



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

JOSÉ CARLOS FRANCISCO DOS SANTOS

**VOCABULÁRIO CONTROLADO EM PERIÓDICOS  
CIENTÍFICOS ELETRÔNICOS:  
UMA PROPOSTA DE CONTROLE DE TERMOS**

JOSÉ CARLOS FRANCISCO DOS SANTOS

**VOCABULÁRIO CONTROLADO EM PERIÓDICOS  
CIENTÍFICOS ELETRÔNICOS:  
UMA PROPOSTA DE CONTROLE DE TERMOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre.  
Linha de Pesquisa: Organização e Representação da Informação e do Conhecimento

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Brígida Maria Nogueira Cervantes

Londrina  
2015

### **Dados Internacionais de Catalogação-na-Publicação (CIP)**

S237v Santos, José Carlos Francisco dos.  
Vocabulário controlado em periódicos científicos eletrônicos: uma proposta de controle de termos / José Carlos Francisco dos Santos. – Londrina, 2015.  
144f. : il.

Orientador: Brígida Maria Nogueira Cervantes.

Dissertação (Mestrado em Ciências da Informação) – Universidade Estadual de Londrina, Centro de Educação, Comunicação e Artes, Programa de Pós-Graduação em Ciência da Informação, 2015.

1. Palavras-chave. 2. Controle de Vocabulário. 3. Ferramenta para Controle de Termos. 4. Vocabulário Controlado. I. Cervantes, Brígida Maria Nogueira. II. Universidade Estadual de Londrina. Centro de Educação, Comunicação e Artes. Programa de Pós Graduação em Ciência da Informação. III. Título.

CDU 025.43.06

JOSÉ CARLOS FRANCISCO DOS SANTOS

**VOCABULÁRIO CONTROLADO EM PERIÓDICOS CIENTÍFICOS  
ELETRÔNICOS:  
UMA PROPOSTA DE CONTROLE DE TERMOS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre.

**BANCA EXAMINADORA**



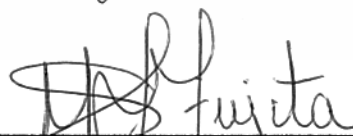
---

Prof<sup>a</sup>.Dr<sup>a</sup>. Brígida Maria Nogueira Cervantes  
UEL – Londrina - PR



---

Prof<sup>a</sup>.Dr<sup>a</sup>. Ana Esmeralda Carelli  
UEL – Londrina - PR



---

Prof<sup>a</sup>. Dr<sup>a</sup>. Mariângela Spotti Lopes Fujita  
Universidade Estadual Paulista – UNESP  
Câmpus de Marília

Londrina, 3 de julho de 2015.

*Dedico à minha esposa Fabiana,  
aos meus filhos: Antonio Carlos e José Eduardo,  
aos meus pais,  
aos meus familiares e amigos.*

## **AGRADECIMENTOS**

A Deus por mais uma etapa cumprida com êxito.

À minha família, pelo apoio durante este período de estudos e ausências, minha esposa Fabiana, meu filho Antonio Carlos e meu filho José Eduardo.

Aos meus pais Sebastião e Terezinha, pelo apoio de sempre quando diz respeito à ampliação dos conhecimentos.

Aos meus irmãos, cunhados, cunhadas, sobrinhos, sogro, sogra, e demais familiares que me apoiaram durante esta etapa.

À professora Brígida Maria Nogueira Cervantes, pela atenção, paciência e dedicação, pelos caminhos que me apontou e pela maneira exemplar com que conduziu a orientação deste trabalho.

Aos amigos do ESAP pelo apoio e compreensão nas minhas ausências.

À professora Neila Estigarribia, diretora do ESAP, por flexibilizar minhas atividades profissionais.

Ao professor Miguel Contani, pelas orientações em minha formação acadêmica e profissional e, em especial, por sua dedicação no atendimento de minhas solicitações durante este programa de Pós-Graduação.

Aos membros da banca, pelas preciosas observações desde o momento de qualificação deste estudo e pelas relevantes contribuições oferecidas.

A todo o corpo docente do Programa de Pós-Graduação em Ciência da Informação – PPCGI pelo conhecimento novo que permitiram adquirir e pelo crescimento que me proporcionaram em todas as oportunidades de contato direto.

Aos amigos da turma PPGCI 2013 e PPGCI 2012, pelo inesquecível convívio e pelo que compartilhamos de muitos momentos importantes de socialização do conhecimento.

Aos amigos e colaboradores de editoração da Revista Informação & Informação, Miguel Magarzo e Tânia Calheiros pelos inestimáveis serviços por eles executados para que fosse concluída, com êxito, a coleta de dados deste estudo.

A todos os que colaboraram, direta e indiretamente, com a minha formação acadêmica.

SANTOS, José Carlos Francisco dos. **Vocabulário controlado em periódicos eletrônicos: uma proposta de controle de termos**. 2015. 144f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Londrina, 2015.

## RESUMO

O vocabulário controlado é uma estrutura desenvolvida para ser aplicada nos processos de representação e recuperação de conteúdos informacionais. Diante da permanente necessidade de obter o controle de vocabulário das palavras-chave de artigos disseminados em periódicos científicos eletrônicos, origina-se a proposta da ferramenta *VCPC tools*, descrita neste estudo. Realizou-se um estudo piloto na Revista Informação & Informação, um periódico da Universidade Estadual de Londrina. O *corpus* foi constituído com dados levantados junto ao suporte periódicos científicos eletrônicos, com foco na representação temática da informação, no âmbito do controle de vocabulário, para indexação de conteúdos por meio da atribuição de palavras-chave. Os procedimentos principais envolviam a coleta de metadados e a validação das palavras-chave com dois vocabulários controlados da área da Ciência da Informação. Foram enunciados quatro requisitos funcionais mapeados conforme o reconhecimento do ambiente de periódicos científicos, considerando como se constituem, a ligação que promovem as palavras-chave ao estarem atribuídas do modo como foram encontradas e que efeito geram para a representação dos conteúdos temáticos. Dentre os resultados obtidos, destacou-se o índice de termos que apresenta também a correlação deste com os respectivos artigos científicos eletrônicos. Concluiu-se que a aplicação do controle de vocabulário nas palavras-chave, atribuídas em periódicos científicos, é fundamental para se obter consistência e relevância visando facilitar a recuperação de conteúdos.

**Palavras-chave:** Ferramenta para controle de vocabulário. Vocabulário controlado. Periódicos científicos eletrônicos.

SANTOS, José Carlos Francisco dos. **Controlled vocabulary in electronic journals: a proposal for term control**. 2015. 144p. Dissertation (Master in Information Science) - Londrina State University, Londrina, 2015.

### **ABSTRACT**

Controlled vocabulary is a structure designed for use in the processes of representation and recovery of information contents. To cope with the permanent need to obtain the vocabulary control of keywords in articles published by scientific electronic journals, were the VCPC tools created as shown in this study. A pilot study has been carried out at the Informação & Informação, a University of Londrina based periodical. Data have been collected in the support of electronic scientific periodicals with emphasis on thematic representation of information, in the field of vocabulary control, for content indexation through the assignment of keywords. The main procedures covered the metadata collection and the validation of keywords with two controlled vocabularies from Information Science area. Four mapped functional requirements have been stated, according to the recognition of the environment of scientific periodicals of how they are constituted, the connections keywords establish with the format they found with and what effect they generate for the representation of thematic content. Among the accomplished results, the index of terms was a prominent one and contained its correlation to the corresponding electronic scientific paper. The conclusion was that the application of the vocabulary control in keywords defined for scientific periodicals is of utmost importance to achieve consistency and relevance when aiming to facilitate content recovery.

**Keywords:** Tools for vocabulary control. Controlled vocabulary. Electronic scientific journals.

## LISTA DE QUADROS

<b>Quadro 1</b> - Funções do tesauro.....	27
<b>Quadro 2</b> - Tipologias de vocabulários controlados .....	31
<b>Quadro 3</b> - Demonstrativo das atualizações nas normas.....	33
<b>Quadro 4</b> - Indicadores colaboradores na construção, aperfeiçoamento e adequação do uso de linguagens documentárias .....	34
<b>Quadro 5</b> - Processo para obter o índice - indexação .....	46
<b>Quadro 6</b> - Estrutura do artigo.....	66
<b>Quadro 7</b> - Histórico das revistas científicas eletrônicas .....	70
<b>Quadro 8</b> - Principais acontecimentos OA .....	72
<b>Quadro 9</b> - Dados de acesso PCE .....	80
<b>Quadro 10</b> - Evolução do número de acessos ao Portal de Periódicos (2004-2013).....	84
<b>Quadro 11</b> - Ações decorrentes dos objetivos do estudo.....	88
<b>Quadro 12</b> - Tipos e funções de metadados baseado em Gilliland-Swetland .....	92
<b>Quadro 13</b> - Elementos <i>Dublin Core</i> .....	94
<b>Quadro 14</b> - Dicionário de dados do projeto do banco de dados .....	111
<b>Quadro 15</b> - Organização dos termos TBCI em tabela .....	115
<b>Quadro 16</b> - Descrições dos termos.....	116
<b>Quadro 17</b> - Estrutura vetor dados TCI .....	117
<b>Quadro 18</b> - Inserção do TBCI na base de dados.....	119
<b>Quadro 19</b> - Dados coleta de metadados.....	121
<b>Quadro 20</b> - Quantidade de palavras-chave por artigos.....	121
<b>Quadro 21</b> - Resultados de compatibilizações .....	123
<b>Quadro 22</b> - Resultados da compatibilização das palavras-chave.....	123
<b>Quadro 23</b> - Quantidade de palavras-chave x palavras-chave compatibilizadas por artigos.....	125

## LISTA DE FIGURAS

<b>Figura 1</b> - Diagrama de blocos: ferramenta .....	19
<b>Figura 2</b> - Metodologia de indexação para determinação das palavras-chave .....	40
<b>Figura 3</b> - Os novos paradigmas na ciência da informação.....	71
<b>Figura 4</b> - Evolução dos periódicos com texto completo (2004-2013).....	83
<b>Figura 5</b> - Diagrama de blocos processo de análise - redução de sufixo .....	98
<b>Figura 6</b> - Exemplos de cálculo do índice contido .....	99
<b>Figura 7</b> - Modelagem de banco de dados .....	101
<b>Figura 8</b> - Exemplo de construção de banco de dados em um SGBD .....	102
<b>Figura 9</b> - Interface do protocolo OAI-PMH do repositório digital .....	107
<b>Figura 10</b> - Parte código fonte da compatibilização de palavras-chave idênticas .....	108
<b>Figura 11</b> - Projeto físico do banco de dados estruturado relacional.....	109
<b>Figura 12</b> - Interface do <i>phpMyAdmin</i> com o banco de dados <i>controlevocabulario</i> .....	110
<b>Figura 13</b> - Interface de importação de metadados .....	120
<b>Figura 14</b> - Interface de análise de metadados - palavras-chave .....	122
<b>Figura 15</b> - Interface de resultados.....	123
<b>Figura 16</b> - Representação total de compatibilizações .....	124
<b>Figura 17</b> - Interface para tratamento manual de palavras-chave .....	126
<b>Figura 18</b> - Interface de acesso ao índice de termos.....	127
<b>Figura 19</b> - Interface da lista de termos .....	127
<b>Figura 20</b> - Interface da lista de artigos .....	128
<b>Figura 21</b> - Interface dos relacionamentos dos termos.....	128
<b>Figura 22</b> - Perfil da Informação & Informação - compatibilização de palavras-chave com o TBCI .....	129
<b>Figura 23</b> - Perfil da Informação & Informação - compatibilização de palavras-chave com o TCI .....	130

## LISTA DE ABREVIATURAS E SIGLAS

ACM	<i>Association of Computing Machinery</i>
APA	Associação de Psicologia Americana
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BIREME	Biblioteca Virtual em Saúde
BOAI	<i>Budapest Open Access Initiative</i>
Capes	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CD-ROM	<i>Compact Disc Read-Only</i>
CI	Ciência da Informação
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CV	Controle de Vocabulário
DC	<i>Dublin Core</i>
E-pub	Grupo de Publicações Eletrônicas em Medicina e Biologia
ER	Entidade Relacionamento
ERP	<i>Enterprise Resource Planning</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FTP	<i>File Transfer Protocol</i>
HAL	<i>Hyper Article en Ligne</i>
HTML	<i>Hypertext Mark-up Language</i>
HTTP	Hypertext Transfer Protocol
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IES	Instituição de Ensino Superior
IIB	<i>Institut International de Bibliographie</i>
JCU	<i>James Cook University</i>
LD	Linguagem Documentária
LN	Linguagem Natural
MTD-BR	Padrão Brasileiro de Metadados de Teses e Dissertações
MySQL	Sistema de Gerenciador de Banco de Dados
OAI	<i>Open Archives Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
oasisbr	Sistema Portal Brasileiro de Acesso Aberto à Informação Científica
OC	Organização do Conhecimento

OI	Organização da Informação
OJS	<i>Open Journal Systems</i>
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana da Saúde
ORC	Organização e Representação do Conhecimento
PC	Periódico Científico
PCE	Periódico Científico Eletrônico
PDF	<i>Portable Document Format</i>
PHP	Linguagem de Programação
PLN	Processamento de Linguagem Natural
PPGCI-UEL	Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina
PubMed	Serviço da Biblioteca Nacional de Medicina Americana
RC	Representação do Conhecimento
RF	Requisito Funcional
RI	Representação da Informação
SciELO	<i>Scientific Eletronic Library Online</i>
SEER	Sistema Eletrônico de Editoração de Revistas
SGBD	Sistema de Gerenciamento de Base de Dados
SGBDR	Sistema de Gerenciamento de Base de Dados Relacional
SGML	<i>Standard Generalized Markup Language</i>
SI	Sistemas de Informação
SOC	Sistemas de organização do conhecimento
SQL	<i>Structured Query Language</i>
TGS	Teoria Geral de Sistemas
TIC	Tecnologias da Informação e Comunicação
URL	<i>Uniform Resource Locator</i>
VCII	Vocabulário Controlado da Informação & Informação
VCPC <i>tools</i>	Nome da ferramenta
www	<i>Web ou World Wide Web</i>
XML	<i>eXtensible Markup Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>14</b>
<b>2</b>	<b>VOCABULÁRIO CONTROLADO COMO INSTRUMENTO PARA A ORGANIZAÇÃO E REPRESENTAÇÃO DE CONHECIMENTO</b> .....	<b>23</b>
2.2	NORMALIZAÇÃO E METODOLOGIA PARA CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS .....	31
2.3	PALAVRAS-CHAVE: ELEMENTO DE COMPATIBILIZAÇÃO.....	37
2.4	PROCESSO DE INDEXAÇÃO.....	42
<b>3</b>	<b>PERIÓDICOS CIENTÍFICOS ELETRÔNICOS: CONTEÚDO E ACESSO</b> .....	<b>55</b>
3.1	COMUNICAÇÃO CIENTÍFICA .....	55
3.2	PERIÓDICO CIENTÍFICO: PRINCIPAIS ABORDAGENS .....	59
3.2.1	Artigos Científicos.....	63
3.3	PERIÓDICO CIENTÍFICO ELETRÔNICO .....	67
3.3.1	Movimento de Acesso Livre.....	72
<b>4</b>	<b>METODOLOGIA</b> .....	<b>86</b>
4.1	CARACTERIZAÇÃO DA PESQUISA .....	86
4.2	ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC <i>TOOLS</i> .....	88
4.2.1	Identificar o Endereço do Repositório Digital do Periódico (1.1).....	88
4.3	ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS .....	102
4.3.1	Identificar os Vocabulários Controlados da Área Especializada (2.1) .	102
4.3.2	Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2) .....	102
4.3.3	Descrever em Linguagem <i>SQL</i> os Termos dos Vocabulários Controlados (2.3).....	103
4.3.4	Importar os Dados dos Termos dos Vocabulários Controlados (2.4) ..	103
4.4	ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS .....	103
4.4.1	Coletar e Armazenar os Metadados (3.1).....	103

4.4.2	Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2) .....	103
4.5	ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE.....	104
4.5.1	Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1) .....	104
4.5.2	Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2) .....	105
<b>5</b>	<b>ANÁLISE E RESULTADOS .....</b>	<b>106</b>
5.1	ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC TOOLS.....	106
5.1.1	Identificar o Endereço do Repositório Digital do Periódico (1.1).....	106
5.1.2	Desenvolver o Projeto da Ferramenta VCPC tools (1.2) .....	107
5.1.3	Desenvolver o Projeto Físico do Banco de Dados (1.3) .....	108
5.1.4	Criar a Base de Dados no Sistema Gerenciador de Banco de Dados (1.4).....	110
5.2	ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS .....	114
5.2.1	Identificar os Vocabulários Controlados da Área Especializada (2.1) .....	114
5.2.2	Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2) .....	115
5.2.3	Descrever em Linguagem SQL os Termos dos Vocabulários Controlados (2.3).....	117
5.3	ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS .....	120
5.3.1	Coletar e Armazenar os Metadados (3.1).....	120
5.3.2	Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2) .....	121
5.3.3	Verificar as Palavras-chave que Remetem à Igualdade dos Termos do Vocabulário Controlado (3.3).....	122
5.3.4	Vincular as Palavras-chave e Termos dos Vocabulários Controlados (3.4).....	124
5.4	ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE .....	125

5.4.1	Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1) .....	125
5.4.2	Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2) .....	126
5.5	ÍNDICE DE TERMOS.....	126
5.6	LIMITAÇÕES DA PESQUISA .....	131
<b>6</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>132</b>
6.1	RECOMENDAÇÕES E SUGESTÕES PARA FUTUROS TRABALHOS .....	135

## 1 INTRODUÇÃO

A origem do objeto de estudo da Ciência da Informação (CI) remonta a uma discussão que enfatiza a gestão do saber. Segundo Fernandes (1995), a proposta do objeto de estudo da CI é realmente a gestão institucional dos saberes e, ainda, as ações exercidas pelas instituições sobre o saber produzido pela sociedade e seus reflexos. Para Borko (1968, tradução nossa), a CI é uma disciplina que dá suporte informacional a várias instituições com procedimentos dedicados à gestão e transmissão do conhecimento, nos quais estão inclusos: os livros (conhecimento embalado); as escolas (repassando o conhecimento adquirido por várias gerações); as bibliotecas (armazenamento e disseminação do conhecimento); os filmes e televisão (forma visual do conhecimento); os jornais (comunicação escrita em campos especializados) e as conferências para comunicação oral da informação.

Diante das definições em torno da CI, torna-se essencial conceituar também duas disciplinas da CI, a Biblioteconomia e a Arquivologia, que estão envolvidas neste estudo. A biblioteconomia é “[...] parte da bibliologia que trata das atividades relativas à organização, administração, legislação e regulamentação das bibliotecas” (CUNHA; CAVALCANTI, 2008, p. 55). Já a arquivologia é uma “[...] disciplina que tem por objetivo o conhecimento dos arquivos e dos princípios e técnicas a serem observados na sua constituição, organização, desenvolvimento e utilização.” (CUNHA; CAVALCANTI, 2008, p. 30-31).

De modo geral, Moreiro González (2011, p. 34) faz menção à nova *web*, que tem melhorado a estrutura de organização do conhecimento por meio de ontologias, taxonomias, e a obrigação de se adaptar às linguagens preexistentes. As deficiências listadas são: a sobrecarga de Informação, a ineficácia das palavras-chave, falta de autoridade literária e/ou a falta de sistemas com tecnologias Processamento de Linguagem Natural (PLN). Essa nova *web* tem proximidades com a realidade dos periódicos científicos eletrônicos e está realmente aparente nesta perspectiva de estudo.

O compartilhamento de informações e conhecimento científico teve alterados seus processos de produção, disseminação, acesso e reprodução. Tais alterações foram proporcionadas a partir da implantação e/ou surgimento das Tecnologias da Informação e Comunicação (TIC) e, conseqüentemente, reelaboraram a dinâmica de organização da informação com novas ferramentas (COSTA; MOURA, 2013). Para

Vogel (2007), a comunicação da informação é possível por meio de uma linguagem, desde que tenha significação para os envolvidos. A circulação refere-se a processos que dependem do uso de uma linguagem que represente a informação de forma organizada e permita uma significação, como a Linguagem Documentária (LD). Essa construção, portanto, requer metodologia com conceitos linguísticos e terminológicos.

Antes mesmo de entrar na integração de conceitos da organização do conhecimento, faz-se importante distinguir os conceitos de conhecimento e de informação. Há distanciamentos e aproximações existentes na literatura e, para este efeito, “[...] entende-se conhecimento como o conjunto de conceitos (unidades do conhecimento) presentes em determinada área temática [...]” (BRANDT; MEDEIROS, 2010, p. 112). Para Currás (2010, p. 19), “Conhecimento pode ser entendido como um processo mental, inteligente, para adquirir saber, o que suporia também um passo intermediário na elaboração de linhas de opinião.” Relaciona-se com a quantidade de informação que tenha utilidade e que é despejada na mente humana com o objetivo de conduzir à produção de conhecimento. Também “[...] entende-se como conhecimento o saber acumulado no tempo, de forma que poderia se estabelecer uma quase-sinonímia ou comparação, com ciência e com cultura.” (CURRÁS, 2010, p. 19).

Para Vital (2010), o conhecimento ocorre quando algo se encontra “[...] no plano das ideias, em um processo de abstração que objetiva compreender o mundo e é individual [...]” (VITAL, 2010, p. 38). É importante mencionar que, na transição do mundo medieval para a modernidade, a constituição em torno do saber científico, aproximadamente nos “séculos XVI e XVII, pensadores como Galileu, Newton e Bacon propõem novas formas de conhecimento do mundo, que identificam no empirismo o núcleo central da experiência científica.” (ALMEIDA; CRIPPA, 2009, p. 114). Brascher e Café (2008) argumentam que os termos conhecimento e informação têm sido utilizados em diferentes contextos da área da CI, e o emprego desses termos “[...] revela falta de clareza quanto à delimitação do conceito [...]” (BRASCHER; CAFÉ, 2008, p. 2). Brascher e Café (2008) completam que “[...] como a informação o conhecimento, a OI e a OC também se inter-relacionam, mas são dois processos distintos”, assim também pode ocorrer na área de pesquisa, “[...] a OI e a OC são próximas e provavelmente compartilhem alguns aspectos teóricos e metodológicos comuns.” (BRASCHER; CAFÉ, 2008, p. 8). Existem dois tipos

diferentes de organização, “[...] um que se aplica às ocorrências individuais de objetos informacionais – [...] organização da informação, e outro que se aplica a unidades do pensamento (conceitos) – [...] organização do conhecimento.” (BRASCHER; CAFÉ, 2008, p. 6).

No que tange à OI, ela “[...] compreende [...] a organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções [...]”. Nesse contexto, tem-se “[...] a organização da informação nas bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos.”. Consequentemente será produzido o processo de RI, que envolve “[...] o conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo [...].” (BRASCHER; CAFÉ, 2008, p. 6).

Por outro lado, a informação deriva de uma evolução histórica que vem recebendo várias definições nos diversos meios de comunicação. O termo informação, segundo McGarry (1999), tornou-se popular com o nascimento da imprensa no século XV, uma vez que os acontecimentos passaram a ser descritos como informação. Os conceitos permitidos à expressão informação podem partir de um quase sinônimo de termo e alcançar o sentido de uma complexa descrição externa dos fatos, composta de termos que produzam efeitos no receptor, algo capaz de reduzir incertezas. Para bibliotecários, a informação implica a produção de termos ou conteúdo de um documento. A informação, de um modo geral, costuma ser vista como um mapa que as pessoas produzem em seu intelecto, capaz de conduzir um processo de aprendizagem necessário para a tomada de decisões (McGARRY, 1999).

Ainda nesta vertente, “A maneira de organizar o conhecimento, tendo em vista seu acesso universal, pressupõe o emprego da tecnologia em ampla escala.” Este contexto não é somente para supor o caminho a ser percorrido pelo livro universal no futuro, mas inserir no processo de criação o aparato tecnológico. Este livro poderá ser criado utilizando, como recurso, os “[...] repertórios universais, organizados pelo IIB, a instituição que materializa os ideais, metodologias, técnicas e tecnologias preconizada por Otlet.” Contudo Otlet visualiza os repertórios como sendo um progresso do livro, “[...] em direção ao ‘livro universal’, no qual todo o saber, na sua essência, estivesse objetivado, atualizado e organizado de maneira eficiente e facilmente acessível.” (SANTOS, 2007, p. 59).

Diante do contexto informacional, as publicações eletrônicas de periódicos

científicos, disponíveis na web, vêm aumentando consideravelmente no âmbito da CI e demais áreas especializadas. As discussões sobre a organização do conhecimento (OC) e sua representação na web são alvo de pesquisas e estudos da Biblioteconomia, como também da CI. A representação do contexto dos documentos, de maneira que seja uma representação fidedigna do conteúdo, sempre foi e será uma tarefa árdua e complexa, portanto se torna um grande desafio para os profissionais da área de Informação. Como consequência, esta tarefa de representar o contexto de documentos tem reflexo direto nos usuários, que têm a expressa necessidade de recuperar conteúdos de qualidade e de acordo com sua expectativa de leitura.

A representação do conhecimento (RC) é uma dimensão inerente ao controle de vocabulário (CV). Há periódicos científicos (PC) que utilizam tesouros eletrônicos para RC, e o fazem por meio de um instrumento documentário para padronização dos termos. Outros disponibilizam nuvem de *tags* para apresentar o conteúdo do periódico de forma hierarquizada.

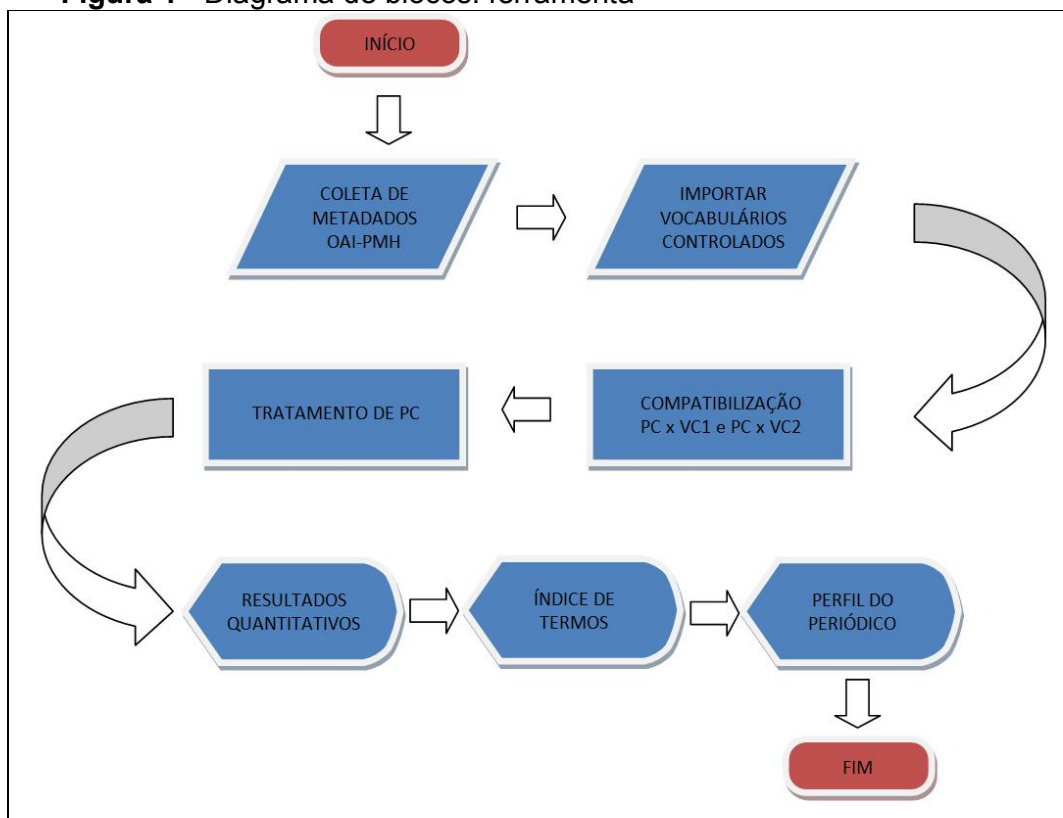
Dias e Cervantes (2012) relatam, em sua pesquisa descritiva com abordagem quali-quantitativa realizada nos periódicos científicos eletrônicos na área de Biblioteconomia e CI, que diante da questão “[...] Você considera importante o uso de instrumentos documentários para a representação das palavras-chave dos artigos? [...]” constatou-se que “[...] Apenas um periódico eletrônico considerou que não há necessidade de utilizar tais instrumentos documentários para a elaboração de palavras-chave [...]”. Dos 22 periódicos pesquisados que responderam à questão “[...] Qual a linguagem de representação utilizada para atribuição das palavras-chave? [...]”, 86% (19) afirmam que a linguagem de representação para a atribuição de palavras-chave é livre. O sentido desse termo livre é ambivalente, podendo implicar liberdade de usar qualquer critério ou não usar nenhum (DIAS; CERVANTES, 2012, p. 13).

Percebe-se que os periódicos utilizam a linguagem livre para a construção de palavras-chave e, na grande maioria, dão extrema importância para esta representação. Portanto cabe ressaltar a ineficiência da linguagem livre como representação das palavras-chave diante de diversos fatores, como sintaxe dos termos (ora com plural, ora sem plural, termos compostos separados por um traço ou não), semântica dos termos e sua organização, que levam a uma estrutura de palavras-chave inconsistente na base de dados dos periódicos.

Destaca-se, nesse ponto essencial, a área de estudo desta pesquisa que é a Ciência da Informação e sua subárea, a Organização e Representação do Conhecimento. O objeto de estudo deste trabalho visa proporcionar contribuições no contexto de organização do conhecimento, no que diz respeito ao processo de atribuição de conceitos ao documento analisado, aqui trabalhado com o suporte periódico científico eletrônico e as palavras-chave a ele atribuídas. Portanto a delimitação da temática desta pesquisa está centrada no conceito de Brascher e Café (2008) sobre a Organização e Representação do Conhecimento que, sendo fruto de um processo de análise de domínio com visão consensual da realidade, é a abstração do mundo real.

As dificuldades de visualizar as palavras-chave no momento em que vão sendo atribuídas são expressivamente grandes, em meio à quantidade de possibilidades, tanto de expressões diretas como de termos derivados constantes nas bases de dados de periódicos científicos eletrônicos. Quando se pensa em realizar uma busca de todas as palavras-chave, é necessário criar uma expressão de busca, porém, se não há conhecimento das palavras-chave existentes na base de dados, dificulta-se a implementação da expressão de busca. Exatamente nesse contexto reside a relevância de buscar as palavras-chave sem deixar de conhecer as palavras-chave que estão na base, surge assim o interesse de investigação desta pesquisa, que é encadeada pela problemática: Como realizar o controle do vocabulário das palavras-chave atribuídas aos artigos científicos disponíveis em periódicos científicos eletrônicos?

Portanto esta pesquisa intitulada “Vocabulário controlado em periódicos científicos eletrônicos: uma proposta de controle de termos” teve, como proposição, o desenvolvimento de uma ferramenta para a realização do controle de vocabulário com a utilização de vocabulários controlados para a compatibilização das palavras-chave atribuídas aos artigos contidos nos periódicos científicos eletrônicos, como um aporte teórico-metodológico. Por meio da Figura 1, é possível visualizar a análise sistêmica da ferramenta.

**Figura 1 - Diagrama de blocos: ferramenta**

Fonte: Elaborado pelo autor.

As palavras-chave têm um papel importante para a recuperação, porém nem todos os periódicos têm dado a atenção necessária para este viés. Em muitos periódicos, as orientações/normas de atribuição das palavras-chave não estão registradas de forma explícita; alguns periódicos contam com profissionais que auxiliam no processo de escolha de palavras-chave que representem o conteúdo de forma equivalente ao artigo, porém este é um procedimento com alto custo para a instituição editora do periódico e conseqüentemente compromete a sua saúde financeira. Comumente são vistas, nas regras de submissão, orientações para os autores tomarem como base e executar a atribuição, da melhor forma, como indício para uma representação fidedigna de seu texto. Este é um fato recorrente que ainda é pouco explorado cientificamente (GONÇALVES, 2008).

Para Miguéis e Neves (2013), palavras-chave se distanciam do termo de indexação, porque o termo de indexação tem o resultado centrado no processo de análise conceitual que faz a tradução para o VC, e por seu turno, a atribuição de palavras-chave é livre, sendo escolhidas pelo autor, editor ou até mesmo geradas de

maneira automática. Portanto palavra-chave é uma palavra considerada mais representativa no texto, já o termo é uma palavra ou expressão atribuída ao texto para ser incluído nos sistemas de informação (CUNHA; CAVALCANTI, 2008).

Reconhecer a dificuldade com os resultados de busca a determinadas produções científicas, por intermédio de palavras-chave, é o primeiro fator de motivação essencial no sentido de serem melhorados os critérios de formulação da RC por meio das palavras-chave. Situa-se, principalmente, nessa condição, a relevância de estudar formas de aperfeiçoar o uso de instrumentos documentários para RC, sempre se voltando para a importância de sua recuperação.

De acordo com Gil Leiva e Alonso Arroyo (2005), cabe reiterar a importância das palavras-chave atribuídas de maneira livre pelo autor do artigo científico. Referidos autores, em estudo comparativo de palavras-chave e descritores dos artigos avaliados, constataram que, em média, 32% das palavras-chave são exatamente iguais e, em média, 31% poderiam ser convertidas em descritores com pequenas correções de padronização, como plurais, ortografias, entre outras.

As principais justificativas deste estudo estão norteadas na busca incessante de soluções da OC e RC em diversos âmbitos das áreas especializadas e são a razão e importância desta pesquisa, que está vinculada na produção de opções para fundamentar melhores práticas na OC. As contribuições estão refletidas na produção de resultados de uma pesquisa com procedimentos para aplicação nos periódicos científicos eletrônicos, estas com impacto na sociedade do conhecimento, que tem como foco o usuário (pesquisador e autores).

É importante salientar que os pesquisadores e autores poderão se beneficiar das conclusões para a escolha dos termos a serem atribuídos para os seus próprios artigos, o que fará com que os usuários possam ter melhor acesso aos conteúdos indexados nos periódicos científicos. Por outro lado, legitima-se a iniciativa em decorrência das contribuições diretas para a discussão desta temática e aporte para o desenvolvimento de uma proposta de instrumentos de organização e visualização das palavras-chave contidas nas bases de dados dos periódicos eletrônicos. Na realidade atual, as palavras-chave dos artigos dos periódicos não são padronizadas, e isso torna impossível a visualização de maneira organizada.

Outro elemento importante de motivação é o fato de este estudo integrar o projeto de pesquisa coordenado pela professora orientadora desta pesquisa, vinculado ao Programa de Pós-Graduação em Ciência da Informação da

Universidade Estadual de Londrina (PPGCI-UEL). Também é importante salientar que uma grande motivação está no contexto de integração de conhecimentos de duas áreas, a Tecnologia da Informação e a Ciência da Informação (formação do autor), para proporcionar contribuições relevantes a essas e às demais áreas especializadas. Ainda uma relevância a destacar deste estudo são as áreas em que poderá ser aplicado. Inicialmente será posto em prática na área da Ciência da Informação como estudo piloto na revista Informação & Informação do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, e a perspectiva é de que esse teste possa avaliar as configurações para serem adaptadas e aplicadas em qualquer área especializada do conhecimento.

Diante da problemática deste projeto e dos vínculos do estudo com a linha de pesquisa, teve-se como objetivo geral implantar o controle de vocabulário, tendo como base as palavras-chave atribuídas no periódico científico eletrônico Informação & Informação, para aprimorar a recuperação dos conteúdos temáticos.

Os objetivos específicos foram:

- Desenvolver a ferramenta VCPC *tools*;
- Coletar e armazenar os vocabulários controlados;
- Compatibilizar as palavras-chave com os termos dos vocabulários controlados;
- Tratar, manualmente, as palavras-chave não compatibilizadas automaticamente.

Esta pesquisa caracterizou-se como exploratória e descritiva, com uma abordagem mista qualitativa e quantitativa. O trabalho encontra-se estruturado em seis tópicos, além das referências.

O tópico 2, vocabulário controlado como instrumento para a organização e representação de conhecimento, trata da linguagem documentária, da normalização e metodologia para construção de vocabulários controlados, das palavras-chave: elemento de compatibilização, e do processo de indexação.

O tópico 3, periódicos científicos eletrônicos: conteúdo e acesso, aborda o seu surgimento com a comunicação científica, os aspectos históricos e os conceitos do periódico científico, suas principais abordagens, os artigos científicos e suas características e estrutura, o periódico científico eletrônico e suas fases históricas, o movimento de acesso livre.

O tópico 4, metodologia, trata da caracterização da pesquisa, das etapas da proposta metodológica, que são: Etapa 1 - desenvolver a ferramenta VCPC *tools*; Etapa 2 - coletar e armazenar os vocabulários controlados; Etapa 3 - compatibilizar as palavras-chave com os termos dos vocabulários controlados; e Etapa 4 – tratar, manualmente, as palavras-chave não compatibilizadas automaticamente .

No tópico 5, análise e resultados, são apresentados os resultados das etapas metodológicas e suas ações: Etapa 1 - desenvolver a ferramenta VCPC *tools*: identificar o endereço do repositório digital do periódico (1.1), desenvolver o projeto da ferramenta VCPC *tools* (1.2), desenvolver o projeto físico do banco de dados (1.3), criar a base de dados no sistema gerenciador de banco de dados (1.4); Etapa 2 - coletar e armazenar os vocabulários controlados: identificar os vocabulários controlados da área especializada (2.1), estruturar os dados dos termos dos vocabulários controlados (2.2), descrever, em linguagem SQL, os termos dos vocabulários controlados (2.3), importar os dados dos termos dos vocabulários controlados (2.4); Etapa 3 - compatibilizar as palavras-chave com os termos dos vocabulários controlados: coletar e armazenar os metadados (3.1), verificar as palavras-chave que são idênticas com os termos dos vocabulários controlados (3.2), verificar as palavras-chave que remetem à igualdade dos termos do vocabulário controlado (3.3), vincular as palavras-chave e termos dos vocabulários controlados (3.4); Etapa 4 - tratar manualmente as palavras-chave não compatibilizadas automaticamente: verificar as palavras-chave que tenham termos correspondentes nos vocabulários controlados (4.1), inserir a palavra-chave como termo no vocabulário controlado da revista (4.2). Também abordam-se os resultados disponibilizados aos usuários da revista por meio do índice de termos e as limitações da pesquisa.

O tópico 6, considerações finais, destaca as inferências sobre os resultados e as análises em correlação com os objetivos e proposição do estudo, assim como apresenta recomendações e sugestões para futuros trabalhos e fornece as referências que fundamentaram esta dissertação.

## 2 VOCABULÁRIO CONTROLADO COMO INSTRUMENTO PARA A ORGANIZAÇÃO E REPRESENTAÇÃO DE CONHECIMENTO

O entendimento de Brandt e Medeiros (2010), quanto à organização do conhecimento, é o de que ela está ligada ao processo de análise conceitual de um domínio do conhecimento. A estruturação e execução desses processos geram uma representação do conhecimento, “[...] que será então usado para a organização da informação desse domínio de conhecimento produzida.” (BRANDT; MEDEIROS, 2010, p. 112). Brascher e Café (2008) argumentam nessa direção e complementam: “A organização do conhecimento [...] visa à construção de modelos de mundo que se constituem em abstrações da realidade.” (BRASCHER; CAFÉ, 2008, p. 6). Currás (2010, p. p. 25-26) afirma que “Na maioria dos casos ‘sistema de classificação’ – temática – figura como sinônimo da organização do conhecimento.”

Diante desse contexto, agem os sistemas de organização do conhecimento que são requeridos, por meio dos sistemas que executam a recuperação de documentos, “[...] base de dados eletrônicos, catálogos coletivos e informatizados, bancos de dados, repositórios, portais, gerenciadores de publicações eletrônicas [...]” Esses documentos são tratados nos suportes impresso e eletrônico, ambos armazenados em unidades de informação, sendo as “[...] bibliotecas tradicionais ou em bibliotecas digitais ou virtuais disponíveis na internet.” (BOCCATO, 2011b, p. 167). Os referidos Sistemas de Organização do Conhecimento (SOC) remetem aos instrumentos que geram as estruturas semânticas de conhecimento que é qualificado em um determinado domínio, e as estruturas semânticas estabelecem relações entre elas (BOCCATO, 2011b; BRASCHER; CAFÉ, 2008).

Nessa perspectiva do SOC, vale ressaltar os instrumentos documentários instaurados por Otlet, quando do seu dedicado trabalho no *Institut International de Bibliographie* (IIB), com todo o mérito, na execução das atividades para integração da teoria e prática num incessante esforço “[...] para a consolidação de novas metodologias para a análise e síntese do conhecimento, visando à sua circulação.” Neste seguimento, Otlet tem “[...] a crença de que a universalização do acesso ao conhecimento seria o caminho para a paz mundial. Esses aspectos são indissociáveis da proposta de criação do campo da documentologia ou bibliologia.” (SANTOS, 2007, p. 55).

As estruturas semânticas de conhecimento de Brascher e Café (2008) são

contempladas também por Boccato (2011b, p. 167) e complementa que a sua função é desenvolvida “[...] pelos sistemas de classificação, listas de cabeçalhos de assuntos, taxonomias, tesouros, ontologias [...]”. Para a construção e atualização dessas estruturas, são necessárias metodologias específicas, dependendo da especificidade do uso ou determinada aplicação (BOCCATO, 2011b). Os instrumentos “[...] de organização, representação e recuperação da informação, tais como os sistemas de classificação, as listas de cabeçalhos de assuntos, os tesouros, as taxonomias, as ontologias [...]” colaboram para a “[...] institucionalização social da ORC [...]”. Em geral, esses instrumentos se relacionam com a organização conceitual e caminham para um controle de vocabulário, o que se pode considerar como vocabulários controlados (MOREIRA et al., 2015, p. 2).

Barité Roqueta (2011) ressalta que a expressão “sistemas de organização do conhecimento” é o termo que compreende o conjunto de ferramentas destinadas para realização de classificação e indexação. Assim esse termo é geral e abrangente, tanto para as linguagens criadas para indexar e classificar, como para outros instrumentos nascidos para servir de referência conceitual.

As estruturas ou instrumentos instanciados para a organização do conhecimento têm, como objetivo, a representação do conhecimento por meio dos sistemas de organização do conhecimento e, nesse âmbito, inserem-se os conceitos pertinentes de representação da informação e do conhecimento. Novellino (1996) denomina como representação da informação, que por sua vez é composta por dois processos basilares, um deles sendo “[...] análise de assunto de um documento e a colocação do resultado desta análise numa expressão linguística”, o segundo, “[...] atribuição de conceitos ao documento analisado.” (NOVELLINO, 1996, p. 38).

Nesse segundo processo, tem-se “[...] uma linguagem documentária, instrumento de padronização da indexação, a qual visa garantir que indexadores de um mesmo sistema ou sistemas afins usem os mesmos conceitos para representar documentos semelhantes.” (NOVELLINO, 1996, p. 38). Por outro lado, funciona como um instrumento de comunicação, quando possibilita a troca de informações referentes a um vocabulário entre os próprios profissionais de indexação (NOVELLINO, 1996, p. 38). Dos conceitos de linguagem documentária, trata o próximo subtópico.

## 2.1 LINGUAGEM DOCUMENTÁRIA

Linguagem documentária ou linguagem de indexação, assim denominada na literatura, “[...] é um conjunto controlado de termos dotados de regras sintáticas e semânticas [...]”, com o objetivo de representar os “[...] conceitos significativos dos assuntos dos documentos durante a indexação na fase de tradução e durante a busca na representação do assunto de interesse do usuário.” (FUJITA, 2011, p. 36). Cunha e Cavalcanti (2008, p. 227) sintetizam a linguagem documentária como “[...] termos, símbolos e regras preestabelecidos para indicação/registro de assuntos constantes de documentos.” Diante da função que a linguagem documentária tem de representação de conteúdos documentários com vistas à recuperação, ela pode ser considerada “[...] como mecanismo de mediação, o que significa que não representa o próprio texto a ponto de substituí-lo.” (FUJITA, 2011, p. 40).

Salienta-se a importante consideração em torno das “[...] novas linguagens documentárias requeridas pelos documentos eletrônicos e sua gestão, sem por isto modificar a estrutura básica dos pré-existentes.” (MOREIRO GONZÁLEZ, 2011, p. 14). Fujita (2011) menciona duas formas de acesso às informações organizadas por assunto em base de dados, e elas ocorrem: “[...] pela linguagem natural, que não possui controle de vocabulário, e pela linguagem documentária, que propicia o controle de vocabulário.” (FUJITA, 2011, p. 42).

Para o processo de tratamento temático da informação, as “[...] linguagens de indexação, instrumentos que fornecerão os termos padronizados para representar o assunto ou assuntos identificados nos documentos analisados.” Há dois tipos de linguagem de indexação que são principais: “[...] linguagens alfabéticas e linguagens simbólicas. Os principais tipos de linguagens alfabéticas que existem são as listas de cabeçalhos de assuntos e os tesouros.” Já os cabeçalhos são de uso mais geral, em bibliotecas públicas ou com grande abrangência de assuntos. Por outro lado, “[...] os tesouros [...] são voltados para coleções especializadas, seja em determinado campo do conhecimento, seja em qualquer área orientada para uma missão.” Para representar os assuntos dos documentos por meio de símbolos, usam-se as linguagens simbólicas, que por sua vez limitam-se aos sistemas de classificação bibliográfica (DIAS; NAVES, 2013, p. 13-14).

A finalidade do tratamento temático do conteúdo é “[...] uma descrição do conteúdo para uma representação condensada do que está expresso no texto a fim

de propiciar uma acessibilidade temática.” (FUJITA, 2011, p. 38). As operações que esse tratamento utiliza são: “[...] análise, síntese e representação para realizar a classificação, a indexação, a catalogação de assuntos e a elaboração de resumos [...]”. Desse modo, obtêm-se diversos tipos de resultados referentes ao nível de condensação, por exemplo, o resumo “[...] apresenta um nível de condensação menor que os termos de indexação ou número de classificação.” (FUJITA, 2011, p. 38).

Novellino (1996) define que a lista foi desenvolvida estritamente para instrumento padronizador da indexação, ao passo que o tesouro foi pensado “[...] como instrumento facilitador da comunicação dentro do sistema, padronizando as linguagens de indexação e de recuperação, a partir da terminologia da área representada.”, diante da necessidade de explicitar-se “[...] a organização de determinadas áreas de assunto conduziu ao tesouro facetado e ao classauro.” (NOVELLINO, 1996, p. 40). Dentre as vantagens da aplicação do tesouro está a “[...] a possibilidade de expressar o conjunto completo de relações associativas entre conceitos e não apenas relações genéricas.” Por outro lado, tem-se o que pode ser considerado familiar aos indexadores e usuários, como os termos utilizando a linguagem natural de um tesouro, diferentemente das notações de qualquer sistema de classificação (NOVELLINO, 1996, p. 41).

Os tesouros, que são linguagens documentárias, têm, em sua composição, os “[...] descritores relacionados entre si, semanticamente e genericamente, que podem ser combinados ou não na indexação.” Assim procuram representar, no primeiro momento, os conceitos e, posteriormente, especificam suas relações. Esses descritores são “[...] desenvolvidos para sistemas pós-coordenados de recuperação da informação. Constituem-se como linguagem documentária, controlada, pós-coordenada e combinatória.” (FUJITA, 2011, p. 35). Na norma NBR 12676, é dado como equivalente ao termo preferido, que é o “termo utilizado consistentemente na indexação para representar um conceito; algumas vezes conhecido como descritor.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

Zavitoski (2001, p. 36) apresenta as funções do tesouro com suas respectivas descrições, demonstradas no Quadro 1.

**Quadro 1 - Funções do tesauro**

FUNÇÕES	DESCRIÇÃO	EXPLICAÇÃO
Principal	Compatibilização	“[...] compatibilizar os diversos termos usados na linguagem natural com um termo específico, que defina mais apropriadamente o conceito que se deseja representar [...]”
Segunda	Remissiva	“[...] sistema de remissiva, onde termos sinônimos e quase-sinônimos, bem como as variações lexicais, vão compor um vocabulário que remeterá os usuários aos termos controlados do tesauro.”
Terceira	Normalizadora	“[...] Quando seleciona o termo que melhor representa o conceito de um domínio de especialidade. Ao representar [...] somente um termo deve ser usado [...] - o termo preferido [...]”
Quarta	Ferramenta	“[...] Auxiliar na compreensão geral das áreas do conhecimento, por possibilitar a elaboração de mapas semânticos [...]”

Fonte: Elaborado pelo autor, fundamentado em Zavitoski (2001, p. 36).

Zavitoski (2001) aponta que a macro-organização das Linguagens Documentárias é lógico-hierárquica, enquanto que no tesauro a configuração interna é lógico-semântica. A estrutura básica do tesauro é composta pelas relações hierárquicas, o termo superordenado “[...] representa uma categoria, classe ou o todo de que o termo subordinado é membro ou parte.” Essas relações são expressas por TG (termo genérico) e TE (termo específico). As relações genéricas estabelecem ligações entre a classe e seus respectivos membros. Já as partitivas são as relações todo/parte e as não hierárquicas denominadas de associativas (ZAVITOSKI, 2001, p. 38-40).

Novellino (1996) afirma que “Nem sempre as linguagens documentárias foram pensadas como instrumentos de indexação e recuperação. Inicialmente, elas tinham como objetivo apenas padronizar as entradas de assunto [...].” A partir destes instrumentos (lista de cabeçalhos de assuntos, vocabulário livre e controle de vocabulário aplicado a lista de termos) utilizados na representação da informação, as linguagens documentárias tiveram, como consequência, a missão de mostrar “[...] ao usuário a estrutura da linguagem de representação [...]” que “[...] deu origem aos tesouros, tesouros facetados e classauros.” (NOVELLINO, 1996, p. 39). Por outro lado, Currás (2010, p. 87) afirma que “[...] uma lista de termos de indexação, ou um cabeçalho de assunto serão classificações documentárias pós-controladas.”

Para Boccato (2011a), “as listas de cabeçalhos de assunto são linguagens pré-coordenadas, de estrutura associativa ou combinatória de palavras ou expressões, que têm a finalidade de representar os conteúdos documentários.” Elas têm regras “[...] específicas para as formas de entrada dos cabeçalhos, do uso de abreviaturas e, geralmente, arrolam palavras correspondentes a diversas áreas do

conhecimento.” (BOCCATO, 2011a, p. 13).

Fujita (2011) argumenta que os critérios para se definir uma coordenação são estabelecidos “[...] pelo momento da combinação dos termos: se realizada durante a descrição, a linguagem é pré-coordenada e, se realizada durante a estratégia de busca, a linguagem é pós-coordenada.” (FUJITA, 2011, p. 43-44).

Para que a representação da informação seja adequada é necessário o cumprimento de um requisito de elaboração das linguagens documentárias. Assim “as linguagens documentárias devem ser elaboradas com base na terminologia de uma área, com categorias e subcategorias delimitadas, a partir de relações lógico-semânticas de termos claros e bem definidos [...]” (BOCCATO, 2011a, p. 15-16).

Na construção das linguagens documentárias, observa-se a contribuição da Terminologia “[...] por meio de princípios norteadores na identificação de áreas científicas, na determinação de conceitos e termos, na consistência das relações lógico-semânticas.” Portanto o meio de atuação das linguagens documentárias está relacionado com o “[...] universo de significações livres da língua, enquanto a Terminologia se preocupa com as significações conceituais e relacionais provenientes do discurso, denominadas termos.” (BOCCATO, 2011a, p. 16).

Os conceitos representados por termos na construção de linguagens documentárias são decorrentes de dois elementos essenciais, que são “as unidades da linguagem natural e das linguagens de especialidades [...]”. Diante disso, temos a linguagem documentária sendo um “[...] um canal de comunicação social [...]”, com grande valoração, os “[...] conceitos representados por termos devem refletir a cultura do indivíduo, do ambiente em que ele está inserido e da área de conhecimento a que ele corresponde.” (BOCCATO, 2011a, p. 16).

Para Novellino (1996), “as listas de cabeçalhos de assunto foram construídas para instrumentalizar a indexação de assuntos de documentos, que seriam registradas em fichas catalográficas para compor o catálogo alfabético de assuntos.” Essa é uma maneira de organizar o conhecimento do universo em disciplinas nas unidades de informação em seu acervo (NOVELLINO, 1996, p. 39).

Termo, para Cunha e Cavalcanti (2008, p. 360), é “[...] palavra (ou expressão) empregada para a inclusão temática de um item em sistemas de informação e para a recuperação posterior [...]”, que equivale a palavra, rótulo semântico. Já para a norma internacional ANSI/NISO Z39:19-2005, termo é uma ou mais palavras que designam um conceito. Pode ser composto por um termo de

entrada ou termo pré-coordenado (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

Na norma internacional ANSI/NISO Z39:19-2005, o termo de indexação é a representação de um conceito numa linguagem de indexação, geralmente na forma de um substantivo ou um sintagma nominal (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). Termos, cabeçalhos de assuntos e a combinação cabeçalhos e subcabeçalhos são exemplos de termos de indexação, também chamados de descritores. A NBR 12676 afirma que termo de indexação é “Representação de um conceito sob uma das seguintes formas: a) termo derivado da linguagem natural, de preferência um nome ou uma locução nominal; b) símbolo de classificação.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

O descritor tem atuação também no termo não preferido, que para a NBR 12676 é o “sinônimo ou quase sinônimo do termo preferido (descritor); também conhecido como ‘não-descritor’.” Assim não se faz atribuição aos documentos, mas utiliza-se “[...] como remissiva no índice, para instruir o usuário (Use ou Ver) na procura do termo preferido.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

Novellino (1996) argumenta que o vocabulário livre tem as características dos primeiros sistemas pós-coordenados. “Atribuíam-se quantos termos isolados fossem necessários para descrever determinado documento cabendo ao usuário coordená-los no momento da busca.” (NOVELLINO, 1996, p. 39). Um problema decorrente encontrado é a ausência de um instrumento para padronização nas linguagens dos usuários e produtores de informação, isto é, a “multiplicidade de termos para representar um mesmo conceito, descontextualização dos termos em relação ao assunto total do documento e também da área de domínio da qual fazia parte.” (NOVELLINO, 1996, p. 40).

O controle de vocabulário é utilizado para aumentar a eficácia do armazenamento de informações, dos sistemas de recuperação, dos sistemas de navegação na *Web* e de outros ambientes que precisam identificar como localizar o conteúdo desejado por meio de algum tipo de descrição utilizando uma linguagem. O objetivo principal do controle de vocabulário é alcançar consistência na descrição de objetos de conteúdo e facilitar a recuperação (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS

ORGANIZATION, 2005).

Para Smit e Kobashi (2003), “O controle de vocabulário é um recurso para organizar e recuperar documentos – e informações – com consistência, gerando, conseqüentemente, confiança no sistema.” (SMIT; KOBASHI, 2003, p. 14). É importante diferenciar dois momentos no controle e no uso controlado de vocabulário: construção e utilização. Na construção, existem variáveis que devem ser levadas em consideração, como “Política da instituição como um todo (sua missão e objetivos); Atividades/funções envolvidas na documentação a ser organizada (linguagem de especialidade)”. Na utilização do vocabulário controlado, as variáveis são: “Política de acesso à informação – Política de indexação; Público-alvo.” (SMIT; KOBASHI, 2003, p. 18).

Vocabulário controlado visa à uniformidade de armazenagem e facilidade de recuperação, por meio de um “conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como no da recuperação.” O objetivo é aproximar as linguagens do pesquisador com o indexador (CUNHA; CAVALCANTI, 2008, p. 378).

Novellino (1996) argumenta sobre a volta do “[...] uso do controle de vocabulário”, em que se passam “a usar conceitos pré-coordenados [...]”. A partir daí inicia-se o uso das listas de termos autorizados, “[...] que continham registros de decisões tomadas, em relação à indexação, como modelos para os indexadores.” (NOVELLINO, 1996, p. 40). Essa é lista de termos que já não são mais aplicados somente para indexação, mas também na recuperação do conhecimento. “Surgiram então os tesouros, que adicionaram a este relacionamento entre os termos de indexação, outros, visando instrumentalizar não só a representação mas também a busca da informação.” (NOVELLINO, 1996, p. 40).

No **Quadro 2**, são apresentadas as tipologias de vocabulários controlados que Novellino (1996) trata como instrumento utilizado para a tomada de decisão no momento da indexação. Smit e Kobashi (2003) afirmam que existem vários tipos de vocabulários controlados e acrescentam que “[...] a tipologia se estabelece em função do princípio de organização dado aos termos do vocabulário [...]”. Complementam, ainda, que esta tipologia se estabelece “[...] desde a organização por ordem alfabética [...], sistemas mais rígidos (planos de classificação) até chegar, historicamente, a sistemas com uma organização menos draconiana (os tesouros).” (SMIT; KOBASHI, 2003, p. 35).

## Quadro 2 - Tipologias de vocabulários controlados

TIPOLOGIAS DE VOCABULÁRIOS CONTROLADOS E SUAS APLICAÇÕES EM ÁREAS DE ESPECIALIDADE	
TIPOLOGIAS	APLICAÇÕES
<b>Listas</b>	Listas são usadas para exibir pequenos conjuntos de termos que são utilizados para finalidades estritamente definidas, como uma lista de opções ou uma lista de comandos para baixar conteúdos da Web.
<b>Redes de sinônimos</b>	Redes de sinônimos são com frequência utilizadas como “pano de fundo” para melhorar a recuperação, especialmente em um ambiente em que a indexação não utiliza um controle de vocabulário e/ou não há indexação para busca por texto completo.
<b>Taxonomias</b>	Taxonomias são criadas e utilizadas, muitas vezes, em aplicações de indexação para navegação na Web. Devido a sua (normalmente simples) estrutura hierárquica, são eficazes na condução dos usuários até os termos mais específicos em um determinado domínio.
<b>Tesauros</b>	Tesauros são instrumentos de controle terminológico utilizados em sistemas de informação para representar a linguagem natural de documentos, de indexadores e de usuários, num sistema de linguagem documentária, usada na indexação e recuperação de informações de um determinado ramo do conhecimento.

Fonte: Cervantes (2009, p. 61).

A partir dos conceitos, das aplicações e das tipologias de vocabulários controlados, discorre-se sobre a normalização e metodologia para construção de vocabulários controlados, no próximo subtópico.

### 2.2 NORMALIZAÇÃO E METODOLOGIA PARA CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS

A construção de um vocabulário controlado é um processo complexo que engloba uma equipe de trabalho. A verificação da existência de tal vocabulário apresenta algumas dificuldades: temática recente e com número razoável de vocabulários controlados; os vocabulários controlados construídos por empresas privadas não são de domínio público. Para contornar essa última dificuldade, sugerem os autores a cooperação entre as instituições que desenvolvem os vocabulários controlados como meio de intercâmbio. É de grande importância a diferenciação da nomenclatura controle de vocabulário e vocabulário controlado, sendo que o primeiro está ligado ao processo e o segundo é o meio, um instrumento (SMIT; KOBASHI, 2003, p. 19).

A norma ANSI/NISO Z39.19, em sua versão atual (quarta edição), é

verdadeiramente construída por gigantes<sup>1</sup>. A primeira edição (1974) foi elaborada pela *Subcommittee 25 on Thesaurus Rules and Conventions of the American National Standards Committee Z39*, revisada em 1980. Em 1988, inicia-se um processo de ampliação, lançado em 1993. Em 1998, a norma foi revisada pelo ciclo normal, ou seja, periodicidade de revisão de cinco em cinco anos, e nela seria tratada a informação eletrônica.

Segundo Boccato (2011b), a norma britânica BS 8723 (British Standard 8723) teve, em seu desenvolvimento, duas versões, ambas com a denominação de *Guidelines for the establishment and development of monolingual thesauri*. Tinha como foco a construção de tesouros monolíngue com ênfase nos “[...] relacionamentos conceituais, simbologias e abreviaturas empregadas, formas de apresentação e gestão” (BOCCATO, 2011b, p. 173).

A norma internacional ISO 25964, em sua primeira parte publicada em 2011, – *ISO 25964-1:2011 - Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval* (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011), traz recomendações para a construção e manutenção de tesouros destinados para a recuperação da informação. Isso é extensível ao uso de vocabulários com o propósito na recuperação da informação para qualquer tipo de recursos informacionais e independentemente do suporte utilizado (texto, som, imagem, vídeo), incluídos bases e portais de conhecimento, textos, bases de dados bibliográficos, museus ou coleções de multimídias. Fornece também um modelo de dados e o formato recomendado para importar e exportar dados de tesouros. É aplicável em tesouros monolíngues e multilíngues. A segunda parte está centrada na interoperabilidade entre tesouros e tesouros e outros vocabulários controlados - *ISO 25964-2:2013 - Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies* (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013).

O **Quadro 3** apresenta as principais atualizações nas normas ANSI/NISO Z39.19-2005, BS 8723 e ISO 25964 para a construção de vocabulários controlados.

---

<sup>1</sup> Expressão usada na norma “*Built on the shoulders of giants*”

**Quadro 3 - Demonstrativo das atualizações nas normas**

OCORRÊNCIAS	ANSI/NISO Z39:19-2005	BS 8723	ISO 25964
Título	Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.	Structured vocabularies for information retrieval-guide.	Information and documentation – Thesauri and interoperability with other vocabularies.
Estrutura	Onze seções e seis apêndices: - ampliação e atualização da norma ANSI/NISO Z39:10-2003.	Cinco partes: - a parte 2 substitui a norma BS 5723-1987; - a parte 4 substitui a norma BS 6723-1987.	Duas partes: - a parte 10 corresponde às partes 1, 2 e 5 da BS 8723; - a parte 2 corresponde às partes 3 e 4 BS 8723; - a ISSO 25964 substitui as normas ISO 2788-1986 e ISO 5964-1985.
Vocabulários controlados	Listas, anéis de sinônimos, taxonomias, redes semânticas e tesouros.	Sistemas de classificação, sistemas de classificação de negócios para gerenciamento de registros, redes semânticas, taxonomias, listas de cabeçalhos de assunto, ontologias.	Sistemas de classificação, sistemas de classificação de negócios para gerenciamento de registros, listas de cabeçalhos de assunto, taxonomias, mapas conceituais, redes semânticas, terminologias/banco de termos, listas de autoridades de nomes e anéis de sinônimos.
Pontos de atualização/ ampliação	- interoperabilidade entre vocabulários controlados; - aplicação da análise facetada na construção de tesouros; - vocabulários controlados multilíngues; - definição de protocolos e formatos necessários para o intercâmbio de dados.	- interoperabilidade entre vocabulários controlados; - aplicação da análise facetada na construção de tesouros; - vocabulários controlados multilíngues; - definição de protocolos e formatos necessários para o intercâmbio de dados.	- interoperabilidade entre vocabulários controlados; - aplicação da análise facetada na construção de tesouros; - vocabulários controlados multilíngues; - definição de protocolos para o intercâmbio de dados.

Fonte: Boccato (2011b, p. 175).

A necessidade do controle de vocabulário é decorrente de duas características básicas da linguagem natural: 1) duplicidade de nomenclatura para um mesmo conceito – exemplo: SOC/Sistemas de Organização do Conhecimento; 2) duplicidade de conceito para uma mesma palavra – exemplo: manga (camisa), manga (fruta). Para a obtenção do vocabulário controlado, são necessários três principais métodos: 1) definição do escopo dos termos; 2) uso do relacionamento equivalente para a ligação de termos com sinônimos e quase sinônimos; 3) distinção entre homógrafos (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

Diante da necessidade caracterizada pela ANSI/NISO Z39:19-2005 e que se converte nos principais métodos de obtenção do vocabulário controlado, passa-se a discutir esses procedimentos encapsulados nesses métodos que, por sua vez, têm suas especificidades. No Quadro 4, apresentam-se os Indicadores colaboradores na

construção, aperfeiçoamento e adequação do uso de linguagens documentárias.

**Quadro 4** - Indicadores colaboradores na construção, aperfeiçoamento e adequação do uso de linguagens documentárias

Indicador	Descrição
1	Construção do vocabulário tomando-se por base as linguagens de especialidades das áreas científicas e a linguagem de busca do usuário, tendo em vista a compatibilidade entre a linguagem adotada pelo sistema e a de busca do usuário.
2	Incorporação de novos termos com o intuito de promover a atualização da linguagem que se fizer necessária, por meio de coleta em fontes de informação formais (dicionários especializados, glossários técnico-científicos, diretórios, entre outros) e informais (formulários de sugestões de assuntos preenchidos pelos usuários, catálogo e listas de assuntos locais elaborados pela biblioteca).
3	Eleição de termos expressivos que tenham por objetivo a clareza na designação do assunto.
4	Revisão da tradução de termos existentes na linguagem, objetivando a devida correspondência conceitual que deve ocorrer em relação à terminologia das áreas científicas nacionais.
5	Eliminação das ambiguidades causadas pela homonímia e polissemia advindas da linguagem natural: adoção de termos qualificadores agregados ao termo preferido de modo que se definam diferentes aspectos, conceitos e pontos de vista abordados pelo autor sobre o assunto do documento. Os termos qualificadores possibilitam a especificidade na representação e na recuperação da informação. O uso de singular e plural também é um recurso auxiliar para a distinção entre termos homógrafos.
6	Incorporação de termos específicos, a fim de conseguir a especificidade exigida do tratamento de conteúdos documentários para a recuperação precisa da informação.
7	Controle de sinônimos: controle efetivo dos termos sinônimos, quase sinônimos e das variantes lexicais (ortografia, singular – plural, nome completo divergente da abreviatura) evitando-se a dispersão temática e proporcionando maior exatidão na indexação e busca por assunto.
8	Estabelecimento das relações lógico-semânticas: a construção das relações hierárquicas, de equivalência e não hierárquicas tem por finalidade a indicação dos relacionamentos semântico-conceituais entre os termos, propiciando a consistência na representação e na recuperação. O incremento das relações não hierárquicas (associativas) promove uma aproximação maior com a linguagem de busca do usuário, relevando a função comunicativa que a linguagem deve ter.
9	Verificação da sintaxe dos cabeçalhos de assuntos compostos referentes à ordem das ideias que os compõem.
10	Incorporação de notas de escopos nos cabeçalhos de assunto.
11	Intensificação da função comunicativa da linguagem.
12	Representação de conceitos por meio de termos, visando à obtenção de cabeçalhos mais consistentes.

Fonte: Boccato (2011a, p. 27-29).

Para Lancaster (1995), a sinonímia remete a palavras ou frases que são exatamente equivalentes, ou seja, verdadeiros sinônimos, os mais comuns são as abreviaturas e acrônimos. Smit e Kobashi (2003) afirmam que, nos casos de “[...]”

sinonímia simples, trata-se, em suma, de fazer uma opção, remetendo da expressão não adotada para a adotada.” Quando há uma legislação ou alguma normatização que assegure a terminologia que deve ser aplicada na caracterização de remissivas para variantes, em situações que não exista, coloca-se em prática o conhecimento cultural local (SMIT; KOBASHI, 2003, p. 25). Quanto aos quase sinônimos, Lancaster (1995) afirma que eles contêm uma ligação entre os extremos denominada *continuum* e são tratados como sinônimos. Para Smit e Kobashi (2003), a quase sinonímia é algo próximo da sinonímia, “[...] porém mais operacional.” Um exemplo pode ser referenciado pela licença-gestante e licença-maternidade, “[...] designa momentos subsequentes na vida da mulher: primeiro a gestação e depois a maternidade.” (SMIT; KOBASHI, 2003, p. 26).

O controle de vocabulário geralmente “[...] tem sido exercido por intermédio de: seleção de termos preferidos; distinção de homógrafos; estabelecimento formal das relações entre os termos” (FUJITA, 2011, p. 42). Para Smit e Kobashi (2003, p. 26), a homonímia é “[...] mesma expressão para atividades diferentes [...]”. Lancaster (1995) define como uma cadeia de caracteres com significados diferentes. Para Lancaster (1995), Smit e Kobashi (2003) e Austin e Dale (1993), quando ocorre um homógrafo, é determinado um qualificador e inserido, entre parênteses, depois do termo. Um exemplo clássico pode ser o termo manga, que pode ser MANGA (fruto da mangueira); MANGA (parte do vestuário onde se enfia o braço) (AUSTIN; DALE, 1993).

A adoção de termos compostos em vocabulário controlado é designada por “[...] expressões formadas por mais de uma palavra, geram uma série de dúvidas relacionadas à manutenção, ou não, da expressão como um todo, quanto à ordem a ser adotada entre os termos”. Diante da complexidade, quanto às decisões na adoção de termos compostos, algumas regras podem ser seguidas de maneira geral. A primeira é “[...] aproximar-se, tanto quanto possível, da linguagem consolidada pela legislação e da linguagem utilizada pela instituição [...]”, a segunda é “[...] valorizar a coerência interna do vocabulário (ou seja, manter a mesma decisão ao longo de todo o vocabulário) [...]”, a terceira e última é “[...] diferenciar a nomeação de atividades das ações que as modificam.” (SMIT; KOBASHI, 2003, p. 29).

Austin e Dale (1993) concordam explicitando a complexidade do tratamento de termos compostos. Sugerem a aplicação da fatoração do termo composto que

poderá ser por meio de duas técnicas, fatoração semântica e fatoração sintática. A primeira trata de representar um termo composto por um conceito simples ou definições, e pode ser aplicada tanto em termos simples como compostos, por exemplo, “TERMÔMETROS podem ser expressos pela combinação de três termos: TEMPERATURA + MEDIDA + INSTRUMENTOS”. Já a segunda fatoração “[...] é aplicada a termos compostos, por exemplo, termos que podem ser objeto de uma análise morfológica em componentes separados, cada um podendo ser aceito, por si só, como termo indexador.” Por exemplo: “CONSTRUÇÃO DE EDIFÍCIOS pode ser expressa por: EDIFÍCIOS + CONSTRUÇÃO” (AUSTIN; DALE, 1993, p. 34, grifo do autor).

Diante desse contexto, a compatibilização é o principal processo para efeito deste estudo. Portanto existem diversos tipos de compatibilização, termo muito utilizado na tecnologia para determinar a equivalência das características de dois ou mais equipamentos, e o modo em que os sistemas operam sem necessidade de conversão ou outro procedimento para se tornar intercambiável. A compatibilidade de dados é o modo em que se trabalha com a mesma estrutura de dados para ser possível essa troca de dados. Por outro lado, tem-se a compatibilidade de linguagens documentárias, que é a “[...] capacidade de diversas linguagens documentárias de expressar um mesmo termo ou termo equivalente”, porém para existir este procedimento precisa ser na mesma temática e com normas similares, geralmente aplicadas às normas internacionais (CUNHA e CAVALCANTI, 2008, p.95).

Para Zavitoski (2001, p. 36), compatibilização compreende agrupar os diversos termos da linguagem natural com um termo específico “[...] que defina mais apropriadamente o conceito que se deseja representar [...]” (vide Quadro 1, f. 27). Além disso, Boccato e Torquetti (2012, p. 84) relatam em sua pesquisa que existe uma diversidade de termos aplicados para essa vertente, são eles: “[...] compatibilidade, integração, unificação, convertibilidade, interoperabilidade, entre outros [...]”, com o mesmo conceito e inseridos em suas respectivas épocas de uso.

O próximo subtópico trata das palavras-chave como elemento de compatibilização, para chegar a essa finalidade, passa-se pelo processo de compatibilidade dessas palavras-chave com o vocabulário controlado.

### 2.3 PALAVRAS-CHAVE: ELEMENTO DE COMPATIBILIZAÇÃO

As palavras-chave ou *keywords* tiveram seu surgimento “[...] como termos de acesso nos primeiros índices automatizados existentes nas Bibliografias especializadas, hoje denominadas de bases de dados.” (FUJITA, 2004, p. 258). A palavra-chave “[...] é uma representação do conteúdo significativo do texto [...]” e é requerida no momento da necessidade de recuperação, ou seja, na formulação da estratégia de busca (FUJITA, 2004, p. 258).

A atribuição das palavras-chave aos artigos científicos é de extrema importância para a recuperação, porém nem todos os periódicos tornam claras as diretrizes para tal processo. “Em algumas bases de dados e periódicos especializados, pode-se contar com o auxílio de profissionais que farão a escolha de quais palavras-chave são mais adequadas para representação [...]” (GONÇALVES, 2008, p. 6-7). No entanto essa estrutura torna-se onerosa, e a partir daí “[...] é comum observar nas regras para submissão de artigos orientações para que o próprio autor designe as palavras-chave para seu texto, embora esse processo não seja o mais adequado.” (GONÇALVES, 2008, p. 6-7). Acrescenta o autor que essa temática ainda é pouco explorada na ciência e de suma importância, uma vez que se trata de um elemento obrigatório nos artigos.

Fujita (2004, p. 258) aborda o conceito da palavra-chave que está vinculada aos mecanismos de busca de informação em bases de dados e lembra que, desse modo, as palavras-chave são solicitadas por ferramentas de busca “[...] que deverão apresentar compatibilidade com as palavras-chaves representativas dos textos e, a partir disso, recuperar o documento demandado pela busca por palavras-chaves.” Fujita (2004, p. 258).

Diante desse contexto, é possível utilizar a palavra-chave como indexação nos motores de busca e categorizações, já que é uma representação sucinta do texto. É por meio delas que o leitor pode julgar o texto como relevante para seu estudo. Outra aplicação é a utilização para medir a similaridade entre os documentos. Por outro lado, há uma grande quantidade de documentos que ainda não possuem palavras-chave designadas em sua estrutura. Sendo assim, considerando essa grande quantidade de documentos que não têm palavras-chave e a necessidade da cognição aplicada ao processo que os envolve, resolver a situação torna-se uma tarefa árdua. Portanto seria desejável a automatização de tal

processo por meio da integração da aprendizagem de máquina e o PLN (ERCAN; CICEKLI, 2007, p. 1705).

Gil Leiva e Alonso Arroyo (2005, tradução nossa) afirmam que o produto dos profissionais de indexação não deve ser substituído pelas palavras-chave atribuídas pelos autores, mas as palavras-chave são uma importante fonte de alimentação para as análises dos artigos científicos nos processos de indexação, assim como uma fonte importante para início das metodologias de uma indexação automática. As vantagens atribuídas ao uso das palavras-chave são: aumentar o acesso ao conteúdo dos documentos; constituir-se mais um meio de representar além dos já existentes (título e resumo); “[...] traduzir o pensamento dos autores; e [...] manter o contacto com a realidade da prática quotidiana, acompanhando a evolução científica e tecnológica refletida pelos documentos.” (MIGUÉIS; NEVES, 2013, p. 117).

Gonçalves (2008, p. 11) relata que uma simples semelhança de termos não significa que eles sejam correspondentes ou sinônimos. “No entanto, a coincidência de palavras pode apontar para a existência de um vocabulário comum que não estaria evidente apenas com a leitura dos resumos [...]”, diante deste contexto, as “[...] palavras-chave podem ser vistas como um complemento [...]”. Nesta conjuntura, nas bases de dados, as palavras-chave podem ser utilizadas como metadado descritivo, “[...] especialmente em bases que não possuem vocabulário controlado. Assim, o fato de haver palavras-chave coincidentes afetará os resultados de busca em maior ou menor escala” (GONÇALVES, 2008, p. 11). Relata ainda o autor que com uma análise mais profunda de palavras-chave pode ser possível a descoberta do tipo de pesquisa (teórica, empírica) e o contexto do estudo.

Entende-se que as palavras-chave são um ou mais termos que representam os conceitos principais tratados no artigo. Constituem um ponto de referência para sua localização e recuperação exaustiva sem ruído nos resultados. Aqui se situa a importância da atribuição das palavras-chave (ARENAS et al., 2005, p. 60, tradução nossa). As palavras-chave são importantes para os autores, editores, indexadores, profissionais que fazem resumos e para as pessoas que pesquisam informações na *Web* e em bases de dados. Sem dúvida, haveria maior coerência entre autores e revistas dentro das diferentes disciplinas se fossem usados os termos padronizados dentro de várias categorias. Com isso a probabilidade da recuperação de informação poderia ser mais exata em todas as diferentes disciplinas (HARTLEY; KOSTOFF, 2003, p. 435).

Para Cunha e Cavalcanti (2008, p. 274), palavra-chave é definida como “[...] palavra significativa encontrada no título de um documento, no resumo ou no texto [...]” e é usada no procedimento que caracteriza o conteúdo temático, utilizada para criar índices, catálogos. Para Aquino e Aquino (2013, p. 229-230), “O termo palavras-chave refere-se às palavras mais importantes do texto científico utilizadas pelos autores, para indexação [...]”.

Para a NBR 6028, palavra-chave é a “Palavra representativa do conteúdo do documento, escolhida, preferentemente, em vocabulário controlado” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003b, p. 1), e essa mesma fonte também define como deve ser estruturada: “As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave, separadas entre si por ponto e finalizadas também por ponto” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003b, p. 2).

Para Ercan e Cicekli (2007, p. 1705), “palavras-chave pode ser considerado como um breve resumo de um texto. Por isso, é possível considerá-los como um conjunto de frases semanticamente, cobrindo a maior parte do texto.” O resumo pode descrever mais o texto com um nível maior de detalhamento que as palavras-chave. Embora o resumo de um texto seja capaz de fornecer mais informações sobre o texto do que palavras-chave, o resumo pode não ser apropriado para determinadas aplicações diante da complexidade de sua estrutura materializada nas sentenças. “Palavras-chave não são substitutos para sumarização, mas representações de síntese alternativas que poderiam ser consumidos por outros aplicativos com mais facilidade.” (ERCAN; CICEKLI, 2007, p. 1705).

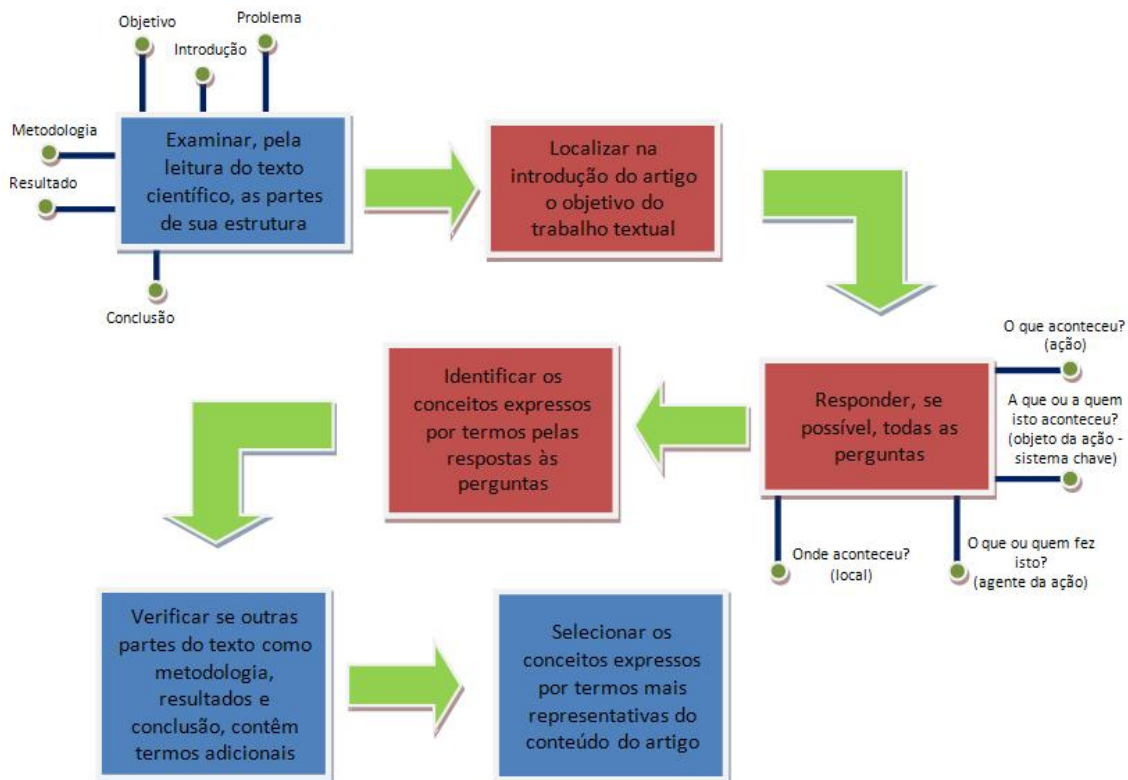
Na norma internacional ANSI/NISO Z39:19-2005, palavra-chave é definida como sendo a palavra que aparece no texto, configurada por meio da linguagem natural, e é considerada a mais significativa para indexar ou recuperar o texto (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

Para Gonçalves (2008), a palavra-chave pode ser retirada de qualquer local do documento, seja do título, do resumo ou do próprio texto como um todo, portanto não tem implicação no controle de vocabulário. “Sua atribuição é livre, podendo ser definida pelos autores do documento, por editores, e até ser geradas automaticamente.” (GONÇALVES, 2008, p. 5-6). Fujita (2004, p. 261) assevera que a determinação de uma palavra-chave é parte do processo de indexação

documentária. Desse modo, é resultante de um grau maior de redução do texto original, portanto existe aí uma “[...] perda de informação semântica que poderá ser minimizada com uma representação totalmente voltada para o conteúdo do texto.” (FUJITA, 2004, p. 261).

Na Figura 2, é apresentado um fluxograma de processos para realizar a aplicação da metodologia de indexação para determinação das palavras-chave.

**Figura 2** - Metodologia de indexação para determinação das palavras-chave



Fonte: Elaborado pelo autor, fundamentado em Fujita (2004, p. 267).

Com referência à metodologia ilustrada na Figura 2, Fujita (2004, p. 267) orienta que os processos b, c e d (vermelho) devem ser integrados “[...] de forma a sistematizar a exploração da estrutura textual do artigo científico com a identificação de conceitos por meio de questionamento e, assim facilitar, para o autor-leitor [...]” a identificação de palavras-chave que representem o conteúdo do texto de maneira significativa.

Utilizar uma linguagem documentária para tomar a decisão de escolha de um descritor ou cabeçalho de assunto pode ser uma medida sensata que

consequentemente deduz “[...] a diversidade e a ambiguidade do vocabulário e estabelece uma uniformidade de representação dos termos selecionados pelo indexador para descrever o assunto dos documentos [...]” (FUJITA, 2011, p. 36). Essa ilustração é devida ao cenário que os autores se utilizam de diversas palavras para expressar um ponto de vista, como também o usuário em sua expressão de busca poderá convergir para diversas palavras, com o mesmo sentido (FUJITA, 2011, p. 36).

O descritor é elemento “[...] de uma linguagem documentária, que pode ser empregado para representar um texto em sistemas de informação. Traduz os conceitos (os assuntos) contidos no texto; identificador do documento.” (CUNHA; CAVALCANTI, 2008, p. 119). Diante desse contexto, é importante observar que a palavra-chave tem seu significado decorrente do uso de linguagens documentárias, desse modo é o resultado de uma representação “[...] mediante compatibilização com linguagens documentárias submetida aos termos identificados e selecionados durante a análise de assunto do artigo pelo autor-leitor.” (FUJITA, 2004, p. 270). Portanto aproxima-se dos conceitos de descritores.

A aplicação dos descritores são as mais diversas, Fujita (2011) apresenta alguns exemplos. A autora afirma que na indexação a atuação está na etapa de tradução, o que permite que “[...] a representação do conteúdo do documentário feita durante a análise e síntese seja ‘traduzida’ como os descritores da linguagem documentária.” (FUJITA, 2011, p. 40). A formação de uma linguagem documentária se dá por meio do vocabulário e da sintaxe. O “[...] vocabulário refere-se à relação dos descritores usados para a identificação do conteúdo de um documento [...]”; a sintaxe faz referência “[...] às regras utilizadas para a combinação dos descritores usados para identificação de um documento.” (FUJITA, 2011, p. 36).

Em síntese, considera-se que há uma aproximação entre palavras-chave atribuídas aos textos pelos autores e as que são produtos do processo de indexação, uma vez que se trata da seleção de expressões significativas do texto. No entanto observa-se um distanciamento no elemento que executa esse processo e sua denominação, se realizado pelo autor, denomina-se atribuição de palavras-chave; enquanto pelo profissional de indexação ou sistema (no processo automático), denomina-se seleção de expressões significativas do texto. Além disso a palavra-chave é extraída do próprio texto, já o termo é uma expressão selecionada para representar tematicamente o texto.

O próximo subtópico trata do processo de indexação e sua integração com as ferramentas tecnológicas.

## 2.4 PROCESSO DE INDEXAÇÃO

Neste subtópico considera-se o processo de indexação, realizado de modo manual e/ou automático, enfatizando a inclusão de recursos computacionais que possibilitam transformações expressivas nos quesitos de recuperação da informação. Nesse contexto, abordam-se os conceitos de sistemas, banco de dados, metodologia de desenvolvimento de *software*, entre outros aspectos relevantes na modelagem de um sistema de informação.

A propagação entre indexação e resumos está cada vez mais evidente. Por um lado, há os termos que poderão ser escritos sem um miniresumo, do outro, o texto de resumos que podem ser mantidos em uma base de dados de modo acessível por um sistema. A partir desse cenário, é possível construir combinações de palavras para realizar uma expressão de busca. De certa forma, ambos os mecanismos visam a “[...] uma descrição clara e de boa qualidade do conteúdo do documento, mas também em criar um registro que seja uma representação eficaz para fins de recuperação.” (LANCASTER, 2004, p. 7).

Para a NBR 12676, indexação é o “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2). Por sua vez, a norma internacional ANSI/NISO Z39:19-2005 trata a indexação como um método pelo qual os termos ou cabeçalhos de assuntos de um vocabulário controlado são selecionados, por um computador ou um humano, para representar os conceitos ou atributos do conteúdo de um objeto. Os termos podem ou não aparecer no conteúdo do objeto.

Cunha e Cavalcanti (2008, p. 193) conceituam o termo indexação como “Representação do conteúdo temático de um documento por meio dos elementos de uma linguagem documentária ou termos extraídos do próprio documento (palavras-chave, frases-chave).” Gil Leiva e Fujita (2012, p. 72) resumem a indexação em “[...] um processo executado nos objetos suscetíveis de serem representados e nas solicitações dos usuários para, em última análise, satisfazer as necessidades de informação.” Para o autor supracitado, os produtos da indexação são: “[...] palavra-

chave, descritor ou cabeçalho de assunto [...]”, complementa que esse produto não é associado a nenhum ponto específico do documento (GIL LEIVA; FUJITA, 2012, p. 74).

Para Gomes e Gusmão (1983), a indexação inicia-se após a leitura do texto de maneira a identificar o conteúdo, objetivo, abordagem, vocabulário do autor e ênfase de tópicos que desencadeiam dois processos dentro da indexação temática: análise conceitual e tradução. Outro fator importante a ser observado pelo indexador, no que se refere a controle de vocabulário, é a diferença de indexar palavras ou assuntos. Gil Leiva e Fujita (2012, p. 72) destacam que o objetivo geral de indexação é “[...] o armazenamento da informação para atender às necessidades de informação. Portanto, a indexação e recuperação são duas faces da mesma moeda.”

O principal propósito da indexação é o fornecimento da informação, no momento preciso, ao usuário de forma “[...] eficiente e econômica [...]” (CARNEIRO; 1985, p. 222). Para esse processo, é necessária a utilização de instrumentos normativos, bem como instrumentos linguísticos: Linguagem Natural (LN) é a reunião de sinais reconhecidos pelo homem (indexação livre) e Linguagem Documentária (LD) é linguagem artificial com conjuntos de símbolos e regras específicas (FEITOSA, 2006).

No momento de escolha de termos de entrada “[...] deve-se buscar consistência, isto é, para uma noção deve-se adotar sempre o mesmo termo, ou expressão [...]” (GOMES; GUSMÃO, 1983, p. 19-20), tornando-a remissiva. Duas perguntas são importantes para se tirar a “prova real” de que a informação foi corretamente indexada: “[...] 1) - Se eu tivesse interessado neste assunto, como procuraria? 2) - Se eu tivesse interessado neste assunto, esta informação seria satisfatória? [...]” (GOMES; GUSMÃO, 1983, p. 19-20).

A seleção de termos de indexação enfatiza que o indexador deve se atentar para algumas práticas. Em primeiro plano, “[...] usar os descritores cabíveis já existentes na linguagem de indexação utilizada [...]”; posteriormente, “[...] para os termos que representam novos conceitos, deve-se verificar sua precisão e aceitabilidade em instrumentos de referência [...]”. Esses instrumentos podem ser exemplificados como: “[...] dicionários e enciclopédias de autoridade reconhecida nas suas especialidades; tesouros, especialmente os elaborados de acordo com as ISO 2788 ou ISSO 5964; tabelas de classificação” (ASSOCIAÇÃO BRASILEIRA DE

NORMAS TÉCNICAS, 1992, p. 3).

Gil Leiva e Fujita (2012, p. 71) asseveram que as “[...] etapas de indexação são a análise dos documentos e as questões para a seleção dos conceitos explícitos ou implícitos [...]”, (GIL LEIVA; FUJITA, 2012, p. 71), o que, por outro lado, “[...] implica um esforço de avaliação e condensação do conteúdo [...]” (GIL LEIVA; FUJITA, 2012, p. 74). Ainda dentro das etapas, inclui-se “[...] o armazenamento destas palavras-chave como estão”, que também pode realizar sua conversão para uma linguagem controlada (GIL LEIVA; FUJITA, 2012, p. 74). Os estágios da indexação, segundo a norma NBR 12676, são compostos por: “[...] a) exame do documento e estabelecimento do assunto de seu conteúdo; b) identificação dos conceitos presentes no assunto; c) tradução desses conceitos nos termos de uma linguagem de indexação.” Na realização desses estágios, “[...] deve-se contar com a ajuda de instrumentos de indexação, tais como: tesouros, códigos de classificação, cabeçalhos de assuntos, etc.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

O principal critério para a seleção de conceitos “[...] deve ser sempre o valor de um conceito para a expressão e recuperação do assunto do documento.” Assim o indexador deve considerar sempre, na escolha dos conceitos, “[...] as consultas que podem ser feitas ao sistema de informação.” Mediante este critério, o indexador deverá “[...] escolher os conceitos que forem considerados os mais apropriados para uma determinada comunidade de usuários [...]”. Também o indexador precisa “[...] adaptar tanto os instrumentos de indexação como os próprios procedimentos em função da retroalimentação obtida através dos pedidos de informação.” Uma atenção especial deve ser dispensada para esta adaptação que, por sua vez, “[...] não deve alterar a estrutura ou a lógica da linguagem de indexação.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 3).

Gil Leiva e Fujita (2012, p. 76) observam que a indexação pode ser um “[...] processo intelectual ou automático que visa obter um conjunto de unidades conceituais que representam o objeto analisado integralmente.” O resultado, que é a representação do objeto a ser indexado, pode ser descrito em linguagem natural, denominada palavras-chave. Desse modo, “[...] significa que elas foram tomadas literalmente dos objetos analisados e, portanto, é o resultado da indexação (*indización*).” Essas palavras-chave selecionadas passam por uma “[...] normalização e controle antes de serem armazenadas no banco de dados. Esta

modificação é feita para termos sistemas de informação coerentes e mais eficazes.” (GIL LEIVA; FUJITA, 2012, p. 76).

Lancaster (2004) afirma que “a indexação de assuntos envolve duas etapas principais: 1. Análise conceitual, e 2. Tradução” (LANCASTER, 2004, p. 8-9). Análise conceitual é a “[...] fase inicial do ato classificatório que consiste em determinar os principais assuntos ou conceitos relacionados com o documento objeto de análise.” (CUNHA; CAVALCANTI 2008, p. 14). A tradução é a “[...] Mudança de representação para outra, sem que seu conteúdo seja afetado; documento derivado.” (CUNHA; CAVALCANTI, 2008, p. 366).

Para a atribuição de termos de indexação, “não se deve estabelecer nenhum limite arbitrário ao número de termos ou descritores que possam ser atribuídos a um documento.” Esse processo deve ser relacionado com a necessidade do usuário e com a quantidade de informação do conteúdo do documento. A consequência dessa imposição “[...] pode levar a alguma perda de objetividade na indexação e à distorção da informação que poderia vir a ser de valor na recuperação.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 3).

Gomes e Gusmão (1983) relatam, em sua pesquisa, que no Brasil não existia uma tradição de indexar periódicos, assim como uma representação neste tipo de serviço que envolvia fatores de qualidade, e o pesquisador, por sua vez, não dispunha das ferramentas necessárias para suas pesquisas. No ambiente internacional, é observado que as diversas associações científicas estão na etapa da organização dos serviços bibliográficos especializados, mas cabe ressaltar a falta de índice, que embora não gere um problema, leva essas associações a se esforçar para que os editores especializados possam publicar o seu índice.

Não é difícil identificar as mudanças que decorrem da afirmação dos autores supracitados (considerando o ano de publicação) até o momento atual no que se refere à indexação de periódicos no Brasil. Segundo Packer (2011, p. 32), “O número de periódicos brasileiros indexados, internacionalmente, varia entre os diretórios e índices bibliográficos devido às diferentes políticas e critérios de cobertura, mas a presença brasileira é sempre destacada e consistente.” Retrata-se aqui o atual momento da indexação dos periódicos científicos brasileiros, quando o autor supracitado ainda discorre, em sua pesquisa, que a quantidade de periódicos científicos brasileiros indexados é maior que os periódicos científicos indexados da África do Sul e também da Coreia do Sul.

No **Quadro 5**Quadro 1, são apresentados os processos para se obter o índice (*indexación*) e a indexação (*indización*), segundo Gil Leiva e Fujita (2012, p. 77). Para esclarecimento dos termos *indexación* e *indización*, Gil Leiva e Fujita (2012, p. 77) esclarecem que “em espanhol, convivem as palavras *indexación* e *indización* e, às vezes, são entendidas como sinônimos.” Contudo *indexación* refere-se a “[...] elaboração automática de entradas ou índices em um documento diante da marcação de palavras que não sejam vazias [...]”. A *indización* refere-se a “[...] todos os itens obtidos manual ou automaticamente, que representam o conteúdo de um documento.” (GIL LEIVA; FUJITA, 2012, p. 77).

**Quadro 5** - Processo para obter o índice - indexação

Palavras-chave	Controle do vocabulário	Indexación	Indización
Selecionar do próprio texto as palavras ou frases que melhor representam o conteúdo do mesmo e armazená-las da mesma maneira no campo destinado para esse fim na base de dados.	Uma vez selecionadas as palavras ou frases que melhor representam o conteúdo do texto, são “filtradas” para uma linguagem de indexação para evitar sinonímia, polissemia, etc. na base de dados. Dessa maneira, se consegue um vocabulário controlado e normalizado para ser introduzido na base de dados correspondente.	Selecionar ou marcar automaticamente todas as palavras do texto que não estejam num antídicionário, ou seja, as que não são, por exemplo, verbos, adjetivos, preposições, artigos, conjunções, pronomes, etc. para que possam ser utilizadas mais tarde na recuperação do documento.	Selecionar as palavras ou frases que melhor representam o conteúdo do texto e, normalmente, são “filtradas” por uma linguagem de indexação para conseguir um controle de vocabulário apresentado na base de dados.

Fonte: Gil Leiva e Fujita (2012, p. 77).

No que tange à natureza dos índices, a indexação de materiais textuais sempre deve atender às expectativas do leitor, assim como a natureza dos documentos pode interferir no processo de construção de um índice, e esse é um fator determinante de um vocabulário controlado. Neste contexto, há três elementos aqui considerados sistema de comunicação: autores, usuários e indexadores. As implicações são as mais diversas possíveis, quando se trata da natureza do conhecimento e suas formas de registro; porém, para o processo construtivo de índices, dois aspectos são relevantes: metodologias e vocabulário. Nesta perspectiva, surgem dois tipos de índice: interno e externo. O índice interno é aquele que acompanha a obra, sendo uma fonte primária, e o externo é o utilizado como um

guia de assunto de alguma área do conhecimento - fonte secundária (GOMES; GUSMÃO, 1983).

O índice de assuntos é tratado, na NBR 12676, como a “Listagem alfabética ou sistemática de assuntos que indica a posição de cada assunto num documento ou numa coleção de documentos.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2). Gil Leiva e Fujita (2012) contextualizam os índices como listas “[...] que servem para conectar uma palavra ou frase do texto ao lugar exato do mesmo.”, sua elaboração se dá mediante a seleção de palavras ou frases do texto e suas localizações (GIL LEIVA; FUJITA, 2012, p. 74).

Historicamente, o índice foi utilizado primeiramente pelos romanos e denominou-se uma revelação. Na armazenagem e recuperação da informação, o índice, como instrumento, tem sua origem “[...] quando o homem ao se preocupar em tornar mais acessível a informação registrada[...]”, de alguma forma, ordenou-a (GOMES; GUSMÃO, 1983, p. 12-14). Gil Leiva e Fujita (2012) definem que o índice tem como objetivo estabelecer uma ligação do objeto com sua localização, isto é, quando se trata de documentos, o processo de elaboração de um índice é “[...] extrair palavras ou frases do seu contexto, classificá-las e indicar o local do documento onde se localiza cada palavra ou frase. Por isso, procura-se estabelecer uma conexão exata entre dois pontos.” (GIL LEIVA; FUJITA, 2012, p. 72).

Um exemplo de aplicação de índices pode ser dado por Eusébio ao estruturar os Evangelhos em seções enumeradas (texto da Escritura). Posteriormente, temos a obra “Apothegmata” que traz “[...] uma lista de provérbios gregos sobre tópicos teológicos [...]” (GOMES; GUSMÃO, 1983, p. 12-14) ordenados alfabeticamente, no século VI. Portanto a indexação passa a ser objeto de estudos profundos, uma vez que se torna um sistema de recuperação de informação que tem sua atuação marcada como filtro que, conseqüentemente, será avaliado pelo usuário tão somente nos resultados, não importando os caminhos percorridos pela técnica (GOMES; GUSMÃO, 1983).

Também historicamente foi no *The Reader's Guide to Periodical Literature*, 1901, que foi proporcionado, pelos preceitos da publicação das *Rules for a Dictionary Catalog de Cutter*, em 1876, “[...] um índice de assuntos e autores, numa só seqüência, com uma coleção consistente de cabeçalhos de assuntos e uma forma regular e completa de referências.” Dessa forma foi oferecida, de maneira autêntica, a ordem e não truques verbais. “Os assuntos não dependiam das

palavras-chave do título.” (FERRIDAY, 1974, p. 107).

O autor supracitado contextualiza que “a principal função de um índice de periódico é indicar a localização de itens específicos.” Portanto a pesquisa mais ampla e completa não é o objetivo do usuário diante desses índices, mas não se podem desconsiderar as exceções, “[...] especialmente no caso de periódicos em que a seção de resumos constitui uma característica importante, ou de periódicos que são fonte de informação para determinado assunto.” (FERRIDAY, 1974, p. 97). A determinação de um índice para um periódico está no ato de revisar a evolução de um grupo de periódicos, por exemplo, o *Poole's Index to Periodical Literature* utilizou o índice usual, trazendo, à frente, as palavras-chave por meio da inversão do título. Falhas graves são decorrentes desse procedimento, exemplo: “Um artigo sobre ‘Drenagem em Calcutá’ aparece em ‘Calcutá’, sem referência em ‘Drenagem’, apesar de haver outro artigo sobre ‘Drenagem em Londres’ indexado ‘Drenagem’, mas sem aparecer em ‘Londres’.” (FERRIDAY, 1974, p. 106-107).

O índice de um periódico é definido por um tipo que é exigido de acordo com “[...] os fatores determinantes o assunto, as características especiais, a natureza dos pensamentos que comunica o número e o tipo de usuário a que se destina.” Posteriormente, devem ser levadas em consideração as características dos que podem ser “[...] a) periódicos noticiosos; b) revistas de ficção, que geralmente não exigem índice; c) revistas de sociedades eruditas; d) periódicos de ciências em geral e tecnologia; e) revista técnica e comerciais; e f) periódicos de resumos e críticas.”. Também se leva em consideração a temporalidade desses índices, geralmente anuais, ou em alguns casos trimestrais. Na construção deve ser levado em consideração que “[...] a paginação do periódico recomeça em cada número – e esta é uma prática invariável em jornais – é importante, ao indexar cada item, indicar além da referência à página, o número ou o mês e a data do fascículo.” (FERRIDAY, 1974, p. 98).

O índice poderá ter influência “[...] dependendo de outras variáveis como assunto, tipo de usuário a que se destina, natureza do pensamento que é comunicado, etc.[...]” (GOMES; GUSMÃO, 1983, p. 32), ou seja, o periódico poderá assumir algumas características, e isto é uma ingerência no índice. A comunicação das pesquisas sendo original e/ou resenha de livros, novas publicações, obituários, perfazem a composição da natureza dos textos, que têm diversas classificações, segundo Gomes e Gusmão (1983, p. 32), em periódicos de: “[...] comunicação

original; [...] características múltiplas; [...]” A classificação dos periódicos de características múltiplas, por sua vez, se subdivide em periódicos: “[...] de instituições e sociedades científicas e culturais; [...] gerais para a ciência (ou a tecnologia); [...] técnicos e comerciais; [...] em que os resumos constituem uma característica significativa; [...] de resumos e de revisões da literatura.” (GOMES; GUSMÃO, 1983, p. 32).

O recuo no tempo se fez necessário para resgatar a atuação dos índices e suas inferências no periódico. Agora, em um salto para atualidade, atualizam-se os suportes, com vistas às teorias iniciais. Portanto, com o aumento de recursos computacionais e diminuição dos valores para armazenamento, torna-se cada vez maior a disponibilidade de textos completos em base de dados, o que traz dificuldades para a recuperação. Os aspectos intelectuais da recuperação têm reduzido a utilização de palavras-chave e frases-chave combinadas com operadores booleanos ou de proximidade. Técnicas de indexação automática ou semiautomática (baseadas em processos estatísticos, linguísticos e inteligência artificial) têm sido pesquisadas e implementadas em textos completos que não passaram por uma indexação manual (FEITOSA, 2006).

Fujita (2012, p. 19) argumenta que o “[...] conceito de armazenamento com as tecnologias atuais se modificou, pois os novos documentos digitais propiciam acesso ao texto sem que seja necessária a indexação.” Porém a indexação mantém sua existência por se tratar de um mecanismo de extrema importância para a organização e representação de documentos, em especial no meio tecnológico e científico, em que se torna necessária a especificidade de seleção de conteúdos. Consequentemente “[...] essa situação de disponibilidade digital alterou o comportamento informacional no acesso e recuperação exigindo novas formas de orientação de acesso à informação.” (FUJITA, 2012, p. 19).

As atividades como catalogação, classificação e indexação são comuns aos profissionais de informação, e a tecnologia da informação sempre colaborou para melhorar a tarefa de identificação, localização e acesso a documentos impressos. Com o advento da *internet*, os registros vão se tornando digitais, outros são concebidos digitalmente, e podem ser facilmente acessados com o conhecimento de suas respectivas *Uniform Resource Locator* (URL), diante disso, os procedimentos de localização e acesso tornaram-se triviais (MARCONDES, 2001).

Gil Leiva e Fujita (2012) complementam que o “[...] desenvolvimento da

*internet* supôs a difusão e a popularização de conceitos (palavras chave, descritores, vocabulário controlado, tesouros, classificações temáticas) [...]”. Por outro lado, acontece também essa popularização das “[...] técnicas (classificação, indexação, indexação automática) e das práticas (elaboração de tesouros ou de classificações) [...]”, que até então eram conhecidas somente pelos bibliotecários, documentalistas e arquivistas. Portanto “[...] tudo isso criou um universo da indexação na *web*.” (GIL LEIVA; FUJITA, 2012, p. 105).

Com o advento das tecnologias, os conceitos vêm sofrendo alterações. Com a inserção da informática nos processos informativo-documentários, “desaparece o caráter subjetivo, pessoal, do conhecimento. Fala-se, inclusive, de base do conhecimento, quando, na realidade, trata-se somente de uma base ou armazém de dados acumulados na memória do computador.” (CURRÁS, 2010, p. 19). Para Gomes (2009), esta inserção de tecnologias é utilizada como elemento essencial porque traz novos avanços no que tange ao tratamento da informação e, conseqüentemente, faz surgir contribuições para a organização do conhecimento.

O tratamento da indexação automática tem sido desenvolvido por meio de diversas técnicas, tanto para pequenas coleções como para grandes acervos. Essas técnicas podem ser de modalidade “[...] baseada na ocorrência dos termos [...]”; “[...] indexação de palavras não significativas; uso de sufixos; e extensão da busca [...]” (FEITOSA, 2006, p. 26). Em geral as técnicas de indexação automáticas e semiautomáticas são “[...] baseadas em processos estatísticos, linguísticos ou baseados na inteligência artificial [...].” (FEITOSA, 2006, p. 22).

Diante disso, tem-se as Tecnologias de Informação e Comunicação (TIC) que “[...] contribuem para o desenvolvimento de processos e instrumentos aplicados ao tratamento da informação documentária.” A utilização das tecnologias “[...] possibilita o desenvolvimento de softwares para a construção, manutenção e gerenciamento de linguagens documentárias, bem como a automação da atividade de indexação [...].” Com este aparato tecnológico aplicado na representação e recuperação da informação, visa-se à otimização dos processos e dos recursos. Desse modo, “[...] a temática indexação automática e semiautomática são desenvolvidas com o intuito de agilizar a indexação de fontes de informação [...]” na perspectiva de usuários exigentes para a recuperação da informação de acordo com a sua necessidade (LIMA; BOCCATO, 2009, p.1).

Elucida-se aqui, portanto, a utilização de todo aparato de sistemas de

informação implementados em um suporte tecnológico. Desse modo, é importante conhecer os conceitos em torno de sistemas de informação. A existência de sistemas vem de longa data, na Antiguidade, desde as formas de registros de Jacó e Labão do controle de criação e comercialização de ovelhas, os registros egípcios e babilônicos das transações financeiras, que caracterizam um sistema.

Assim os estudos do alemão *Ludwig Von Bertalanffy* sobre a Teoria Geral de Sistemas surgem para a produção de conceitos no que tange à questão científica dos sistemas, em 1950 (REZENDE; ABREU, 2008). Cunha e Cavalcanti (2008, p. 340) conceituam sistemas como “conjunto de elementos inter-relacionados de forma lógica.” A concepção de sistema está integrada com a produção de informações, portanto sistemas de informação podem ser considerados todos os suportes que geram informações.

No meio empresarial, os sistemas de informação têm como objetivo auxiliar nos processos de tomada de decisões. A decomposição dos sistemas de informação se dá por meio dos processos de entrada, processamento e saída. A evolução dos sistemas de informação contribuiu para o aumento da sua complexidade. Diante disso, o desenvolvimento de sistemas de informação é integrado à tecnologia da informação, que por sua vez são os recursos tecnológicos para geração e uso da informação (REZENDE; ABREU, 2008).

Cunha e Cavalcanti (2008, p. 344) definem sistemas de informação como “[...] grupo lógico de subsistemas e dados ou informação, necessários para suprir as necessidades de informações de uma comunidade, grupo ou processo.” Por outro lado a tecnologia da informação é a “[...] aquisição, processamento, armazenamento e disseminação da informação vocal, pictórica, textual e numérica, através da combinação da informática e das telecomunicações.” (CUNHA; CAVALCANTI, 2008, p. 356).

Nos componentes da tecnologia da informação, estão incluídos: “[...] hardware e seus dispositivos e periféricos; software e seus recursos; sistemas de telecomunicações; gestão de dados e informações.” (REZENDE; ABREU, 2008, p. 52). O *software* são os programas que são executáveis por um computador independentemente de sua arquitetura, conteúdos, informações descritivas, tanto no suporte impresso quanto na forma virtual, que abrange todos os formatos de mídia eletrônica. Em todos os aspectos da vida humana, o *software* já se tornou parte integrante na convivência das pessoas (PRESSMAN, 2011). Rezende e Abreu

(2008) complementam que *software* ou aplicativo é um conjunto de instruções ou comandos executados por um computador para resolver determinados problemas.

Os procedimentos para o desenvolvimento de *software* são um processo de engenharia, porém não são fabricados no sentido clássico. Em todo desenvolvimento de um *software* deve-se, primeiramente, concentrar-se em um esforço para a compreensão do problema. As informações técnicas para o desenvolvimento de *software* são fornecidas pelos métodos da engenharia de *software*. “Os métodos envolvem uma ampla gama de tarefas, que incluem: comunicação, análise de requisitos, modelagem de projeto, construção de programa, testes e suporte.” (PRESSMAN, 2011, p. 40).

Os requisitos funcionais “[...] são as condições ou capacitação que devem ser contempladas pelo *software*, geralmente necessitadas pelo cliente e/ou usuário para resolver um problema ou alcançar um objetivo.” (REZENDE; ABREU, 2008, p. 9). As funções e atividades do sistema de informação são definidas e elaboradas com base em relatos dos usuários, utilizados para especificar o projeto de um sistema de informação, suas funções, desempenho, interfaces, restrições, entre outras descrições de conformidade com a metodologia de desenvolvimento de *software* (REZENDE; ABREU, 2008).

O desenvolvimento de um *software* é modulado, constam dois módulos principais: interface e armazenamento de dados. Na interface se concentram os comandos, operações, cálculo; faz-se o acesso dos dados que estão armazenados em uma base de dados e os apresentam para o usuário. Portanto, para o funcionamento de um *software* que necessita de armazenamento de informações, torna-se indispensável a modelagem da base de dados. Essa modelagem leva em consideração o tipo de sistema gerenciador de banco de dados (SGBD) (HEUSER, 2004), que é uma ferramenta para gerenciamento de base de dados, executa o armazenamento e a recuperação de modo eficiente dos dados e informações ao invés de armazenar em coleções de arquivos (RAMAKRISHNAN; GEHRKE, 2008). Entre os SGBD disponíveis no mercado estão os relacionais, que são os mais utilizados nos *softwares*. A técnica mais conhecida para SGBD relacionais é a abordagem entidade-relacionamento. Esta abordagem foi desenvolvida em 1976 por Peter Chen, e os conceitos centrais são: entidade, relacionamento, atributo, generalização/especialização e entidade associativa (HEUSER, 2004).

Entidade é “[...] qualquer objeto distinguível que deva ser representado no

banco de dados [...]." (DATE, 2004, p.11); já para Ramakrishnan e Gehrke (2008, p. 24), é o "objeto do mundo real distinguível de outros objetos"; e para Heuser (2004, p. 23), é o "[...] conjunto de objetos da realidade modelada sobre os quais deseja-se manter informações no banco de dados".

O relacionamento é "uma associação entre duas ou mais entidades." (RAMAKRISHNAN; GEHRKE, 2008, p. 25); para Heuser (2004, p. 24), é o "[...] conjunto de associações entre entidades." Cada relacionamento possui a cardinalidade definida como o "[...] número (mínimo, máximo) de ocorrências de entidade associadas a uma ocorrência da entidade em questão através do relacionamento." (HEUSER, 2004, p. 26).

O atributo é a descrição de uma entidade (RAMAKRISHNAN; GEHRKE, 2008). Heuser (2004, p. 34) define atributo como "[...] dado que é associado a cada ocorrência de uma entidade ou de um relacionamento." Ramakrishnan e Gehrke (2008, p. 24) complementam que a "[...] escolha de atributos reflete o nível de detalhes com os quais desejamos representar as informações sobre as entidades."

Por meio do conceito da generalização/especialização, é possível "[...] atribuir propriedades particulares a um subconjunto das ocorrências (especializadas) de uma entidade genérica." Integrada a esse conceito está a ideia de herança de propriedades das entidades (HEUSER, 2004, p. 39). Portanto é "[...] o processo de identificar subconjuntos de um conjunto de entidades que compartilham algumas características distinguíveis" (RAMAKRISHNAN; GEHRKE, 2008, p. 32).

Em relação à entidade associativa, Heuser (2004, p. 43) argumenta que "[...] nada mais é que a redefinição de um relacionamento, que passa a ser tratado como se fosse também uma entidade". Salienta-se que, no modelo relacional, "[...] os dados são representados por meio de linha em tabelas (tuplas em relações), e essas linhas podem ser interpretadas diretamente como proposições verdadeiras." (DATE, 2004, p. 13). Essas proposições podem representar um conjunto de atributos e seus conteúdos (linhas), por exemplo: *O identificador 1 é referente ao termo da palavra-chave Sistemas de Informação*; destaca-se que *identificador* e *palavra-chave* são os atributo de uma entidade. A tupla/linha é o conteúdo *1* e *Sistemas de Informação*. (exemplo elaborado pelo autor fundamentado em Date, 2004).

Nesta contextualização de sistemas de informação, tecnologias da informação, engenharia de software, banco de dados e os diversos conceitos e ferramentas para o desenvolvimento de *software*, os metadados têm fundamental

importância, até mesmo internamente, no SGBD, para descrição das tabelas criadas, seus atributos, proprietários da tabela, controles de acesso. No próximo tópico discute-se o periódico científico eletrônico.

### 3 PERIÓDICOS CIENTÍFICOS ELETRÔNICOS: CONTEÚDO E ACESSO

Este tópico aborda os aspectos inerentes aos conceitos em torno dos periódicos científicos eletrônicos. Em cada item deste tópico, é feito, inicialmente, uma contextualização histórica com os principais pontos regulatórios e os conceitos a partir dos autores pesquisados; realiza-se uma discussão entre as argumentações, de maneira a formar o conhecimento necessário para subsidiar o objeto desta pesquisa. Dentre os diversos tipos de comunicação científica abordados, o foco se concentra no periódico científico eletrônico.

#### 3.1 COMUNICAÇÃO CIENTÍFICA

O nascimento do PC está diretamente correlacionado com o início da comunicação científica, que por sua vez não tem, ao certo, a definição de quando tudo começou.

Para Meadows (1999, p. 7), “Os periódicos científicos surgiram na segunda metade do século XVII devido a várias razões.” Entre elas, estão algumas “[...] específicas (como a expectativa de seus editores de que teriam lucro); algumas, gerais (como a crença de que para fazer novos descobrimentos era preciso que houvesse um debate coletivo).” Mueller (2000b, p. 73) confirma também que “os periódicos científicos apareceram no século XVII na Europa, em uma época marcada por mudanças em toda a sociedade, inclusive no campo científico.” (MUELLER, 2000b, p. 73).

A origem da comunicação científica foi ao encontro da propagação da imprensa escrita. O marco desse surgimento é, no entanto, difícil de determinar. Meadows (1999) define que o surgimento da comunicação científica está vinculado ao início das pesquisas científicas e que essa comunicação, informada de diferentes maneiras, tinha a fala e a escrita como as formas mais relevantes. Diante da caracterização do surgimento da revista científica por Meadows (1999), faz-se importante transcrever o conceito de comunicação científica, que, segundo Lara (2006, p. 139), é o “processo que envolve a construção, comunicação e uso do conhecimento científico com o objetivo de promover sua evolução.”

A contribuição para a tradição da pesquisa comunicada de maneira escrita

[...] são as obras dos gregos, tendo à frente Aristóteles [...]”. Os debates que se realizavam eram registrados por meio de manuscritos e conservados de maneira precária, “[...] copiados repetidas vezes, influenciaram primeiro a cultura árabe e depois a Europa ocidental.” (MEADOWS, 1999, p. 3). A comunicação científica “compreende canais formais e informais utilizados pelos cientistas tanto para comunicar os resultados de sua pesquisa, como para se informar sobre os resultados alcançados por outros pesquisadores.” (LARA, 2006, p. 139). O informal utiliza-se de meios informais e provisórios, o que normalmente ocorre de maneira mais pessoal ou pesquisa ainda sem conclusão, por exemplo, trabalhos em congressos. O formal se refere aos canais formais e oficiais como os periódicos, livros, dentre eles, os periódicos têm importância maior para a ciência (MUELLER, 2000a).

Lara (2006, p. 411) complementa afirmando que “os aspectos formais compreendem o processo de disseminação e transferência da informação, viabilizados pelos encontros científicos e pelas publicações; os informais, principalmente, os contatos.” Esse sistema tem sua atuação centrada no campo científico de forma ativa, como uma espécie de jogo em que são estabelecidas relações de dependência entre os participantes, cujo objetivo é traçado de acordo com o interesse de cada envolvido e “[...] que envolve práticas e processos de comunicação formais [...].” (LARA, 2006, p. 411).

A comunicação científica, no presente momento, conta com uma diversidade de instigações e novas perspectivas. Juntam-se às publicações impressas “[...] jornais científicos on-line, fóruns de discussão, sistemas de *open archives* e *open Access* [...].” A partir dessas implicações, surge a necessidade de mais pesquisas e aprofundamentos teóricos sobre as novas tendências da comunicação científica (CORTES, 2006, p. 53).

Diante deste contexto, há as atividades de comunicação científica que passam a ser denominadas como sistema de comunicação científica. Nesse sistema, incluem-se os processos de comunicação formal e informal e seus suportes de publicações. Com o advento da tecnologia, são perceptíveis grandes mudanças nesse sistema, em decorrência da integração das redes eletrônicas e dos computadores, a comunicação torna-se mais eficiente e ultrapassa barreiras geográficas, hierárquicas e financeiras (MUELLER, 2000a).

Lara (2006, p. 411) define o sistema de comunicação científica como

“Sistema que compreende a construção, a ordenação, a circulação e o uso dos registros científicos.” Compreende, também, nesse contexto, um processo de interação social que se pode inferir como mecanismo de troca de informações entre cientistas e, conseqüentemente, um processo de comunicação científica de resultados de pesquisas também para a sociedade (LARA, 2006, p. 411). Considerado como um processo social interativo, pressupõe o intercâmbio de informações entre cientistas, o controle mútuo através dos processos de comunicação das pesquisas e o compartilhamento do conhecimento científico com a sociedade. “Embora o sistema de comunicação científica exista desde a Antiguidade, considera-se que os primeiros periódicos científicos publicados no séc. XVII sejam os precursores do modelo moderno de comunicação científica.” (LARA, 2006, p. 139).

Para Castro (2006, p. 58-59), o fluxo da comunicação científica tradicional é formado por etapas decorrentes uma da outra, executadas em um período longo de tempo e dependentes entre si. Na primeira etapa do fluxo, “[...] os autores redigem trabalhos resultantes de pesquisa ou de reflexão teórica. Terminada a redação, os autores submetem os trabalhos a uma editora ou revista científica para publicação [...]” (CASTRO, 2006, p. 58-59). Na próxima etapa deste fluxo, que é a revisão por pares, há um procedimento de avaliação por membros da comunidade científica que tenham, em sua especialidade, o tema do trabalho a ser avaliado. Na internet, as etapas decorrem concomitantemente e com espaço de tempo regular (CASTRO, 2006).

Diante desse tempo regular do fluxo da comunicação científica, tem-se o crescimento exponencial dos periódicos. Com isso, as revistas de resumos tendem a acompanhar esse crescimento, seguindo numa curva paralela a dos periódicos. Isso acarreta um volume em duplicidade de informações científicas e, a partir daí, surge um novo problema nesses últimos anos: “[...] qual a melhor maneira de localizar informações em periódicos de resumos [...]” As revistas de resumos são aquelas que “[...] trazem versões condensadas de artigos publicados em revistas científicas [...]” Essa modalidade de revista, por sua vez, representa parcialmente a literatura primária e está inserida na literatura secundária. Fora concebida no século XIX para melhorar os mecanismos de recuperação referenciados, àquela época, da literatura primária (MEADOWS, 1999, p. 31).

Sobre o fato de que a literatura primária e a secundária advêm das fontes de

informação, Pinheiro (2006, p. 3) esclarece: "Entre as fontes primárias estão os artigos de periódicos, os anais de congressos e eventos científicos, relatório de pesquisa, patentes, dissertações e teses etc." Já publicações secundárias compreendem "[...] as bibliografias, os dicionários e enciclopédias, os manuais, as publicações ou periódicos de indexação e resumos, artigos de revisão, catálogos etc.". A autora ainda faz referência às "fontes ou recursos de informação eletrônicos, que tanto abrangem fontes primárias, secundárias e terciárias, disponíveis eletronicamente na internet, quanto novas fontes construídas especificamente para meio eletrônico." (PINHEIRO, 2006, p. 3).

A comunicação científica configura-se em modelos, portanto são "Modelos que propõem ver o processo de comunicação científica ao longo do tempo através dos binômios centrais - construção/geração, comunicação/disseminação, uso/acesso do conhecimento científico." (LARA 2006, p. 403). Dentro desse contexto, tem-se o modelo clássico de comunicação científica que é dedicado ao processo de geração do conhecimento científico, "[...] o modelo que considera a constituição de um sistema de informação da ciência voltado à identificação de problemas da disseminação, e um modelo emergente, ou aberto, centrado no acesso à informação como pressuposto do uso da informação." (LARA 2006, p. 403). Portanto destaca-se que, no modelo emergente ou aberto, o foco está na visibilidade (LARA 2006).

Moreno e Márdero Arellano (2005, p. 78-81) explanam que "[...] a contribuição dos novos formatos digitais ao aperfeiçoamento dos processos tradicionais de editoração científica está visível nos novos modelos de comunicação científica, que começam a ser vislumbrados nestas duas últimas décadas. [...]" Com as contribuições tecnológicas no modelo de comunicação científica, "[...] o fluxo do processo editorial sofre uma modificação onde a impressão passa a ser a última fase antes da distribuição, podendo ser lida antes de ser impressa. [...]", conseqüentemente, a reprodução de documentos digitais independe de quantidades de impressões, porém tem necessidade de novos suportes e ferramentas para sua preservação, o que inviabilizada sua comercialização (MORENO; MÁRDERO ARELLANO, 2005, p. 78-81).

Com essa modalidade, o acesso aos documentos depende de uma rede e a disponibilidade dela. Portanto é importante salientar que, neste modelo de comunicação, o periódico científico eletrônico gera alterações "[...] no fluxo tradicional, na medida em que o número de etapas entre o início e a conclusão e

disseminação do conteúdo do artigo diminui consideravelmente. Como consequência, também há redução do tempo para se ter acesso ao mesmo [...]” (MORENO; MÁRDERO ARELLANO, 2005, p. 78-81).

Após esta contextualização do periódico científico na comunicação científica, o próximo subitem versa sobre as principais abordagens do periódico.

### 3.2 PERIÓDICO CIENTÍFICO: PRINCIPAIS ABORDAGENS

Anteriormente ao marco, o surgimento do PC, a ciência era dominada somente por filósofos que faziam as comunicações e publicações de seus materiais à comunidade por meio de cartas, nas quais era possível demonstrar o estudo sobre determinado assunto. O PC surge decorrente de uma necessidade de comunicação rápida e confiável de pesquisas, e este acontecimento se dá durante a ciência moderna (MEADOWS, 1999; MUELLER, 2000b).

Meadows (1999) e Burke (2003) relatam que, em 1660, surgem a *Philosophical Transactions da Royal Society*, em Londres, e o *Journal des Savants*, em Paris. O *Journal des Savants* fornecia mecanismos de divulgação de notícias e acontecimentos da República das Letras na Europa. Burke (2003) argumenta que o conhecimento em específico da academia era disseminado por meio de revistas cultas, publicadas mensalmente ou a cada dois meses, e que “[...] a revista culta podia ser um bom negócio é demonstrado pelo fato de que o *Journal des Savants* foi pirateado em Amsterdã e ‘Colônia’, e sua fórmula foi imitada em Roma, Veneza, Leipzig e outras cidades.” (BURKE, 2003, p. 152-153). Meadows (1999) complementa que a origem da *Royal Society* se deu ao fim de 20 anos de guerra, durante a restauração da monarquia em Londres, com o surgimento de pequenos grupos para debater questões relacionadas à filosofia, não entrando no mérito de assuntos polêmicos. Os membros da *Royal Society* faziam viagens para coleta de informações e construíam resumos da literatura publicada – este processo marca o surgimento da revista científica.

Price (1976) relata que não se deve esquecer que, mesmo no século XVII, o objetivo principal da *Royal Society* e do *Journal des Savants* era “[...] o controle e a assimilação de obras e cartas, já então demasiadamente numerosas para que um indivíduo pudesse acompanhá-las em sua leitura e correspondência diárias”, e não a publicação de novos artigos (PRICE, 1976, p. 9). “Em fins do século XVII, Amsterdã

era a sede da publicação das *Nouvelles de la République des Lettres* publicada por Henri Desbordes, e sua rival *Bibliothèque Universelle et Historique*, editada por Jean Leclerc.” (BURKE, 2003, p. 152-153).

Numa tentativa de comprimir a crescente realidade das publicações no século XVIII, surgem, aos poucos, e de forma tímida, os periódicos especializados e as resenhas, “[...] e que dizer de Sir Humphrey Davy, cujo hábito era jogar fora os livros que acabava de ler, baseado no princípio de que ninguém jamais teria tempo de ler a mesma obra duas vezes?”. A literatura especializada conta com a dedicação dos cientistas que se consideram inundados por esta avalanche de pesquisas que acontece em cada década e aumenta exponencialmente. “[...] A ciência foi sempre moderna; sempre explodindo sobre os indivíduos, continuamente no limiar de sua revolução expansiva [...].” (PRICE, 1976, p. 9).

Le Coadic (2004) afirma que, devido ao surgimento de novos conhecimentos, “[...] implantou-se progressivamente, um conjunto de estruturas que visam a proporcionar status científico e social à ciência da informação. 1) As revistas científicas são uma dessas estruturas.” (LE COADIC, 2004, p. 23). Existem, entre os autores, especificidades em relação aos termos periódico científico e revista científica. Meadows (1999, p. 8) traz o conceito de cada um: o termo periódico “[...] entrou em uso comum na segunda metade do século XVIII e se refere a qualquer publicação que apareça a intervalos determinados e contenha diversos artigos de diferentes autores. [...]”, levando em consideração a aplicação em sua pesquisa. Por sua vez, o autor conceitua o termo revista também quanto ao emprego em sua publicação, “[...] como uma maneira abreviada de nos referirmos a uma coletânea de artigos científicos escritos por diferentes autores. Conjuntos desses artigos são reunidos a intervalos, impressos, encadernados e distribuídos sob um título único. [...].” (MEADOWS, 1999, p. 7).

Em consonância com a definição de periódico de Meadows (1999), Lara (2006) complementa que o termo científico é todo “tipo de comunicação formal organizada em fascículos ou números, publicados segundo uma periodicidade [...].” Essa periodicidade pode ser “[...] definida, por tempo indeterminado, com tiragem e disseminação amplas, contendo predominantemente artigos científicos.” (LARA, 2006, p. 405).

Fundamentado nas definições de Meadows (1999) e Lara (2006) para periódicos e revistas científicas, pode-se inferir que as diferenças existem, porém a

maioria dos autores trata como sinônimos, quando é acrescido de uma específica atuação – a científica. Portanto há abordagens como periódicos científicos e revistas científicas; no entanto, para efeito deste estudo, será adotado o termo periódico científico.

Nesse tipo de comunicação, “dois processos [...] são utilizados: um processo escrito – formal – e um processo oral – informal. A formalização da comunicação científica data de mais de trezentos anos” (LE COADIC, 2004, p. 33). Esta necessidade do novo conhecimento e suas novas formas de disseminação é decorrente da necessidade de comunicação dos resultados de pesquisa entre os cientistas. “A ciência mudava de status: de atividade privada tornava-se, como vimos antes, uma atividade social” (LE COADIC, 2004, p. 33). O cientista passou a ser um trabalhador como qualquer outro de uma cadeia produtiva, em que lhe são exigidas produtividade e competitividade. Utilizando o jargão do meio empresarial, o cientista, para se manter no mercado, deve mostrar resultados. Além dessa preocupação, a comunicação deve permitir o reconhecimento pelos demais cientistas e o recebimento de contribuições. Isso proporcionará outros caminhos de pesquisa, portanto “[...] a ‘comunicabilidade’ é a característica principal da produção científica, pois permitirá o reconhecimento do cientista pelos pares e lhe garantirá sucesso na comunidade científica.” (LE COADIC, 2004, p. 33).

As revistas científicas surgiram com o objetivo de “[...] documentar as opiniões, ideias e resultados dos debates acadêmicos, comuns tanto nas sociedades científicas [...]” das áreas exatas e médicas como das áreas humanistas. Inicialmente, eram poucos artigos com resumos de processos e resultados científicos (GONÇALVES; RAMOS; CASTRO, 2006, p. 166-167). O marco na construção da comunicação científica foi pautado no surgimento da revista científica, em razão de necessidades como a troca de experiências e divulgação dos resultados de pesquisas. É importante e de grande relevância essa criação, ainda que o livro seja um meio de transmissão de informação (WEITZEL, 2006).

No Brasil, poucos periódicos comunicavam letras e artes, e os que existiam tiveram um ciclo de vida muito pequeno. Esse acontecimento ocorreu desde a chegada da Corte até a década de 1930, momento em que estavam sendo construídas e passando por formação, as questões política e cultural. Também contribui com tal fato a estrutura administrativa, educacional e científica não estar ainda latente. Os periódicos de cunho político não tiveram escassez. O termo

utilizado, naquela época, para o que hoje se define como periódico científico, tinha outras formas de escrita, como: “[...] revista literária, jornal de cultura, jornal de ciências e artes, e, principalmente, jornal literário [...]” Portanto existe uma dificuldade de identificação dos periódicos científicos no Brasil. Numa revisão histórica, os termos utilizados para definição de PC da época estão relacionados ao momento cultural e científico vivido (FREITAS, 2006, p. 57).

Contudo a importância dos periódicos científicos aumenta, desde que a Ciência e a Tecnologia passam a ser compreendidas como fundamentais no desenvolvimento sustentável dos países, assim como em proporcionar o bem-estar da população. Portanto, para essa consequência, os periódicos se tornam canais de comunicação científica majoritários na disseminação de resultados de pesquisas (RODRIGUES et al., 2011). Os resultados de pesquisas que não vão a público por meio destes mecanismos de propagação científica têm sua importância menosprezada e ainda não recebem o devido reconhecimento dos especialistas. A revista científica, então, pode ser considerada o meio formal da comunicação científica, e os artigos nela contidos são a publicação definitiva dos resultados de pesquisa, que serão lidos e até mesmo citados (GONÇALVES; RAMOS; CASTRO, 2006).

Durante os anos de reinado de Dom Pedro I, tem-se imigração de franceses e, com este acontecimento, “as condições políticas para o ramo melhoram significativamente. Aliadas as condições políticas, começam a surgir condições materiais para seu desenvolvimento.” Máquinas de tipografias são importadas e com isto há “[...] intensa produção de pasquins [...]” Todas essas implicações políticas tiveram influência nas instituições educacionais e científicas, essas instituições, porém, não se firmaram, devido à própria cultura científica daquela época (FREITAS, 2006, p. 62).

Cortes (2006) entende que as edições de livros e periódicos científicos trazem alguns efeitos benéficos juntamente com um conjunto de características como “Facilidade de reprodução do texto original, permitindo sua distribuição e utilização em diversos locais; - Redução dos custos de difusão; - Alta possibilidade de retenção e documentação por parte do receptor [...]”. (CORTES, 2006, p. 46). Em outras palavras, o material impresso torna-se indelével, desde que o receptor tenha todos os cuidados no armazenamento e gerenciamento desse material. Por outro lado, torna fácil o relacionamento entre pesquisas e proporciona um avanço do

conhecimento em várias temáticas, reforça a eficiência da comunicação, eliminando os ruídos e distorções do conhecimento, próximo da totalidade. “Aumento significativo da velocidade de difusão; - Criação de jornais e revistas científicas, incrementando a troca de ideias e incentivando o debate.” (CORTES, 2006, p. 46). Conseqüentemente, o conhecimento é atingido em maiores proporções, tanto geograficamente quanto número de pesquisadores.

Como parte integrante do periódico científico estão os artigos científicos, assunto de que trata o próximo item.

### 3.2.1 Artigos Científicos

As revistas científicas permanecem inalteradas até 1900, quando ainda não se tem nenhum artigo científico de acordo com os modelos atuais, portanto “[...] a transformação do artigo científico em sua forma atual não se completou senão há cerca de um século atrás.” Em épocas anteriores a essa transformação, o que se tinha eram publicações científicas em formato de notícias, descritas de forma simples com as observações realizadas e já publicadas, como uma espécie de revisão. Por outro lado, publicações monográficas que também existiam poderiam ter se firmado em livro se contassem com meios avançados de impressão e distribuição. Portanto “a diferença não consiste apenas na extensão - se muito curtos, são cartas; se muito longos, monografias” (PRICE, 1976, p. 41).

Segundo Cunha e Cavalcanti (2008, p. 32), artigo científico é o “texto escrito que foi aprovado para publicação ou publicado num periódico científico.” A NBR 6022 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003a, p. 2) trata o artigo científico como “parte de uma publicação com autoria declarada, que apresenta e discute ideias, métodos, técnicas, processos e resultados nas diversas áreas do conhecimento”. Já o artigo de revisão é “parte de uma publicação que resume, analisa e discute informações já publicadas”. Trata também do artigo original, que é definido como “uma publicação que apresenta temas ou abordagens originais.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003a, p. 2). Em outro ponto, a NBR 6022 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003a, p. 3) classifica o artigo em “[...] original (relatos de experiência de pesquisa, estudo de caso, etc.)” e de revisão.

Marconi e Lakatos (2006) definem os artigos científicos como “[...] pequenos

estudos, porém completos, que tratam de uma questão verdadeiramente científica, mas que não se constituem em matéria e um livro.” Devido à sua forma reduzida de apresentação de resultados dos estudos, tornam-se um diferencial dos demais tipos de trabalho científico. Por serem completos, permitem ao leitor reaplicar a experiência por meio da “[...] descrição da metodologia empregada, do processamento utilizado e resultados obtidos.” A partir da conclusão de um trabalho, seja documental, bibliográfico ou de campo, é necessária a sua publicação. Essa publicação acontece “em revistas ou periódicos especializados e formam a seção principal deles.”, isso “[...] proporciona não só a ampliação de conhecimentos como também a compreensão de certas questões.” (MARCONI; LAKATOS, 2006, p. 263).

Segundo Rodrigues (2011), um artigo é “[...] um texto científico escrito por um ou mais autores, com a finalidade de divulgar e tornar conhecidos, através de sua publicação em revistas, jornais ou periódicos especializados [...]” O tornar conhecido é expor as “[...] ideias, os métodos, as técnicas, os processos e os resultados de estudos e pesquisas.” (RODRIGUES, 2011, p. 8). O autor ainda faz menção de que o artigo pode ser original ou de revisão. É original “[...] quando apresenta temas e abordagens inéditas, resultantes de pesquisa científica, como relatos de caso, comunicação ou notas prévias.” É de revisão “[...] quando analisa, discute e resume um assunto já publicado.” (RODRIGUES, 2011, p. 8).

Marconi e Lakatos (2006) complementam ainda que “[...] os artigos podem ser de três tipos: argumento teórico, artigo de análise e artigo classificatório.” O argumento teórico é um “[...] tipo de artigo que apresenta argumentos favoráveis ou contrários a uma opinião.” O artigo de análise caracteriza-se como aquele em que “[...] o autor faz análise de cada elemento constitutivo do assunto e sua relação com o todo.” (MARCONI; LAKATOS, 2006, p. 263). No classificatório, “o autor, nesse caso, procura classificar os aspectos de um determinado assunto e explicar suas partes.” (MARCONI; LAKATOS, 2006, p. 264). Além desta caracterização dos artigos, salienta-se que Pinheiro (2006) insere o artigo na fonte primária e na secundária como artigos de revisão.

A elaboração do trabalho científico se dá “[...] por processo que lembra o de coser ou a maneira de acomodar as peças de um quebra-cabeças, ligando-se umas às outras.” Então o artigo científico tem sua matéria prima centrada no que se valem os cientistas, para que assim possa, a partir de outros trabalhos que serão citados, compor as referências de um novo artigo que está em desenvolvimento. Poder-se-ia,

portanto, fazer uma analogia com a família humana, com exceção ao “[...] fato de que, em vez de serem necessários dois pais para produzir um filho, fazem-se necessários cerca de doze pais diversos [...]” e que, com diversas combinações, pode ser produzido um filho por ano (PRICE, 1976, p. 117).

Para o autor supracitado, “se um homem trabalha e o principal resultado de sua pesquisa é conhecimento, algo que deve ser dado a público, para que ele possa reclamar-lhe a paternidade, dizemos que esse homem está dedicado à ciência.” Por outro lado, considera-se que se o resultado desse trabalho gera “[...] uma coisa, um produto químico, um processo, algo para ser comprado e vendido [...]” (PRICE, 1976, p. 117). A dedicação desse humano está estritamente focada na tecnologia; porém, na atual situação, não são considerados somente os resultados, mas também a matéria prima utilizada na produção desses produtos científicos.

Os cientistas disseminavam os resultados de pesquisas às quais se dedicavam por meio de publicações especializadas, primordialmente livros. Com a necessidade de rápida atualização, passaram a ler artigos, posteriormente cartas dirigidas ao editor (PRICE, 1976, p. 118). Uma ressalva importante para a disseminação da pesquisa por meio de artigos científicos, tratando-se do cenário atual, é ser um dos mecanismos com abrangência rápida, seja geográfica ou tematicamente. No século XIX, Charles Babbage, na Inglaterra, e Nathaniel Bowditch, nos Estados Unidos, deploravam, amargamente, o descaso pela nova era científica que justamente explodia sobre eles (PRICE, 1976, p.9).

Consequentemente, um passo após a elaboração do trabalho científico, é observar a estrutura do artigo científico que segundo a NBR 6022 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003a, p. 3) é composto por três partes, as quais são apontadas por meio do Quadro 6.

### Quadro 6 - Estrutura do artigo

5.1 Elementos pré-textuais	
Os elementos pré-textuais são constituídos de:	
	<ul style="list-style-type: none"> <li>a) título, e subtítulo (se houver);</li> <li>b) nome(s) do(s) autor(es);</li> <li>c) resumo na língua do texto;</li> <li>d) palavras-chave na língua do texto.</li> </ul>
5.2 Elementos textuais	
Os elementos textuais constituem-se de:	
	<ul style="list-style-type: none"> <li>a) introdução;</li> <li>b) desenvolvimento;</li> <li>c) conclusão.</li> </ul>
5.3 Elementos pós-textuais	
Os elementos pós-textuais são constituídos de:	
	<ul style="list-style-type: none"> <li>a) título, e subtítulo (se houver) em língua estrangeira;</li> <li>b) resumo em língua estrangeira;</li> <li>c) palavras-chave em língua estrangeira;</li> <li>d) nota(s) explicativa(s);</li> <li>e) referências;</li> <li>f) glossário;</li> <li>g) apêndice(s);</li> <li>h) anexo(s).</li> </ul>

Fonte: Elaborado pelo autor, de acordo com a ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003a, p. 3).

Importante observar, diante da apresentação no Quadro 6, a composição dos elementos pré-textuais que, por sua vez, são elementos que têm o objetivo de representar o conteúdo do artigo científico. Entre esses mecanismos de representação está elencado, no item 5.1, o objeto de pesquisa deste estudo, que são as palavras-chave, e cabe salientar sua relevância.

Price (1976, p. 42) faz o seguinte questionamento quanto ao protótipo do artigo científico: “Se o protótipo do artigo científico moderno é mais um instrumento social do que uma técnica para acumular '*quanta*' de informação, a que força poderosa se deve atribuir seu aparecimento e conservação?”. Com toda certeza, a resposta dessa questão está centrada na conservação da propriedade intelectual. Diante desta realidade, encontra-se uma necessidade dos cientistas de trazer, para si, o novo conhecimento.

O que se buscou identificar, nesta etapa do trabalho, foram as diversas manifestações no periódico científico desde seu surgimento. Foram repassadas algumas nomenclaturas que caracterizam os vários tipos de publicações em

evolução: periódicos especializados que continham resultados de pesquisas dos cientistas; resenhas que cumpriam o papel de publicar o que já foi publicado; publicações especializadas que são compostas por livros, artigos e cartas dirigidas ao editor; publicações científicas no modelo de notícias como forma de revisão de outras publicações; publicações monográficas que poderiam ser transformadas em livros. No Brasil, são várias as denominações para periódico científico: revista literária, jornal de cultura, jornal de ciências e artes, jornal literário.

Em meio às diversas evoluções do periódico científico, o próximo item abordará o periódico científico com o suporte computacional, o que origina o periódico científico eletrônico, suas características fundamentais que ultrapassam os limites do papel e integram-se com recursos como hipertexto, *links* e multimídias, ou seja, um formato de comunicação científica dentro de um modelo de processos eletrônicos de editoração.

### 3.3 PERIÓDICO CIENTÍFICO ELETRÔNICO

As transformações tecnológicas têm influenciado na evolução das revistas científicas, a partir da década de 1960, com o tratamento bibliográfico realizado por computador. Além do armazenamento e edição de textos complexos e com grande quantidade de informações, surgem também as redes de telecomunicações e o “[...] mercado de informações on-line [...]”, com a criação de revistas eletrônicas e bases de dados de textos completos (GONÇALVES; RAMOS; CASTRO, 2006, p. 168-169). Esta afirmação também é feita por Meadows (1999), que ainda adiciona que o computador visto como “devorador de números” (MEADOWS, 1999, p. 33) poderia ser utilizado no trabalho com informações alfabéticas.

Os textos são caracteres codificados em um sistema binário, cuja quantidade de *bits* é decorrente do código utilizado. A transmissão desses textos se dá por meio de sinais elétricos, de modo que, de acordo com o volume de informações, serão centenas de caracteres, o que resultará em milhares de *bits* (LE COADIC, 2004, p. 91-92). Para Meadows (1999), são várias as vantagens proporcionadas pelo computador, e um dos objetivos na utilização dessas vantagens era encontrar, na literatura, os itens que satisfaziam a necessidade dos usuários; o uso clássico a ser obtido foi a busca por palavras-chave, que o autor citado acima define como: “[...] palavra ou frase que sirva para identificar o conteúdo do item

respectivo; isto é, exclui vocábulos como o, de, etc.” (MEADOWS, 1999, p. 33).

Com seu surgimento, a eletrônica “[...] quase se traduziu na transição dos suportes materiais para suportes imateriais [...]”. Como decorrência, esse acontecimento reforçou as tendências da informática e da comunicação à distância por meio das telecomunicações. “De multiplexação, amplificação e armazenamento de enormes volumes de informações ocorrem sem cessar e, às vezes, nos fazem duvidar da cordialidade da nova sociedade da informação!” (LE COADIC, 2004, p. 6).

Mueller (2000b) menciona a mudança profunda do acesso à informação decorrente do surgimento das redes de internet, que após 1994, tiveram uma grande propagação. Essa fase pode ser considerada como uma etapa de transição em que se tem a interação entre o sistema de publicação tradicional e o sistema de publicação eletrônico. Ambos os sistemas se diferenciam por ser o tradicional muito rígido e o eletrônico, aberto. Essa interação cresce mais com a inclusão do periódico eletrônico que, por sua vez, traz consigo as características dos periódicos tradicionais. Essa circunstância se dá em um período instável.

Não é difícil de identificar a facilidade e a flexibilidade dos documentos eletrônicos, em especial, os hipertextos. Neles, as informações são armazenadas em uma rede de nós e podem conter gráficos, áudio, vídeo. Os nós possuem ligações que permitem o vínculo de todas as mídias contidas no documento. “Os vínculos unem os verbetes: do texto lido aos textos a ler, da ilustração ao trecho de música... Sempre se pode modificá-los, ao contrário do documento impresso.” (LE COADIC, 2004, p. 59). Portanto, para esse tipo de documento eletrônico, surgiu o *Hypertext Mark-up Language* (HTML), que “[...] é o hipertexto ativo na tela da internet. As fichas tornam-se páginas da Rede, as pilhas, sítios, e os botões, apontadores.” Diante de tantas funcionalidades, esta linguagem de hipertexto tornou-se pouco extensível para descrever todos os tipos de documentos. “Surgiu, então, um novo tipo de linguagem *eXtensible Markup Language* (XML), derivado da *Standard Generalized Markup Language* (SGML).” (LE COADIC, 2004, p. 59).

O periódico científico eletrônico (PCE) é “[...] um formato eletrônico de publicação [...]” que se utiliza dos benefícios de obter, em seu conteúdo, recursos como imagens, áudio, vídeo, *links* que direcionam para outros documentos. Usualmente é disponibilizado, por meio da Internet, com o suporte de hipertexto que pode conceder, ao leitor, o acesso às referências utilizadas na referida publicação.

Uma exceção na disponibilidade: além da internet, o periódico científico eletrônico também pode estar no suporte de CD-ROM (LARA, 2006, p. 406). Mueller (2000b) complementa que o PCE é o termo utilizado para os periódicos que fazem uso de equipamentos eletrônicos para obtenção de acesso, coaduna com o conceito de Lara (2006) e o classificam em duas categorias denominadas: *online* e em CD-ROM.

Mueller (2000b) corrobora com Gonçalves, Ramos e Castro (2006) ao tratar essas características dos formatos dos periódicos científicos eletrônicos em uma franca evolução e concorda com Gonçalves, Ramos e Castro (2006, p. 169) que relatam a utilização do termo eletrônico para diversos tipos de publicação. As revistas eletrônicas não têm as mesmas características e nem os mesmos formatos, isto em se tratando de um meio comum de comunicação. Tais características podem ser associadas em três grupos: 1) “[...] cópias exatas das versões impressas, normalmente apresentadas em arquivos em formato PDF [...]”; 2) textos com “[...] recursos adicionais da publicação eletrônica (navegação hipertextual e multimídia)”; 3) matérias “[...] criadas e publicadas exclusivamente em meio eletrônico [...]”, mesmo com a utilização mínima de recursos hipertextual e multimídia (GONÇALVES; RAMOS; CASTRO, 2006, p. 169).

Mueller (2000b) explica que as fontes de informações úteis para os pesquisadores e cientistas, além do periódico eletrônico, podem ser “[...] os boletins ou *newsletters*, listas de discussões ou *listserves*, sítios de editoras e outras tantas [...]” que, por não terem as características próximas aos periódicos, não farão parte da análise deste trabalho.

As características do periódico científico eletrônico são semelhantes às dos periódicos impressos, “[...] inclusive na periodicidade e na maneira de identificar volumes e fascículos, especialmente aqueles que são apenas versão eletrônica de um periódico existente em formato tradicional” (MUELLER, 2000b, p. 82-83).

A crescente quantidade de informação está diretamente relacionada com os tipos em que ela se manifesta, no formato eletrônico, como também a velocidade e a eficiência das redes de transmissão, que, por sua vez, gera uma economia para o sistema de comunicação, aí está a origem da revista eletrônica oportunamente ao que se tem hoje (LE COADIC, 2004, p. 101). Le Coadic (2004) traça uma linha do tempo para o surgimento das revistas eletrônicas, refletido em três momentos principais, apresentados no Quadro 7.

**Quadro 7 - Histórico das revistas científicas eletrônicas**

Momento	Órgão financiador	Localização	Denominação	Descrição geral
Primeira 1978 a 1980	National Science Foundation	EUA, New Jersey Institute of Technology	EIES (Eletronic Information Exchange System)	“[...] abrangia correio eletrônico (informal <i>news-latter</i> ), conferência com auxílio de computador ( <i>unreferred paper fair</i> ), um boletim dirigido por um editor com ajuda de especialistas e um bloco de notas.” (LE COADIC, 2004, p. 101).
Segunda 1980 a 1984	<i>Department of Research and Development da British Library</i>	Grã-Bretanha Universidades de <i>Birmingham e Loughborough</i>	<i>Blend (Birmingham and Loughborough Eletronic Network Development)</i>	“[...] levou à edição da revista <i>Computer Human Factors</i> , que chegou a publicar cerca de 50 artigos e contou com uns sessenta assinantes.” (LE COADIC, 2004, p. 101).
Terceira 1984 a 1987		França <i>Direction des Bibliothèques, des Musées et de l'information Scientifique et Technique (DBMIST)</i>	<i>JOURNAL REVUE</i>	“[...] incluía uma revista e um jornal.” Ao mesmo tempo, réplicas eletrônicas de revistas em papel foram desenvolvidas pela Royal Chemical Society, Elsevier, Cornell University, American Chemical Society e American Medical Association. Surgiram newsletters eletrônicas, como <i>Online Chronicle</i> , <i>Videotech</i> , etc.” (LE COADIC, 2004, p. 101-102).

Fonte: Elaborado pelo autor, fundamentado em Le Coadic (2004, p. 101-103).

A partir daí, outros acontecimentos foram se replicando como, por exemplo: “[...] o projeto *Electronic Journals on SuperJANET* reúne nove editores científicos em torno do *Institute of Physics* [...]”. Depois, tem-se uma proliferação de organizações científicas que oferecem a revista científica distribuída por meio da internet, por exemplo, a “[...] *American Society for Biochemistry and Molecular Biology* e o *Journal of Biological Chemistry – online version* [www.jbc.org/].” (LE COADIC, 2004, p. 103). Por outro lado, aparecem as organizações que visam à venda de informações, como a *Elsevier*, *Springer*. Essa oferta de informação tem, como contrapartida, vários benefícios, como o acesso a recentes publicações por meio do serviço de alerta de citações. “Daí a reação dos cientistas no sentido de fazer com que a literatura científica mundial volte a ser um recurso público.” (LE COADIC, 2004, p. 103).

Um exemplo dado por Gonçalves, Ramos e Castro (2006, p. 169) ilustra o início da publicação eletrônica: “Na América Latina, algumas iniciativas pioneiras tiveram lugar no início da década de 1990: no Brasil, a do *Journal of Venomous Animals and Toxins*, e do Grupo de Publicações Eletrônicas em Medicina e Biologia (E-pub) [...]”, as quais tiveram início com a publicação no formato eletrônico (GONÇALVES; RAMOS; CASTRO, 2006, p. 169). Tem-se “[...] no México, a coleção

de periódicos eletrônicos Artemisa, produzida pela *Red Nacional de Colaboración en Información y Documentación en Salud*, inicialmente em CD-ROM e hoje on-line.” (GONÇALVES; RAMOS; CASTRO, 2006, p. 169).

Na Figura 3, Le Coadic (2004) apresenta o antigo e o novo paradigma da ciência e da tecnologia.

**Figura 3** - Os novos paradigmas na ciência da informação

	O ANTIGO PARADIGMA		O NOVO PARADIGMA
CIÊNCIA	trabalho individual	→	trabalho coletivo
	acervo	→	fluxo
	orientado para bibliotecário	→	orientado para usuário
TECNOLOGIA	papel		→ Elétron

Fonte: Le Coadic (2004, p. 108).

Do ponto de vista ético, as publicações de pesquisas custeadas pelo governo, ou seja, com recursos públicos, deveriam ser de acesso aberto. Porém, no sistema de comunicação, não é o que ocorre: geralmente, os usuários devem pagar para ter acesso aos resultados de pesquisas, sejam elas realizadas com recursos públicos ou privados. Essa é somente uma pequena proporção que relaciona acesso livre versus direitos autorais dos autores, os quais são entregues aos editores. Em muitos casos, o autor efetua o pagamento de uma taxa para ter sua produção avaliada pelos pares e possivelmente publicada (KURAMOTO, 2006).

Diante do aumento dos valores das assinaturas de periódicos científicos, inicia-se um processo de desuso do conhecimento científico por meio deles, o que ocasiona uma diminuição ao acesso, bem como sua audiência. Essa cobrança bloqueou o livre acesso ao fluxo do conhecimento científico e aos resultados de pesquisas, portanto impediu o crescimento da ciência. Por outro lado e em decorrência deste cenário, inúmeros cientistas estão em busca de formas alternativas para o desenvolvimento da ciência (MARCONDES; SAYAO, 2002), o que faz surgir o movimento de acesso livre de que trata o próximo item.

### 3.3.1 Movimento de Acesso Livre

Surge o movimento de acesso livre, que vem ganhando proporções mediante as declarações e manifestos de instituições do mundo todo, por exemplo: Bethesda, Budapeste, Berlim, Brasil – Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). O movimento de acesso livre não se pauta somente no discurso, mas também em suportes tecnológicos e ferramentas, por meio do modelo *open archives*. Esse modelo estabelece padrões de interoperabilidade entre os repositórios digitais (KURAMOTO, 2006). A interoperabilidade é a palavra-chave para a criação dos arquivos abertos (TRISKA; CAFÉ, 2001 p. 93). No Quadro 8, apresentam-se os principais acontecimentos do *Open Archives*.

**Quadro 8 - Principais acontecimentos OA**

<b>Data</b>	<b>Acontecimento</b>
10/1999	Lançamento da <i>Open Archives Initiative</i> , pela Convenção de Santa Fé
2001	Carta aberta da <i>Public Library of Science (PLoS)</i>
14/02/2002	Iniciativa de Budapeste para o Acesso Aberto
30/10/2002	Carta ECHO
11/04/2003	Declaração de Bethesda
27/08/2003	<i>Association fo Learned and Professional Society Publishers (ALPSP)</i>
22/10/2003	Declaração de Berlim sobre o Livre Acesso ao Conhecimento
11/2003	Declaração de Princípios do Wellcome Trust em apoio à edição em livre acesso
4/12/2003	Posicionamento do InterAcademy Panel sobre o acesso à informação científica
5/12/2003	Declaração do <i>Internacional Federation of Libraries Association (Ifla)</i> sobre o livre acesso à literatura científica e aos documentos da pesquisa
12/12/2003	Declaração de Princípios da Cúpula Mundial sobre a Sociedade da Informação (SMSI)
15/01/2004	Declaração de Valparaíso
30/01/2004	Declaração da <i>Organization for Economic Cooperation and Development (OECD)</i> sobre o acesso aos dados da pesquisa financiada por fundos públicos
16/03/2004	Princípios de Washington D. C. para o Livre Acesso à Ciência
30/07/2004	Publicação do relatório do comitê do Parlamento Britânico sobre edição científica
13/09/2005	Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica
26/09/2005	Declaração de Salvador sobre Acesso Aberto: A Perspectiva dos Países em Desenvolvimento. Declaração de Salvador – Compromisso com a Equidade
12/2005	Carta de São Paulo
05/2006	Declaração de Florianópolis
23/05/2007	PL – 1120/2007 - Dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências.
17/10/2007	Políticas locais de OA para o Brasil e outros países de língua portuguesa

Fonte: Adaptado de Kuramoto (2006).

A partir daí, surgem os arquivos abertos (*Open Archives*), cujo mecanismo possibilita que os cientistas divulguem os resultados de suas pesquisas. O primeiro modelo que surgiu foi “[...] a coleção de pré-prints publicações (*preprints*) e artigos eletrônicos em física, que abrange cerca de 150 mil consultas por dia ([HTTP://arxiv.org](http://arxiv.org)).” Na CI, é possível visualizar dois exemplos: “*Eprints in Library and Information Science* e *Archive Science Information Science Communication* (SIC) (<http://eprints.rclis.org>; <http://archivesic.ccsd.fr>).” (LE COADIC, 2004, p. 103-104).

É importante definir os termos *Open Archives Initiative* (OAI) e o movimento de acesso livre, “A OAI é uma iniciativa que surgiu com a Convenção de Santa Fé, em 1999. O Movimento de Acesso Livre surgiu com a Declaração de Budapest em 2001.” Como ambos os movimentos estão trabalhando para um objetivo, que é o acesso livre, estão inclusos no modelo OA que é baseado no modelo *open access*. Algumas dificuldades são encontradas na fase de transição do modelo tradicional para o modelo OA (WEITZEL, 2006, p. 103).

O IBICT, ao desenvolver o apoio ao manifesto de acesso livre à informação, esboça uma política nacional de acesso livre. Um exemplo prático é a implantação da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que já inicia um processo referenciado no modelo *open archives*, tanto o acesso em nível nacional quanto o acesso local na Instituição de Ensino Superior (IES) (KURAMOTO, 2006). O acesso livre é o mecanismo de disponibilizar a literatura de cunho acadêmico ou científico na internet, com a permissão dada para “[...] qualquer utilizador ler, baixar arquivos, copiar, distribuir, imprimir, pesquisar, indexar, fazer *links* ou referenciar o texto integral dos documentos.” Cabe aos autores procurar preservar a integridade do trabalho. Os princípios e planejamentos de implantação são disponibilizados no documento *Budapest Open Access Initiative* (BOAI) (LARA, 2006, p. 389).

O movimento *open archives* origina-se nos EUA no *Laboratório Nacional de Los Alamos*, na década de 1990, o qual implantou o repositório digital *arXiv*. A experiência de criar um repositório digital foi decorrente das dificuldades de acesso às publicações, em razão do custo muito alto, e nem sempre as respostas às necessidades de pesquisas davam retorno em tempo hábil (KURAMOTO, 2006). Lara complementa relatando que o movimento *open archives* são “arquivos ou repertórios de trabalhos científicos em formato digital (*ePrints*) disponibilizados para o acesso público via *ftp* ou *http*.” Portanto são incluídos nestes repositórios trabalhos

científicos em suas versões preliminares que foram submetidas à avaliação de pares no processo tradicional de publicação de artigos científicos. Com isso, torna-se um modelo alternativo e equitativo para disseminação de materiais científicos, instigando a discussão e a troca de conhecimento de maneira rápida (LARA, 2006, p. 390-391).

A OAI é uma organização com esforços explícitos e transitórios, empenhada em explorar novas aplicações no âmbito das tecnologias e suas aplicabilidades, compreender o universo e a cultura das comunidades adeptas ao movimento e, desse modo, adequar-se às evoluções, à missão e a sua própria organização. O objetivo da OAI é promover a interoperabilidade entre os provedores de dados por meio de padrões e, com este objetivo, proporcionar maior disponibilidade da comunicação científica (OPEN ARCHIVES INITIATIVE, 2014).

Gonçalves, Ramos e Castro (2006) coadunam com a perspectiva de a OAI possibilitar a interoperabilidade de dados. Por outro lado, Gonçalves, Ramos e Castro (2006) e Packer e Meneghini (2006) complementam que a OAI pode aumentar a visibilidade das revistas científicas e favorecer essa tendência da comunicação científica que potencializa mais possibilidades de publicações internacionais. Lara (2006) complementa que o OAI “[...] possui código compartilhado para campos de metadados (exemplo: “*date*”, “*author*”, “*title*”, “*journal*” etc), de forma que os textos completos, mesmo em diferentes formatos, são interoperáveis entre si.” (LARA, 2006, p. 391).

O modelo *open archives*, em sua estrutura tecnológica, fornece o devido suporte e conta com um protocolo de comunicação que permite a coleta de metadados, denominado de *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). Suas funções estão relacionadas com o processo de coleta de metadados dos provedores de dados. O provedor desse serviço de coleta deve utilizar ferramentas de *harvester* que realizam a coleta de dados dos provedores de dados que, por sua vez, devem estabelecer um padrão de metadados, geralmente os provedores utilizam o *Dublin Core*. Porém existem outros padrões, por exemplo, a BDTD que utiliza o padrão de metadados MTD-BR – Padrão Brasileiro de Metadados de Teses e Dissertações. Portanto o protocolo OAI – PMH pode se adequar aos vários padrões de metadados. Esse padrão tem influências na interoperabilidade; no entanto estabelecer um padrão de metadados é proporcionar um alto nível de interoperabilidade entre sistemas de repositórios digitais (KURAMOTO, 2006; MARCONDES; SAYAO, 2002).

Marcondes e Sayao (2002) abordam o funcionamento técnico do OAI-PMH. São seis os comandos de comunicação com o provedor de dados, a partir do programa robotizado do provedor de serviços: *identify* (dados administrativos, políticas), *ListSets* (estrutura do repositório), *ListMetadataFormats* (formato de metadados do repositório), *ListIdentifiers* (identificadores de registros), *ListRecords* (metadados no formato do repositório) e *GetRecords* (metadados de um registro específico). Essa estrutura é a base da comunicação para que ocorra a interoperabilidade entre os sistemas e a possível coleta de dados.

Esses repositórios digitais ou *e-prints* não têm um conselho editorial para avaliar as produções e surgiram como alternativa do sistema de comunicação científica tradicional. Os *papers* eram, no entanto, moderados por meio de um pacote de *software* que fazia o papel de filtro e garantia que seu escopo fazia parte desse repositório. Nesse contexto, se fez importante a atuação dos leitores que, por sua vez, poderiam escrever comentários e o autor poderia responder a esta interação com uma nova versão (KURAMOTO, 2006). Para Weitzel (2006), esta modalidade de comunicação científica surge para inibir a proliferação de artigos para somente cumprir formalidades e aumentar a consistências dos artigos, que se materializam por meio dos arquivos abertos, denominados *preprints*. A distribuição do conhecimento, neste contexto, apoia-se na confiança da tecnologia e rompe barreiras de tempo e geográfica.

Os *preprints* são os artigos que ainda não foram publicados oficialmente, essa modalidade de comunicação científica vem ganhando espaço e importância em algumas áreas do conhecimento. Por outro lado, o periódico tradicional demora mais tempo até ser avaliado e citado pelos leitores. Os *preprints* são depositados utilizando o modelo de livre acesso. Vale ressaltar que o artigo depositado pode ser publicado, pode ter alterações significativas para ser publicado, ou pode não ser publicado, ou seja, aceito ou rejeitado pela editores. Os artigos depositados nessa base, geralmente, não passam por avaliação; ainda que tenha esta limitação ou ineficiência, essas bases são muito bem aceitas e vêm sendo utilizadas cada vez mais (LARA, 2006; MARCONDES; SAYAO, 2002; MUELLER, 2000b). Lara (2006) complementa que “os pesquisadores são encorajados a fazer o auto-arquivamento das duas versões, relacionadas entre *links* nos arquivos *ePrints*.” (LARA, 2006, p. 396).

Triska e Café (2001) esclarecem que “O auto-arquivamento refere-se ao

direito de o próprio autor enviar o seu texto para publicação sem intermédio de terceiros.” O objetivo dessa modalidade de divulgação é tornar o material científico disponível mais rapidamente, coibindo o monopólio de editoras científicas, porque até então era poder delas desencadear tal processo de comunicação científica (TRISKA; CAFÉ, 2001 p. 92). Corroborando com Mueller (2000b), Silva, Ramos e Noronha (2006) complementam que esse procedimento de comunicação científica, por meio de *open archives*, tem formado círculos restritos de editores e autores, e de maneira alternativa, disseminado resultados de pesquisas ou saberes manifestados. A característica principal do *open archives* é de ser uma opção não comercial.

É grande o impacto dos padrões de tecnologias da *Open Archives Initiative* nos sistemas de informação em Ciência e Tecnologia, “[...] afetando substancialmente a maneira como bibliotecas e centros de documentação desempenham funções tradicionais, como seleção, aquisição, registro/tratamento técnico/indexação/classificação e disseminação.” Ainda a cooperação é colocada em um patamar diferenciado (MARCONDES; SAYAO, 2002, p. 44).

Por outro lado, Triska e Café (2001) relatam sobre o processo de revisão entre os pares, o qual, a partir da filosofia de trabalho dos arquivos abertos, torna o processo com mais “[...] transparência das críticas e sugestões feitas aos textos eletrônicos depositados no repositório.” Portanto, diante deste cenário, toda a comunidade passa a ter acesso aos procedimentos de avaliação do texto científico, juntamente com todas as revisões sofridas com base nos pareceres dos pares. “Assim, quebra-se o conceito de revisão sigilosa feita por um comitê científico, caracterizando mais uma vez o aspecto democrático desta iniciativa.” (TRISKA; CAFÉ, 2001, p. 93).

Tudo isto se soma a bons exemplos, como é o caso do Sistema Portal Brasileiro de Acesso Aberto à Informação Científica (OASISBR, 2014), que “[...] é um mecanismo de busca multidisciplinar que permite o acesso gratuito à produção científica de autores vinculados a universidades e institutos de pesquisa brasileiros.” O foco do sistema é a busca em fontes de informação de língua portuguesa. Esse sistema é resultado de esforços do IBICT, que une aqui a busca integrada de “[...] revistas científicas, repositórios institucionais, repositórios temáticos, bibliotecas digitais de teses e dissertações e outras fontes de informação de natureza científica e tecnológica ou academicamente orientada.” Tal projeto está em constante evolução e crescimento de acessos e conteúdos disponíveis (OASISBR, 2014).

Dentre os quesitos da OAI, fica clara a inserção do periódico científico eletrônico no modelo do OA, bem como as estratégias para dar acesso, da maneira mais aberta possível, às publicações científicas, além de proporcionar a interoperabilidade de dados para tornar o periódico científico eletrônico mais visível à comunidade científica. Portanto o periódico científico será objeto de discussão do próximo item deste trabalho que, posteriormente, passará a tratar a modalidade eletrônica.

Com todas essas transformações, as contribuições das novas TIC são visíveis e, com este suporte, o periódico científico eletrônico tem condições de ser visualizado nas diversas áreas do conhecimento. Ainda que este suporte traga os efeitos benéficos de acesso ao periódico via *web*, com uma ampla divulgação perante os índices dos buscadores, isso se torna algo imensurável diante da crescente utilização do Sistema Eletrônico de Editoração de Revistas (SEER) (SOUZA, 2010). Surge então, no SEER, a perspectiva de gerenciamento de periódico científico eletrônico junto de um “[...] modelo alternativo de publicação do conhecimento científico para ampliar o acesso, a preservação e o impacto das pesquisas e dos resultados daí provenientes [...]” (SOUZA, 2010, p. 1).

Souza (2010) e Mueller (2000b) relatam que o periódico científico eletrônico está entre as formas mais práticas, rápidas e econômicas de disseminação do conhecimento. Diante deste cenário, existe no Brasil a mobilização para um aumento da visibilidade dos artigos produzidos nacionalmente, “[...] dentro de uma política de acesso livre e destacamos aqui o Sistema Eletrônico de Editoração de Revistas (SEER) [...]”. Esse sistema customizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), a partir do programa *Open Journal Systems* (OJS), contribui para a inserção dos periódicos brasileiros no cenário de pesquisa internacional por meio do acesso aberto (SOUZA, 2010, p. 2). Mueller (2000b) complementa que estão surgindo várias propostas de comunicação científica, porém nem todas poderão ser implementadas. Duas merecem destaque especial, que são: os periódicos eletrônicos e as bases eletrônicas de *preprints*, por terem domínio sobre os meios tradicionais.

Deste modo, “o SEER/OJS é um software livre, de código aberto, que permite a criação de uma revista ou um portal de revistas. Fornece facilidades para criação de revistas eletrônicas disponibilizadas na *Web*.” A automatização do processo de editoração eletrônica de revistas é uma funcionalidade de grande

relevância. Entre os diversos serviços disponibilizados no SEER/OJS, considerados como “[...] facilidades opcionais que ofertam serviços específicos”, estão os *plugins*. “Esses plugins devem ser ativados para apoiar os processos de editoração da revista, integrando-se perfeitamente à ferramenta.” (BRITO; GUEDES; SHINTAKU, 2013, p. 25).

O SEER é um sistema recomendado pela CAPES, o “[...] processo editorial no SEER permite uma melhoria na avaliação da qualidade dos periódicos e uma maior rapidez no fluxo das informações”. Por outro lado, a aceitação do SEER foi grande e intensa pela comunidade brasileira de editores, diante da facilidade do uso das suas respectivas adaptações do processo editoração. Com o SEER, o cenário brasileiro de comunicação científica passou a ter os padrões editoriais internacionais de periódicos e ainda 100% eletrônicos. “O Sistema de Eletrônico de Editoração de Revistas (SEER) é resultado da prospecção tecnológica realizada pelo IBICT”. O SEER surgiu em 2003, o primeiro periódico brasileiro a implantar o SEER foi a revista *Ciência da Informação*. A partir de então, o SEER foi disponibilizado para os periódicos brasileiros (IBICT, 2015).

As primeiras revistas eram desenvolvidas “[...] a partir de aplicativos como as conferências eletrônicas (NEWS), listas de discussão (LISTSERV).” Deste modo, os grupos com as atividades regulares conseguiram concretizar as revistas eletrônicas. “Hoje, o programa HAL (*Hyper Article en Ligne*) oferece uma interface que permite aos autores depositar seus trabalhos.” (LE COADIC, 2004, p. 104).

A influência na crescente melhoria com o advento da tecnologia a favor da comunicação torna também progressiva a quantidade de informação para ser disseminada, influenciando a criação de novos procedimentos e instrumentos para melhorar a qualidade de publicação, de modo que a criação do processo editorial é lançada para suprir essa necessidade. O processo editorial de uma revista, segundo Gonçalves, Ramos e Castro (2006, p. 172), é composto pelas etapas: “[...] recebimento de trabalhos para publicação, pré-seleção dos trabalhos, seleção de revisores, encaminhamento e acompanhamento do processo da revisão por pares [...]”. Depois desses processos finalizados, é necessário dar continuidade às etapas com “[...] contato com o(s) autor(es) sobre comentários dos revisores, aprovação ou rejeição para publicação, revisão de texto, revisão gráfica e publicação [...]”. (GONÇALVES; RAMOS; CASTRO, 2006, p. 172).

Diante de uma gama de perspectivas e modalidades envolvidas dentro do

sistema de comunicação científica, é importante colocar-se perante as tendências centradas nos periódicos científicos, sendo que "[...] dentre as principais [...] perspectivas [...], destacam-se: a publicação totalmente eletrônica, explorando cada vez mais recursos hipertextuais e multimídia [...]" (GONÇALVES; RAMOS; CASTRO, 2006, p. 185-187). Em consonância a essas tendências estão "[...] o gerenciamento eletrônico do fluxo editorial, a valorização do artigo como unidade informacional, a interoperabilidade entre os sistemas e programas de revistas eletrônicas existentes e o acesso aberto à informação [...]" (GONÇALVES; RAMOS; CASTRO, 2006, p. 185-187). Salientando que a interoperabilidade entre os sistemas torna-se indispensável para a visibilidade das revistas científicas com o aumento de acessos, o que trará uma valorização ao fluxo da comunicação científica por parte dos produtores.

Esta situação de evolução tecnológica contribuiu para o chamado dilema do artigo científico, assim denominado pelo autor, por se tratar de uma imprecisão no momento de construir o estado da arte de um determinado assunto, devido à grande quantidade de publicações recuperadas. Contudo, esse dilema foi iniciado logo após a proposta de Vannevar Bush ser colocada em prática para contribuir com a construção do conceito de hipertexto que proporia Theodor Holm Nelson. O termo hipertexto foi apresentado na conferência anual da *Association of Computing Machinery* (ACM), por Ted Nelson em 1965, que posteriormente desenvolveu o sistema *Xanadu*, denominado um sistema literário que daria o acesso a muitas obras em formato digital, o qual poderia ser alterado e, a partir desta edição do texto, poderia sofrer as avaliações e novas implementações, de maneira ágil e fácil. Todo esse experimento serviu de base para a *Web* ou *World Wide Web* que seria proposta por Tim Berners-Lee (CORTES, 2006).

Rocha e Silva (2012, p. 1) argumentam sobre as duas categorias de periódicos científicos: com *peer review* e sem *peer review*. No Quadro 9, expõem-se os dados apresentados.

**Quadro 9** - Dados de acesso PCE

Instituição	Total de periódicos	Com <i>peer review</i>
CAPES	6.000	480 (8%)
Ulrich's	300.000	27.000 (9%)

Fonte: Elaborado pelo autor, fundamentado em Rocha e Silva (2012, p. 1).

Isto mostra que o cenário nacional está proporcional ao internacional. Um outro fator de implicação é a visibilidade. Até o século passado, os artigos eram totalmente invisíveis, muito pouco citados, e aqui é um marco divisório para a visibilidade, pois, a partir no novo milênio, com a *Scientific Eletronic Library Online* (SciELO) e PubMed, o cenário passa ser outro (ROCHA E SILVA, 2012, p. 1).

Gonçalves, Ramos e Castro (2006), Rocha e Silva (2012), Mueller (2000b) afirmam que a SciELO é uma biblioteca de revistas científicas com textos completos, uma base de dados bibliográficos. Rocha e Silva (2012) complementam, apenas como demonstrativo antes do marco divisório de visibilidade, que o fator de impacto não chegava a 0,500; após 2004, duas revistas nacionais atingiram o fator de impacto acima de 1,000. Além da quantidade de downloads que apontam, em 1999, 100 mil, e em 2010, 100 milhões. Já Gonçalves, Ramos e Castro (2006) adicionam que a SciELO é também um sistema com indicadores bibliométricos.

A SciELO foi “[...] a primeira base de dados de textos completos de revistas científicas eletrônicas [...]”, teve suas atividades iniciada em junho de 1998, porém “[...]concebida em 1995, por meio de uma parceria entre a BIREME/OPAS/OMS - Centro Latino-americano e o Caribe de Informação em Ciências da Saúde e a FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo [...]”, e conta com o apoio do CNPq (GONÇALVES; RAMOS; CASTRO, 2006, p. 170).

Diante da grande quantidade de títulos de revistas científicas, surge a necessidade de criar um índice para que assim sejam possíveis de recuperação e acesso, com mais facilidade, os artigos científicos de cada revista científica. “Em 1879, foi criado o *Index Medicus*, seguido em 1907 pelo *Chemical Abstracts*, e em 1921, pelo *Biological Abstracts*”. Esse mecanismo criado serviu de alicerce para as bases de dados bibliográficos (GONÇALVES; RAMOS; CASTRO, 2006, p. 168). Portanto a indexação de periódicos com suas diversidades geográficas e/ou temáticas é denominada índice bibliográfico, existe a atuação no cenário mundial, os

índices internacionais, e em uma região ou país específico, os índices nacionais (PACKER; MENECHINI, 2006, p. 246).

Lara (2006, p. 399) complementa a definição de índice bibliográfico como: “Coleção de registros de elementos de dados bibliográficos que identificam os conteúdos de um conjunto de publicações científicas, selecionadas de acordo com critérios pré-estabelecidos [...]”, esses critérios podem ser a região geográfica, área temática, data, tipo, qualidade da publicação. É permitida a exploração de todos os registros recuperados ou uma porção, com a finalidade de recuperar o conteúdo do registro ou listar por diversos elementos de dados para cálculos bibliométricos (LARA, 2006, p. 399).

No século XX, cresce o número de submissões de trabalhos e, portanto, torna-se necessária uma seleção de trabalhos que envolve a avaliação por pares da metodologia e resultados/discussões. Em geral, são avaliados por “[...] dois ou mais especialistas da área, membros da própria comunidade científica [...]”. Também se pode considerar como processo educacional para todos os avaliadores envolvidos neste processo. Na avaliação, alguns aspectos formais são comuns aos periódicos, como: periodicidade e pontualidade, duração, normalização, trabalho editorial, difusão e indexação, endogenia, indicadores bibliométricos. Assim como os aspectos de conteúdo: caráter científico, revisão por pares e corpo editorial (GONÇALVES; RAMOS; CASTRO, 2006, p. 172). Lara (2006, p. 411) define revisão por pares como o “[...] processo de validação do mérito e do método científico dos trabalhos enviados para publicação em periódicos, executado por pares da comunidade científica.”

Ainda que sejam elencadas inúmeras vantagens para o periódico científico eletrônico, tem-se uma não aceitação pela comunidade científica, se comparado ao periódico impresso de maneira tradicional. É visivelmente essa a situação por parte dos autores ao planejar a submissão de sua pesquisa; já para o leitor, esta preocupação é menor.

Por um lado, a ciência tenta resolver um problema que é a comunicação das publicações dos periódicos por meios tradicionais ser lenta, cara, com recursos ultrapassados, passível de muitas críticas; por outro, ao contrário dos problemas já citados, o periódico científico eletrônico resolve a maior parte dessas situações. Então, por que não ser aceito? (MUELLER, 2000b).

A difusão das pesquisas por meio dos artigos científicos teve uma

significativa melhoria, que veio para tornar prática, rápida e facilitada a discussão entre os pares. Porém nem toda melhoria pode acabar com todos os problemas; podem deixar de existir alguns, mas irão surgir outros. Como se observa, “[...] um deles é que o acesso às publicações não é de todo democratizado ou generalizado, uma vez que os custos de distribuição crescem conforme a distância geográfica [...].” E ainda a dependência de órgãos do governo ou da própria instituição a que as revistas estão vinculadas. Para assegurar sua sobrevivência financeira, as revistas encontram sérias implicações nos quesitos “[...] periodicidade, longevidade, número de exemplares e a distribuição [...].” (CORTES, 2006, p. 47-48).

Outro fator de relevância que faz com que a comunidade científica ainda se distancie da modalidade de comunicação científica por meio do periódico científico eletrônico é o fato de não inspirar confiança na substituição da versão impressa, que tem a função de registro primário, assim como dificuldades de acesso. Não são decorrentes do fator tecnologia esses dois quesitos problemáticos, confiança e acesso, mas sim dos demais fatores como o “[...] hábito da comunidade científica, interesses de editoras comerciais e questões tais como precariedade de algumas bibliotecas, regiões ou países, que não dispõem de recursos materiais e humanos adequados.” (MUELLER, 2000b, p. 84-85).

O fator comercial das editoras é um dos mais complexos problemas existentes na propagação dos periódicos científicos eletrônicos de acesso livre, uma vez que elas necessitam do lucro no mercado de conhecimento. Os assinantes principais como grandes bibliotecas efetuavam pagamentos de valores altíssimos pela assinatura do periódico tradicional, cabe notar que essas editoras já vêm acompanhando e implantando essa transição – do meio tradicional para o eletrônico. O acesso, nesse caso, não será livre, exigindo o pagamento pela assinatura (MUELLER, 2000b, p. 84-85).

Por outro lado, tem-se a qualificação da produção intelectual realizada pela CAPES, denominada Qualis, que é um “[...] conjunto de procedimentos utilizados pela CAPES para estratificação da qualidade da produção intelectual dos programas de pós-graduação.” Esse procedimento foi criado de acordo com as necessidades de avaliação dentro do sistema federal de ensino. A coleta é realizada por meio de informações inseridas em sistema próprio de coleta e, em seguida, divulgada por meio de uma lista nos meios de comunicação utilizados pelos programas de pós-graduação e suas respectivas informações. O estrato indicativo de qualidade é

definido por A1, A2, B1, B2, B3, B4, B5, C, sendo que A1 é o mais elevado índice de qualidade e C, o mais baixo índice. Essa estratificação “[...] é realizada de forma indireta. Dessa forma, o Qualis afere a qualidade dos artigos e de outros tipos de produção, a partir da análise da qualidade dos veículos de divulgação, ou seja, periódicos científicos.” A consulta a essa fonte pode ser realizada por meio do aplicativo WebQualis (CAPES, 2014a, p. 1).

O Portal de Periódicos da Capes foi originado do Programa para Bibliotecas de Instituições de Ensino Superior, a partir de onde se cria o Programa de Apoio à Aquisição de Periódicos. “O Programa está na origem do atual serviço de periódicos eletrônicos oferecido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) à comunidade acadêmica brasileira.” (CAPES, 2014b, p.1).

Na Figura 4, apresenta-se a quantidade de periódicos com texto completo entre 2004 e 2013, e é perceptível um considerável crescimento.

**Figura 4 - Evolução dos periódicos com texto completo (2004-2013)**



Fonte: Portal de periódicos CAPES (2014?).

A partir da quantidade de periódicos com texto completo apresentada na Figura 4, infere-se o total de periódicos cadastrados no Portal de Periódicos da Capes, da área de Ciência da Informação, que é 99. Diante desses dados significativos demonstrados, é possível observar a evolução do número de acessos no Quadro 10, que demonstra o exponencial crescimento, induzido a partir de disponibilidade de conteúdo; portanto é possível fazer a seguinte inferência: para

8.156 periódicos, há 13.099.471 acessos, e para 37.073 periódicos, 44.420.626, logo a proporção de crescimento da quantidade de periódicos é 354,55%, e o crescimento de acessos é de 239,10%.

**Quadro 10** - Evolução do número de acessos ao Portal de Periódicos (2004-2013)

Evolução do Número de Acessos ao Portal de Periódicos (2004-2013)		
Ano	Texto completo	Bases referenciais
2004	13.099.471	13.763.661
2005	13.754.226	18.975.465
2006	15.000.000	32.000.000
2007	18.058.420	38.538.452
2008	21.222.922	39.591.556
2009	23.386.833	41.642.827
2010	25.367.166	42.025.639
2011	34.231.457	42.107.835
2012	39.470.709	45.200.535
2013	44.420.626	56.524.022

Fonte: Portal de Periódicos CAPES (CAPES, 2014a).

Os periódicos tendem a ter prestígio e reconhecimento a partir de sua visibilidade, a qual “[...] é uma característica que é devida ao entendimento de que ciência só se realiza se publicada [...].” (PACKER; MENEHINI, 2006, p. 238-245). Essa visibilidade ocorre em duas dimensões: “[...] ser referência (de qualidade e credibilidade) no âmbito de uma disciplina ou área temática a ser indexada [...]”. Nesta mesma dimensão, a condição de ser indexado é a capacidade de ser acessado para servir de respostas aos chamados realizados tradicionalmente, caracterizados por “[...] pesquisas bibliográficas realizadas em índices e bases de dados bibliográficas e mais recentemente, também nos buscadores da internet [...].” (PACKER; MENEHINI, 2006, p. 238-245).

Este contexto é determinado pela agregação de metadados que contêm as referências básicas do conteúdo dos artigos como “[...] nome do autor, título,

resumo, palavras-chave que descrevem o conteúdo e, mais recentemente, o *link* para o texto completo [...].” Esse é um aspecto importante para que seja possível a visibilidade do artigo na internet: quanto mais bem estruturada for a designação dos metadados, melhor será a visibilidade (PACKER; MENEHINI, 2006, p. 238-245). Portanto, também “como meio de garantir sua visibilidade, os periódicos devem ser indexados em índices nacionais e internacionais.” (LARA, 2006, p. 405).

Em suma, este tópico buscou situar o periódico científico eletrônico, que passou pelo caminho da comunicação científica e seus preceitos e evoluções históricas, até a estrutura atual de artigos científicos em periódicos científicos eletrônicos. Passa-se à discussão do próximo tópico que é a metodologia deste estudo.

## 4 METODOLOGIA

Os procedimentos metodológicos adotados para a execução deste estudo serão, a seguir, descritos e explicitados, bem como justificada a razão de sua escolha em função dos objetivos específicos e da proposta de conceber um sistema para implantação do controle de vocabulário de palavras-chave atribuídas para subsidiar o processo de indexação e a recuperação dos conteúdos em um periódico científico eletrônico.

A finalidade de tornar visível a metodologia utilizada é fazer emergir a organização do todo e das partes, assim como associar as abordagens, as técnicas de coleta de dados, as análises e a prospecção dos resultados às inferências permitidas pela análise realizada.

### 4.1 CARACTERIZAÇÃO DA PESQUISA

Caracteriza-se esta como uma pesquisa exploratória e descritiva. A opção por exploratória se deu por essa modalidade ser compatível com o principal objetivo deste tipo de pesquisa, que é “[...] desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores.” (GIL, 2008, p. 27). A classificação como descritiva é definida para este estudo, pois ela é a “[...] descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis.” (GIL, 2008, p. 28).

A abordagem mista qualitativa e quantitativa foi decidida por ter sido entendida como requisito da estratégia de realizar a análise dos dados com prioridade na perspectiva qualitativa e posteriormente na quantitativa. Cuidou-se para que a análise de dados no processo de métodos mistos estivesse firmemente relacionada com o tipo de resultado buscado na pesquisa escolhida e conforme os procedimentos descritos, na etapa de projeto. Portanto realiza-se a análise “[...] tanto dentro da técnica quantitativa (análise descritiva e numérica inferencial) como da técnica qualitativa (descrição e texto temático ou análise de imagem) e, muitas vezes, entre as duas técnicas.” (CRESWELL, 2007, p. 222). Com a amplitude oferecida pela abordagem qualitativa e quantitativa, que se delinearão ao longo da

apresentação e análise de dados, constrói-se a convicção de sua efetividade para a abrangência das finalidades desta pesquisa.

O *corpus* é construído com dados levantados junto ao suporte periódicos científicos eletrônicos, com foco na representação temática da informação, no âmbito do controle de vocabulário, para indexação de conteúdos por meio da atribuição de palavras-chave. Adota-se, como procedimento, localizar e estruturar as palavras-chave já existentes nos periódicos científicos, considerando como se constituem, a ligação que promovem ao estarem atribuídas do modo como foram encontradas e que efeito geram para a representação dos conteúdos temáticos.

Dada a finalidade desta pesquisa, em que se pretende que o conhecimento gerado esteja associado à criação de uma proposta metodológica, ocorre uma demanda por validação, ou seja, o teste da proposta desenvolvida. A metodologia de estudo de caso foi então adotada como recurso estratégico para testar a aplicabilidade da proposta metodológica em questão. A opção por esta forma de organização leva em conta o que se quer responder: “[...] as questões ‘como’ e ‘por que’ são mais explanatórias e provavelmente levam ao uso dos estudos de caso, pesquisas históricas e experimentos como métodos de pesquisa preferidos.” (YIN, 2010, p. 30). Há uma conexão dessas duas modalidades de questão com os vínculos operacionais que necessitam ser planejados.

A questão-problema que orienta esta pesquisa indaga sobre como realizar o controle do vocabulário das palavras-chave atribuídas aos artigos científicos disponíveis em periódicos científicos eletrônicos. A análise para resolvê-la beneficia-se dos parâmetros de Yin (2010, p. 41), que afirma não ser essa proposição foco de somente “[...] um tipo de pesquisa qualitativa, usando uma mistura de evidência quantitativa e qualitativa.” Assim, os estudos de caso não têm a necessidade de “[...] incluir a evidência observacional direta e detalhada marcada pelas outras formas de pesquisa qualitativa.” (YIN, 2010, p. 41).

A técnica de coleta de dados constitui a utilização de instrumentos para esta etapa do trabalho, e surge, aqui, uma preocupação no que tange aos registros por meio de técnicas que possam trazer confiabilidade à mensuração dos dados (BARROS; LEHFELD, 1990). Portanto a técnica de coleta de dados foi aplicada com o auxílio de mecanismos automatizados, por meio de coleta de metadados utilizando o protocolo OAI-PMH do repositório digital do periódico científico eletrônico. Tais procedimentos foram implementados na ferramenta resultante da proposta

metodológica deste estudo. Serão elas descritas, etapa por etapa, conforme as ações a serem desenvolvidas a partir dos objetivos específicos da pesquisa. No Quadro 11, tais ações estão apresentadas, de forma sistematizada, para uma visão de conjunto.

#### Quadro 11 - Ações decorrentes dos objetivos do estudo

Objetivo geral: implantar o controle de vocabulário, tendo como base as palavras-chave atribuídas no periódico científico eletrônico Informação & Informação, para aprimorar a recuperação dos conteúdos temáticos.	
Objetivos específicos	Ações
Desenvolver a ferramenta VCPC <i>tools</i> – Etapa 1.	<ul style="list-style-type: none"> <li>• Identificar o endereço do repositório digital do periódico (1.1);</li> <li>• Desenvolver o projeto da ferramenta VCPC <i>tools</i> (1.2),</li> <li>• Desenvolver o projeto físico do banco de dados (1.3);</li> <li>• Criar a base de dados no sistema gerenciador de banco de dados (1.4).</li> </ul>
Coletar e armazenar os vocabulários controlados – Etapa 2.	<ul style="list-style-type: none"> <li>• Identificar os vocabulários controlados da área especializada (2.1);</li> <li>• Estruturar os dados dos termos dos vocabulários controlados (2.2);</li> <li>• Descrever, em linguagem SQL, os termos dos vocabulários controlados (2.3);</li> <li>• Importar os dados dos termos dos vocabulários controlados (2.4).</li> </ul>
Compatibilizar as palavras-chave com os termos dos vocabulários controlados – Etapa 3.	<ul style="list-style-type: none"> <li>• Coletar e armazenar os metadados (3.1);</li> <li>• Verificar as palavras-chave que são idênticas com os termos dos vocabulários controlados (3.2);</li> <li>• Verificar as palavras-chave que remetem à igualdade dos termos do vocabulário controlado (3.3);</li> <li>• Vincular as palavras-chave e termos dos vocabulários controlados (3.4).</li> </ul>
Tratar manualmente as palavras-chave não compatibilizadas automaticamente – Etapa 4.	<ul style="list-style-type: none"> <li>• Verificar as palavras-chave que tenham termos correspondentes nos vocabulários controlados (4.1);</li> <li>• Inserir a palavra-chave como termo no vocabulário controlado da revista (4.2).</li> </ul>

Fonte: Elaborado pelo autor.

## 4.2 ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC *TOOLS*

### 4.2.1 Identificar o Endereço do Repositório Digital do Periódico (1.1)

A ação de identificação e localização do endereço do repositório digital da revista, uma vez que o endereço deste repositório não se encontra visível nas páginas da revista, é pré-requisito para estruturar as demais fases do processo de coleta de dados. Realizada por meio de várias pesquisas nos buscadores e de testes utilizando o validador de protocolos OAI-PMH para verificar a consistência dos

endereços coletados, a coleta de dados aferiu, dentre esses endereços, os que eram possíveis para localização do repositório. Cabe ressaltar que o endereço não está claramente disponível, o que resulta em várias tentativas para sua identificação. Conseguida essa identificação e validação do endereço do repositório, foram efetuadas as diversas requisições para alcançar a resposta por meio da estrutura de um arquivo XML. A partir desta ação foi possível obter informações para o desenvolvimento da ferramenta *VCPC tools*, a seguir descrita.

#### 4.2.2 Desenvolver o Projeto da Ferramenta *VCPC tools* (1.2)

O projeto da ferramenta *VCPC tools* foi desenvolvido na linguagem de programação *PHP*, por se tratar de uma linguagem livre para *web* e permitir uma perfeita integração com o sistema gerenciador de banco de dados *MySQL*. A ferramenta executa os seguintes procedimentos, aqui tratados de requisitos funcionais, enumerados com a sigla RF:

- RF01 Coleta de metadados;
- RF02 Análise de Metadados e Vocabulário controlado;
- RF03 Tratamento Palavras-chave;
- RF04 Resultados.

Os requisitos funcionais serão descritos nos próximos itens.

##### - RF01 Coleta de metadados

Neste requisito funcional, o principal objetivo é coletar os metadados do periódico e armazená-los na base de dados. Cunha e Cavalcanti (2008, p. 246) definem metadados como “[...] informação que descreve a estrutura dos dados e sua relação com outros, p.ex.: uma etiqueta no registro de uma base de dados, que indica o campo que contém o nome do autor.”.

Sayão (2010) relaciona os metadados com a catalogação, e complementa que existem outras razões importantes para a existência deste conceito, como “[...] facilitar a descoberta de informações relevantes, seja no ambiente da biblioteca, seja no ambiente *web*.” O *Dublin Core Metadata Element Set* é “[...] uma das mais importantes iniciativas na área de metadados, cujo objetivo essencial é apoiar a

descoberta de recursos no extenso e fragmentado universo web [...]", e é um excelente exemplo de aplicação, "[...] que apesar da sua riqueza informacional não foi pensado especificamente para a recuperação de informação." (SAYÃO, 2010, p. 2). Breitman (2005, p. 16) complementa que "o tradicional cartão de biblioteca é uma forma de metadado, da mesma forma que qualquer item de um catálogo é representado por um código de produto."

Não existe um consenso entre os autores, porém a maioria divide em três categorias os metadados: "[...] metadados descritivos, metadados estruturais e metadados administrativos." Essa divisão é importante para compreender mais facilmente os tipos de informação a que se referem, embora esses conceitos não sejam delineados com precisão (SAYÃO, 2010, p. 5).

Metadados descritivos compreendem a descrição física e temática e são "[...] a face mais conhecida dos metadados, [...] descrevem um recurso com o propósito de descoberta e identificação; [...] tais como título, autor, resumo, palavras-chave e identificador [...]" Os metadados estruturais "[...] são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados.", pode-se citar, como exemplo, a vinculação das páginas de um livro que são digitalizadas em separado, sendo que este vínculo se dá de forma ordenada para estruturar um capítulo. Por outro lado, os metadados administrativos "[...] fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais.", neles são incluídos os dados de criação do recurso, assim como o motivo de sua elaboração; "Nessa categoria, estão metadados técnicos que explicam as especificidades e dependências técnicas do recurso; incluem-se também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso." (SAYÃO, 2010, p. 5).

Dias e Naves (2013) argumentam que os metadados são registros bibliográficos e catalográficos. O primeiro se relaciona com a descrição física, e o segundo, com descrição temática. Para efeito deste estudo, serão considerados os metadados descritivos, com ênfase na descrição temática em que se observam os termos utilizados para identificação do assunto do documento, em específico as palavras-chave dos PCE's.

Breitman (2005, p. 17) sintetiza, no quadro 6, os tipos e funções dos metadados, e é importante salientar que "[...] a utilização de metadados não é novidade nem foi introduzida por pesquisadores da Web Semântica, pois trata-se de

um conceito que vem sendo aplicado há centenas de anos [...]” pelos profissionais da ciência da informação. Por outro lado, o emprego de metadados “[...] faz com que a adoção de vocabulários controlados, deixada de lado durante décadas por catalogadores, seja fundamental para garantir a comunicação entre diversas aplicações.”

Campos (2007, p. 28) complementa que a necessidade de utilizar metadados se tornou cada vez mais intensa e complexa para aumentar a reflexividade e a transparência. “Cada vez mais, metadados, além de referirem-se aos objetos informacionais primários, voltam-se para a caracterização, descrição e representação dos próprios sistemas [...]” (CAMPOS, 2007, p. 28). Esses sistemas são as “[...] regras e processos que representam, organizam e tratam esses objetos.”. Campos (2007, p. 28) esclarece que “a reflexividade é um ato de auto-orientação e descrição que acarreta melhor visibilidade dos objetos informacionais e mais versatilidade em seu uso.” No Quadro 12, apresentam-se os tipos de metadados.

**Quadro 12** - Tipos e funções de metadados baseado em Gilliland-Swetland

Tipo	Definição	Exemplos
Administrativo	Metadados utilizados na gerência e na administração de recursos de informação	Aquisição de informação Registro de direitos e reprodução Documentação dos requisitos legais de acesso Informação de localização Critérios de seleção para a digitalização Controle de versão
Descritivo	Metadados utilizados para descrever e identificar recursos de informação	Registro de catalogação Auxílio para a procura de informação Indexes especializados Utilização de hiperlinks entre recursos Anotações
Preservação	Metadados relacionados ao gerenciamento dos recursos de informação	Documentação sobre a condição física dos recursos Documentação sobre as ações tomadas de modo a preservar as versões físicas e digitais dos recursos, e.g., atualização e migração
Técnica	Metadados relacionados a funcionalidades do sistema e como seus metadados se comportam	Documentação sobre hardware e software Informação relativa à digitação, e.g., formatos, compressão, rotinas de escalonamento Registro do tempo de resposta do sistema Autenticação de dados, e.g., senhas e criptografia
Utilização	Metadados relacionados ao nível e ao tipo de utilização dos recursos	Registro de exibição Registro do uso e dos usuários dos recursos Reutilização do conteúdo e informação relativa a multiversiamento

Fonte: Breitman (2005, p. 17).

Os metadados podem ser “[...] agrupados em estruturas abstratas conhecidas como esquemas ou formatos de metadados, que são conjuntos de elementos criados com fins específicos [...]”. Esquemas são desenvolvidos e, na maioria, diferentes, com suas especificidades para determinadas aplicações, isto é, incluem limites para tratar somente o contexto a que se referem. Por exemplo: i) *MODS (Metadata Object Description Schema)* é um esquema bibliográfico proveniente do MARC 21; ii) *EAD (Encoded Archival Description)* designado para a área de Arquivologia; iii) *LOM (Learning Object Metadata)* esquema desenvolvido para gerenciar, avaliar e localizar objetos de aprendizagem; iv) *MPEG Multimedia Metadata* utilizado na representação de objetos multimidiáticos (SAYÃO, 2010, p. 6).

A iniciativa do *Dublin Core* “[...] surgiu durante um workshop em Dublin, Onio, a representação proposta ficou conhecida como Dublin Core. Desde aquele workshop tem havido um crescente interesse em elementos [...] para descrever [...]”, de forma fácil e simplificada, os recursos (BREITMAN, 2005, p. 18). Estes esquemas também denominados de padrão, por Campos (2007, p. 23), dão “[...] condição necessária para o entendimento das representações (estruturas, convenções, sintaxe e significado consensual dos metadados).” Este padrão é o resultado de uma convenção, de maneira cooperativa, que fará cumprir os compromissos ora assumidos para padronizar a representação do objeto. A partir daí é possível obter a “[...] comunicação efetiva, reutilização e melhor compartilhamento de componentes, diminuição de custos e menos retrabalho.” Neste contexto, relaciona-se com a interoperabilidade, que é “[...] a habilidade para transferir e utilizar informações entre sistemas com eficiência e uniformidade, exigindo padronização e flexibilidade em certo nível.” (CAMPOS, 2007, p. 23).

O *Dublin Core*, esquema de metadados, é um tanto especial por se tratar de um esquema que não tem nenhuma aplicação específica, sendo assim pode ser aplicado em qualquer domínio ou objeto informacional. Outra característica importante é a sua composição, possui poucos elementos essenciais (o core), possíveis de serem modelados em vários formatos implementandos, dessa forma, uma língua comum aos metadados. Tem-se, assim, um dos elementos primordiais para estabelecer a interoperabilidade. Por outro lado, ressalta-se o fato de o *Dublin Core* ser didático a ponto de permitir que o próprio autor faça a descrição de sua publicação. “Não obstante, o esquema possui uma estrutura simples e flexível e pode ser aplicado a recursos complexos; além do mais, pode ser representado através de sintaxes variadas [...]”. Alguns exemplos mostram como são possíveis a codificação nas linguagens de marcação HTML ou XML e seguir a estrutura proposta pelo *Resource Description Framework – RDF*, padronizando, desta forma, “[...] o intercâmbio e o reuso.” (SAYÃO, 2010, p. 6).

Foi criado o Dublin Core Metadata Initiative (DCMI – [www.dublincore.org](http://www.dublincore.org)) que se dedica “[...] à disseminação e à adoção de padrões de metadados e ao desenvolvimento de vocabulários especializados para a descrição de recursos, de modo a viabilizar sistemas mais inteligentes de recuperação da informação.” (BREITMAN, 2005, p. 18).

No Quadro 13, são apresentados os elementos que compõem o padrão

Dublin Core.

**Quadro 13 - Elementos Dublin Core**

Elemento (PT)	Elemento (EN)	Descrição
Assunto	<i>Subject</i>	O tópico abordado pelo trabalho
Título	<i>Title</i>	Nome do objeto
Criador	<i>Creator</i>	Pessoas responsáveis pelo conteúdo intelectual do objeto
Descrição	<i>Description</i>	Descrição do conteúdo do objeto
Editor	<i>Publisher</i>	Agente ou agência responsável por disponibilizar o objeto
Outro agente	<i>Contributor</i>	Pessoas que fizeram contribuições significativas para o objeto
Data	<i>Date</i>	Data da publicação
Tipo de Objeto	<i>Type</i>	Gênero do objeto, se ficção, novela, poema, ou dicionário
Formato	<i>Format</i>	Manifestação física do objeto. Exemplos são arquivos executáveis, do tipo texto ou PDF
Identificador	<i>Identifier</i>	Cadeia ou número utilizado para identificar unicamente aquele objeto
Relacionamento	<i>Relation</i>	Relacionamento com outros objetos
Fonte	<i>Source</i>	Outros objetos, eletrônicos ou físicos, dos quais este foi derivado (caso seja aplicável)
Linguagem	<i>Language</i>	Linguagem do conteúdo intelectual
Cobertura	<i>Coverage</i>	Localizações espaciais e durações temporais características do objeto
Direitos	<i>Rights</i>	Informação sobre os direitos acerca do objeto

Fonte: Breitman (2005, p. 18).

Desse modo, o RF01 – Coleta de metadados - executará uma requisição ao protocolo OAI-PMH: `?verb=ListIdentifiers&metadataPrefix=oai_dc`, o qual retorna os identificadores de todos os documentos deste repositório. Cabe frisar que, conforme a quantidade de registros nos arquivos XML de retorno, passa a constar a *tag resumptionToken*. No caso, a cada 500 registros, é gerado um novo arquivo XML, o qual é legível somente por meio da citação do conteúdo da *tag resumptionToken* na requisição. Dessa forma, a próxima requisição para obter os identificadores dos documentos, a partir do 501º, se dará pela seguinte requisição: `?verb=ListIdentifiers&resumptionToken=conteúdo_da_tag`, em que consta “conteúdo\_da\_tag” e se refere ao conteúdo da *tag resumptionToken* presente no arquivo XML anterior. E assim por diante, prosseguindo até a *tag resumptionToken* não apresentar mais resultado algum, o que significa o fim dos registros. O RF01 faz o registro de todos esses identificadores em uma variável denominada de *array* (vetor), exceto os registros de identificadores com o atributo *status="deleted"* da *tag*

*header*. A cada registro de identificador, é realizada uma nova requisição ao protocolo OAI-PMH, abaixo descrita.

A requisição seguinte ao protocolo OAI-PMH: *?verb=GetRecord&identifier=identificador&metadataPrefix=oai\_dc*, é realizada para todos os identificadores dos documentos, “identificador” significa o identificador de cada documento acima descrito. Esta requisição retorna os registros dos metadados de cada documento, seguindo o padrão de metadados *Dublin Core* (DC). De posse desses registros, é chegada a fase de inserir, no banco de dados, os descritivos de cada metadado, ou seja, o conteúdo de cada *tag* do arquivo XML, em resposta à requisição enviada ao protocolo OAI-PMH.

A ferramenta recupera do arquivo XML os seguintes conteúdos: doi contido na *tag <dc:identifier>*; uri contido na *tag <dc:identifier>*; títulos (eng, esp e port - caso exista) contidos na *tag <dc:title xml:lang="pt-BR">* e o identificador contido na *tag <identifier>*. Em seguida, executa uma busca, na base de dados, para verificar a existência do artigo/documento, utilizando, como parâmetros, os dados extraídos do arquivo XML. Em caso de localização do registro, será exibida a mensagem de sua existência e o respectivo identificador. Caso não seja encontrado, passa-se pelas fases seguintes, isto é, a inclusão do artigo/documento na base de dados.

Primeiramente, é extraído o idioma a partir da *tag <dc:language>*, após, verifica-se a existência do idioma e, caso não exista na base de dados, deverá ser inserido. Depois, recuperam-se as seguintes *tags*: área de conhecimento *<dc:subject xml:lang="pt-BR">*; editora *<dc:publisher xml:lang="pt-BR">*; revista *<dc:source xml:lang="pt-BR">*. A *tag* “revista” contém outras informações que precisam ser separadas e inseridas em seu respectivo campo na base de dados, são elas: nome da revista, volume, número, ano e paginação. Todos os processos de inserção passam, antes, por verificação da existência de seu registro que é, portanto, executado para área do conhecimento, editora, edição e revista.

Em seguida, inicia-se a inserção dos artigos/documentos, sendo que alguns dados já foram extraídos nos processos anteriores como, por exemplo, doi, identifier, uri, título, paginação. O resumo será obtido da *tag <dc:description xml:lang="pt-BR">*. Em caso de existência em outras línguas, o resumo é inserido no campo correspondente à língua. O formato é extraído da *tag <dc:format>*, já o tipo de documento é obtido da *tag <setSpec>*.

A inserção dos autores na base de dados é realizada a partir da extração do conteúdo da *tag* `<dc:creator>`, onde constam o autor e sua afiliação; é executada a separação para serem inseridos ambos em seus respectivos campos e tabelas específicas. Da mesma forma, verifica-se a existência de cada registro inserido na base, neste caso, os autores e a suas afiliações.

Posteriormente, trata-se das palavras-chave, que são extraídas da *tag* `<dc:subject xml:lang="pt-BR">`. Geralmente, elas vêm separadas por um ponto e vírgula, porém alguns registros encontram-se separados por vírgula, outros com ponto final. Realiza-se um tratamento para a separação de cada palavra-chave, verifica-se a sua existência e, em caso de não existir na base de dados, ocorre a sua inserção.

As palavras-chave serão coletadas nas seguintes seções da revista: artigos, pontos de vista, relatos de experiência, comunicações, resenhas de livros/mídia e teses e dissertações, excluindo-se editoriais e expedientes.

Observa-se que a sequência de inserção dos dados executada tem o intuito de não comprometer a integridade do banco de dados, ou seja, registros “filhos” sem registros “pais”. Portanto, inicia-se o processo de inserção nas tabelas que não têm nenhuma dependência, por conseguinte, com as demais que possam ter algum vínculo.

#### - RF02 Análise de Metadados e Vocabulário controlado

O objetivo deste requisito funcional é verificar as palavras-chave que remetem à igualdade dos termos com os vocabulários controlados. Aqui são aplicados os mecanismos para a identificação de sinonímias e quase sinonímias por meio da verificação das palavras-chave que têm termos idênticos ou aproximados (variações como singular, plural, termo composto, entre outras mínimas alterações que podem dar o mesmo significado). Neste processo, o foco é a sintaxe das palavras-chave comparadas aos vocabulários controlados. Para a compatibilização, utilizam-se os seguintes mecanismos de comparação:

- 1) idênticas - comparação de cada palavra-chave com a busca do respectivo termo do vocabulário controlado, caracterizando, aqui, a compatibilização de palavras-chave exatamente idênticas com um termo do vocabulário controlado. Nos

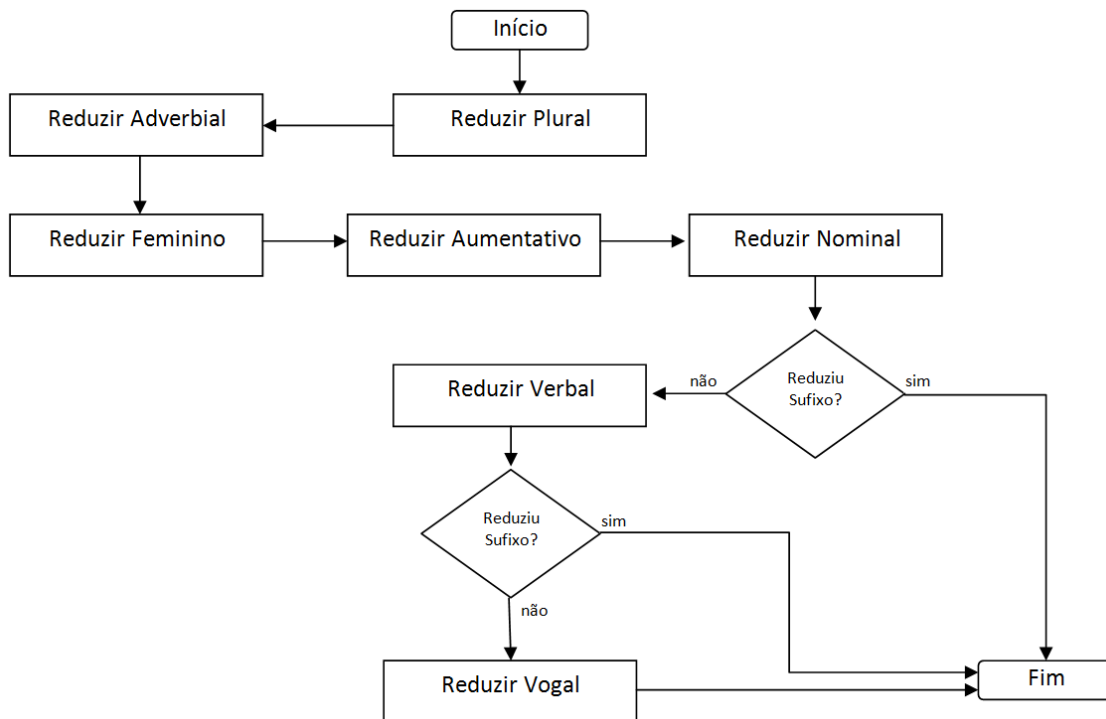
casos de compatibilidade, será estabelecido um vínculo entre a palavra-chave e o termo. O processo é contabilizado nos resultados de compatibilização.

2) idênticas retirando as palavras vazias - exclusão das palavras-chave, palavras vazias, e o resultado inserido em um novo campo do banco de dados. O mesmo procedimento é realizado com os termos do vocabulário controlado. Alguns exemplos de palavras vazias: *a, e, o, u, de, da, ainda*. Este procedimento é fundamentado por fazer parte dos processos do primeiro módulo da proposta de automatização da indexação de Gil Leiva (1999). A seguir, realizam-se os testes comparativos, desconsiderando os caracteres maiúsculos, minúsculos e acentuação. Em caso de existir o termo compatível no vocabulário controlado, estabelece-se o seu vínculo com a palavra-chave.

3) idênticas retirando os plurais - exclusão dos plurais das palavras-chave. Nesta fase, utilizam-se os resultados da exclusão das palavras vazias executadas na segunda compatibilização. São tratados os plurais, utilizando o conjunto de regras e exceções de Orengo; Buriol e Coelho (2007) que propõem melhorias para o algoritmo de *stemming* em língua portuguesa RSPL, proposto por Orengo e Huyck (2001) - Removedor de Sufixos. Retira-se o plural das palavras-chave e gravam-se os resultados em um novo campo no banco de dados, para preservar o registro original. O mesmo procedimento é aplicado aos termos do vocabulário controlado. Posteriormente, verifica-se a compatibilidade; caso exista, é estabelecido o vínculo, e este processo é recursivo a todas as demais compatibilizações.

4) idênticas retirando os sufixos - redução de sufixos empregando regras e exceções de Orengo, Buriol e Coelho (2007), utilizadas na compatibilização anterior. Inicia-se com a redução do plural, o qual foi executado na terceira compatibilização. Em seguida, executam-se as seguintes reduções: adverbial, do feminino, do aumentativo e nominal. Todos esses procedimentos são executados independentemente de ter havido a redução ou não. Caso não ocorra a redução nominal, executam-se as próximas reduções que são a verbal e de vogais. Contudo, a redução de vogais somente será realizada se não houver redução verbal. Na Figura 5, apresenta-se o diagrama de blocos que representa, graficamente, o processamento das compatibilizações 2, 3 e 4.

**Figura 5** - Diagrama de blocos processo de análise - redução de sufixo



Fonte: Adaptado de Orengo, Buriol e Coelho (2007).

5) idênticas por meio do índice contido - execução a partir do índice contido que será calculado e armazenado no banco de dados entre cada palavra-chave e cada termo do vocabulário controlado. A partir do propósito de Gil Leiva (1999) de medir o índice de consistência dos termos em sua proposta de automatização da indexação, surgiu a iniciativa de desenvolver o índice contido, porém com o propósito de compatibilizar, caso alcance os parâmetros determinantes de equivalência de termos. A fórmula do cálculo do índice contido é composta por:

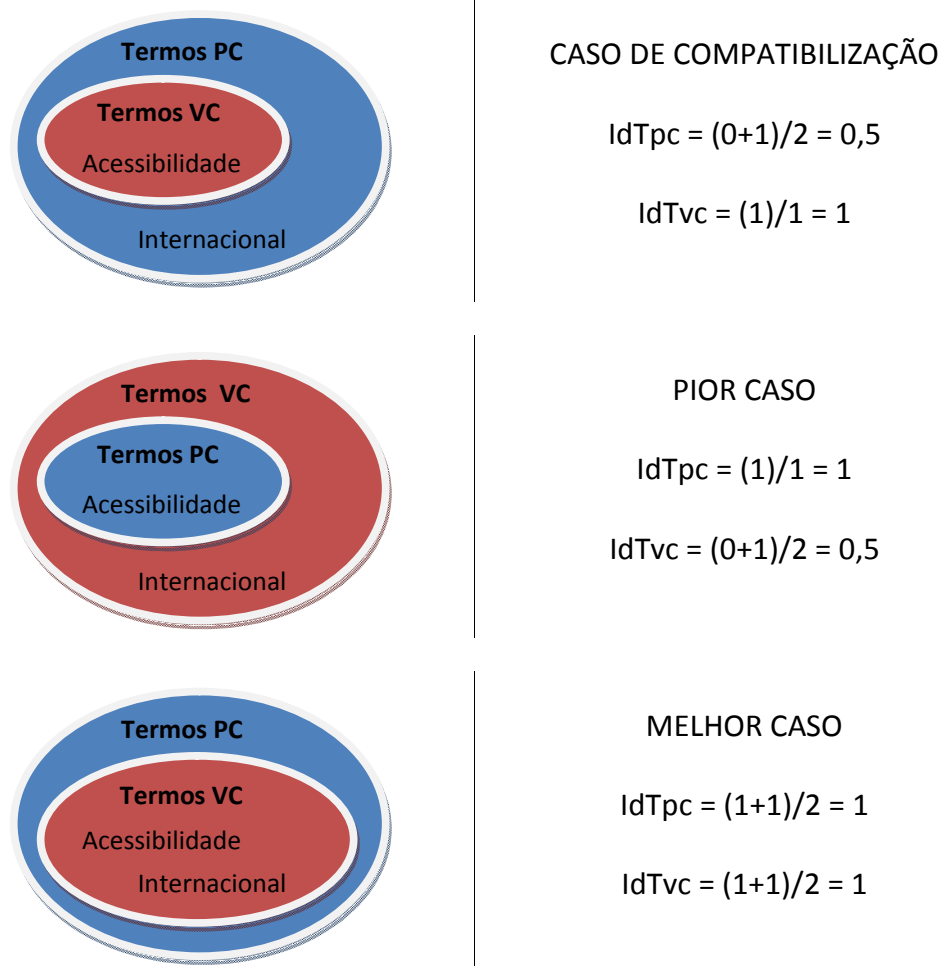
$$IdTpc = (n1+n2+n3....nn)/n$$

Onde  $IdTpc$  é o índice de termos de palavras-chave que está contido nos termos do vocabulário controlado;  $n1$  terá valor 1, se a palavra-chave possuir algum termo idêntico ou parte dele no vocabulário controlado.

$$IdTvc = (n1+n2+n3....nn)/n$$

Onde  $IdTvc$  é o índice de termos do vocabulário que está contido nos termos das palavras-chave. Os índices calculados somente serão armazenados no banco de dados se o  $IdTvc$  for igual a 1 e  $IdTpc$  for maior que 0,39. Após o armazenamento, será verificado o maior  $IdTpc$  e será compatibilizada a palavra-chave com o termo do vocabulário controlado. Na Figura 6, mostra-se a realização do cálculo.

**Figura 6** - Exemplos de cálculo do índice contido



Fonte: Elaborado pelo autor.

As compatibilizações são realizadas considerando-se os mecanismos de comparação para os vocabulários controlados. Na possibilidade de existir mais de um vocabulário controlado para a realização das compatibilizações, a comparação ocorre com ambos os vocabulários controlados, entretanto será considerado o

primeiro como o principal, isto é, a palavra-chave somente será validada com o segundo vocabulário caso não tenha sido validada com o primeiro.

#### - RF03 Tratamento Palavras-chave

O objetivo do RF03 é realizar o tratamento das palavras-chave não compatibilizadas com os vocabulários controlados, com a intervenção do humano. Neste processo, o foco é na semântica das palavras-chave para identificar os possíveis relacionamentos entre palavras-chave e termos dos vocabulários controlados. Está centrado em três operações: 1) compatibilização com termo correspondente no primeiro vocabulário controlado; 2) compatibilização com o segundo vocabulário controlado; 3) inserção em vocabulário controlado da própria revista.

O VCPC *tools* lista todas as palavras-chave não compatibilizadas, o usuário seleciona a palavra-chave a ser compatibilizada e aciona a busca textual. Em seguida, são apresentados os termos candidatos para a compatibilização, conforme os vocabulários controlados. O usuário poderá selecionar, nestas duas listas, os termos correspondentes. Caso não esteja listado entre os termos candidatos, o usuário poderá solicitar a listagem completa de termos dos vocabulários controlados. Se não for possível a compatibilização, o usuário poderá optar em cadastrar este termo no vocabulário controlado da revista. O vocabulário controlado da revista será um conjunto complementar de termos considerados termos preferidos, em um vocabulário controlado, e que ainda não constam dos vocabulários controlados utilizados para as compatibilizações. Neste momento, em que a VCPC *tools* tem seus processos gerenciados por um humano, a VCPC *tools* entra no modo de aprendizagem, isto é, ao registrar a correspondência de um termo, serão inseridos, na base de dados, os registros pertinentes a esta aprendizagem.

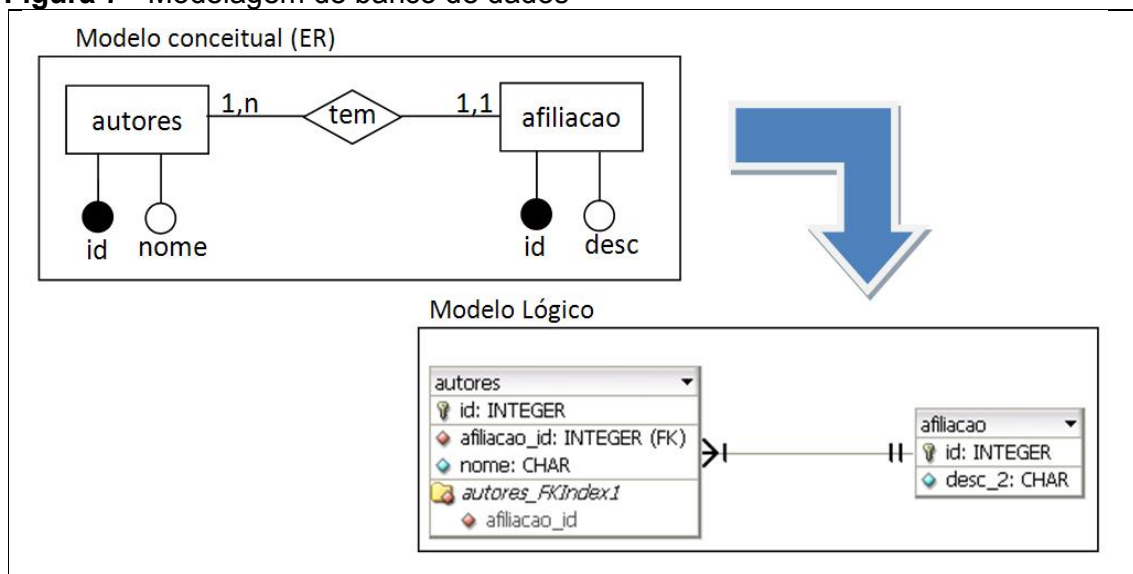
#### - RF04 Resultados

O objetivo do RF04 é mostrar os resultados quantitativos dos processamentos referentes à compatibilização de palavras-chave por processos de comparação.

#### 4.2.3 Desenvolver o projeto físico do banco de dados (1.3)

Neste planejamento, está contido o desenvolvimento do projeto físico do banco de dados – base estruturada relacional. Segundo Heuser (2004), o projeto do banco de dados passa por dois estágios: modelo conceitual – modelagem das necessidades de armazenamento por meio de um diagrama de entidade relacionamento, apresentado um modelo na Figura 7 – e o modelo lógico – modelagem do diagrama entidade relacionamento obtido para um SGBD específico, apresentado na Figura 7.

**Figura 7 - Modelagem de banco de dados**



Fonte: Elaborado pelo autor.

#### 4.2.4 Criar a Base de Dados no Sistema Gerenciador de Banco de Dados (1.4)

A construção do banco de dados no sistema de gerenciamento de banco de dados é realizada por meio da linguagem *SQL*, utilizando o comando *CREATE TABLE*. Na Figura 8, é apresentado um modelo de construção das tabelas modeladas na Figura 7. A base de dados foi criada no SGBD *MySQL*, escolhido por se tratar do mesmo sistema gerenciado de banco de dados utilizado no sistema SEER.

**Figura 8 - Exemplo de construção de banco de dados em um SGBD**

```
CREATE TABLE afiliacao (  
  id INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  desc_2 CHAR NULL,  
  PRIMARY KEY(id)  
);
```

```
CREATE TABLE autores (  
  id INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  afiliacao_id INTEGER UNSIGNED NOT NULL,  
  nome CHAR NULL,  
  PRIMARY KEY(id),  
  INDEX autores_FKIndex1(afiliacao_id)  
);
```

Fonte: Elaborado pelo autor.

## 4.3 ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS

### 4.3.1 Identificar os Vocabulários Controlados da Área Especializada (2.1)

Para a identificação de vocabulários controlados, considera-se a disponibilização eletrônica, com estrutura de fácil acesso aos termos e seus relacionamentos. A consistência do vocabulário controlado é fundamental para a decisão da escolha, assim como a data de publicação, ou seja, a mais atualizada possível.

### 4.3.2 Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2)

Refere-se à estruturação dos termos dos vocabulários controlados, em que se verificam a modalidade, o formato e a estrutura dos termos para traçar um planejamento de como organizar os termos em tabelas, planilhas, arquivos textos ou outras formas que deem suporte eficaz e eficiente para terceira ação.

#### 4.3.3 Descrever em Linguagem SQL os Termos dos Vocabulários Controlados (2.3)

Nesta atividade, estão incluídas as tarefas para a descrição dos dados em linguagem SQL. Essa linguagem é a forma padronizada de manipulação de dados no banco de dados. Portanto, nesta atividade, utiliza-se o comando *INSERT* para descrição dos termos dos vocabulários controlados e seus relacionamentos.

#### 4.3.4 Importar os Dados dos Termos dos Vocabulários Controlados (2.4)

Esta atividade é a importação, para o banco de dados, dos termos e seus relacionamentos dos vocabulários controlados. Este processo é simples; acontece por meio de execução dos comandos SQL no banco de dados.

### 4.4 ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS

#### 4.4.1 Coletar e Armazenar os Metadados (3.1)

A importação dos metadados do periódico para a base de dados é o processo de realizar as requisições ao protocolo OAI-PMH e tratar o retorno da requisição, por meio do arquivo XML. Este processo faz parte da modelagem do RF01 da ferramenta proposta. Portanto, nesta etapa, se dá a execução dessa funcionalidade.

#### 4.4.2 Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2)

A verificação das palavras-chave idênticas a um termo do vocabulário controlado importa na comparação de igualdade; este processo será executado pelo RF02 em seu mecanismo de compatibilização de palavras-chave idênticas.

#### 4.4.3 Verificar as Palavras-chave que Remetem à Igualdade dos Termos do Vocabulário Controlado (3.3)

Esta verificação acontece mediante a constatação da igualdade dos termos por meio da retirada de palavras vazias, plurais, sufixos e do índice contido. Os processos aqui executados são modelados no RF02, nas compatibilizações a partir da retirada de palavras vazias, em seguida, retirada dos plurais e, na sequência, a retirada dos sufixos. Por último se realiza o cálculo e compatibilizam-se as palavras-chave com um termo do vocabulário controlado.

#### 4.4.4 Vincular as Palavras-chave e Termos dos Vocabulários Controlados (3.4)

Os vínculos são estabelecidos entre as palavras-chave e os termos do vocabulário controlado após a execução das ações 3.1, 3.2 e 3.3, em específico nas ações 3.2 e 3.3, que verificam a compatibilização das palavras-chave com os termos do vocabulário controlado. Esta atividade é modelada no RF02, quando da execução da ferramenta.

### 4.5 ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE

#### 4.5.1 Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1)

Esta atividade é realizada no RF03 Tratamento Palavras-chave, e os mecanismos manuais são: a) identificação da palavra-chave não compatibilizada; b) realizar buscas em recursos terminológicos da área especializada; c) analisar os sinônimos ou termos relacionados; d) localizar, nos vocabulários controlados, o termo correspondente à palavra-chave; c) registrar por meio do RF03.

#### 4.5.2 Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2)

Esta ação é realizada nos procedimentos do RF03 Tratamento Palavras-chave, a partir das atividades da ação 4.1, e, em caso de insucesso na localização do termo correspondente nos vocabulários controlados, a partir da verificação nos recursos terminológicos, opta-se por inserir, no vocabulário controlado da revista, a palavra-chave, caso tenha características para ser considerado um termo. Nas palavras-chave não consideradas como termo apropriado, não será realizado nenhum tratamento.

## 5 ANÁLISE E RESULTADOS

A análise dos dados é um momento importante da pesquisa, porque com base nela é que serão enunciadas as respostas objetivadas no estudo, e serão postas em prática as inferências do próprio pesquisador, num processo de interpretação e busca da significação oferecida pelos dados. A forma e a condução deste procedimento são, em diversos casos, planejadas antes mesmo da etapa de coleta de dados (BARROS; LEHFELD, 1990, p. 87). O termo análise significa “[...] buscar sentido mais explicativo dos resultados da pesquisa [...].”(BARROS; LEHFELD, 1990, p. 87).

Os resultados aqui serão apresentados em conformidade com os fundamentos lançados como guia na execução dos procedimentos metodológicos: 4.2 Etapa 1 - Desenvolver a ferramenta VCPC tools; 4.3 Etapa 2 - Coletar e armazenar os vocabulários controlados; 4.4 Etapa 3 - Compatibilizar as palavras-chave com os termos dos vocabulários controlados; 4.5 Etapa 4 - Tratar manualmente as palavras-chave não compatibilizadas automaticamente. Logo ao término da apresentação dos resultados das etapas de que decorrem cada uma das ações, será apresentado, no item 5.5, o Índice de termos, no qual são expostos o índice de termos agrupados por letra inicial de cada termo e o perfil da revista selecionada para este estudo.

### 5.1 ETAPA 1 - DESENVOLVER A FERRAMENTA VCPC TOOLS

Os resultados mais expressivos desta etapa poderão ser visualizados nos resultados obtidos nas etapas 3 e 4, uma vez que o produto dessas etapas é a ferramenta VCPC tools.

#### 5.1.1 Identificar o Endereço do Repositório Digital do Periódico (1.1)

Os resultados desta ação é o reconhecimento do ambiente por meio da análise de funcionamento do protocolo OAI-PMH para a coleta de metadados. Apresenta-se na Figura 9 a interface desse protocolo.

Figura 9 - Interface do protocolo OAI-PMH do repositório digital

## OAI 2.0 Request Results

[Identify](#) | [ListRecords](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. More information about this XSLT is at the [bottom of the page](#).

Datestamp of response	2014-10-01T11:04:21Z
Request URL	[REDACTED]

Request was of type Identify.

Repository Name	[REDACTED]
Base URL	[REDACTED]
Protocol Version	2.0
Earliest Datestamp	2009-01-29T21:48:22Z
Deleted Record Policy	persistent
Granularity	YYYY-MM-DDThh:mm:ssZ
Admin Email	[REDACTED]

### OAI-Identifier

Scheme	oai
Repository Identifier	[REDACTED]
Delimiter	:
Sample OAI Identifier	[REDACTED]

### Unsupported Description Type

The XSL currently does not support this type of description.

Fonte: OAI-PMH – Revistas UEL – Universidade Estadual de Londrina.

A escolha para a aplicação do protótipo da ferramenta *VCPC tools* foi a Revista Informação & Informação, por ser um periódico científico da área da CI e possuir indicativo de qualidade no *Qualis* da Capes, atualmente no extrato B1. A Revista Informação & Informação está vinculada ao Programa de Pós-Graduação em Ciência da Informação do Departamento de Ciência da Informação da Universidade Estadual de Londrina; lançada em 1996, em formato impresso até 2002, depois migrou todos os seus fascículos para o formato eletrônico.

Atualmente, a revista publica edições quadrimestrais, com foco nos estudos da CI e áreas de interface. Portanto o resultado principal desta ação é a identificação do endereço do protocolo OAI-PMH do repositório digital da revista.

#### 5.1.2 Desenvolver o Projeto da Ferramenta *VCPC tools* (1.2)

O resultado desta ação somente é possível ser apresentado por meio das interfaces da *VCPC tools* e pelo código fonte utilizado na programação, que se torna expressivamente grande. As interfaces da *VCPC tools* serão apresentadas juntamente com os itens 5.3 e 5.4 para melhor visualização. Por outro lado, apresenta-se, na Figura 10, parte do código fonte, que trata da compatibilização de palavras-chave idênticas com os termos do vocabulário controlado.

**Figura 10 - Parte código fonte da compatibilização de palavras-chave idênticas**

```

43 //compatibilização PC/VC idênticas
44 function comp1(){
45     @mysql_connect("127.0.0.1", "root", "") or die (mysql_error ());
46     mysql_select_db("controlevocabulario") or die(mysql_error());
47     mysql_set_charset ("charset=utf8");
48     mysql_query("SET NAMES 'utf8'");
49     mysql_query('SET character_set_connection=utf8');
50     mysql_query('SET character_set_client=utf8');
51     mysql_query('SET character_set_results=utf8');
52     $strSQL = "SELECT id_palavraschave, termo FROM palavraschave";
53     $rs = mysql_query($strSQL) or die(mysql_error());
54     $contcompvc_ing=0;
55     $contcompvc_esp=0;
56     $contcompvc=0;
57     while ($row = mysql_fetch_array($rs)){
58         $strSQLvc = "SELECT Termo, id_termo FROM tbr_termos WHERE Termo LIKE '". $row['termo']. "'";
59         $rsv = mysql_query($strSQLvc) or die(mysql_error());
60         $rowvc = mysql_fetch_array($rsv);
61         if ($rowvc['Termo'] != "") {
62             $strSQLinsert = "UPDATE palavraschave SET mdcomp_tbr = 'IDENTICO', tbr_termos_id_te
63             | WHERE id_palavraschave = ". $row['id_palavraschave'];
64             $rsinsert = mysql_query($strSQLinsert) or die(mysql_error());
65             $strSQLuse = "SELECT * FROM tbr_termos, tbr_use WHERE tbr_use.id_termo_use = tbr_te
66             | $rsvuse = mysql_query ($strSQLuse);
67             $rowuse = mysql_fetch_array ($rsvuse);
68             if ($rowuse ['Termo'] != ''){
69                 insere_vcii($rowuse['Termo'], $row['id_palavraschave'], $rowuse['Id_ter
70             | }
71             | }
72             | else{
73                 insere_vcii($rowvc['Termo'], $row['id_palavraschave'], $rowvc['id_termo'].
74             | }
75             | }
76             | $contcompvc++;
77     }

```

Fonte: Elaborado pelo autor utilizando o *software Komodo Edit 8*

Na linha 44 da Figura 10, é declarada a função de compatibilização de palavras-chave idênticas, e nas linhas 45 a 51 efetua-se o estabelecimento de conexão com a base de dados. Nas linhas 54 a 56, são zeradas as variáveis de controle da quantidade de palavras-chave compatibilizadas. Da linha 57 à linha 76 (no código fonte, a função finaliza na linha 112 – que na **Erro! Fonte de referência não encontrada.** é apresentada até a 76), consta o *loop* que realiza a busca de um termo do vocabulário controlado idêntico a cada palavra-chave inserida na base de dados, e é estabelecido seu vínculo com o termo recuperado.

### 5.1.3 Desenvolver o Projeto Físico do Banco de Dados (1.3)

O resultado desta ação é o projeto físico do banco de dados. Na Figura 11, é apresentado o diagrama refinado com as tabelas, campos, tipos de dado de cada campo, relacionamentos entre as tabelas com suas respectivas cardinalidades. Foi utilizado o *software DBDesigner*.



As tabelas *afiliacao*, *areasconhecimento*, *artigos*, *artigos\_has\_autores*, *artigos\_has\_palavraschave*, *autores*, *edicao*, *editoras*, *genero*, *idiomas*, *pais*, *palavraschave* e *revistas* foram utilizadas para o armazenamento dos resultados da coleta de metadados. Logo as demais tabelas são complementares para o funcionamento da ferramenta VCPC *tools*, de acordo com a normalização de banco de dados, inclusive as tabelas que foram modeladas para receber os vocabulários controlados.

#### 5.1.4 Criar a Base de Dados no Sistema Gerenciador de Banco de Dados (1.4)

Os resultados desta ação não são expressivamente visuais, portanto, para efeito demonstrativo, apresenta-se a interface do *phpMyAdmin*, que demonstra as tabelas instanciadas no banco de dados, na Figura 12. No Quadro 14, é apresentado o dicionário de dados, que é um descritivo de cada campo das tabelas.

**Figura 12 - Interface do *phpMyAdmin* com o banco de dados *controlevocabulario***

Tabela	Registros	Tipo	Agrupamento (Collation)	Tamanho	Suspensão
afiliacao	225	MyISAM	utf8_unicode_ci	18.1 KB	-
areasconhecimento	8	MyISAM	utf8_unicode_ci	2.5 KB	-
artigos	437	MyISAM	utf8_unicode_ci	502.3 KB	-
artigos_has_autores	741	MyISAM	utf8_unicode_ci	7.5 KB	-
artigos_has_palavraschave	1,199	MyISAM	utf8_unicode_ci	11.5 KB	-
autores	540	MyISAM	utf8_unicode_ci	29 KB	-
edicao	44	MyISAM	utf8_unicode_ci	3.1 KB	-
editoras	1	MyISAM	utf8_unicode_ci	2 KB	-
genero	0	MyISAM	utf8_unicode_ci	1 KB	-
idiomas	1	MyISAM	utf8_unicode_ci	2 KB	-
pais	0	MyISAM	utf8_unicode_ci	1 KB	-
palavraschave	834	MyISAM	utf8_unicode_ci	263.5 KB	-
resultados	4,959	MyISAM	utf8_unicode_ci	127.7 KB	-
revistas	1	MyISAM	utf8_unicode_ci	1 KB	-
tbr_cat	2,020	MyISAM	utf8_unicode_ci	40.5 KB	-
tbr_indices_pc	78	MyISAM	utf8_unicode_ci	2.3 KB	-
tbr_pg	48	MyISAM	utf8_unicode_ci	3.2 KB	-
tbr_te	898	MyISAM	utf8_unicode_ci	8.9 KB	-
tbr_termos	1,814	MyISAM	utf8_unicode_ci	503.5 KB	-
tbr_tg	910	MyISAM	utf8_unicode_ci	9 KB	-
tbr_tr	1,837	MyISAM	utf8_unicode_ci	17.1 KB	-
tbr_up	769	MyISAM	utf8_unicode_ci	7.8 KB	-
tbr_use	769	MyISAM	utf8_unicode_ci	7.8 KB	-
tci_bt	1,693	MyISAM	utf8_unicode_ci	29.4 KB	13.5 KB
tci_facetas	14	MyISAM	utf8_unicode_ci	1.6 KB	-

Fonte: *phpMyAdmin* (2015).

**Quadro 14 - Dicionário de dados do projeto do banco de dados**

<b>TABELA: afiliação</b>		
<b>ATRIBUTO</b>	<b>TIPO</b>	<b>DESCRIÇÃO</b>
id_afiliacao	int(11)	Chave de identificação da afiliação dos autores
Instituição	varchar(255)	Nome da instituição dos autores
<b>TABELA: areasconhecimento</b>		
id_areasconhecimento	int(11)	Chave de identificação da área de conhecimento
Nome	varchar(255)	Nome da área de conhecimento
<b>TABELA: artigos</b>		
id_artigo	int(11)	Chave de identificação do documento/artigo
idiomas_id_idiomas	int(11)	Chave estrangeira de identificação do idioma do artigo
edicao_id_edicao	int(11)	Chave estrangeira de identificação da edição do artigo
Título	text	Título do documento/artigo língua portuguesa
titulo_eng	text	Título do documento/artigo em língua inglesa
titulo_esp	text	Título do documento/artigo em língua espanhola
resumo	text	Resumo do documento em língua portuguesa
resumo_eng	text	Resumo do documento em língua inglesa
resumo_esp	text	Resumo do documento em língua espanhola
data	date	Data do documento
formato	varchar(20)	Formato do documento disponível eletronicamente
uri	varchar(255)	Uniform Resource Identifier do documento/artigo
doi	varchar(255)	Digital Object Identifier System do documento/artigo
paginacao	varchar(10)	Página do documento na edição do periódico
setspec	varchar(20)	Identificação do tipo de documento
set	varchar(20)	Identificação do tipo de documento
identifier	varchar(200)	Identificação do artigo no protocolo OAI-PMH
<b>TABELA: artigos_has_autores</b>		
autores_id_autores	int(11)	Chave estrangeira código do autor
artigos_id_artigos	int(11)	Chave estrangeira código do artigo
<b>TABELA: artigos_has_palavraschave</b>		
artigos_id_artigo	int(11)	Chave estrangeira identificação do artigo
palavraschave_id_palavraschave	int(11)	Chave estrangeira identificação da palavra-chave
<b>TABELA: autores</b>		
id_autores	int(11)	Chave de identificação do autor
pais_id_pais	int(11)	Chave estrangeira de identificação do país do autor
afiliacao_id_afiliacao	int(11)	Chave estrangeira de identificação da afiliação do autor
nome	varchar(255)	Nome do autor
<b>TABELA: edição</b>		
id_edicao	int(11)	Chave de identificação da edição
volume	varchar(10)	Volume da edição
numero	varchar(10)	Número da edição
ano	varchar(10)	Ano de publicação da edição
revistas_id_revistas	int(11)	Chave estrangeira da revista
<b>TABELA: editoras</b>		
id_editoras	int(11)	Chave de identificação da editora
nome	varchar(255)	Nome da editora
localizacao	varchar(255)	Localização da editora
<b>TABELA: gênero</b>		
id_genero	int(11)	Chave de identificação do gênero
nome	varchar(100)	Nome do gênero
<b>TABELA: idiomas</b>		
id_idiomas	int(11)	Chave de identificação de idiomas
nome	varchar(100)	Nome do idioma
<b>TABELA: pais</b>		
id_pais	int(11)	Chave de identificação do país
Nome	varchar(255)	Nome do país
Sigla	varchar(20)	Sigla do país
<b>TABELA: palavraschave</b>		
id_palavraschave	int(11)	Chave de identificação da palavra-chave
termo	varchar(255)	Termo da palavra-chave em língua portuguesa
termo_eng	text	Termo da palavra-chave em língua inglesa
termo_esp	text	Termo da palavra-chave em língua espanhola
vci termos_id_termo	int(11)	Chave estrangeira do vocabulário controlado da I&I
tbr termos_id_termo	int(11)	Chave estrangeira do TBCI
tci termos_id_termo	int(11)	Chave estrangeira do TCI
mdcomp_tci	varchar(10)	Modo de compatibilização com o TCI
mdcomp_tbr	varchar(10)	Modo de compatibilização com o TBCI
stem_spv	text	Termo sem palavras vazias
stem_spl	text	Termo sem plural
stem_ssuf	text	Termo sem sufixo

Continua

Continuação

<b>TABELA: resultados</b>		
Qtdepcii	int(11)	Quantidade de palavras-chave iguais constatadas
qttotalpc	int(11)	Quantidade de palavras-chave importadas
qtdepccompvc_ident_TBR	int(11)	Quantidade de palavras-chave compatibilizada com o TBCI por igualdade dos termos
qtdepccompvc_pv_TBR	int(11)	Quantidade de palavras-chave compatibilizada com o TBCI por igualdade dos termos sem as palavras vazias
qtdepccompvc_pl_TBR	int(11)	Quantidade de palavras-chave compatibilizada com o TBCI por igualdade dos termos sem palavras vazias e sem plural
qtdepccompvc_suf_TBR	int(11)	Quantidade de palavras-chave compatibilizada com o TBCI por igualdade dos termos sem palavras vazias, sem plural e sem sufixo.
qtdepnaocompvc_TBR	int(11)	Quantidade de palavras-chave não compatibilizada com o TBCI
qtdeartigos	int(11)	Quantidade de artigos/documentos importado
dataimport	datetime	Data e hora da importação
set1	varchar(20)	Tipo de documento/artigo importado
id	int(11)	Chave de identificação do processo de importação
qtdepccompvc_id_TBR	int(11)	Quantidade de palavras-chave compatibilizada com o TBCI por meio do índice contido.
qtdepccompvc_ident_TCI	int(11)	Quantidade de palavras-chave compatibilizada com o TCI por igualdade dos termos
qtdepccompvc_pv_TCI	int(11)	Quantidade de palavras-chave compatibilizada com o TCI por igualdade dos termos sem palavras vazias
qtdepccompvc_pl_TCI	int(11)	Quantidade de palavras-chave compatibilizada com o TCI por igualdade dos termos sem palavras vazias e sem plural
qtdepccompvc_suf_TCI	int(11)	Quantidade de palavras-chave compatibilizada com o TCI por igualdade dos termos sem palavras vazias, sem plural e sem sufixo.
qtdepccompvc_id_TCI	int(11)	Quantidade de palavras-chave compatibilizada com o TCI por meio do índice contido.
qtdepnaocompvc_TCI	int(11)	Quantidade de palavras-chave não compatibilizada com o TCI por meio do índice contido.
<b>TABELA: revistas</b>		
id_revistas	int(11)	Chave de identificação da revista
areasconhecimento_id_areasconhecimento	int(11)	Chave estrangeira de identificação da área de conhecimento
editoras_id_editoras	int(11)	Chave estrangeira de identificação da editora
titulo	varchar(255)	Título da revista
<b>TABELA: t_corresp</b>		
tbr_termos_id_termo	int(11)	Chave estrangeira de identificação do TBCI
tci_termos_id_termo	int(11)	Chave estrangeira de identificação do TCI
termo	text	Termo correspondente a termo do TBCI ou TCI
<b>TABELA: tbr_cat</b>		
Id_termo	int(11)	Chave estrangeira de identificação do termo
Id_PG_Cat	varchar(10)	Chave estrangeira de identificação da categoria
<b>TABELA: tbr_indices_pc</b>		
palavraschave_id_palavraschave	int(11)	Chave estrangeira de identificação da palavra-chave
tbr_id_termo	int(11)	Chave estrangeira de identificação do termo
idtpc	float	Índice contido de termo palavra-chave
idtv	float	Índice contido de termo vocabulário controlado
<b>TABELA: tbr_pg</b>		
Id_PG	varchar(10)	Chave de identificação do Plano Geral de Classificação
Desc_PG	text	Descrição do Plano Geral de Classificação
<b>TABELA: tbr_te</b>		
Id_termo	int(11)	Chave de identificação do termo
Id_termo_te	int(11)	Chave de identificação do termo específico
<b>TABELA: tbr_termos</b>		
Id_termo	int(11)	Chave de identificação do termo
Termo	text	Termo em língua portuguesa
Ing	text	Termo em língua inglesa
Esp	text	Termo em língua espanhola
NE	text	Nota explicativa
stem_spv	text	Termo sem palavra vazia
stem_spl	text	Termo radicalizado – sem plural
stem_ssuf	text	Termo radicalizado – sem sufixo
<b>TABELA: tbr_tg</b>		
Id_termo	int(11)	Chave estrangeira de identificação do termo
Id_termo_tg	int(11)	Chave estrangeira de identificação do termo genérico

Continua

Continuação

<b>TABELA: tbr_tr</b>		
Id_termo	int(11)	Chave estrangeira de identificação do termo
Id_termo_tr	int(11)	Chave estrangeira de identificação do termo relacionado
<b>TABELA: tbr_up</b>		
Id_termo	int(11)	Chave estrangeira de identificação do termo
Id_termo_up	int(11)	Chave estrangeira de identificação do termo usado para
<b>TABELA: tbr_use</b>		
Id_termo	int(11)	Chave estrangeira de identificação do termo
Id_termo_USE	int(11)	Chave estrangeira de identificação do termo use
<b>TABELA: tci_bt</b>		
id_termo	int(11)	Chave estrangeira de identificação do termo
id_termo_bt	int(11)	Chave estrangeira de identificação do termo genérico
<b>TABELA: tci_facetas</b>		
id_faceta	int(11)	Chave de identificação da faceta
faceta	text	Descrição da faceta
<b>TABELA: tci_indices_pc</b>		
palavraschave_id_palavraschave	int(11)	Chave estrangeira de identificação da palavra-chave
tci_id_termo	int(11)	Chave estrangeira de identificação do termo
idtpc	float	Índice contido de termo palavra-chave
idtvpc	float	Índice contido de termo vocabulário controlado
<b>TABELA: tci_nt</b>		
id_termo	int(11)	Chave de identificação do termo
id_termo_nt	int(11)	Chave de identificação do termo específico
<b>TABELA: tci_rt</b>		
id_termo	int(11)	Chave estrangeira de identificação do termo
id_termo_rt	int(11)	Chave estrangeira de identificação do termo relacionado
<b>TABELA: tci_termos</b>		
id_termo	int(11)	Chave de identificação do termo
sn	text	Nota explicative
termo	text	Termo
stem_spv	text	Termo sem palavra vazia
stem_spl	text	Termo radicalizado – sem plural
stem_ssuf	text	Termo radicalizado – sem sufixo
<b>TABELA: tci_termos_faceta</b>		
id_termo	int(11)	Chave estrangeira de identificação do termo
id_faceta	int(11)	Chave estrangeira de identificação da faceta
<b>TABELA: tci_uf</b>		
id_termo	int(11)	Chave estrangeira de identificação do termo
id_termo_uf	int(11)	Chave estrangeira de identificação do termo usado para
<b>TABELA: tci_use</b>		
id_termo	int(11)	Chave estrangeira de identificação do termo
id_termo_use	int(11)	Chave estrangeira de identificação do termo use
<b>TABELA: vcii_termos</b>		
id_termo	int(11)	Chave de identificação do termo
termo	text	Termo
ne	text	Nota explicativa
estado	varchar(20)	Destinado à avaliação, quando for sugerido pelo autor
loc_np	text	Indica se são localidades ou nomes pessoais
te	text	Termo Específico
tg	text	Termo Geral
tr	text	Termo Relacionado
up	text	Usado para
id_termo_use	int(11)	Chave estrangeira de identificação do termo use
tbr_pg_id_pg	text	Chave estrangeira de identificação do plano geral de classificação
tbr_id_termo	int(11)	Chave estrangeira de identificação do termo do TBCI
tci_id_termo	int(11)	Chave estrangeira de identificação do termo do TCI

Fonte: Elaborado pelo autor.

## 5.2 ETAPA 2 - COLETAR E ARMAZENAR OS VOCABULÁRIOS CONTROLADOS

Nesta etapa os resultados mais visuais somente são possíveis de visualizar por meio de consultas ao banco de dados e na execução da VCPC *tools*. Porém as tomadas de decisão para se alcançar a importação dos vocabulários controlados são de extrema importância.

### 5.2.1 Identificar os Vocabulários Controlados da Área Especializada (2.1)

O resultado desta ação é a definição pela utilização de dois vocabulários controlados consolidados e disponíveis eletronicamente, que serão exigidos na execução da funcionalidade de análise de metadados (especificamente as palavras-chave) com vocabulário controlado da VCPC *tools* (RF02). São eles:

- PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 2014. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/tesouro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>>. Acesso em: 12 nov. 2014. (TBCI)
- MOREIRA, M. P. **Tesouro em Ciência da Informação**. Belo Horizonte: PUC Minas. Instituto de Informática; UFMG, [2006]. Disponível em: <<http://icei.pucminas.br/ci/tci/index.php>>. Acesso em: 9 mar. 2015 (TCI)

Já para a funcionalidade RF03, além do TBCI e TCI, optou-se pela utilização dos recursos terminológicos para consulta de termos, tais como:

- Universidade Federal do Espírito Santo. **Dicionário Eletrônico de Terminologia em Ciência da Informação (DeltCI)**. Vitória: UFES, 2007-2008.
- ARRUDA, Susana Margaret de; CHAGAS, Joseane. **Glossário de Biblioteconomia e Ciências afins: português – inglês**. Florianópolis: Cidade Futura, 2002.
- SANTOS, Gildenir Carolino; Ribeiro, Célia Maria. **Acrônimos, siglas e termos técnicos: arquivística, biblioteconomia, documentação e informática**. Campinas: Átomo, 2012.

- CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

### 5.2.2 Estruturar os Dados dos Termos dos Vocabulários Controlados (2.2)

Para realizar o procedimento de importação para o TBCI, na base de dados, foi necessário organizar os termos do TBCI a partir da conversão do arquivo com a extensão pdf (publicada pelos autores) para a extensão txt, importado posteriormente para a ferramenta *Microsoft Excel*. Após essa importação, os termos foram organizados, um por um, bem como suas respectivas relações ou funções, em formato de tabela, conforme apresentado no Quadro 15.

**Quadro 15** - Organização dos termos TBCI em tabela

ID	Termo	USE	ING	ESP	UP	TG	TE	TR	NE	CAT
17	acesso às publicações	acesso ao documento								6.2 Transferência e Acesso à Informação
18	acesso bibliográfico	acesso ao documento								6.2 Transferência e Acesso à Informação
19	acesso livre			acceso libre al documento	acesso aberto	acesso		acesso à informação	Acesso à informação científica ...	6.2 Transferência e Acesso à Informação

Fonte: Elaborado pelo autor utilizando o *software Microsoft Excel* fundamentado em Pinheiro e Ferrez (2014).

As relações e/ou funções de cada termo apresentadas no Quadro 15, por siglas, são descritas no Quadro 16 com suas respectivas explicações.

**Quadro 16** - Descrições dos termos

Descritivo – TBCI	Descritivo – TCI	Descrição
USE	USE	“O termo que se segue é o termo preferido quando se deve escolher entre sinônimos ou quase-sinônimos.” (AUSTIN, 1993, p.15)
UP	UF	“Usado para: o termo que se segue é um sinônimo ou um quase-sinônimo do termo preferido.” (AUSTIN, 1993, p.15)
ING	-	Termo em língua inglesa
ESP	-	Termo em língua espanhola
TG	BT	“Termo Genérico: o termo que se segue refere-se a um conceito com conotação mais ampla.” (AUSTIN, 1993, p.15)
TR	RT	“Termo Relacionado: o termo que se segue está associado, mas não é nem um sinônimo, nem um termo genérico ou específico.” (AUSTIN, 1993, p.15)
TE	NT	“Termo Específico: o termo que se segue refere-se a um conceito com conotação mais específica.” (AUSTIN, 1993, p.15)
NE	SN	“Nota explicativa: uma nota que se junta a um termo para indicar seu significado específico dentro de uma linguagem de indexação.” (AUSTIN, 1993, p.15)
CAT	FACETA	Classificação dos termos em categorias ou facetas.

Fonte: Elaborado pelo autor.

O campo NE, por se tratar de um campo com múltiplas linhas, requereu que, em alguns termos, fosse realizada a junção manual dessas linhas para uma célula da tabela do *Microsoft Excel*. Por meio da fórmula *SUBSTITUIR* do *Microsoft Excel*, foi possível eliminar as siglas de identificação constantes nos termos, USE, UP, ING, ESP, TG, TR, TE, NE, CAT. Posteriormente, foi utilizada a fórmula *ARRUMAR* para retirada de espaçamentos redundantes.

Para o processo de importação do TCI, foi desenvolvido um protótipo para automatizar esta ação. Esse protótipo coleta os dados em páginas *HTML* e foi implementado na linguagem *PHP*. Para a construção desse protótipo, fez-se o reconhecimento do ambiente de coleta de dados, de como estão estruturadas as informações dos termos. Diante desse contexto, foi constatado que todos os termos do vocabulário controlado estão sendo listados em ordem alfabética e ligados a seus respectivos descritivos, por meio de um *link*, e separados por letras iniciais de cada termo.

Portanto o protótipo de importação dos termos do TCI faz a coleta de todos os *links*, letra por letra, e faz o acesso a cada um deles para proceder à coleta dos relacionamentos de cada termo. Esse processo se dá por meio do registro de todo o

conteúdo de cada arquivo *html* e da localização das *tags* que contêm o conteúdo necessário a este processo. A partir dessa localização, são realizadas a retirada dos *links* e a limpeza das *tags* da linguagem *html* que está neles contido. Este mesmo processo é executado quando se faz o acesso à página dos relacionamentos e funções de cada termo. Esta página está estruturada por meio de tabela. Internamente, o protótipo organiza-se, com todos os dados descritivos, em uma estrutura de variável chamada *array* (vetor), apresentada no Quadro 17.

**Quadro 17 - Estrutura vetor dados TCI**

ÍNDICE RAÍZ	ÍNDICE DO TERMO	ÍNDICE DE RELAÇÃO	ÍNDICE DE TERMOS RELACIONADOS	CONTEÚDO
		[Termo]	[0]	AACR
[0]	[0]	[SN]	[0]	Anglo-American Cataloguing Rules - É um conjunto de regras que descrevem o conteúdo contido em registros de catálogo de biblioteca. A segunda edição deste trabalho é chamada de Chamado AACR2 ou AACR2R. A primeira edição foi concluída em 1967, a segunda, em 1979, e a revisão, em 1988.
		[UF]	[0]	Anglo-American Cataloguing Rules
		[BT]	[0]	Norma de catalogação bibliográfica
		[RT]	[0]	Norma internacional
		[Faceta]	[0]	BI - Informação e operações em biblioteca
ESCRITA DO VETOR: Array ( [0] => Array ( [0] => Array ( [Termo] => Array ( [0] => AACR ) [SN] => Array ( [0] => Anglo-American Cataloguing Rules - É um conjunto de regras que descrevem o conteúdo contido em registros de catálogo de biblioteca. A segunda edição deste trabalho é chamada de Chamado AACR2 ou AACR2R. A primeira edição foi concluída em 1967, a segunda em 1979, e a revisão em 1988. ) [UF] => Array ( [0] => Anglo-American Cataloguing Rules ) [BT] => Array ( [0] => Norma de catalogação bibliográfica ) [RT] => Array ( [0] => Norma internacional ) [Faceta] => Array ( [0] => BI - Informação e operações em biblioteca ) ).				

Fonte: Elaborado pelo autor.

### 5.2.3 Descrever em Linguagem SQL os Termos dos Vocabulários Controlados (2.3)

Para alcançar a descrição em linguagem SQL dos termos do TBCI, algumas atividades foram necessárias. Foi adicionada, para os termos, uma identificação. Nos campos USE, UP, TG, TR, TE e CAT, tem-se, em sua estrutura, um ou mais termos especificados para cada termo preferido e, por isso, foi necessário estabelecer um vínculo com o termo preferido por meio do seu identificador. Nos campos USE, UP, TG, TR e TE, que são termos que constam no rol de termos já identificados juntamente com os termos preferidos, foi necessário localizar cada um e adicionar o identificador para estabelecer os vínculos. Por exemplo, o termo AACR1, com seu identificador 1, tem em seu campo USE o termo Código de Catalogação Anglo-Americano cujo identificador é 380, portanto foi gerada uma coluna com o identificador do termo principal e outra coluna com o identificador do

termo USE. Para este processo, foi desenvolvido um algoritmo para realizar o procedimento de maneira mais rápida e com risco zero de erros.

Nesse algoritmo foram utilizados os seguintes vetores: um vetor com os termos preferidos e não preferidos, outro com seus identificadores e um vetor com os termos. No exemplo utilizado do campo USE, o algoritmo percorre todo o vetor de termos preferidos e não preferidos até localizar o termo do campo USE. Ao localizar, gera-se o comando de *insert* utilizando a linguagem SQL, própria de manipulação de dados em sistemas gerenciador de banco de dados, *INSERT INTO TBR\_USE (Id\_termo, Id\_Termo\_use) VALUES (1, 380);*. Este processo se repete para os demais campos.

No campo CAT, é realizado um tratamento diferenciado. Pelo fato de ele se referir ao Plano Geral de Classificação, foi utilizada a fórmula *ESQUERDA* do *Microsoft Excel*, para retirar a descrição e deixar o identificador próprio da categoria. A partir desse tratamento, foi utilizada a função *CONCATENAR* para gerar o SQL para inserção no banco de dados; por exemplo, *INSERT INTO TBR\_CAT (Id\_termo, Id\_PG\_Cat) VALUES (1, "2.1");*. Como a referência a este campo CAT é o Plano Geral de Classificação, foi necessário proporcionar a inserção deste plano no banco de dados. Para isso, o Plano Geral de Classificação foi importado no *Microsoft Excel*, desse modo, aplicou-se a fórmula *ESQUERDA* para separar o seu identificador da descrição, o mesmo processo realizado no campo CAT. Para obter a descrição de cada item deste plano, aplicou-se a fórmula *DIREITA*. Novamente, com o uso da fórmula *CONCATENA*, foi possível obter o SQL de inserção do Plano Geral de Classificação, por exemplo, *INSERT INTO TBR\_PG (Id\_PG, Desc\_PG) VALUES ('1', 'Epistemologia da Ciência da Informação');*.

Por fim, utilizou-se a fórmula *CONCATENA* do *Microsoft Excel* para gerar o SQL de inserção de cada termo como, por exemplo, *INSERT INTO TBR\_TERMOS (Id\_termo, Termo, Ing, Esp, Ne) VALUES (1, 'AACR1', "", "", "");*. Todas as descrições em linguagem SQL foram salvas em arquivos do tipo txt. A descrição em linguagem SQL dos termos TCI é gerada automaticamente pelo protótipo de coleta de dados em páginas *HTML*, a partir da estrutura Quadro 17.

### 5.2.4 Importar os Dados dos Termos dos Vocabulários Controlados (2.4)

Para obtenção da importação no banco de dados dos termos do TBCI, resultado desta ação, foram executadas as seguintes atividades: 1) Execução de todos os *SQLs*, para inserção do Plano Geral de Classificação na tabela denominada *TBR\_PG*; 2) Execução de todos os *SQLs*, para inserção dos termos na tabela *TBR\_TERMOS*; 3) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *USE* na tabela *TBR\_USE*; 4) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *UP* na tabela *TBR\_UP*; 5) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *TG* na tabela *TBR\_TG*; 6) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *TE* na tabela *TBR\_TE*; 7) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *TR* na tabela *TBR\_TR*; 8) Execução de todos os *SQLs*, para inserção dos vínculos dos termos principais com o campo *CAT* na tabela *TBR\_CAT*;. No Quadro 18, são apresentados os exemplos de inserção de cada tabela na base de dados e a ordem descrita acima.

**Quadro 18 - Inserção do TBCI na base de dados**

ORDEM DE INSERÇÃO	TABELA	EXEMPLO SQL
1	<i>TBR_PG</i>	<i>INSERT INTO TBR_PG (ID_PG, DESC_PG) VALUES ('1', 'EPISTEMOLOGIA DA CIÊNCIA DA INFORMÇÃO');</i>
2	<i>TBR_TERMOS</i>	<i>INSERT INTO TBR_TERMOS (ID_TERMOS, TERMO, ING, ESP, NE) VALUES (1, 'AACR1', ", ", "');</i>
3	<i>TBR_USE</i>	<i>INSERT INTO TBR_USE (ID_TERMOS, ID_TERMOS_USE) VALUES (1, 380);</i>
4	<i>TBR_UP</i>	<i>INSERT INTO TBR_UP (ID_TERMOS, ID_TERMOS_UP) VALUES (6, 7);</i>
5	<i>TBR_TG</i>	<i>INSERT INTO TBR_TG (ID_TERMOS, ID_TERMOS_TG) VALUES (6, 387);</i>
6	<i>TBR_TE</i>	<i>INSERT INTO TBR_TE (ID_TERMOS, ID_TERMOS_TE) VALUES (9, 141);</i>
7	<i>TBR_TR</i>	<i>INSERT INTO TBR_TR (ID_TERMOS, ID_TERMOS_TR) VALUES (9, 10);</i>
8	<i>TBR_CAT</i>	<i>INSERT INTO TBR_CAT (ID_TERMOS, ID_PG_CAT) VALUES (1, "2.1");</i>

Fonte: Elaborado pelo autor.

Por sua vez, para obter a importação do TCI, foi realizada, pelo protótipo, a verificação da existência, na base de dados, de cada termo e seus respectivos relacionamentos, caso não existisse, foi realizada a sua inserção. Os termos foram inseridos na tabela *tci\_termos* com os campos *id\_termo*, *termo* e *sn*. Já os descritivos, cada um em suas respectivas tabelas do banco de dados, ou seja: *UF* na tabela *tci\_uf*, *BT* na tabela *tci\_bt*, *NT* na tabela *tci\_nt*, *RT* na tabela *tci\_rt*, *Facetas* na tabela *tci\_facetas*.

### 5.3 ETAPA 3 - COMPATIBILIZAR AS PALAVRAS-CHAVE COM OS TERMOS DOS VOCABULÁRIOS CONTROLADOS

Para alcançar os demais resultados afixados nos objetivos deste trabalho, executaram-se os processos da *VCPCtools*, que envolvem a apresentação dos resultados a partir da próxima ação desta etapa.

#### 5.3.1 Coletar e Armazenar os Metadados (3.1)

Para alcançar os resultados desta ação, foi necessário obter o resultado da ação 1.1 Identificação do endereço do repositório digital do periódico, que está descrito nos resultados 5.1.1. O endereço do protocolo OAI-PMH da revista foi utilizado nesta ação, que é a importação dos metadados por meio da interface ilustrada na Figura 13. Os dados foram coletados nos metadados dos artigos publicados, desde o primeiro volume no ano de 1996 até o último volume no ano de 2014. Em uma revista científica, constam vários tipos de seções: artigos, pontos de vista, relatos de experiência, resenhas de livros/mídia, teses e dissertações e comunicações, portanto atribuiu-se a expressão artigos como uma designação genérica.

**Figura 13 - Interface de importação de metadados**

Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

A ação foi realizada com sucesso e obteve a importação de 437 registros (documentos), um total de 1199 palavras-chave e 365 palavras-chave idênticas, portanto importadas 834 palavras-chave. No Quadro 19, são apresentados os dados da coleta de metadados com os tipos de documentos.

**Quadro 19 - Dados coleta de metadados**

Tipo de documento	Descrição	Quantidade de documentos	Quantidade de palavras-chave
informacao:ART	ARTIGOS	268	1027
informacao:EDT	EDITORIAL	44	-
informacao:TD	TESES E DISSERTAÇÕES	29	4
informacao:REL	RELATOS DE EXPERIÊNCIA	22	87
informacao:COM	COMUNICAÇÃO	13	40
informacao:RES	RESENHAS DE LIVROS/MÍDIA	34	26
informacao:PTV	PONTOS DE VISTA	3	15
informacao:EXP	EXPEDIENTE	24	-
TOTAL		437	1199

Fonte: Elaborado pelo autor.

Por sua vez, no Quadro 20, é apresentado o agrupamento de artigos que têm as respectivas quantidades de palavras-chave atribuídas, agrupados por categorias de artigos.

**Quadro 20 - Quantidade de palavras-chave por artigos**

Qtde de palavras-chave	qtde de documentos					
	informacao:ART	informacao:TD	informacao:REL	informacao:COM	informacao:RES	informacao:PTV
0	3	27	0	2	23	0
1	2	1	0	0	2	0
2	21	0	2	1	7	0
3	93	1	9	3	1	0
4	70	0	3	6	0	1
5	60	0	6	1	0	1
6	10	0	0	0	0	1
7	8	0	2	0	1	0
8	1	0	0	0	0	0

Fonte: Elaborado pelo autor

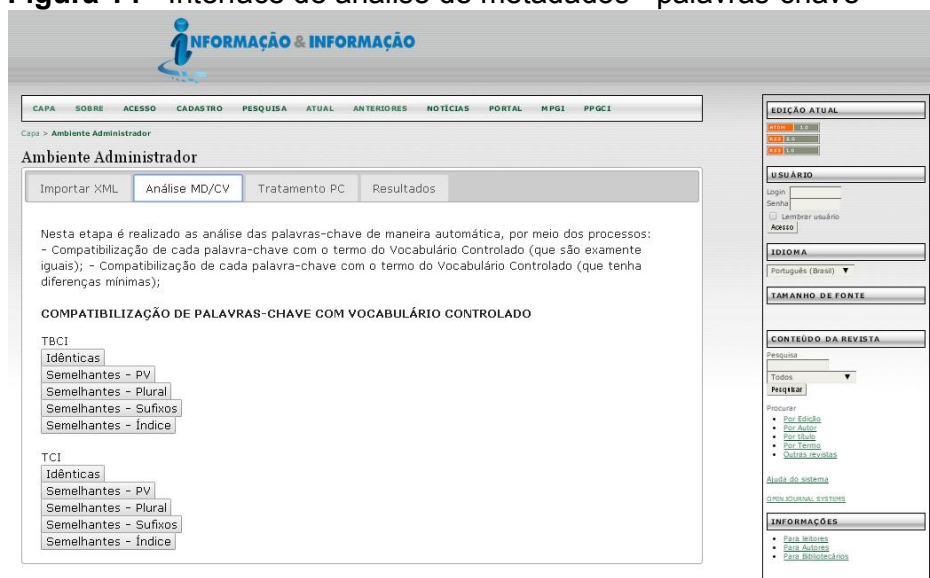
### 5.3.2 Verificar as Palavras-chave que são Idênticas com os Termos dos Vocabulários Controlados (3.2)

Os resultados desta ação e da ação 3.3 são decorrentes da execução das funções da interface da ferramenta VCPC *tools* ilustrada na Figura 14, onde constam todos os procedimentos descritos nos RF02 da metodologia.

Os procedimentos de compatibilização foram realizados na seguinte sequência: 1) TBCI – Idênticas; 2) TBCI – Semelhantes – PV; 3) TBCI – Semelhantes – Plural; 4) TBCI – Semelhantes – Sufixos; 5) TBCI – Semelhantes – índice; 6) TCI – Idênticas; 7) TCI – Semelhantes – PV; 8) TCI – Semelhantes – Plural; 9) TCI – Semelhantes – Sufixos; 10) TCI – Semelhantes – índice. Essa

sequência é a recomendada para que os resultados sejam fidedignos. Como resultado desta ação, foram considerados somente os procedimentos 1 e 6, portanto, no procedimento 1, foram obtidas 142 palavras-chave compatibilizadas, e no procedimento 6, foram obtidas 96 palavras-chave compatibilizadas.

**Figura 14 - Interface de análise de metadados - palavras-chave**



Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

### 5.3.3 Verificar as Palavras-chave que Remetem à Igualdade dos Termos do Vocabulário Controlado (3.3)

Os resultados desta ação são decorrentes da execução das funções da interface da ferramenta VCPC *tools* ilustrada na Figura 14, especificamente: 2) TBCI – Semelhantes – PV; 3) TBCI – Semelhantes – Plural; 4) TBCI – Semelhantes – Sufixos; 5) TBCI – Semelhantes – índice; 7) TCI – Semelhantes – PV; 8) TCI – Semelhantes – Plural; 9) TCI – Semelhantes – Sufixos; 10) TCI – Semelhantes – índice. Foi considerada a sequência recomendada na ação 3.2.

No **Quadro 21**, são apresentados os resultados de cada compatibilização. Observa-se que em nenhuma função ocorreu índice zero de compatibilização.

**Quadro 21 - Resultados de compatibilizações**

FUNÇÃO	TIPO DE COMPATIBILIZAÇÃO	TBCI	TCI
1 e 6	IDENTICA	142	96
2 e 7	IDENTsPV	6	8
3 e 8	IDENTsPL	54	44
4 e 9	IDENTsSUFIXO	6	6
5 e 10	IDENT_ID	74	84

Fonte: Elaborado pelo autor.

No Quadro 22, são apresentados os resultados consolidados das compatibilizações, que são apresentadas pela ferramenta *VCPC tools* na Figura 15.

**Quadro 22 - Resultados da compatibilização das palavras-chave**

	TCI	IDENTICA	IDENTsPV	IDENTsPL	IDENTsSUFIXO	IDENT_ID
TBCI		57	2	15	3	44
IDENTICA	20	45	3	26	2	0
IDENTsPV	0	3	0	5	0	0
IDENTsPL	10	25	5	4	0	0
IDENTsSUFIXO	2	3	0	0	1	0
IDENT_ID	41	9	0	4	0	30

Fonte: Elaborado pelo autor.

**Figura 15 - Interface de resultados**

The screenshot shows the SEER system interface. The main content area displays the following statistics:

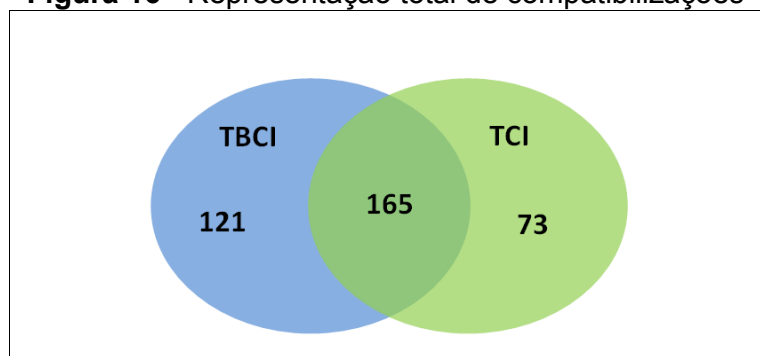
- IMPORTAÇÃO METADADOS - PERIÓDICO**
  - TOTAL DE DOCUMENTOS IMPORTADOS: 437
  - TOTAL DE PALAVRAS-CHAVE IMPORTADAS: 1199
  - TOTAL DE PALAVRAS-CHAVE IDENTICAS: 365
- COMPATIBILIZAÇÃO PALAVRAS-CHAVE x TBCI x TCI**
  - PALAVRAS-CHAVE IDENTICAS TBCI: 57
  - PALAVRAS-CHAVE IDENTICAS (Palavras vazias) TBCI: 2
  - PALAVRAS-CHAVE IDENTICAS (Plural) TBCI: 15
  - PALAVRAS-CHAVE IDENTICAS (Sufixo) TBCI: 3
  - PALAVRAS-CHAVE IDENTICAS (Índice) TBCI: 44
  - TOTAL DE PALAVRAS COMPATIBILIZADAS - TBCI: 121**
  - PALAVRAS-CHAVE IDENTICAS TCI: 20
  - PALAVRAS-CHAVE IDENTICAS (Palavras vazias) TCI: 0
  - PALAVRAS-CHAVE IDENTICAS (Plural) TCI: 10
  - PALAVRAS-CHAVE IDENTICAS (Sufixo) TCI: 2
  - PALAVRAS-CHAVE IDENTICAS (Índice) TCI: 41
  - TOTAL DE PALAVRAS COMPATIBILIZADAS - TCI: 73**
  - PALAVRAS-CHAVE IDENTICAS TBCI e IDENTICAS TCI: 45
  - PALAVRAS-CHAVE IDENTICAS TBCI e IDENTICAS (PV) TCI: 3
  - PALAVRAS-CHAVE IDENTICAS TBCI e IDENTICAS (PL) TCI: 25
  - PALAVRAS-CHAVE IDENTICAS TBCI e IDENTICAS (SUF) TCI: 3
  - PALAVRAS-CHAVE IDENTICAS TBCI e IDENTICAS (ÍNDICE) TCI: 9
  - PALAVRAS-CHAVE IDENTICAS (PV) TBCI e IDENTICAS TCI: 3
  - PALAVRAS-CHAVE IDENTICAS (PV) TBCI e IDENTICAS (PV) TCI: 0

The interface also includes a sidebar with sections for 'EDIÇÃO ATUAL', 'USUÁRIO', 'IDIOMA', 'PARÂMETRO DE FONTE', 'CONTEÚDO DA REVISTA', and 'INFORMAÇÕES'.

Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

Observou-se a necessidade de validar os termos com mais um vocabulário controlado, diante dos resultados dos primeiros testes realizados. Portanto, como é possível observar no Quadro 22 e na Figura 15, tem-se um total de compatibilizações de 121 palavras-chave com somente o TBCI e de 73 palavras-chave com somente o TCI. O total de palavras compatibilizadas foi 359. Na Figura 16, é demonstrada, de forma gráfica, a intersecção dos resultados.

**Figura 16** - Representação total de compatibilizações



Fonte: Elaborado pelo autor.

#### 5.3.4 Vincular as Palavras-chave e Termos dos Vocabulários Controlados (3.4)

Os resultados desta ação estão contidos, internamente, nos processos já executados, porém a relação entre quantidade de palavras-chave atribuídas por artigo e palavras-chave compatibilizadas por artigo é expressamente significativa. Nos totais do Quadro 23, excluem-se os artigos com zero palavras-chave atribuídas e zero palavras-chave compatibilizadas, para que se possa verificar que são mínimas as porcentagens de artigos que não tenham nenhuma palavra-chave compatibilizada de pelo VCPC *tools*. Os artigos caracterizados por ART representam 91,70% sobre o total de artigos que contêm, no mínimo, uma palavra-chave atribuída. Já os caracterizados por TD representam 50% de artigos sobre o total de artigos que contêm, no mínimo, uma palavra-chave atribuída. Esse é o pior caso, porém deve ser levada em consideração a quantidade de artigos enquadrada como TD.

**Quadro 23** - Quantidade de palavras-chave x palavras-chave compatibilizadas por artigos

QTDE DE PALAVRAS-CHAVE	QUANTIDADE DE DOCUMENTOS											
	informacao:ART		informacao:TD		informacao:REL		informacao:COM		informacao:RES		informacao:PTV	
	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP	PC	PC-COMP
0	3	25	27	28	0	0	2	5	23	26	0	0
1	2	69	1	0	0	5	0	2	2	5	0	0
2	21	72	0	1	2	11	1	4	7	3	0	1
3	93	64	1	0	9	2	3	2	1	0	0	1
4	70	27	0	0	3	3	6	0	0	0	1	1
5	60	9	0	0	6	0	1	0	0	0	1	0
6	10	0	0	0	0	1	0	0	0	0	1	0
7	8	1	0	0	2	0	0	0	1	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0	0
<b>TOTAL</b>	<b>265</b>	<b>243</b>	<b>2</b>	<b>1</b>	<b>22</b>	<b>22</b>	<b>11</b>	<b>8</b>	<b>11</b>	<b>8</b>	<b>3</b>	<b>3</b>
%	100	91,70	100	50	100	100	100	72,73	100	72,73	100	100

Fonte: Elaborado pelo autor.

#### 5.4 ETAPA 4 - TRATAR MANUALMENTE AS PALAVRAS-CHAVE NÃO COMPATIBILIZADAS AUTOMATICAMENTE

As palavras-chave não compatibilizadas automaticamente nas ações 3.2 e 3.3 serão tratadas por processo manual, ou seja, com a intervenção humana. Para efeito deste estudo, foi previsto, na ferramenta VCPC *tools*, este tratamento.

##### 5.4.1 Verificar as Palavras-chave que Tenham Termos Correspondentes nos Vocabulários Controlados (4.1)

Os resultados desta ação é a apresentação dos procedimentos a serem realizados na interface do VCPC *tools* para tratamento das palavras-chave, apresentado na Figura 17. Deve ser selecionado a palavra-chave no rol de palavras-chave (listagem disponibilizada pelo VCPC *tools*), e executar a busca. Nessa busca, será mostrada a opção para inserção deste termo no VCII ou a seleção dentre os termos do TBCI e do TCI candidatos, que são escolhidos, automaticamente, por meio de busca textual. Essa busca textual integra as formas de escrita em linguagem natural, porém as palavras com três caracteres não são tratadas, diante da configuração padrão do sistema de gerenciamento de banco de dados.

Para realizar o processo de tratamento das palavras-chave não compatibilizadas, será verificada cada palavra-chave nas ferramentas terminológicas. Utilizam-se termos correspondentes, sinônimos, amplitude da palavra-chave (geral/específico).

**Figura 17 - Interface para tratamento manual de palavras-chave**

CAPIA - AMBIENTE ADMINISTRADOR

CAPIA | SOBRE | ACESSO | CADASTRO | PESQUISA | ATUAL | ANTERIORES | NOTÍCIAS | PORTAL | HPG1 | PPGCI

Ambiente Administrador

Importar XML | Análise MD/CV | **Tratamento PC** | Resultados

Identificador: 194

PALAVRA-CHAVE/TERMO: Análise do Discurso

Inserir palavra-chave no VC do periódico

Termos sugeridos do TBCI	Termos sugeridos do TCI
<p>Listar demais termos do TBCI</p> <ul style="list-style-type: none"> <li><input type="radio"/> análise comparativa</li> <li><input type="radio"/> análise de agrupamentos</li> <li><input type="radio"/> análise de citação</li> <li><input type="radio"/> análise de co-citação</li> <li><input type="radio"/> análise de conglomerados</li> <li><input type="radio"/> análise de coocorrência</li> <li><input type="radio"/> análise de coocorrência de palavras</li> <li><input type="radio"/> análise de coocorrência de termos</li> <li><input type="radio"/> análise de custos</li> <li><input type="radio"/> análise de dados</li> <li><input type="radio"/> análise de domínio</li> <li><input type="radio"/> análise de facetas</li> <li><input type="radio"/> análise de informação na indústria</li> <li><input type="radio"/> análise de macroambiente</li> <li><input type="radio"/> análise de tarefas</li> <li><input type="radio"/> análise do fluxo de trabalho</li> <li><input type="radio"/> análise estatística</li> </ul>	<p>Listar demais termos do TCI</p> <ul style="list-style-type: none"> <li><input type="radio"/> Análise de assunto</li> <li><input type="radio"/> Análise de citação</li> <li><input type="radio"/> Análise de cluster</li> <li><input type="radio"/> Análise de co-citação</li> <li><input type="radio"/> Análise de co-ocorrência</li> <li><input type="radio"/> Análise de conteúdo</li> <li><input type="radio"/> Análise de custo-benefício</li> <li><input type="radio"/> Análise de dados</li> <li><input type="radio"/> Análise de domínio</li> <li><input type="radio"/> Análise de elementos finitos</li> <li><input type="radio"/> Análise de fatores</li> <li><input type="radio"/> Análise de workflow</li> <li><input type="radio"/> Análise de Fourier</li> <li><input type="radio"/> Análise de rede</li> <li><input type="radio"/> Análise de sistema</li> <li><input type="radio"/> Análise de variância</li> <li><input type="radio"/> Análise documental</li> </ul>

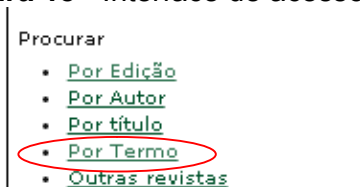
Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

#### 5.4.2 Inserir a Palavra-chave como Termo no Vocabulário Controlado da Revista (4.2)

Os resultados desta ação são demonstrados a partir da tomada de decisão de inclusão da palavra-chave como termo no VCII, em que devem ser verificados os relacionamentos desse termo com o contexto do VCII. Sabe-se que o VCII é composto por termos dos TBCI e TCI que tenham artigos relacionados. Uma possibilidade, ao inserir o termo no VCII, é sinalizar os termos como nomes pessoais ou localidades, uma vez que esses serão apresentados respectivamente caracterizados no índice de termos, apresentado na Figura 19.

### 5.5 ÍNDICE DE TERMOS

A partir de todo o processo de compatibilização automática e do tratamento manual das palavras-chave da ferramenta VCPC *tools*, no ambiente aberto aos usuários do periódico, é possível acessar pelo *link* do portal o *índice de termos*, conforme ilustrado na Figura 18.

**Figura 18 - Interface de acesso ao índice de termos**

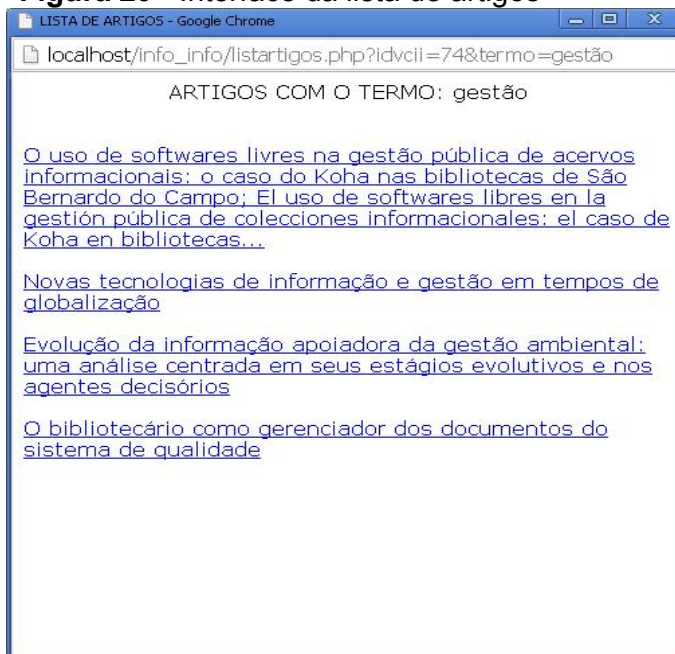
Fonte: Elaborado pelo autor utilizando o *layout* do sistema SEER.

Neste índice, o usuário tem acesso à lista de termos preferidos organizados em ordem alfabética, agrupados por letra inicial dos termos, e também por localidades e nomes pessoais. Na Figura 19 é ilustrada a página onde é possível verificar esta lista. Em cada termo, consta a quantidade de artigos relacionados ao referido termo. Tanto o termo quanto a quantidade de artigos a ele relacionados são *links*, ao clicar no termo, serão mostradas as palavras-chave a ele relacionadas e as devidas compatibilizações com o TBCI e TCI, se for o caso, conforme apresentado na Figura 21. Ao clicar na quantidade de artigos relacionados a este termo, será mostrado o título de cada artigo ligado por meio de *link* ao artigo, caso o usuário queira consultá-lo, ilustrado na Figura 20. Observa-se que nesta lista, somente são apresentados os termos preferidos que tenham registros de vínculos com alguma palavra-chave, que leva a encontrar, no mínimo, um artigo.

**Figura 19 - Interface da lista de termos**

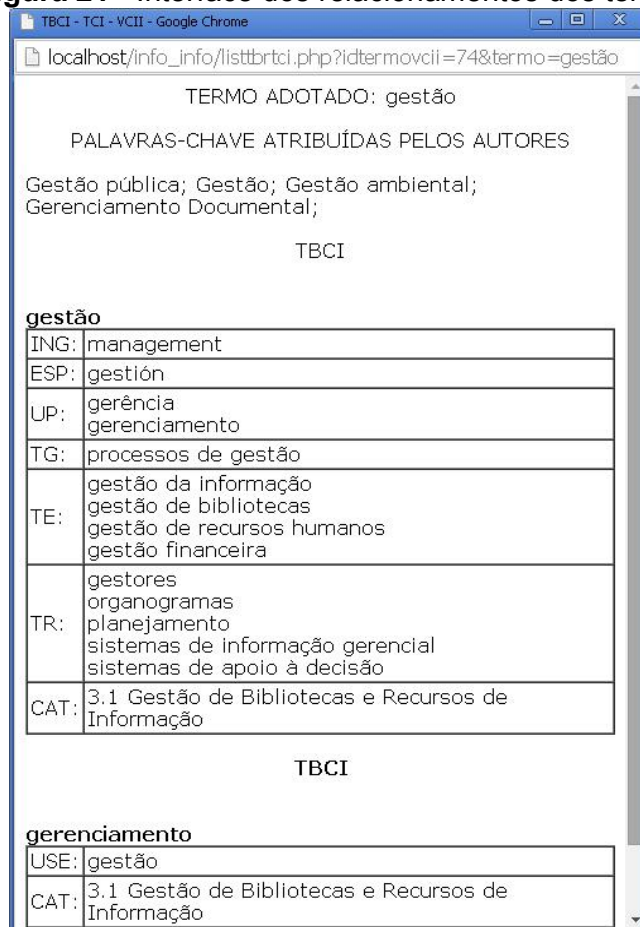
Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

**Figura 20 - Interface da lista de artigos**



Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

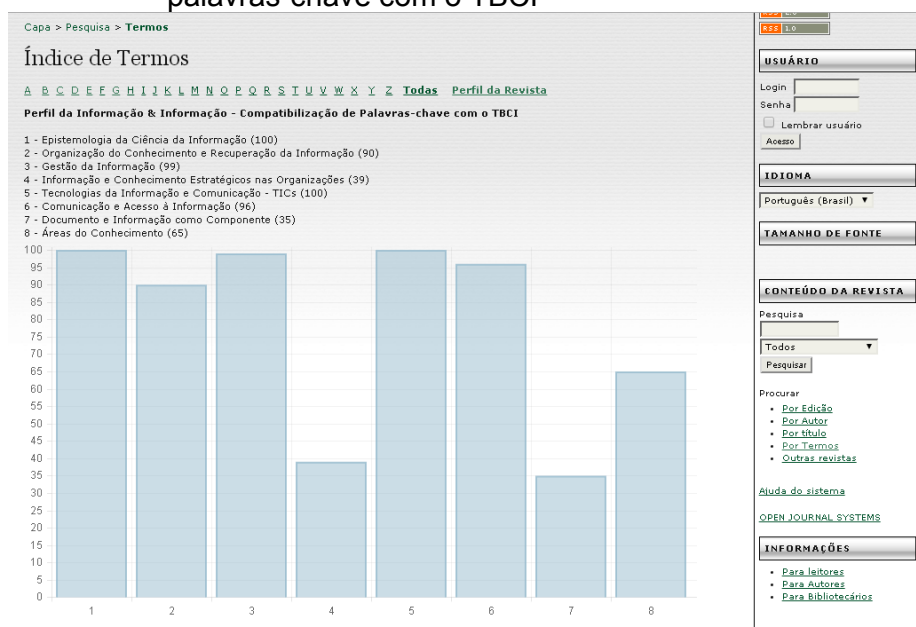
**Figura 21 - Interface dos relacionamentos dos termos**



Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

No ambiente do usuário, é possível acessar o perfil da revista traçado por meio das principais categorias em que enquadram as palavras-chave do Plano de Classificação Geral do TBCI, interface apresentada na Figura 22. O segundo perfil é formado por meio das facetadas do TCI, interface apresentada na Figura 23. Observa-se que, em ambos os perfis, são apresentadas a lista de categorias/facetadas e a quantidade de artigos relacionados por meio da compatibilização de palavras-chave realizada nos processos anteriores, no ambiente do administrador.

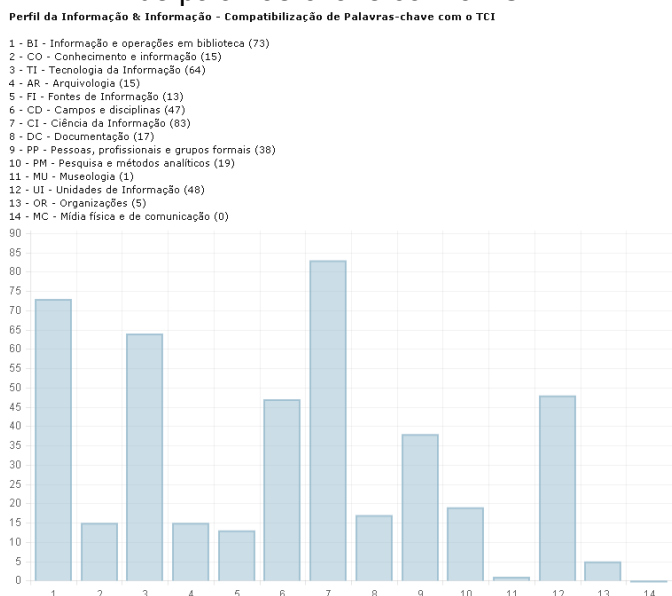
**Figura 22 - Perfil da Informação & Informação - compatibilização de palavras-chave com o TBCI**



Fonte: Elaborado pelo autor utilizando o *layout* do sistema SEER.

Os dados do perfil da revista de acordo com as categorias do TBCI, Figura 22, são: 1 - Epistemologia da Ciência da Informação (100); 2 - Organização do Conhecimento e Recuperação da Informação (88); 3 - Gestão da Informação (99); 4 - Informação e Conhecimento Estratégicos nas Organizações (39); 5 - Tecnologias da Informação e Comunicação - TICs (100); 6 - Comunicação e Acesso à Informação (96); 7 - Documento e Informação como Componente (35); 8 - Áreas do Conhecimento (65). Portanto a revista tem mais publicações nas categorias 1 e 5, com maioria dos registros, na sequência as categorias 3, 6 e 2, em ordem decrescente.

Figura 23 - Perfil da Informação & Informação - compatibilização de palavras-chave com o TCI



Fonte: Elaborado pelo autor, utilizando o *layout* do sistema SEER.

Já os dados do perfil da revista, em conformidade com as facetas do vocabulário controlado TCI, Figura 23, são: 1 - BI - Informação e operações em biblioteca (73); 2 - CO - Conhecimento e informação (15); 3 - TI - Tecnologia da Informação (64); 4 - AR - Arquivologia (15); 5 - FI - Fontes de Informação (13); 6 - CD - Campos e disciplinas (47); 7 - CI - Ciência da Informação (83); 8 - DC - Documentação (17); 9 - PP - Pessoas, profissionais e grupos formais (38); 10 - PM - Pesquisa e métodos analíticos (19); 11 - MU - Museologia (1); 12 - UI - Unidades de Informação (48); 13 - OR - Organizações (5); 14 - MC - Mídia física e de comunicação (0). Portanto, de acordo com este perfil, a revista tem mais publicações em CI - Ciência da Informação, seguida de Informação e operações em biblioteca.

Como é prevista a compatibilização das palavras-chave, o perfil ilustrado nas figuras 22 e 23 é resultado dessa compatibilização, que mantém, na base de dados, o relacionamento estabelecido entre cada palavra-chave e o respectivo vocabulário controlado, que por sua vez tem relacionamentos estabelecidos com as categorias, no caso do TBCI, e nas facetas, no caso do TCI. As palavras-chave têm relacionamentos com os artigos, o que torna possível realizar esta consulta.

## 5.6 LIMITAÇÕES DA PESQUISA

Diante dos objetivos deste estudo e executando os procedimentos metodológicos, observam-se as seguintes limitações de pesquisa:

- Com a falta de materiais que abordam a extração de conteúdos de arquivos XML com padrão de metadados *Doublin Core*, despendeu-se muito tempo para observar em outros materiais outras propostas e modelar para a necessidade do projeto e, conseqüentemente, executar vários testes para atingir os resultados esperados.

- Os processos de interoperabilidade de metadados por meio do OAI-PMH são pouco explorados nos materiais pesquisados com resultados descritos de maneira prática. Os processos ficam integrados com a teoria dos materiais. Portanto, como este estudo trata de conhecer os processos com profundidade para a implementação, gera-se um desgaste em torno de adaptações de outras práticas, em torno de testes para obter os resultados desejados.

- Os vocabulários controlados, por sua vez, encontram-se disponibilizados em diversos formatos eletrônicos, o que tornou oneroso o processo de importação dos termos desses vocabulários controlados para execução da compatibilização nesta pesquisa.

- Com relação aos dados coletados inicialmente, gerou-se um impacto negativo nos resultados pelo fato de constar inconsistência na atribuição de alguns metadados da revista, principalmente em edições antigas. Isso sugere que em tal época não tinha a preocupação com esses metadados, porém a situação foi levada ao conhecimento do editor da revista que, prontamente, e na medida do possível, realizou as correções.

- No processo de retirada de sufixos, as referências desses algoritmos estão conectadas com o processo de recuperação da informação e em outras linguagens de desenvolvimento, portanto demandaram maior proporção de tempo para a reconstrução desses algoritmos na linguagem desenvolvida a VCPC *tools*.

## 6 CONSIDERAÇÕES FINAIS

As publicações científicas estão cada vez mais abundantes e evolutivas, assim como as TIC, contexto diante do qual se tornam cada vez mais complexos os mecanismos de recuperação com algoritmos que tratam da linguagem natural, *tags* atribuídas aos documentos eletrônicos, entre outras organizações por meio de instrumentos para garantir a consistência na busca e recuperação.

Por outro lado, as palavras-chave dos periódicos científicos eletrônicos têm essa finalidade, e o que se observa é a falta de atribuir-lhes o grau de importância para exercer um papel ainda mais exigente, considerando que a tendência do volume de produções é aumentar. Trata-se de um elemento representativo atribuído pelos próprios autores de uma publicação científica. Alude-se uma valorização que deve vir acompanhada de um conhecimento sobre as vantagens para todos, autores e leitores, quando as atribuições de palavras-chave são feitas com precisão, sem improvisos ou deduções de último instante.

A motivação deste estudo centrou-se na busca de tratar a padronização das palavras-chave em sua pertinência ao campo do controle de vocabulário, considerando o alcance daquelas já atribuídas em periódicos científicos, e desenvolver uma metodologia para aferir essa atribuição. A questão problema foi assim enunciada: Como realizar o controle do vocabulário das palavras-chave atribuídas aos artigos científicos disponíveis em periódicos científicos eletrônicos? O objetivo geral foi propor a implantação do controle de vocabulário de palavras-chave atribuídas em um periódico científico eletrônico para subsidiar o processo de indexação e a recuperação dos conteúdos.

O primeiro objetivo específico tratou exclusivamente dos processos de desenvolvimento da VCPC *tools*, que incluíram aqui todas as ações realizadas para colocá-la em funcionamento. A VCPC *tools* foi desenvolvida conforme as metodologias de desenvolvimento proporcionadas pela engenharia de *software*, o que resultou em quatro requisitos funcionais mapeados conforme o reconhecimento do ambiente em que será implantada, em especial a coleta de metadados utilizando o protocolo OAI-PMH do periódico científico eletrônico. Os processos de compatibilização das palavras-chave foram implementados para assegurar o máximo de compatibilização possível por meio da VCPC *tools*. Foram enunciados, portanto, os quatro processos de compatibilização implementados.

Considera-se a VCPC *tools*, com essas facetas explicitadas, uma ferramenta complementar na implantação do controle de vocabulário das palavras-chave atribuídas aos periódicos científicos eletrônicos junto do Sistema Eletrônico de Editoração de Revistas (SEER). Assegura-se a validade desta ferramenta por meio dos procedimentos metodológicos que foram utilizados para executar os processos em um estudo de caso, aplicado na revista Informação & Informação.

Para realização dos testes previstos para a VCPC *tools*, foi necessário ter disponíveis, na base de dados da VCPC *tools*, os termos dos vocabulários controlados para convalidação das palavras-chave, por meio do processo de compatibilização. Nessa ocorrência, situou-se a identificação dos vocabulários controlados caracterizados como TBCI e TCI e, para ambos, foram aplicadas as ações de importação dos termos na base de dados da VCPC *tools*. Do TBCI, utilizaram-se processos manuais para organização dos termos e alguns processos automatizados, por meio de algoritmos, para dar maior rapidez e confiabilidade ao resultado de importação. Para o TCI, o procedimento de importação possibilitou ser totalmente automatizado, devido à organização e à disponibilidade dos termos e seus relacionamentos.

Cabe ressaltar a equivalência da quantidade de termos importados dos vocabulários controlados: o TBCI, 1.814 termos, e o TCI, 1891 termos, o que os torna importantes no processo de compatibilização, por se tratar de um número expressivo de termos na área da Ciência da Informação.

Nos resultados alcançados com o objetivo de compatibilizar as palavras-chave com os termos dos vocabulários controlados, a demanda primeira foi realizar a coleta dos metadados. Evidencia-se a quantidade de palavras-chave exatamente idênticas, 30,44%, fator que instiga a duplicidade de termos das palavras-chave. Observa-se, também, a frequência de palavras-chave atribuídas aos artigos, a maior é concentrada em três palavras-chave atribuídas, na sequência quatro e cinco palavras-chave.

O processo automatizado pela VCPC *tools* de compatibilização é validado a partir dos resultados obtidos. Os índices de compatibilizações foram relativamente surpreendentes e muito expressivos, chegando a um total de 359 palavras-chave compatibilizadas. Cabe ressaltar que, mediante uma relação de frequência de palavras-chave atribuídas aos artigos e a frequência de palavras-chave compatibilizadas, obteve-se um índice médio de 81,19%; portanto, por meio deste

processo automatizado em que foi constituído o índice de termos, significa que 81,19% artigos poderão ser recuperados até este estágio.

Salienta-se que essa média tem, em um dos casos, uma porcentagem baixa decorrente da quantidade de palavras-chave compatibilizadas, sendo uma de duas palavras atribuídas, observado o tipo de seção constante no periódico científico denominado resumos de teses e dissertações (TD). Desse modo, a média é menor.

Das palavras-chave compatibilizadas, é importante esclarecer que se obteve um índice alto em relação à compatibilidade entre palavra-chave x TBCI x TCI, ou seja, das 359 palavras-chave compatibilizadas, 165 são compatíveis com TBCI e TCI. A inferência possível é de que os vocabulários controlados da área da Ciência da Informação utilizados têm grande proximidade dos termos, o que dá mais suporte para o processo de compatibilização.

O objetivo de tratar manualmente as palavras-chave não compatibilizadas automaticamente foi uma grande experiência em relação ao contato com o processo automatizado, no que tange à metodologia. Com a intervenção humana para realizar consultas nos recursos terminológicos, é possível localizar os termos correspondentes a palavra-chave. Diante da quantidade de palavras-chave que determina localidades e nomes pessoais, optou-se em possibilitar a identificação com suas respectivas características. Todo este processo é registrado pela VCPC *tools*. Ressaltam-se, também, nessas palavras-chave não compatibilizadas, termos que não remetem ao conteúdo do artigo, portanto esse se torna obsoleto.

Vale ressaltar que, na VCPC *tools*, foi implementado o registro dos termos correspondentes quando se realiza o tratamento de maneira manual das palavras-chave. Essa decisão no projeto da ferramenta, por uma fatia dos conceitos de aprendizagem de máquina, é importante diante da prospecção de implantação em outros periódicos da área da Ciência da Informação. Justifica-se esta aprendizagem para que o índice de compatibilidade entre palavras-chave e os vocabulários controlados possa ser maior.

A partir do fechamento do processo automatizado de compatibilização das palavras-chave, que é objetivo deste estudo, outros resultados poderão ser disponibilizados aos usuários do periódico científico eletrônico. O índice de termos é a principal visualização de resposta da VCPC *tools*. Esse índice liga-se ao artigo e, ainda, verifica-se a quantidade de artigos relacionados a esse termo. Internamente, o VCPC *tools* constrói um vocabulário controlado da revista com os novos termos

que não existem nos vocabulários controlados utilizados na compatibilização das palavras-chave. Lembrando que o referido vocabulário controlado é um complemento de termos dos vocabulários controlados, portanto é mantido o termo preferido no vocabulário da revista com a devida remissão aos vocabulários controlados - TBCI e TCI.

Ainda de maneira sintetizada por meio das facetas do TCI e das categorias do TBCI, a *VCPC tools* gera, para o usuário, o perfil da revista. Torna-se de extrema importância tanto para os autores, como para usuários e para os futuros autores que poderão submeter suas produções científicas.

Pôde-se comprovar com esta pesquisa que o controle de vocabulário das palavras-chave atribuídas aos artigos científicos eletrônicos da revista Informação & Informação é possível de maneira automática. Porém é necessário o tratamento de maneira manual quando se trata das semânticas das palavras-chave. Vale ressaltar, como já mencionado, que para futuras implantações na área da Ciência da Informação, a ferramenta pode se tornar cada vez mais autossuficiente por meio da aprendizagem de máquina, ou seja, o registro de termos correspondentes.

A aplicação do projeto pode ser adaptada em qualquer área, e aqui se situam a relevância e a amplitude a que esta pesquisa pode remeter, aí também reconhece de que se trata de uma proposição e, como tal, não é definitiva e pode evoluir. As considerações para futuro deste estudo e sua continuidade são inseridas no próximo item.

## 6.1 RECOMENDAÇÕES E SUGESTÕES PARA FUTUROS TRABALHOS

Diante da experiência com este estudo, um ponto importante deveria ser pensado: Propor a inserção da atribuição de palavras-chave utilizando os termos dos vocabulários controlados no ambiente de submissão de trabalhos no Sistema Eletrônico de Editoração de Revistas (SEER), o que permitiria que o autor, no momento de atribuir as palavras-chave a seu artigo, caso não as encontre nos vocabulários controlados, solicitasse a inserção do vocabulário controlado da revista com as devidas fundamentações.

Em consequência, no ambiente do editor, poderiam existir formas de validar este novo termo no vocabulário controlado da revista e, caso fosse necessário, o editor poderia melhorar o termo com as devidas adequações, mediante a utilização

de recursos terminológicos. Para a utilização desse ambiente, é importante construir instruções de manuseio do sistema e os procedimentos de identificação dos melhores termos para representação do conteúdo.

Outro desafio para continuação da pesquisa é a implantação da VCPC *tools* no Sistema Eletrônico de Editoração de Revistas (SEER) de maneira unificada por área, com isto, adotar-se-iam os mesmos vocabulários controlados e seria gerada uma aprendizagem de máquina por meio da diversidade de correspondência de termos. O índice de termos gerado por meio da VCPC *tools* pode ser uma alternativa entre várias. O importante é considerar que a convalidação, as integrações de sistemas e as avaliações de impacto jamais cessem de ser analisadas – as palavras-chave têm uma influência sobre tudo isso, muito maior do que se costuma pensar.

## REFERÊNCIAS

ALMEIDA, M.; CRIPPA, G. De bacon à internet: considerações sobre a organização do conhecimento e a constituição da ciência da informação. **Ponto de Acesso**, Salvador, v. 3, n. 2, ago. 2009. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/3284>>. Acesso em: 27 set. 2014.

AMERICAN NATIONAL STANDARDS INSTITUTE - ANSI; NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. **Z39:19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005.

AQUINO, I. S.; AQUINO, I. S. Análise sobre a forma da escrita de palavras-chave em artigos científicos na área de ciências agrárias publicados no período de 1999-2011. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 18, n. 37, p. 227-238, ago. 2013. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2013v18n37p227>>. Acesso em: 27 abr. 2015.

ARENAS, M. D. et al. Análisis de las palabras-clave de las publicaciones Fisioterapia y Revista Iberoamericana de Fisioterapia y Kinesiología. **Revista Iberoamericana de Fisioterapia e Kinesología**, Madrid, v. 8, n. 2, p. 59-69, 2005.

ARRUDA, S. M.; CHAGAS, J. **Glossário de biblioteconomia e ciências afins**: português - inglês. Florianópolis: Cidade Futura, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS - ABNT. **NBR 12676**: métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

\_\_\_\_\_. **NBR 6022**: informação e documentação: artigo em publicação periódica impressa: apresentação. Rio de Janeiro, 2003a.

\_\_\_\_\_. **NBR 6028**. Informação e documentação: resumo: apresentação. Rio de Janeiro: 2003b.

AUSTIN, D.; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesouros mo-nolíngues**. Tradução de Blanca Amaro de Meto. Brasília: IBICT, 1993.

BARROS, A.; LEHFELD, N. **Projeto de pesquisa**: propostas metodológicas. Petrópolis: Vozes, 1990.

BOCCATO, V. R. C. Linguagem documentária na representação e recuperação da informação pela perspectiva sociocognitiva em ciência da informação. In: BOCCATO, V. R. C.; GRACIOSO, L. S. (Org.). **Estudos de linguagem em ciência da informação**. Campinas: Alínea, 2011a. p. 9-34.

\_\_\_\_\_. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. **InCID**: Revista de Ciência da Informação e

Documentação, Ribeirão Preto, v. 2, n. 1, p. 165-192, jun. 2011b. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42340>>. Acesso em: 26 set. 2014.

BOCCATO, V. R. C.; TORQUETTI, M. C. Interoperabilidade entre Linguagens de Indexação como Recurso de Modelagem de Repertório Terminológico de Coordenadorias de Comunicação Social em Ambientes Universitários: uma proposta metodológica. **Informação & Informação**, [S.l.], v. 17, n. 3, p. 76-101, dez. 2012. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/10800>>. Acesso em: 05 Jul. 2015.

BORKO, H. Information science: what is it? **American Documentation**, Washington, v. 19, n. 1, p. 3, jan. 1968. Disponível em: <<http://pt.scribd.com/doc/533107/Borko-H-v-19-n-1-p-35-1968>>. Acesso em: 25 abr. 2015.

BRANDT, M.; MEDEIROS, M. B. B. Folksonomia: esquema de representação do conhecimento? **Transinformação**, Campinas, v. 22, n. 2, 2010.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENANCIB, 9., 2008. **Anais...** São Paulo: USP, 2008. Disponível em: <<http://www.ancib.org.br/media/dissertacao/1835.pdf>>. Acesso em: 21 out. 2014.

BREITMAN, K. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005.

BRITO, R. F.; GUEDES, M. G.; SHINTAKU, M. **Atribuição de identificadores digitais para publicações científicas: DOI para o SEER/OJS /**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2013.

BURKE, P. **Uma história social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Jorge Zahar, 2003.

CAMPOS, L. F. B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 12, n. 23, p. 16-45, 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16>>. Acesso em: 13 out. 2014.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=13788>>. Acesso em: 20 jan. 2015.

CASTRO, R. C. F. Impacto da Internet no fluxo da comunicação científica em saúde. **Revista de Saúde Pública**, São Paulo, v. 40, p. 57-63, ago. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-89102006000400009](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102006000400009)>. Acesso em: 8 out. 2014.

CERVANTES, B. M. N. **A construção de tesouros com a integração de procedimentos terminográficos**. 2009. 209 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2009.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR - CAPES. **Classificação da produção intelectual**. Disponível em: <<http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual>>. Acesso em: 19 set. 2014a.

\_\_\_\_\_. **Histórico**. Disponível em: <[http://www.periodicos.capes.gov.br/index.php?option=com\\_pcontent&view=pcontent&alias=historico&Itemid=100](http://www.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pcontent&alias=historico&Itemid=100)>. Acesso em: 19 set. 2014b.

CORTES, P. L. Considerações sobre a evolução da ciência e da comunicação científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (Org.). **Comunicação e produção científica**: contexto, indicadores e avaliação. São Paulo: Algelara, 2006. p. 33-56.

COSTA, M. U. P.; MOURA, M. A. A representação da informação em contextos de comunicação científica: a elaboração de resumos e palavras-chave pelo pesquisador- autor = una representación de la información en contextos de comunicación de la ciencia: la preparación de los resúmenes y las. **Informação & Informação**, Londrina, v. 18, n. 3, p. 45-67, out. 2013. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/13958>>. Acesso em: 8 jan. 2014.

CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 2. ed. Porto Alegre: Artmed, 2007.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

CURRÁS, E. **Ontologias, taxonomias e tesouros em teoria de sistemas e sistemática**. Tradução de Jaime Robredo. Brasília: Thesaururs, 2010.

DATE, C. J. **Introdução a sistemas de bancos de dados**. Rio de Janeiro: Elsevier, 2004.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2013.

DIAS, G. D.; CERVANTES, B. M. N. A organização temática da informação em periódicos científicos eletrônicos: atribuição de palavras-chave na Biblioteconomia e Ciência da Informação. In: ENCONTRO NACIONAL EM PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: FIOCRUZ, 2012. Disponível em: <<http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19249.pdf>>. Acesso em: 7 maio 2014.

ERCAN, G.; CICEKLI, I. Using lexical chains for keyword extraction. **Information Processing & Management**, Elmsford, v. 43, n. 6, p. 1705-1714, 2007.

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

FERNANDES, G. C. O objeto da Ciência da Informação. **Informare**: Caderno do Programa de Pós-Graduação em Ciência da Informação, Rio de Janeiro, v. 1, n. 1, p.

25-30, jan./jun.1995. Disponível em: <<http://www.uel.br/pos/ppgci/docs/fernandes.pdf>>. Acesso em: 20 abr. 2015.

FERRIDAY, P. Indexação de periódico. In: KNIGHT, G. N. **Treinamento em indexação, um curso da Society of Indexers**. Tradução de Maria Antonietta Requião Piedade. Rio de Janeiro: Fundação Getúlio Vargas, 1974.

FREITAS, M. H. Considerações acerca dos primeiros periódicos científicos brasileiros. **Ciência e Informação**, Brasília, v. 35, n. 3, dez. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652006000300006&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000300006&lng=pt&nrm=iso)>. Acesso em: 27 abr. 2015.

FUJITA, M. S. L. A política de indexação para representação e recuperação da informação. In: GIL LEIVA, I.; FUJITA, M. S. L. **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012.

\_\_\_\_\_. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Revista Brasileira de Educação Especial**, Marília, v. 10, n. 3, p. 257-272, set./dez. 2004.

FUJITA, M. S. L. Modelos de categorização para a construção de tesouros: metodologia de ensino. In: BOCCATO, V. R. C.; GRACIOSO, L. S. (Org.). **Estudos de linguagem em ciência da informação**. Campinas: Alínea, 2011. Cap. 2.

GIL LEIVA, I. **La automatización de la indexación de documentos**. [S. l.]: Ediciones Trea, 1999.

GIL LEIVA, I.; ALONSO ARROYO, A. La relación entre las palabras clave aportadas por los autores de artículos de revista y su indexación en las bases de datos ISOC, IME e ICYT. **Revista Española de Documentación Científica**, Madrid, v. 28, n. 1, 2005. Disponível em: <<http://redc.revistas.csic.es/index.php/redc/article/view/165>>. Acesso em: 9 set. 2014.

GIL LEIVA, I.; FUJITA, M. S. P. **Aspectos conceituais da indexação in política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GILLILAND, A. J. **Introduction to metadata: setting the stage**. 2008. Disponível em: <[http://www.getty.edu/research/publications/electronic\\_publications/intrometadata/setting.pdf](http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf)>. Acesso em: 9 set. 2014.

GOMES, H. E. Tendências da pesquisa em organização do conhecimento. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 2, n. 1, 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/viewArticle/16>>. Acesso em: 4 set. 2014.

GOMES, H. E.; GUSMÃO, H. R. **Guia prático para a elaboração de índice**. Niterói: APB-RJ, 1983.

GONÇALVES, A. L. Uso de resumos e palavras-chave em ciências sociais: uma avaliação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. 26, p. 1-15, 2008. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78/6636>>. Acesso em: 9 set. 2014.

GONÇALVES, A.; RAMOS, L. M. S. V. C.; CASTRO, R. C. F. Revistas científicas: características, funções e critérios de qualidade. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (Org.). **Comunicação e produção científica**: contexto, indicadores, avaliação. São Paulo: Angellar, 2006. p. 163-190.

HARTLEY, J.; KOSTOFF, R. N. How useful are 'key words' in scientific journals? **Journal of Information Science**, Cambridge, v. 29, p. 433, 2003.

HEUSER, C. A. **Projeto de banco de dados**. 5. ed. Porto Alegre: Sagra Luzzatto, 2004.

IBICT. **Sistema eletrônico de editoração de revistas**. Disponível em: <<http://seer.ibict.br/>>. Acesso em: 1 jun. 2015.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO. **25964-1**: information and documentation - thesauri and interoperability with other vocabularies -- Part 1: thesauri for information retrieval. Geneva, 2011.

\_\_\_\_\_. **25964-2**: information and documentation -- thesauri and interoperability with other vocabularies -- part 2: interoperability with other vocabularies. Geneva, 2013.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência e Informação**, Brasília, v. 35, n. 2, p. 91-102, maio/ago. 2006.

LANCASTER, F. W. **El control del vocabulario en la recuperación de información**. Tradução de Alejandro de La Cueva. Martín: Universitat de València, 1995.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática, Brasília: Brique de Lemos, 2004.

LARA, M. L. G. Glossário: termos e conceitos da área de comunicação e produção científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. **Comunicação e produção científica**: contexto, indicadores e avaliação. São Paulo: Algelara, 2006. p. 387-414.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília: Brique de Lemos, 2004.

LIMA, V. M. A.; BOCCATO, V. R. C. O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 14, n. 1, p. 131-151, abr. 2009. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362009000100010&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100010&lng=en&nrm=iso)>. Acesso em: 5 jun. 2015.

MARCONDES, C. H.; SAYAO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ciência e Informação**, Brasília, v. 31, n. 3, 2002. Disponível em:

<[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652002000300005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000300005&lng=en&nrm=iso)>. Acesso em: 20 set. 2014.

MARCONDES, C. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, jun. 2001. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/225>>. Acesso em: 29 abr. 2015.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2006.

McGARRY, K. **O contexto dinâmico da informação**. Brasília: Briquet de Lemos, 1999.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MIGUÉIS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório científico da Universidade de Coimbra. **Pontodeacesso**, Salvador, v. 7, n. 1, p. 116-131, abr. 2013.

MOREIRA, M. P. **Tesouro em ciência da informação**. Belo Horizonte: PUC Minas. Instituto de Informática; UFMG, [2006]. Disponível em: <http://icei.pucminas.br/ci/tci/index.php>. Acesso em: 9 mar. 2015.

MOREIRA, W. et al. Vocabulário controlado para a representação documentária em arquivos correntes da Unesp. In: SEMINÁRIO CIENTÍFICO ARQUIVOLOGIA E BIBLIOTECANOMIA, 4., 2015, Marília. **Anais...** Marília: Unesp, 2015. Disponível em: <<http://www.marilia.unesp.br/Home/Eventos/2015/seminariodearquivologiaebiblioteconomia/moreiraw.-fujita-m.s.l.-davanzo-l.-piovezan-l.b..pdf>>. Acesso em: 8 jun. 2015.

MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2011.

MORENO, F. P.; MÁRDERO ARELLANO, M. Á. Publicação científica em arquivos de acesso aberto. **Arquivística.net**, Rio de Janeiro, v. 1, n. 1, p. 76-86, jan./jun. 2005. Disponível em: <[www.arquivistica.net](http://www.arquivistica.net)>. Acesso em: 9 out. 2014.

MUELLER, S. P. M. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, B. S.; CEDÓN, B. V.; KREMER, J. M. (Org.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000a.

MUELLER, S. P. M. O periódico científico. In: CAMPELLO, B. S.; CEDÓN, B. V.; KREMER, J. M. (Org.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000b.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v. 1, n. 2, p. 37-45, dez. 1996. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1603>>. Acesso em: 26 set. 2014.

OASISBR. **Sobre o OASISBR**. Disponível em: <<http://oasisbr.ibict.br/sobre-o-oasisbr>>. Acesso em: 21 set. 2014.

OPEN ARCHIVES INITIATIVE. **Mission statement**. Disponível em: <<http://www.openarchives.org/OAI/OAI-organization.php>>. Acesso em: 20 set. 2014.

ORENGO, V. M.; BURIOL, L.; COELHO, A. A study on the use of Stemming for Monolingual Ad-Hoc Portuguese Information Retrieval. In: PETERS, C. et al. **Evaluation of multilingual and multi-modal information retrieval**. Berlin: Springer, 2007. p. 91-98.

ORENGO, V. M.; HUYCK, C. R. A stemming algorithm for the Portuguese language. In: INTERNATIONAL SYMPOSIUM ON STRING PROCESSING AND INFORMATION RETRIEVAL (SPIRE), 8., 2001, Chile. **Anais...** Laguna de San Raphael, Chile, 2001. p. 183-193.

PACKER, A. L. Os periódicos brasileiros e a comunicação da pesquisa nacional. **Revista USP**, São Paulo, n. 89, maio 2011. Disponível em: <[http://rusp.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-99892011000200004&lng=pt&nrm=iso](http://rusp.scielo.br/scielo.php?script=sci_arttext&pid=S0103-99892011000200004&lng=pt&nrm=iso)>. Acesso em: 30 abr. 2015.

PACKER, A. L.; MENECHINI, R. Visibilidade da produção científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (Org.). **Comunicação e produção científica: contexto, indicadores, avaliação**. São Paulo: Angellar, 2006. p. 235-260.

PHPMYADMIN. **About**. Disponível em: <[http://www.phpmyadmin.net/home\\_page/index.php](http://www.phpmyadmin.net/home_page/index.php)>. Acesso em: 9 jun. 2015.

PINHEIRO, L. V. R. Fontes ou recursos de informação: categorias e evolução conceitual. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 1, n. 1, 2006. Disponível em: <<http://revista.ibict.br/pt/cib/index.php/pt/cib/article/view/210/3>>. Acesso em: 16 jun. 2014.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: IBICT, 2014. Disponível em: <<http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/tesouro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510>>. Acesso em: 12 maio 2015.

PRESSMAN, R. **Engenharia de software [recurso eletrônico]: uma abordagem profissional**. Tradução de Ariovaldo Griesi. 7. ed. Porto Alegre: AMGH, 2011.

PRICE, D. S. **A ciência desde a Babilônia**. Tradução de Leônidas Hegenberg e Octanny S. da Mota. Belo Horizonte: Itatiaia, 1976.

RAMAKRISHNAN, R.; GEHRKE, J. **Sistemas de gerenciamento de banco de dados**. 3. ed. São Paulo: McGraw Hill Brasil, 2008.

REZENDE, D. A.; ABREU, A. F. **Tecnologia da informação aplicada a sistemas de informação empresariais**: o papel estratégico da informação e dos sistemas de informação nas empresas. 5. ed. São Paulo: Atlas, 2008.

ROCHA E SILVA, M. Periódicos científicos brasileiros: visibilidade e charme. **Revista da Escola de Enfermagem da USP**, São Paulo, v. 46, n. 1, 2012. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0080-62342012000100001&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342012000100001&lng=en&nrm=iso)>. Acesso em: 19 set. 2014.

RODRIGUES, A. F. **Como elaborar artigos**. São Paulo: Humanitas, 2011.

RODRIGUES, R. S. et al. A publicação de periódicos científicos digitais. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 16, n. 31, p. 1-4, maio 2011. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2011v16n31pi/17780>>. Acesso em: 27 jun. 2014.

ROQUETA, M. B. Sistemas de organização do conhecimento: uma tipologia atualizada. **Informação & Informação**, Londrina, v. 16, n. 2, p. 122-139, dez. 2011. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/9952>>. Acesso em: 6 jun. 2015.

SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação e informática. Campinas, SP: Átomo, 2012.

SANTOS, P. Paul Otlet: um pioneiro da organização das redes mundiais de tratamento e difusão da informação registrada. **Ciência e Informação**, Brasília, v. 36, n. 2, 2007. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652007000200006&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652007000200006&lng=en&nrm=iso)>. Acesso em: 27 set. 2014.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 15, n. 30, p. 1-131, 2010.

SILVA, J. F. M.; RAMOS, L. M. S. V. C.; NORONHA, D. P. Base de dados. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. **Comunicação e produção científica**: contexto, indicadores e avaliação. São Paulo: Algelara, 2006. p. 263-285.

SMIT, J. W.; KOBASHI, N. Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo Do Estado, Imprensa Oficial, 2003. (Projeto como fazer, 10).

SOUZA, J. L. A. Revistas eletrônicas com uso de software livre. **Datagramazero**: Revista de Ciência da Informação, Brasília, v. 11, n. 4, ago. 2010. Disponível em: <[http://www.dgz.org.br/ago10/Art\\_06.htm](http://www.dgz.org.br/ago10/Art_06.htm)>. Acesso em: 10 abr. 2014.

TRISKA, R.; CAFÉ, L. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. **Ciência e Informação**, Brasília, v. 30, n. 3, 2001. Disponível em:

<[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652001000300012&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300012&lng=en&nrm=iso)>. Acesso em: 20 set. 2014.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO. **Dicionário eletrônico de terminologia em ciência da informação (DeltCI)**. Vitória: UFES, 2007-2008.

VITAL, L. P. **Taxonomia como ferramenta para a representação do conhecimento em portais corporativos**. Recife: Universitária da UFPE, 2010.

VOGEL, M. J. M. **A noção de estrutura linguística e de processo de estruturação e sua influência no conceito e na elaboração de linguagens documentárias**. 2007. 124 f. (Dissertação de Mestrado) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo. 2007.

WEITZEL, S. R. Fluxo de comunicação científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. (org.). **Comunicação e produção científica: contexto, indicadores e avaliação**. São Paulo: Algelara, 2006. p. 81-114.

YIN, R. K. **Estudo de caso: planejamento e métodos**. Tradução de Ana Thorell. Porto Alegre: Bookman, 2010.

ZAVITOSKI, M. T. **Exploração do uso do tesauro como instrumento de recuperação da informação**. 2001. Dissertação (Mestrado em Biblioteconomia e Documentação) - Escola de Comunicação e Artes da Universidade de São Paulo, São Paulo, 2001.