membros sejam indicados dentre os membros do corpo grupo de pesquisa E3PI, bem como do corpo técnico. Cabe a Comissão encontrar maneiras de fazer com que todos os assuntos de interesse da instituição se desenvolvam independentemente de atuações individuais, criando mecanismos formais que permitam a participação de todos os interessados.

Responsável	Rene Faustino Gabriel Junior	Versão	1.00
-------------	------------------------------	--------	------

Data	Responsável	Descrição
15/dez.2010	Rene F. Gabriel Junior	Criação do Documento

DISPOSIÇÕES GERAIS

1 FINALIDADE

Estabelecer diretrizes que deverão ser observadas e seguidas ao controle e monitoramento das fontes de informação já indexadas na base Brapci.

2 DEFINIÇÕES E ADESÃO

2.1 As definições utilizadas na presente política têm os significados que lhes são atribuídos no Projeto de Pesquisa “Opções metodológicas em pesquisa: a contribuição da área da informação para a produção de saberes no Ensino Superior” registrado no BANPESQ/THALES/UFPR sob número 2000008165.

2.2 Deverão assinar Termo de Conhecimento à presente política, tornando-se pessoas que tenham acesso a informações relevantes.

2.3 Quaisquer dúvidas acerca das disposições da presente política deverão ser esclarecidas juntamente ao responsável pela administração da base de dados Brapci.

3 OBJETIVO GERAL

Tendo como objetivos principais a otimização do processo de indexação e segurança das informações, a garantia da continuidade dos seus serviços e a preservação do conhecimento.

3.1 Objetivos específicos:

Para uma melhor definição e operacionalização da base, desdobra-se o objetivo geral em outros específicos, definindo que deva:

- a) estabelecer normas para seleção e aquisição de material bibliográfico;
- b) disciplinar o processo de seleção, tanto em quantidade como em qualidade, de acordo com as características de cada biblioteca setorial ;
- c) direcionar o uso racional dos recursos financeiros;
- d) estabelecer os critérios e atribuições para os usuários conforme os perfis definidos para a base;
- e) prever e planejar recursos orçamentários destinados à aquisição;
- f) sugerir fontes para a seleção do material;
- g) definir critérios básicos de seleção;
- h) estabelecer prioridades de aquisição;
- i) estabelecer critérios para avaliação das coleções;
- j) traçar diretrizes para a avaliação da coleção.

4 CONTEXTO

Da concepção inicial de reunir a literatura pertinente da área de CI em um único local que facilitasse a busca e recuperação da informação para pesquisadores, acadêmicos e comunidade em geral até o desenvolvimento do produto final que pudesse ser acessado por todos em um ambiente virtual e coletivo, foram demandados muitos recursos, como tempo e empenho do Grupo E3PI. A concretização dessa ideia foi possível com a aprovação de dois projetos pelo CNPq que garantiu aporte financeiro para aquisição de equipamentos e desenvolvimento da base.

A realização de uma base de dados com escopo temático para área de Ciência da Informação é justificada por estar relacionada com a área de estudo do pesquisador, integrante do Grupo E3PI, da UFPR, composto por docentes e discentes do Mestrado em Ciência, Gestão e Tecnologia da Informação e graduandos do curso de GI. Além disso, contou com uma base de dados pré-estruturada em um banco de dados “antigo” (ambientado no *ProCite5*), no qual a estrutura preexistente possibilitou uma análise, mesmo que de forma parcial, de quais informações seriam necessárias para subsidiar pesquisas futuras.

A concretização da metodologia resultou em um produto final, a Brapci, sendo possível consolidar as etapas de coleta, organização e disponibilização de trabalhos científicos na área de CI publicados em revistas científicas com avaliação Qualis e conseqüentemente sua validação por seus usuários.

Resgatando a história da Brapci, relata-se que a indexação de artigos científicos completou em 2010 onze anos de trabalho com a coleta, seleção e organização de artigos representativos para a área, publicados em revistas científicas. O projeto foi idealizado por Bufrem (2008) e o pelo Grupo E3PI, com intuito de reunir em um único banco todos os trabalhos publicados em revistas relevantes para área da CI, organizados de modo a oferecer possibilidade de se extrair não somente dados quantitativos, mas qualitativos, para análise e desenvolvimento de pesquisas para reconhecimento da área e exercício da crítica da área de CI no Brasil.

Nos três primeiros anos da implantação do projeto (2000-2003), segundo Costa (2009), foram levantados os títulos dos periódicos da área (treze títulos) e, paralelamente, realizada busca no acervo da Biblioteca do Setor de Ciências Sociais Aplicadas (BSCSA) da UFPR, onde foram catalogados os periódicos do acervo físico relacionados à área de CI. Os fascículos não encontrados foram solicitados aos editores ou, após localização no Catálogo Coletivo Nacional de Publicações Seriadas (CCN) do IBICT, a outras bibliotecas, para que fossem preenchidos os dados de acordo com os campos definidos para cadastro na base do *ProCite5*. Nos anos seguintes, compreendidos entre 2004 e 2008, foram realizadas algumas revisões nos critérios da política de títulos ao acervo, possibilitando a incorporação de novos títulos à base. Essa revisão ampliou significativamente a quantidade de títulos selecionados, dos treze iniciais para 27 revistas em 2008, totalizando 4.637 artigos (BUFREM, 2008).

Em meados de 2009 o protótipo foi disponibilizado para ser testado, permanecendo os dois sistemas, o *ProCite5* e a base *online*, por um período de três meses. Com a efetivação do protótipo foram identificados e incorporados à base mais três títulos de publicações, implementando-se o mecanismo de coleta automática de registros (OAI-PHM), que ampliou para 1.984 o número de registros adicionais. Este número significativo de novos registros foi ocasionado pela coleta automática e inclusão de todos os tipos de seções disponibilizados pelo periódico, não se restringindo aos artigos científicos. Enquanto anteriormente no suporte papel eram indexados somente artigos científicos. Os trabalhos coletados dentro da política da base para indexação estão: artigos científicos; comunicações; ensaio; ponto de vista; relato de caso, editorial e revisão de literatura.

A MB tem como diretriz o não descarte ou exclusão dos documentos coletados automaticamente, seja qual for a seção, ficando estes em uma posição estacionária na base para futura utilização ou quando na alteração da política de disponibilização de conteúdo da Brapci. Em dezembro do ano de 2010, a Brapci reunia 23 publicações periódicas vigentes e dez históricas produzidas no Brasil, concentrando 6.632 artigos em CI. O último periódico incorporado a base foi a Revista InCID da Universidade de São Paulo em outubro de 2010.

No protótipo, o primeiro módulo a ser concretizado foi o de manutenção (MM), o que possibilitou as primeiras visualizações sistemáticas e primeiras avaliações dos processos dentro da base. Este módulo foi o que demandou mais trabalho, por ser o que gerar a estruturação do sistema. Na seqüência foram produzidos os módulos público e pesquisador.

A explicitação das informações indexadas é realizada pelo Módulo Público, por questões operacionais e de atualização. Como contraponto aos procedimentos burocráticos e operacionais, o Grupo E3PI optou por disponibilizar artigos “novos” coletados para acesso imediato ao público, mesmo sem passar pelo processo de revisão. Esta opção proporciona à Brapci uma perda relativa da qualidade na indexação, o que é compensado pela sua rápida atualização, chegando a casos de uma publicação estar disponível na base no mesmo dia de sua edição *online*. O tempo máximo de incorporação à Brapci é de sete dias para revistas que utilizam o protocolo OAI-PMH.

Com o protótipo em uso, o processo de validação foi necessário para sua consolidação e efetivação. Essa validação para a área de CC é denominada de “Homologação” e nessa fase os usuários realizam o uso do sistema com dados muito próximos do real, efetuando operações que deveriam ser utilizadas no trabalho rotineiro em diversos processos.

5 COMISSÃO PARA SELEÇÃO DE COLEÇÕES DA BRAPCI

A Brapci é orientada para o atendimento e socialização de conhecimentos em Ciência da Informação, seu gestor é o grupo de pesquisa E3PI da Universidade Federal do Paraná. Este grupo é o responsável por sua manutenção e atualização, bem como a atualização das políticas dispostas neste manual.

Responsável	Rene Faustino Gabriel Junior	Versão	1.01
-------------	------------------------------	--------	------

Data	Responsável	Descrição
10/jul.2009	Tidra V Sorribas	Criação do Documento
15/dez.2010	Rene F. Gabriel Junior	Aplicação do Documento

POLÍTICAS DE CONTROLE E MONITORAMENTO DAS FONTES DE INFORMAÇÃO DA BASE BRAPCI

1 FINALIDADE

Estabelecer diretrizes que deverão ser observadas e seguidas ao controle e monitoramento das fontes de informação já indexadas na base Brapci. Tendo como objetivos principais a otimização do processo de indexação e segurança das informações, a garantia da continuidade dos seus serviços e a preservação do conhecimento.

2 DEFINIÇÕES E ADESÃO

2.1 As definições utilizadas na presente política têm os significados que lhes são atribuídos no Projeto de Pesquisa “Opções metodológicas em pesquisa: a contribuição da área da informação para a produção de saberes no Ensino Superior” registrado no BANPESQ/THALES/UFPR sob número 2000008165.

2.2 Deverão assinar Termo de Conhecimento à presente política, tornando-se pessoas que tenham acesso a informações relevantes.

2.3 Quaisquer dúvidas acerca das disposições da presente política deverão ser esclarecidas juntamente ao responsável pela administração da base de dados Brapci.

3 MONITORAMENTO DA INFORMAÇÃO

3.1 Diretrizes

3.1.1 Racionalizar os processos administrativos e operacionais, como forma de melhoria na qualidade dos serviços prestados.

3.1.2 Monitorar os fascículos das revistas publicados:

As revistas que não são disponibilizadas eletronicamente devem ser solicitadas aos respectivos editores via *e-mail* ou correios à medida que são publicadas. As publicações das revistas ambientadas na Plataforma SEER devem ser monitoradas por meio do sistema de Harvesting implementado na Brapci. Necessitando o cadastro do periódico e a porta de coleta de metadados pelo protocolo OAI-PMH.

O *e-mail* do Projeto Brapci deverá ser acessado diariamente pelo responsável pela indexação (qual é o e-mail da Brapci). As demais revistas disponibilizadas eletronicamente ambientadas em plataformas diferenciadas deverão ser acompanhadas quinzenalmente para o monitoramento de possíveis novos fascículos em seus *websites*.

3.2 RESPONSABILIDADES

Dos responsáveis pela manutenção e atualização da base:

- gerenciar a base Brapci;
- registrar, monitorar e avaliar os processos;
- zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci;

Dos líderes do Projeto Brapci:

- administrar os direitos de acesso das pessoas equipamentos para o desempenho de suas funções;
- zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci;
- prestar orientação às pessoas envolvidas no projeto;
- verificar o cumprimento das normas relacionadas.

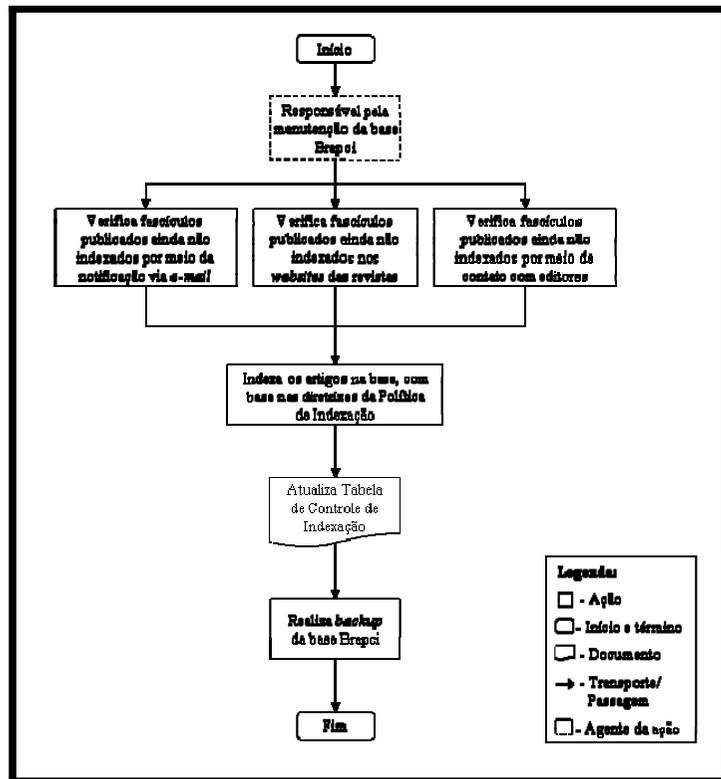
3.3 DISPOSIÇÕES GERAIS

Todos os ativos utilizados para desenvolvimento do projeto Brapci, constituem patrimônio público, da Universidade Federal do Paraná.

Compete aos que utilizam e manuseiam informações, conhecer e cumprir esta Política, para o bom uso das informações.

Compete aos responsáveis por processos que coletam informação: salvaguardar as características da informação fornecida à base.

Compete aos responsáveis que tem a atribuição sobre determinada natureza de informação classificá-la de acordo com os critérios estabelecidos.



QUADRO 1: FLUXOGRAMA DO PROCESSO DE MONITORAMENTO DAS FONTES E INDEXAÇÃO

Fonte: Autores (2011).

4 PERFIS DOS USUÁRIOS DA BASE

Na Brapci existe uma designação de atividade e responsabilidade que cada membro do módulo de manutenção deve exercer. Essa designação é atribuída pelo perfil de cada usuário, estabelecendo-se atividades e relatórios aos quais este poderá ter acesso. Desta forma, as atividades e os processos serão executados por grupos específicos de usuários, com a categorização das respectivas funcionalidades, bem como dos perfis autorizados para cada tarefa. A definição dos perfis dos usuários da base, conforme consenso do Grupo E3PI, é representada no Quadro 1 e tem sido considerada fundamental para definir as atividades ou funcionalidades possíveis de se concretizarem na base.

Perfil dos usuários da base
Coordenador
Editor
Revisor
Indexador
Pesquisador
Visitante

QUADRO 1 - PERFIS DOS USUÁRIOS DA BASE

Fonte: O autor (2010)

Com a definição dos perfis de usuários na base é possível delinear o mapeamento das funções do sistema. Categorizadas as permissões de uso, elaborou-se um processo para indexação e manutenção da Brapci que contempla a possibilidade de ocorrerem atividades simultâneas e online diretamente no sistema (Figura 2).

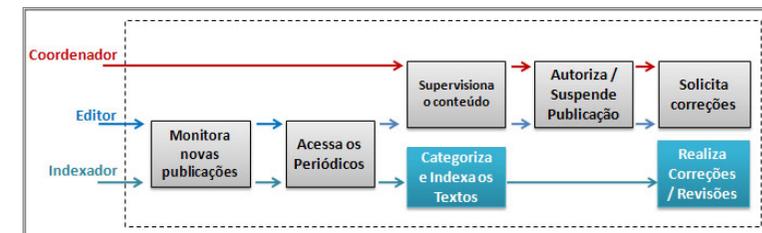


FIGURA 2 - PROCESSO PROPOSTO PARA INDEXAÇÃO E MANUTENÇÃO DA BRAPCI

Fonte: COSTA (2009)

Cada perfil definido pelo Grupo E3PI é autorizado a realizar determinadas atividades dentro da base, conforme descritas no Quadro 6, que apresenta todos os perfis de acesso.

Atividade	Perfil autorizado a realizar
Indexar novo Periódico	Coordenador
Indexar novas Edições	Editor / Indexador
Indexar novos Artigos	Todos
Realizar modificações nos Periódicos	Coordenador / Revisor
Realizar modificações nas Edições	Editor / Revisor / Indexador
Realizar modificações nos Artigos	Editor / Indexador
Excluir Periódicos	Nenhum
Excluir Edições	Nenhum
Excluir Artigos	Editor / Indexador
Autorizar/Suspender Publicação da Edição	Coordenador / Editor
Gerar Relatórios	Todos
Criar Usuários	Coordenador / Editor
Atribuir perfil ao Usuário	Coordenador / Editor
Coletar registros (OAI-PHM)	Todos
Consultar a base (módulo público)	Todos
Selecionar trabalhos e análises	Pesquisador

QUADRO 6 - ATIVIDADES PERMITIDAS A CADA PERFIL DE USUÁRIO
Fonte: Adaptado de COSTA (2009)

5 INDEXAÇÃO DE NOVOS FASCÍCULOS NA BASE

Racionalizar os processos administrativos e operacionais, como forma de melhoria na qualidade dos serviços prestados.

Após o monitoramento das fontes e identificação de novo fascículos publicado, devem ser indexados os artigos na Base Brapci.

Os campos as serem preenchidos na planilha com informações extraídas das revistas, devem estar de acordo com as normas de referência, NBR 6023: informação e documentação: referências – elaboração.

campo 3: a instituição, o depto., o curso, o cargo/função, profissão etc. do(s) autor(es), seguindo a ordem de entrada dos nomes dos autores. Ex.: 1. Profª. Drª.do CBD/ECA/USP. Coordenadora do Núcleo de Produção Científica. 2. Profª. Drª. do CBD/ECA/USP;

campo 4: título original do artigo, sem ponto final;

campo 5: seção do artigo, não indexar resenhas de livros, resenhas, teses/dissertações etc. Ex.: ARTIGO, RELATO DE EXPERIÊNCIA, ARTIGO DE REVISÃO, COMUNICAÇÃO ou PONTO DE VISTA;

campo 10: nome do periódico, destacar em negrito, e o local (cidade) de publicação;

campo 20: data de publicação do artigo;

campo 22: volume do periódico;

campo 24: número do periódico;

campo 25: número de páginas do artigo;

campo 37: contato do autor (endereço, e-mail ou telefone). No caso de vários autores, separar dados com ponto e vírgula;

campo 38: endereço eletrônico do artigo;

campo 40: número do ISSN do periódico;

campo 43: resumo do artigo;

campo 45: palavras-chave do artigo separadas por barras, são respeitados os termos atribuídos pelos autores, sejam provenientes de linguagem natural ou controlada.

A gestão das indexações dos artigos é armazenada automaticamente na tabela de Controle de Indexação, podendo ser consultada pelos gestores.

À medida que a base é modificada, deve ser realizado um *backup*, visando evitar possíveis perdas de informações no caso de ocorrer qualquer irregularidade.

6 SEÇÕES INDEXADAS NA BRAPCI

Na construção da política de incorporação deve-se também analisar quais as seções que serão disponibilizadas para consulta aos usuários, pois as revistas científicas produzem muitas seções que não podem ser consideradas como publicações científicas, muitas vezes publicando documentos, resumos de trabalhos apresentados em eventos, ou coletâneas de teses e dissertações defendidas dentro das instituições mantenedoras. Desta forma, a MB por definição armazena todos os tipos de trabalhos, qualquer que seja sua seção, porém disponibiliza para consulta no módulo público somente as seções definidas na política como visíveis para consulta, conforme os dados apresentados no Quadro 1.

SEÇÃO DE PUBLICAÇÃO	VISÍVEL PARA CONSULTA
Apresentação	Não
Artigo científico	SIM
Artigo Científico Temático	SIM
Artigos completos em anais de congresso	Não
Áudio e Vídeo	Não
Comunicação	SIM
Debate	SIM
Documentos	SIM
Editorial	SIM
Ensaio	SIM
Entrevista	SIM
Expediente	Não
In Memoriam	SIM
Normas de publicação	Não
Opinião	Não
Outros (não processados pela indexação)	Não
Palestra	Não
Ponto de vista	SIM
Prêmio	Não
Recensão	SIM
Relato de caso	SIM
Resenha	SIM
Resumos em anais de congressos	Não
Revisão de literatura	SIM
Sessão de estudos	Não
Teses e Dissertações	Não
Trabalhos acadêmicos	Não
Trabalhos apresentados em eventos	Não
Trabalhos de conclusão de curso	Não

QUADRO 1 - SEÇÕES DAS PUBLICAÇÕES DISPONÍVEIS NA METODOLOGIA BRAPCI
Fonte: O autor (2010)

7 VOCABULÁRIO CONTROLADO NA BRAPCI

A Brapci parte do princípio de que a informação disponível para ser analisada precisa estar organizada, com qualidade e ser precisa. Esta qualidade só pode ser almejada com o controle e padronização dos dados nela inseridos, principalmente com relacionamento de tabelas de dados e padronização nas entradas dos registros. Porém, esta rigidez na padronização não pode ser um entrave na disponibilização de informações, e estas devem estar em constante atualização com a incorporação de novos trabalhos, proporcionando ao usuário pesquisador uma fonte de pesquisa atualizada e fidedigna com a fonte original de informação.

Partindo desta concepção, a MB propõe uma abordagem mista, sem o rigor na coleta de dados com padronização de entradas, entretanto utiliza instrumentos para uma indexação pós-coordenada das informações quando analisadas pelos indexadores no momento de sua revisão. Essa abordagem gera uma perda da qualidade em um primeiro momento, com poucos critérios de padronização na entrada dos dados, mas com a vantagem de manter a base atualizada. O tratamento e padronização das informações são realizados em uma segunda etapa quando ocorre a revisão por um indexador. Cada nova regra criada pelos gestores da base é incorporada, como forma de “ensinar” o sistema a processar informações que coincidam com estas, facilitando e reduzindo trabalho repetitivo e economizando tempo na organização. Um ponto forte da metodologia é de que todos os trabalhos inseridos na base, em algum momento, devem passar por uma revisão e complementação do indexador.

O termos dentro da base são qualificados como:

- descriptores unilexicais e polilexicais, para definir um assunto;
- descriptores cronológicos, para definir um período histórico;
- descriptores geográficos, para definir nomes de cidades, estados, regiões geográficas;
- descriptores onomásticos, para definir nomes e instituições;
- outros, caso o descritor não se categorize em uma das opções acima;
- não analisado, para avaliação do indexador quando incorporado um novo descritor.

A Brapci parte da construção de uma ontologia para que se evidenciem as relações entre conceitos, por meio da construção de um vocabulário controlado, para garantir a convenção, a coesão e a consistência da entrada e manutenção dos dados. Como expresso na Figura 2, o pré-controle ocorre diretamente da linguagem natural, garantindo-se a velocidade de atualização e controle da terminologia, provendo a qualidade aos descritores da base e a possibilidade de definição de novos conceitos.

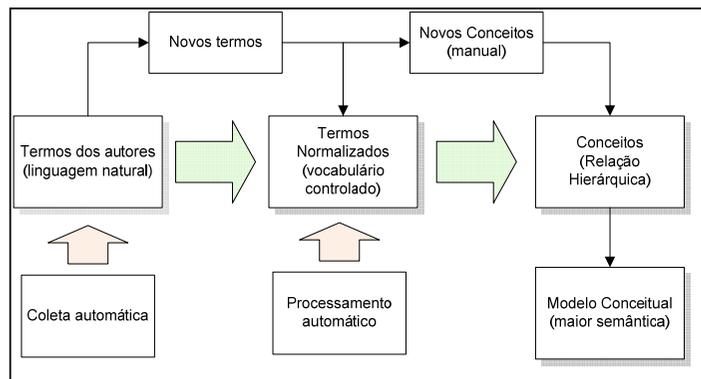


FIGURA 2 - FLUXO DA LINGUAGEM NATURAL PARA CONTROLADA – INDEXAÇÃO PÓS-COORDENADA PROPOSTA PELA METODOLOGIA

Fonte: O autor (2010)

A Brapci realiza uma coleta automática dos descritores (palavras-chave) de todos os trabalhos intelectuais disponíveis de forma automática por meio do protocolo OAI-PHM ou inserida manualmente pelos indexadores. Com a coleta os descritores são comparados aos já existentes na tabela de termos (não controlada), e em caso de não existirem, são incorporados à tabela. A entrada de novos termos é monitorada e analisada, identificando qual sua tipologia e associando-se à descrição de um conceito já existente como remissiva (USE), ou à criação de um novo conceito e indicação como termo preferencial.

8 CONTROLE DE NOMES DE AUTORES

Como no controle dos termos, a Brapci controla a representação do nome dos autores sejam identificados e inseridos na base conforme seu nome registrado no trabalho científico e que seja realizada uma pesquisa *a posteriori*, pela qual todos os autores e todas as formas de grafia sejam agrupados. Em caso de um mesmo autor ser representado de formas diferentes é recomendado o uso das remissivas (USE) como demonstrado no Quadro 2.

NOME CADASTRADO	NOME PADRONIZADO
CAFÉ, Lígia	CAFÉ, Lígia*
CAFÉ, Lígia Maria Arruda	use CAFÉ, Lígia
CAFÉ, Lígia Arruda	use CAFÉ, Lígia
Lígia Arruda Café	use CAFÉ, Lígia
Lígia Maria Arruda Café	use CAFÉ, Lígia
MIRANDA, Lígia Maria Café de	use CAFÉ, Lígia

QUADRO 2 - REMISSIVAS DE NOMES DE AUTOR

Fonte: O autor (2009)

NOTA:* nome definido pela Plataforma Lattes como forma de citação.

Desta forma, toda produção de um autor é concentrada em um único nome padronizado, evitando a dispersão e erros de autoria quando em uma análise deste tipo. A proposta é não alterar o nome pelo qual o trabalho foi representado, mas disponibilizar a informação como foi cadastrada, criando *links* para o agrupamento do autor ensejando a recuperação precisa e análise de autoria.

9 NUVEM DE TAGS

O acoplamento dos termos empregados pelos pesquisadores em suas estratégias de busca é possível porque o sistema armazena todas as formas de consulta do usuário. Este armazenamento possibilita a geração de relatório com a incidência dos termos em ordem decrescente, com limitação de 50 termos. Nesta lista gerada pelas consultas dos pesquisadores (*folksonomia*) são identificadas a maior (f_{max}) e a menor (f_{min}) frequência dos termos. Estes valores são aplicados na fórmula 1 para calcular o tamanho da fonte da letra será grafado o termo.

$$TagSize = 1 + C \frac{\log(f_i - f_{min} + 1)}{\log(f_{max} - f_{min} + 1)}$$

onde:

TagSize: resultado do tamanho da fonte
C: escala máxima do tamanho da fonte
f_i: frequência incidência do termo
f_{min}: menor valor de todos os descritores
f_{max}: pico de incidência do descritor.

Fonte: SINCLAIR; CARDEW-HALL (2008)

Para a base de dados, a nuvem de *tags* não só representa os termos e a temática mais consultada pelos pesquisadores, mas uma forma de disponibilizar conteúdo para ser indexado pelos robôs de busca (*Google, Yahoo, Bing*). A indexação por esses mecanismos aumentou significativamente a visibilidade e o número de acesso da Brapci, devido à múltipla indexação de conteúdo derivada dos termos das *tags*. Este fato não pode ser desprezado conforme demonstrado nas pesquisas de Quadros (2010), pelas quais mais de 75% dos acessos à Brapci são oriundos dos mecanismos de busca e somente 25% de sites referenciais e acesso direto.

10 CONTROLE E REGISTRO DE ACESSO

Com o objetivo de subsidiar a investigação científica, a MB tem como diretriz o controle e monitoramento das consultas realizadas na Base. Esses dados são armazenados de forma contínua e analisados quando solicitados.

Uma das aplicações que utiliza os resultados das consultas é a criação das *Nvens de Tags* da página inicial do módulo público da Brapci. Nesta geração o sistema ordena os cinquenta termos mais consultados e conforme sua incidência aumenta o tamanho de sua letra.

O monitoramento ainda possibilita a visualização das estratégias de busca que os usuários estão empregando na recuperação da informação, bem como a identificação dos temas mais consultados em determinado período.

No registro da estratégia de busca, a MB propõe o registro e dados adicionais do usuário como a data, hora e local (IP) de utilização do sistema, controlando a produção dos indexadores. Ferramentas de monitoração de visitas do site também podem ser empregadas com controle externo. É o caso do *Google Analytics* que cria outros indicadores como origem geográfica de acesso, tempo de permanência no site e índice de rejeição. A incorporação deste instrumento é realizada pela inserção de um código na própria página de consulta do módulo pesquisador.

Responsável	Rene Faustino Gabriel Junior	Versão	1.00
Data	Responsável	Descrição	
15/dez.2010	Rene F. Gabriel Junior	Criação do Documento	

COLETA DE TRABALHOS EM REPOSITÓRIOS OAI-PMH

1 FINALIDADE

Estabelecer diretrizes que deverão ser observadas e seguidas para a coleta de trabalhos em fontes de informações já indexadas na base Brapci. Tendo como objetivos principais a otimização do processo de coleta e incorporação ao corpus, garantindo a atualização e atualização dos dados disponibilizados.

2 DEFINIÇÕES E ADESÃO

2.1 Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

A base estrutural do protocolo OAI-PMH é composta de consultas por meios de “verbos” que um servidor faz a outro, trocando desta forma metadados em linguagem XML, o que possibilita a identificação de novos fascículos e novos trabalhos disponíveis na publicação sem a necessidade de entrar no site do provedor para verificar se houve atualizações. As revistas que publicam trabalhos intelectuais são denominadas de provedor de dados (data provider)

3 MONITORAMENTO DA INFORMAÇÃO

Para definição da política de indexação, a MB define que a base deve trabalhar somente com uma área do conhecimento específico, ou áreas que tenham relação muito próxima. Os procedimentos de aquisição e seleção de novas fontes de informação pautam-se por duas formas:

- automática, por meio do protocolo OAI-PHM, quando o sistema monitora novas publicações diretamente na fonte e as incorpora para indexação;
- manual, para publicações que não possibilitam esta coleta, definindo uma periodicidade para coleta manual de novos trabalhos.
-

3.1 Diretrizes

3.1.1 Racionalizar os processos administrativos e operacionais, como forma de melhoria na qualidade dos serviços prestados.

3.1.2 Monitorar os fascículos das revistas publicados:

As revistas que não são disponibilizadas eletronicamente devem ser solicitadas aos respectivos editores via *e-mail* ou correios à medida que são publicadas.

As publicações das revistas ambientadas na Plataforma SEER devem ser monitoradas por meio do sistema de Harvesting implementado na Brapci. Necessitando o cadastro do periódico e a porta de coleta de metadados pelo protocolo OAI-PMH.

O *e-mail* do Projeto Brapci deverá ser acessado diariamente pelo responsável pela indexação (qual é o e-mail da Brapci).

As demais revistas disponibilizadas eletronicamente ambientadas em plataformas diferenciadas deverão ser acompanhadas quinzenalmente para o monitoramento de possíveis novos fascículos em seus *websites*.

3.2 RESPONSABILIDADES

3.2.1 Dos responsáveis pela manutenção e atualização da base:

3.2.1.1 Gerenciar a base Brapci

3.2.1.2 Registrar, monitorar e avaliar os processos.

3.2.1.3 Zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci.

3.2.2 Dos líderes do Projeto Brapci:

3.2.2.1 Administrar os direitos de acesso das pessoas equipamentos para o desempenho de suas funções.

3.2.2.2 Zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci.

3.2.2.3 Prestar orientação às pessoas envolvidas no projeto.

3.2.2.4 Verificar o cumprimento das normas relacionadas.

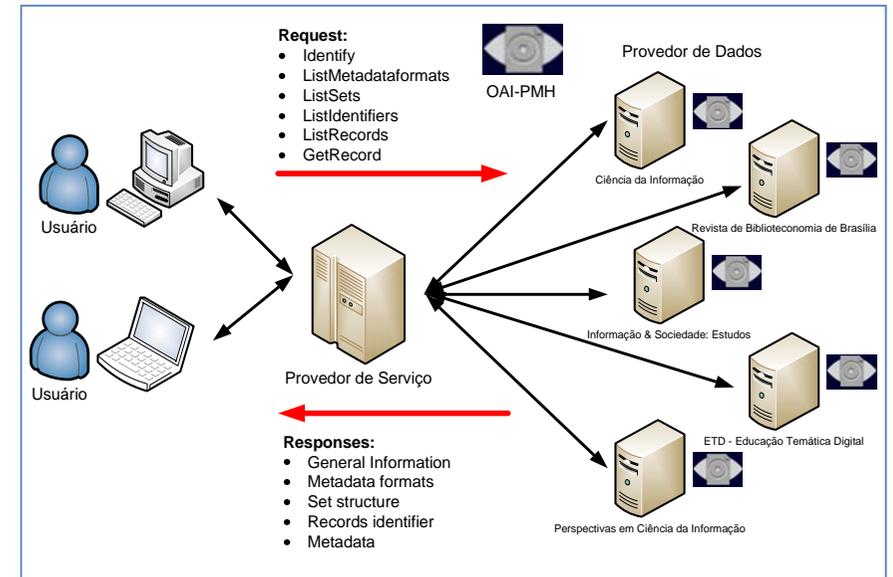
3.3 DISPOSIÇÕES GERAIS

3.3.1 Todos os ativos utilizados para desenvolvimento do projeto Brapci, constituem patrimônio público, da Universidade Federal do Paraná.

3.3.2 Compete aos que utilizam e manuseiam informações, conhecer e cumprir esta Política, para o bom uso das informações.

3.3.3 Compete aos responsáveis por processos que coletam informação: salvar as características da informação fornecida à base.

3.3.4 Compete aos responsáveis que tem a atribuição sobre determinada natureza de informação classificá-la de acordo com os critérios estabelecidos.



Quadro 1: FLUXOGRAMA DO PROCESSO DE MONITORAMENTO DAS FONTES E INDEXAÇÃO

4 INDEXAÇÃO DE NOVOS FASCÍCULOS NA BASE

4.1 Diretrizes

4.1.1 Racionalizar os processos administrativos e operacionais, como forma de melhoria na qualidade dos serviços prestados.

4.1.2 Após o monitoramento das fontes e identificação de novo fascículos publicado, devem ser indexados os artigos na Base Brapci.

4.2 Os campos a serem preenchidos na planilha com informações extraídas das revistas, devem estar de acordo com as normas de referência, NBR 6023: informação e documentação: referências – elaboração.

campo 3: a instituição, o depto., o curso, o cargo/função, profissão etc. do(s) autor(es), seguindo a ordem de entrada dos nomes dos autores. Ex.: 1. Profª. Drª. do CBD/ECA/USP. Coordenadora do Núcleo de Produção Científica. 2. Profª. Drª. do CBD/ECA/USP;

campo 4: título original do artigo, sem ponto final;

campo 5: seção do artigo, não indexar resenhas de livros, resenhas, teses/dissertações etc. Ex.: ARTIGO, RELATO DE EXPERIÊNCIA, ARTIGO DE REVISÃO, COMUNICAÇÃO ou PONTO DE VISTA;

campo 10: nome do periódico, destacar em negrito, e o local (cidade) de publicação;

campo 20: data de publicação do artigo;

campo 22: volume do periódico;

campo 24: número do periódico;

campo 25: número de páginas do artigo;

campo 37: contato do autor (endereço, e-mail ou telefone). No caso de vários autores, separar dados com ponto e vírgula;

campo 38: endereço eletrônico do artigo;

campo 40: número do ISSN do periódico;

campo 43: resumo do artigo;

campo 45: palavras-chave do artigo separadas por barras, são respeitados os termos atribuídos pelos autores, sejam provenientes de linguagem natural ou controlada.

4.3 Após indexação dos artigos deve ser preenchida a tabela de Controle de Indexação (Apêndice B) para atualização da mesma.

4.4 À medida que a base é modificada, deve ser realizado um *backup*, visando evitar possíveis perdas de informações no caso de ocorrer qualquer irregularidade.

4.5 RESPONSABILIDADES

4.4.1 Dos responsáveis pela manutenção e atualização da base:

4.4.1.1 Gerenciar a base Brapci

4.4.1.2 Registrar, monitorar e avaliar os processos.

4.4.1.3 Atentar para a qualidade da indexação para uma satisfatória recuperação da informação por parte dos usuários. Os elementos a serem observados remetem-se ao compromisso, responsabilidade, esforço e competência da equipe visando a atualização periódica dos dados, dentro dos padrões estabelecidos, controle de qualidade e segurança dos dados.

4.4.1.4 Responder pelas informações indexadas, visando sua confiabilidade.

4.4.1.5 Zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci.

4.5 Dos líderes do Projeto Brapci:

4.5.1 Administrar os direitos de acesso das pessoas equipamentos para o desempenho de suas funções.

4.5.2 Zelar pelo cumprimento destas políticas e pela segurança das informações da base Brapci.

4.5.3 Prestar orientação às pessoas envolvidas no projeto.

4.5.4 Verificar o cumprimento das normas relacionadas.

4.6 DISPOSIÇÕES GERAIS

4.6.1 Todos os ativos utilizados para desenvolvimento do projeto Brapci, constituem patrimônio público, da Universidade Federal do Paraná.

4.6.2 Compete aos que utilizam e manuseiam informações, conhecer e cumprir esta Política, para o bom uso das informações.

4.6.3 Compete aos responsáveis por processos que coletam informação: salvaguardar as características da informação fornecida à base.

4.6.4 Compete aos responsáveis que tem a atribuição sobre determinada natureza de informação classificá-la de acordo com os critérios estabelecidos.

Ícone	Situação do fascículo de cada revista
	Trabalhos coletados e não processados (indexação)
	Trabalhos indexados e prontos para primeira revisão
	Trabalhos com a primeira revisão pronta, encaminhado para segunda revisão
	Trabalhos conferidos e finalizados

REFERÊNCIAS

- BARROS, L. A. **Curso básico de terminologia**. São Paulo: EDUSP, 2004.
- BUFREM, L. S. Práticas de organização e divulgação da produção intelectual em ciência da informação no Brasil. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2008.
- BUFREM, L. S.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.
- CABRÉ, M. T. **La terminología**: teoría, metodología, aplicaciones. Barcelona : Ed. Antártida; Empúries, 1993.
- CABRÉ, M. Teresa. **La terminología**: representación y comunicación. Barcelona: Institut Universitari de Lingüística Aplicada, 1999. 369 p.
- CARNEIRO, Marília Vidigal. Diretrizes para uma política de Indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.
- COSTA, F. D. O. Opções metodológicas em pesquisa: modelização do novo sistema gerenciador da base brapci. In: EVINCI, 17., 2009, Curitiba. **Anais...** Curitiba: Universidade Federal do Paraná.
- FREITAS, J. L. **Proposta de metodologia para a recuperação da produção científica em Ciência da Informação em uma base de dados**. 2010. Trabalho de Conclusão de Curso (Graduação em Ciência e Gestão da Informação) - Universidade Federal do Paraná, 2010.
- FREITAS, J. L.; BUFREM, L. S.; GABRIEL JUNIOR, R. F. 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, jul./dez. 2010.
- OPEN ARCHIVES INITIATIVE. **OAI para principiantes**: tutorial en línea del Open Archives Forum. Disponível em <<http://travesia.mcu.es/portalnjb/jspui/html/10421/1823/intro.htm>>. Acesso em 13 mar. 2010.
- STIVAL, K. Opções metodológicas em pesquisa: validação do novo sistema gerenciador da base Brapci. In: EVINCI, 17., 2009, Curitiba. **Anais...** Curitiba: Ed. da UFPR, 2009.