



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

POLÍTICA DE INDEXAÇÃO AUTOMÁTICA PARA
DOCUMENTAÇÃO DIGITAL DO ACORDO DE BASILEIA II: O
CASO DO BANCO DO BRASIL

GEORGEA ALLIEVI

Projeto apresentado à Faculdade de
Ciência da Informação da
Universidade de Brasília como parte
dos requisitos para obtenção do
título de Mestre em Ciência da
Informação.

Professora Orientadora: Prof. Dra. Dulce Maria Baptista

Brasília, DF 2013



FOLHA DE APROVAÇÃO

Título: “Política de indexação automática para documentação digital do Acordo de Basileia II: O caso do Banco do Brasil.”

Autor (a): Georgea Allievi Frizon

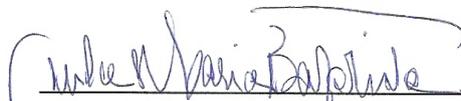
Área de concentração: Gestão da Informação

Linha de pesquisa: Organização da Informação

Dissertação submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação da Faculdade em Ciência da Informação da Universidade de Brasília como requisito parcial para obtenção do título de **Mestre** em Ciência da Informação.

Tese aprovada em: 18 de junho de 2012.

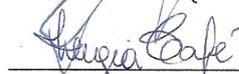
Aprovado por:



Profª Dr.ª Dulce Maria Baptista
Presidente (UnB/PPGCINF)



Prof. Dr. Rogério Henrique de Araújo Júnior
Membro Interno (UnB/PPGCINF)



Profª Dr.ª Ligia Maria Arruda Café
Membro Externo (UFSC)



Prof. Dr. Renato Tarciso Barbosa de Sousa
Suplente (UnB/PPGCINF)

"O saber a gente aprende com os mestres e os livros. A sabedoria, se aprende com a vida e com os humildes"

(Cora Coralina)

Agradecimentos

A meu avô Umberto Allievi (*in memoriam*), pelo exemplo de amor ao trabalho.

À minha mãe Ivone, que semeou em mim o gosto pelos estudos e pelo conhecimento.

Ao Sérgio, meu companheiro amado, por seu apoio inestimável e por estimular a superação dos meus limites.

À minha orientadora Professora Dulce Maria Baptista, por me conduzir neste processo de construção do conhecimento e pelo prazer das nossas longas conversas.

Ao Professor Rogério Araújo Jr e a Professora Lígia Café, por terem participado da banca e pelas importantes contribuições para esta pesquisa.

POLÍTICA DE INDEXAÇÃO AUTOMÁTICA PARA DOCUMENTAÇÃO DIGITAL DO ACORDO DE BASILEIA II: O CASO DO BANCO DO BRASIL

RESUMO

A pesquisa é um estudo, no âmbito da Ciência da Informação, sobre a representação da informação bancária. O objetivo é identificar os aspectos relacionados com a indexação de documentos digitais de conteúdo textual, que devem ser considerados na elaboração de uma política de indexação. A metodologia utilizada foi baseada em estudo de caso. Foi proposta uma política de indexação voltada à representação de documentos digitais produzidos na instituição financeira Banco do Brasil e relacionados ao Acordo de Basileia II.

Palavras chave: representação da informação; indexação automática; política de indexação.

AUTOMATIC INDEXING POLICY FOR DIGITAL DOCUMENTATION OF BASEL II: CASE OF BANCO DO BRASIL

ABSTRACT

Presents research a study, in scope of Information Science, about representation of bank information. The goal is to identify issues related to the indexing digital document unstructured textual content, which should be considered in developing a policy of indexing. The methodology was based on case study. Proposed an indexing policy focused on representation of digital documents produced in the financial institution Banco do Brasil and related to Basel II.

Keywords: representation of information; automatic indexing; policy of indexing.

Lista de Figuras

Figura 1 – Os Pilares do Acordo de Basileia II	2
Figura 2 - Contexto informacional do Acordo de Basileia II.....	3
Figura 3 - Ciclo vital da Informação	9
Figura 4 - Processo de indexação.....	17
Figura 5 – Métricas de avaliação da recuperação de informação	19
Figura 6 - Diagrama representativo do processo de indexação automática.....	23
Figura 7 - Etapas do processo de indexação automática do TCDF	28
Figura 8 - Fluxo do sistema recuperação da informação.....	32
Figura 9 – Critérios para uma política de indexação	40
Figura 10 – Itens de uma política de indexação	42
Figura 11 - Acordo de Basileia II	46
Figura 12 – Elementos que compõe as diretrizes da política de indexação	58
Figura 13 - Estrutura organizacional do Comitê de Gestor de Informações	59
Figura 14 - Recursos TIC para suportar o processo de indexação automática	63
Figura 15 - Processo de indexação	66

Lista de Tabelas

Tabela 1 - Noção de documento	13
Tabela 2 – Correntes teóricas da indexação	16
Tabela 3 - Exemplo da lista de termos alterados no pós-processamento	29
Tabela 4 - Política de indexação - papéis e responsabilidades.....	61

Lista de Abreviaturas

BACEN	Banco Central do Brasil
BIS	<i>Bank for International Settlements</i>
BB	Banco do Brasil
CDT	Centro de Apoio ao Desenvolvimento Tecnológico
CI	Ciência da Informação
CMN	Conselho Monetário Nacional
DIRIS	Diretoria de Gestão de Riscos
OC	Organização do Conhecimento
OI	Organização da Informação
POS	Part of Speech
SRI	Sistema de recuperação da informação
TCDF	Tribunal de Contas do Distrito Federal
TI	Tecnologia da informação
TIC	Tecnologia da informação e da comunicação
URI	Unidade de Relações com Investidores
UnB	Universidade de Brasília

Sumário

Lista de Figuras	vi
Lista de Tabelas	vii
Lista de Abreviaturas	viii
1 Introdução	1
1.1 Justificativa.....	4
1.2 Problema da pesquisa.....	6
1.3 Objetivos	6
1.3.1 Objetivo Geral.....	6
1.3.2 Objetivos Específicos	6
2 Referencial Teórico	7
2.1 Organização e Representação da informação	7
2.2 Documento como objeto informacional	12
2.3 Análise da Informação.....	13
2.4 Indexação como processo para representação da informação	16
2.4.1 Tópicos conceituais sobre indexação	16
2.4.2 Indexação automática	21
2.4.3 Indexação automática aplicada: Projeto TCDF	27
2.5 Sistemas de Recuperação da Informação.....	30
2.6 Gestão de documentos digitais	35
2.7 Elementos da Política de indexação.....	37
2.8 O Acordo de Basileia II.....	43
2.8.1 O Banco do Brasil e o Acordo de Basileia II.....	47
2.10 Conclusão do referencial teórico	50
3 Procedimentos metodológicos	52
3.1 Definição do escopo da pesquisa.....	52
3.2 Passos metodológicos.....	52
3.3 Coleta de dados	53
4 Descrição e análise dos dados	54
5 Proposta de política de indexação para documentos digitais no Banco do Brasil	57
5.1 Diretrizes para política de indexação.....	58
5.2 Conclusão	66
6 Considerações finais	67
7 Referências bibliográficas	69
APÊNDICE	80

1 Introdução

A transição de uma sociedade industrial, baseada na mecanização em larga escala dos processos produtivos, para sociedade da informação, baseada no uso intensivo da informação, acarretou, entre outros aspectos, a explosão informacional. Esse fenômeno se manifestou a partir da segunda metade do século XX, tornando complexo o tratamento do volume cada vez maior de documentos, motivando a adoção de tecnologias para apoiar a gestão do ciclo da informação.

Um exemplo da complexidade do tratamento da informação pode ser verificado nas instituições financeiras. Não por acaso, esse tipo de instituição foi pioneira no tratamento informatizado da informação, que embora não tenha substituído inteiramente o tratamento manual, agilizou, em grande medida, todos os processos relacionados à informação, reduzindo os custos operacionais e minimizando os elementos comprometedores da eficiência operacional da gestão bancária.

O Banco do Brasil é uma instituição financeira brasileira, estatal, constituída na forma de sociedade de economia mista, com participação da União brasileira em 68,7% das ações. Foi o primeiro banco a operar no país e, hoje, é a maior instituição financeira do Brasil, com reconhecida função social e com competência para lidar com os negócios financeiros, estando presente na maioria dos municípios do país.

No Banco do Brasil, são gerados diariamente cerca de 1.000.000 (um milhão) de documentos digitais, referentes apenas à atividade de cadastro de cliente, o que corresponde a apenas uma parcela do volume total dos documentos produzidos nos processos de negócio. (BANCO DO BRASIL, 2012).

Muito embora a instituição faça uso de documentos em suporte papel, vinculados principalmente aos negócios com clientes, fornecedores ou por exigência legal, a grande massa documental produzida atualmente corresponde a documentos em suporte digital.

Os documentos em suporte digital podem ser duas origens: digitalizados ou nativamente digitais. Conforme a definição do CONARQ (2009), o documento digital é aquele codificado em dígitos binários, produzido, processado e armazenado por um sistema computacional. O documento digitalizado é aquele que resulta do processo de conversão de um documento em qualquer suporte ou formato para o formato digital, por meio de dispositivo apropriado, por exemplo, um escâner.

Embora a informatização dos processos referentes à produção, ao armazenamento, a organização, a representação e a recuperação de documentos tenha contribuído para a diminuição dos custos operacionais, para o aumento da carteira de clientes e para ampliação do portfólio de produtos e serviços do Banco do Brasil, ainda existem problemas por resolver, notadamente relacionados com a representação e recuperação da informação contida nos documentos em suporte digital. Esses problemas são críticos porque podem, em determinados casos, significar prejuízo financeiro, risco a imagem da instituição ou perda de mercado para a concorrência.

Na perspectiva da informação produzida em suporte digital no Banco do Brasil, optou-se por delimitar o escopo desta pesquisa ao estudo dos documentos relacionados à regulamentação do Acordo de Basileia II – Pilar III, acerca do domínio das informações para o fortalecimento da supervisão bancária e o estímulo a maior transparência na divulgação das informações ao mercado.

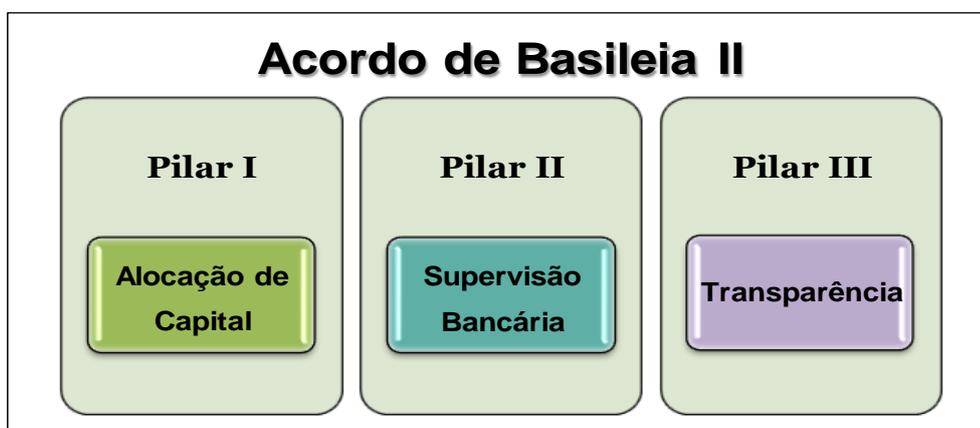


Figura 1 – Os Pilares do Acordo de Basileia II

Fonte: Elaboração própria

O caráter regulatório do Acordo de Basileia II (BACEN, 2013), requer que as instituições financeiras produzam registros documentais que possibilitem a avaliação da real condição financeira da instituição, proporcionando maior transparência e incentivando a implementação de sistemas mais evoluídos para o controle dos riscos presentes nas operações financeiras. Determina também a divulgação destas informações para subsidiar os processos de auditoria interna e externa, bem como para subsidiar os processos decisórios dos seus investidores.

A Figura 2 ilustra a disseminação do Acordo de Basileia II. As diretrizes são estabelecidas pelo *Bank for International Settlements* (BIS) que são internalizadas nos países por seus respectivos Bancos Centrais. No caso do Brasil, o Banco Central (BACEN) normatiza como o Acordo de Basileia II deve ser cumprido pelas instituições financeiras nacionais. O Banco do Brasil como uma dessas instituições, adota modelos internos de controle dos riscos de mercado, crédito e operacional e produz as evidências documentais sobre sua exposição a riscos para divulgação aos cliente internos e externos.



Figura 2 - Contexto informacional do Acordo de Basileia II

Fonte: Elaboração própria

No Banco do Brasil grande parte destas informações está disponível em documentos digitais do tipo: relatórios contábeis e financeiros, projeções financeiras, metodologias de cálculo, instruções normativas, controles internos, resultado de avaliações, notas técnicas, políticas e estratégias de gestão de riscos, manuais de procedimentos operacionais, entre outros.

Observou-se, no contexto da documentação digital, a ausência de uma estratégia corporativa que oriente e padronize a organização e representação da informação.

A partir da caracterização desse problema, a presente pesquisa propõe a elaboração de diretrizes para uma política de indexação automática voltada à representação de documentos digitais relacionados ao Acordo de Basiléia II – Pilar III, no domínio do Banco do Brasil e da sua comunidade de usuários.

A exploração do referencial teórico fundamenta-se na literatura de Ciência da Informação, por meio de levantamento bibliográfico e seleção de trabalhos cientificamente relevantes que viessem a apoiar e expandir o entendimento sobre o problema da pesquisa. Em linhas gerais, fazem parte do escopo do levantamento bibliográfico assuntos como: organização e representação da informação, extração de informação, análise da informação, linguagens documentárias, gestão documental, indexação automática. Adicionalmente foram abordados assuntos complementares a esta pesquisa como: sistemas de informação bancária e Acordo Basileia II.

1.1 Justificativa

A criação de documentos nos ambientes digitais, apoiada pelas tecnologias da informação e da comunicação (TIC), tem possibilitado acesso à informação através de diferentes fontes de dados. Em decorrência dessa realidade, surgem novos problemas, dentre os quais, os relacionados à ineficiência da representação da informação. Circunstâncias essas, que exigem dos usuários grande esforço de análise das informações recuperadas pelo sistema informatizado, até que consigam localizar a informação que atenda a sua necessidade, caracterizando assim a baixa precisão nos resultados das buscas.

Baeza-Yates e Ribeiro-Neto (2011) ressaltam que os objetos informacionais recuperados pelos sistemas de informação nem sempre são suficientes para satisfazer a necessidade do usuário, ou seja, uma consulta pode conduzir a um resultado equivocado das intenções originais do usuário.

Esta situação pode ser causada por inconsistências na organização e representação da informação.

Lancaster (2004) considera que a solução para os problemas na recuperação da informação consiste em procurar aproximar as necessidades de informação da representação do conteúdo do documento, porque quanto melhor a representação dos documentos mais eficazes serão seus pontos de acesso.

No âmbito da Ciência da Informação, a linha de pesquisa Organização e Representação da Informação estuda os processos relacionados ao tratamento da informação, propondo instrumentos de mediação entre os objetos informacionais e os usuários dos sistemas de informação.

Nessa perspectiva, a indexação automática é um processo de representação que possibilita melhorar a recuperação de informação. Segundo Araújo JR. (2007), a indexação é um processo de representação do conteúdo dos documentos e um elemento fundamental para o processo de busca e recuperação da informação e, será considerada satisfatória se representar com fidedignidade o conteúdo dos documentos.

No caso das instituições, financeiras recuperação da informação é um fator determinante da eficiência dos seus processos. As consultas dos usuários consistem, normalmente, em pesquisa por informações para resolver questões específicas e em curto prazo, como por exemplo, a consulta destinada a atender as necessidades informacionais de um determinado cliente, ou a consulta solicitada por gestores para apoiar a tomada de decisões. Observa-se que a parte da documentação bancária referente, entre outros assuntos, às atividades administrativas e regulatórias, é constituída de documentos digitais, escritos em linguagem natural e que apresentam baixa especificidade na sua localização e recuperação devido a falta de tratamento para organização e representação do seu conteúdo.

De acordo com Lancaster (2004), as organizações deveriam produzir suas próprias diretrizes de indexação, segundo o seu contexto de atuação e o público de interesse.

Essa pesquisa é motivada pela necessidade de definição de diretrizes que possam embasar a criação futura de uma política de indexação automática, voltada para representação da informação dos documentos digitais em formato textual, produzidos nos processos relacionados ao Acordo Basileia II no seu Pilar III, com vistas à representação da informação regulatória, no contexto do Banco do Brasil.

1.2 Problema da pesquisa

Esta pesquisa pretende responder a seguinte questão:

Quais os requisitos para indexação automática de documentos digitais em formato textual devem ser considerados na elaboração de uma política de indexação?

1.3 Objetivos

1.3.1 Objetivo Geral

Elaborar diretrizes para uma política de indexação automática para documentos digitais em formato textual e de natureza bancária que atendem as exigências do Acordo de Basileia II no Banco do Brasil.

1.3.2 Objetivos Específicos

- Levantar na literatura os elementos essenciais para uma política de indexação;
- Verificar como ocorre a indexação de documentos digitais produzidos no Banco do Brasil para atender as exigências do Acordo de Basileia II – Pilar III;
- Levantar no Acordo de Basileia II – Pilar III os requisitos que orientam a elaboração de uma política de indexação.

2 Referencial Teórico

A Ciência da Informação fornece o aporte teórico e metodológico adequado à organização e representação da informação, que abrange, dentre outros, o processo de indexação.

Explorou-se a literatura contida em artigos, periódicos, teses, dissertações, capítulos de livros e *web sites* referentes aos temas: organização e representação da informação, indexação automática, gestão da informação, gestão de documentos, política de indexação e Acordo de Basileia II.

2.1 Organização e Representação da informação

A Ciência da Informação é segundo Le Coadic (2004), uma ciência social que estuda as propriedades gerais da informação: a natureza, a gênese e os efeitos, e analisa os seus processos de construção, comunicação e uso.

Saracevic (1996) enfatiza a Ciência da Informação como o campo dedicado ao estudo dos problemas da comunicação do conhecimento no contexto social, institucional ou individual do uso e das necessidades de informação.

Na concepção de Novellino (1996), a Ciência da Informação estuda os fenômenos subjacentes à informação, com a finalidade de criação de instrumentos e o estabelecimento de metodologias que viabilizem a transferência das informações, que podem ser aplicados, conforme Tarapanoff (2008), nos contextos organizacionais, sociais e individuais, buscando soluções para a disseminação da informação e do conhecimento.

A clássica definição de Borko (1968), de que a Ciência da Informação se ocupa da investigação das propriedades e do comportamento da informação, do fluxo informacional e dos meios de processamento para o acesso e uso, atribuindo a característica de uma ciência interdisciplinar, porque deriva de, e se relaciona com vários outros campos do conhecimento.

Segundo Le Coadic (2004), o caráter interdisciplinar da Ciência da Informação ocorre pelas suas intersecções com outras disciplinas que também concentram seus estudos na informação e seus componentes. Tarapanoff (2006) por sua vez, aponta que as duas relações interdisciplinares mais fortes da Ciência da Informação são com a Biblioteconomia e com a Ciência da Computação.

A forte interação da Ciência da Informação com a Ciência da Computação é reconhecida também por Saracevic (1996), quando declara que dentre as características da existência e evolução da Ciência da Informação está à ligação indissociável com a Ciência da Computação. Apesar de tratarem do mesmo objeto de estudo – a informação – a Ciência da Computação estuda os algoritmos computacionais que transformam as informações, enquanto a Ciência da Informação trata da natureza da informação, sua comunicação e uso.

Para Le Coadic (2004), a Ciência da Computação, tem por objeto de estudo a concepção dos produtos, sistemas e serviços que apoiam a Ciência da Informação na construção, comunicação, armazenamento e uso da informação.

A integração dos recursos tecnológicos com as metodologias de tratamento e organização da informação ocorre por meio dos sistemas de informação, que segundo Robredo (2003), tratam os fenômenos relacionados ao ciclo da informação nos processos de captura, armazenamento, processamento, distribuição e uso.

Segundo Floridi (2002), no ciclo da informação é formado pela: gênese, organização, recuperação e comunicação e pelos estágios que seguem desde a criação até a sua utilização e possível descarte. Pode-se observar na Figura 3 que a atividade de indexação constitui um dos estágios da organização da informação.



Figura 3 - Ciclo vital da Informação

Fonte: Floridi (2002)

Para Bräscher e Café (2008) a organização da informação trata de descrever as características físicas e de conteúdo dos objetos informacionais e produz, como resultado, a representação da informação. No ciclo de vida da informação os processos de organização e representação tratam da explicitação do conteúdo dos documentos para atender as necessidades de informação dos usuários e aperfeiçoar os resultados das estratégias de busca.

Segundo Silva e Fujita (2004), a representação da informação não altera o documento, porém cria novas informações a respeito dele, a partir de metodologias formalizadas de reconhecimento temático e descritivo do seu conteúdo.

O conjunto das representações temáticas e descritivas do documento formam os pontos de acesso, que são armazenados no sistema de informação independentes do documento original. Segundo Lancaster (2004), é mediante esses pontos de acesso que o documento será localizado e recuperado, durante a busca num sistema de informação.

Observa-se que não há um consenso entre os pesquisadores da Ciência da Informação acerca da definição do processo organização e representação da informação e, muitas vezes este se confunde com a definição de organização e representação do conhecimento.

Alvarenga (2003) não faz distinção entre a organização do conhecimento e a organização da informação: "A organização da informação ou conhecimento [...] compreenderia um processo de representação que se destina prioritariamente à recuperação eficaz por parte dos usuários." A autora subdivide o processo de representação em dois tipos: representação primária, que se encontra na instância do processo cognitivo de expressão dos pensamentos em um suporte documental, e representação secundária, quando são identificados os conceitos dos registros primários e escolhidos os pontos de acesso que garantem a representação do conteúdo do documento para futura recuperação.

Na definição de Guimarães (2003) organização e representação da informação ocorrem durante a análise documental, que consiste de um conjunto de procedimentos de natureza analítico sintética, visando à aplicação de critérios de natureza descritiva (física) ou temática (de conteúdo) aos objetos informacionais, para permitir que sejam recuperados pelos usuários.

Segundo Galvão (2003), a representação da informação tanto explícita os dados necessários para identificação dos autores e do contexto de produção da informação (local e data), quanto explícita, de forma concisa, os conceitos informacionais presentes no documento, por meio de índices e resumos.

Dahlberg (2006) diferencia os conceitos de Organização do Conhecimento (OC) e Organização da Informação (OI). A OC significa a construção de sistemas conceituais e a OI é o relacionamento de unidades desse sistema conceitual com objetos de informação.

Bräsher e Café propõem uma delimitação conceitual para organização e representação da informação, como segue:

A organização da informação é, portanto, um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. (BRÄSHER; CAFÉ 2008)

Segundo Bräsher e Café (2008), a organização da informação é o processo, e a representação da informação é o produto documentário resultante desse processo e inferem que a organização da informação se aplica a objetos informacionais (documentos) individuais e a organização do conhecimento se aplica às unidades do pensamento (conceitos).

Os processos de organização e representação informação são estruturados, conforme afirma Taylor e Joudrey (2009), em torno das seguintes atividades:

- reconhecimento de todos os tipos de objetos de informação e como eles são disponibilizados;
- acesso às diferentes fontes de informações;
- percepção da natureza do trabalho contido nos objetos de informação;
- reunião sistemática dos objetos de informação em bancos de dados, arquivos, diretórios, internet, etc.;
- produção de listas dos objetos de informação, de acordo com regras padrão de citação;
- criação dos metadados que facilitem o acesso ao objetos informacional;
- provimento dos meios para a localização de cada objetos informacional, sua localização física ou seu endereço eletrônico.

Acrescenta-se também a afirmação de Araújo Jr. (2007), de que a representação da informação se tornará mais efetiva se houver um conhecimento prévio, por parte dos analistas da informação, quanto aos requisitos informacionais dos usuários, o que influenciará positivamente na precisão da resposta na recuperação da informação.

No que diz respeito à organização e representação da informação, com o objetivo de explicitação da descrição física e do conteúdo dos objetos informacionais, Novellino (1996) afirma que envolve duas etapas: a análise do assunto e sua transcrição para uma expressão linguística em substituição a

uma entidade linguística longa e complexa (o texto do documento) por sua descrição abreviada que enfatiza as informações essenciais para recuperação.

Como não há uma forma única para compreensão da realidade, a organização e representação da informação estão implícitas no domínio do sistema de informação para estruturar do conhecimento a partir das propriedades dos documentos. Organiza-se para recuperar, ou seja, o resultado das buscas dos usuários depende dos procedimentos de organização e representação empregados nos sistemas de informação.

Os documentos tornam explícito o conhecimento organizacional e possibilitam o compartilhamento das informações que se concentram em indivíduos ou unidades de trabalho.

2.2 Documento como objeto informacional

Para Le Coadic (2004), o documento é o termo genérico utilizado para designar os objetos portadores de informação, onde o conhecimento é registrado por meio de um sistema de signos gráficos, icônicos, sonoros ou visuais.

A compreensão de documento como informação registrada é, segundo Miranda e Simeão (2002), o objeto de estudo da Ciência da Informação. Esses estudos centram-se na análise da massa documental, seja no suporte convencional ou digital, para compreender sua natureza, ciclo e uso da informação.

Para Frohmann (2008) *apud* Lara (2010), o documento é o meio de materialização da informação, sendo que essa característica de materialidade é importante para o entendimento dos aspectos públicos e sociais da informação.

Lara (2010) analisou os discursos de influentes autores da Ciência da Informação para identificar quais são as concepções acerca de documento e informação, conforme apresentado na Tabela 1. O levantamento realizado pela autora possibilita concluir que documento e informação não podem ser definidos de modo isolado, mas um relativamente ao outro.

OTLET	BUCKLAND	BELKIN, BROOKES, WERSIG, INGWERSEN	CAPURRO, HJØRLAND
Suporte de dados; Receptáculo de ideias; Meio de transmissão do pensamento	Não se refere a documento, mas em informação.	Não se refere a documento, mas em informação.	Não se refere a documento, mas em informação.

Tabela 1 - Noção de documento

Fonte: Adaptado de Lara (2010)

Para Alvarenga (2003), o meio digital constitui-se em um espaço sem precedentes para produção, representação, recuperação de documentos textuais, sonoros, iconográficos, multimídia e, passou a requerer novos elementos facilitadores de sua representação e recuperação. Isso resultou em uma nova estrutura informacional permitindo que o leitor/receptor utilize um fluxo multidirecional, *on-line* e sem mediador para acesso a várias fontes de informação disponíveis, selecionando-as segundo seu critério de prioridade e relevância.

A partir do entendimento de que o documento é o meio de materialização da informação, para sua representação é necessário a realização das atividades de análise da informação e tradução do conteúdo para linguagens documentárias.

2.3 Análise da Informação

A descrição condensada do documento recebe, segundo Bräscher e Café (2008), várias denominações: análise documentária, análise temática, análise de assunto, descrição do conteúdo, tratamento temático da informação e análise da informação.

Segundo Baptista, Araújo Jr. e Carlan (2010), a análise da informação compreende um conjunto de práticas que servem à representação condensada do conteúdo de um objeto informacional, a partir da síntese e tratamento da informação, atendendo com precisão as necessidades dos usuários.

Baptista, Araújo Jr. e Carlan (2010) observam que a análise da informação apresenta similaridade com o ciclo da informação, por também contemplar as etapas de descrição, representação e organização, armazenamento e recuperação. No caso da análise da informação, os autores atribuem outra denominação ao conjunto de etapas a que é submetido o documento: ações preliminares, ações de descrição, ações de representação e ações complementares.

Segundo Guimarães (2003), o processo de análise documentária, ou seja, análise da informação ocorre em dois momentos: na etapa de análise e na etapa de síntese, quando são selecionados os conceitos de maior expressividade para condensação (o resumo) e para conversão dos termos originalmente expressos em linguagem natural para termos padronizados típicos de linguagem documentária (a indexação).

Silva e Fujita (2004) afirmam que o objetivo da análise da informação é o tratamento temático dos documentos que resulta na indexação, classificação e elaboração de resumos, considerando as diferentes finalidades de recuperação da informação.

A transformação no ambiente organizacional do conhecimento tácito em explícito requer, segundo Baptista, Araújo Jr. e Carlan (2010), esforços para análise da informação. Para este fim, a análise da informação pode ser abordada sob três perspectivas:

- **Dimensão conceitual:** aborda os conceitos e definições, as hierarquias associativas entre os termos e conceitos, o controle terminológico e a tipologia da informação, com o objetivo de compatibilizar a linguagem natural, em que é expressa a necessidade informacional do usuário, com a linguagem documentária dos sistemas de informação.
- **Dimensão estratégica:** formas de implementação de políticas e estratégias que determinam, por exemplo, os documentos a serem incluídos nos repositórios, o público de usuários atendido e as normas para representação física e descritiva dos documentos.

- **Dimensão operacional:** procedimentos que transformam o documento em informação disponível, fazendo uso das linguagens documentárias para viabilizar a representação descritiva, e a representação condensada do conteúdo temático. Os produtos são a classificação, a indexação e os resumos.

Sendo esta pesquisa motivada pela necessidade de diretrizes para uma política de indexação aplicável aos documentos digitais produzidos nos processos da instituição financeira, entende-se que a análise da informação será abordada na dimensão estratégica. Para garantir conforme Baptista, Araújo Jr. e Carlan (2010) a qualidade da representação e, por conseguinte da recuperação da informação pela comunidade de usuários.

No que se refere à representação do conteúdo informacional, Guimarães (2003) evidencia a necessidade da análise da informação, sendo que o contexto é o delimitador do sentido dos termos, da seleção dos conceitos e das relações semânticas. Uma vez que nas organizações um mesmo termo poderá ter significados diferentes para as diferentes comunidades de usuários, a representação deve estar embasada no contexto de uso e não apenas pelo contexto linguístico.

Em relação aos documentos inseridos num sistema de informação, são citadas as considerações de Araújo Jr. (2007) *apud* Guimarães (2003), segundo o qual, os descritores atribuídos ou extraídos dos documentos serão mais fidedignos ao representarem tematicamente o seu conteúdo se forem observadas as três dimensões: a do documento, a do usuário e a da organização.

Dentre os procedimentos metodológicos para análise da informação difundidos na Ciência da Informação, destaca-se no âmbito desta pesquisa, o processo de indexação, para identificação dos termos representativos do conteúdo e intermediação entre as buscas dos usuários e os documentos armazenados nos sistemas de informação. Autores como Lancaster (2004) e Fujita (2003) afirmam que a indexação é a etapa mais importante da representação da informação.

2.4 Indexação como processo para representação da informação

Gil Leiva (2008) afirma que a indexação é um processo regulador de valor e de eficiência em um sistema de informação. Qualquer objeto informacional é passível de ser indexado, ou seja, reduzido a representações conceituais que facilitem o seu armazenamento e recuperação.

A indexação eficiente deve considerar a representação da temática do documento e também as necessidades de informação da comunidade de usuários.

2.4.1 Tópicos conceituais sobre indexação

Na literatura da Ciência da Informação, é possível identificar as três principais correntes teóricas que norteiam as pesquisas sobre indexação. A tabela abaixo sintetiza essas correntes:

<i>Corrente francesa</i>	<i>Corrente espanhola</i>	<i>Corrente inglesa</i>
A indexação está inserida na análise documentária, como produto resultante da fase de representação.	Análise documentária está dividida em: análise descritiva ou bibliográfica (física) e do conteúdo.	Análise documentária e indexação são processos idênticos que compreendem as etapas: análise do conteúdo e tradução em linguagem de indexação.

Tabela 2 – Correntes teóricas da indexação

Fonte: Silva e Fujita (2004)

Na definição de Lancaster (2004), o processo de indexação consiste em decidir do que trata um objeto informacional e atribuir-lhe um rótulo que represente esta decisão, independente se o objetivo seja organizar documentos em estantes, ou registros em catálogos, índices impressos ou bases de dados eletrônicas.

Segundo Gil Leiva (2008), a indexação é um processo de tratamento técnico documental, intelectual ou automático, para obtenção de um conjunto

de unidades conceituais que representem integralmente um objeto ou uma necessidade informacional, conforme ilustrado na figura a seguir:

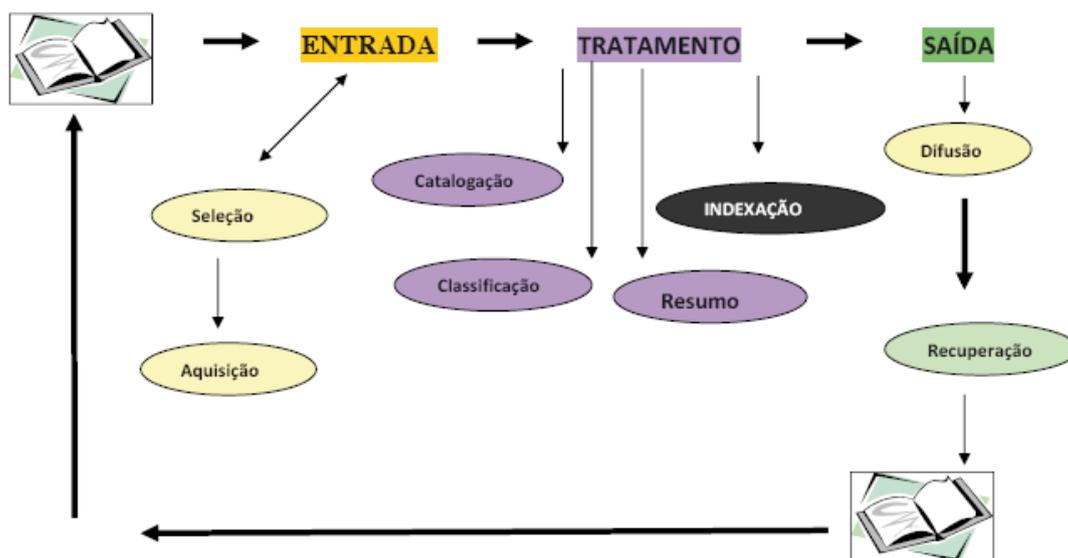


Figura 4 - Processo de indexação

Fonte: Gil Leiva (2012)

Segundo Lancaster (2004), o processo de indexação constitui-se de duas etapas: análise conceitual e tradução. Durante a etapa de análise conceitual ocorre a identificação e segmentação dos conceitos e, na etapa de tradução, os conceitos e termos selecionados são representados de maneira padronizada, por meio das linguagens documentárias.

Encontra-se na literatura da área outras propostas para as etapas do processo de indexação. Entretanto Souza e Alvarenga (2004) argumentam que, independente do número de etapas, a indexação visa identificar o assunto de um documento e representá-lo por meio das linguagens documentárias, entre outras, com o intuito de intermediação entre o documento e o usuário para recuperação da informação.

Os termos extraídos do documento sofrem, na maioria das vezes, uma transformação consciente, isto é, uma normalização antes do seu armazenamento no sistema de informação. A normalização favorece a

eficiência e a coerência nos sistemas de informação, evitando ter num mesmo sistema diferentes formas para denominar a mesma coisa. (GIL LEIVA, 2008)

Para atender a essa finalidade, Lancaster (2004) defende o uso de vocabulários controlados, que são essencialmente uma lista de termos autorizados e organizados em forma de estrutura semântica, destinados ao controle de sinônimos, a diferenciar homógrafos, a reunir ou ligar termos que possuem entre si uma relação hierárquica ou associativa.

Gil Leiva (2008) apresenta as diferentes abordagens para realização da indexação:

- **Indexação centrada no documento:** representação fiel do documento desconsiderando o contexto e as necessidades de informação dos usuários;
- **Indexação centrada no usuário:** os conceitos extraídos do documento são convertidos em termos controlados por uma linguagem documentária, levando em conta os conhecimentos dos usuários e suas necessidades de informação;
- **Indexação centrada no domínio:** se fundamenta no conhecimento profundo da organização sobre os aspectos da análise do domínio, das necessidades dos usuários, papéis desempenhados ou adotados pelos profissionais da informação e a análise do documento.

Lancaster (2004) propõe que a indexação pode ser realizada em níveis, conforme a quantidade de termos selecionados:

- **Indexação seletiva:** emprega uma quantidade menor de termos a fim de abranger somente o conteúdo temático principal do documento;
- **Indexação exaustiva:** emprega termos suficientes para abranger todo o conteúdo temático do documento, proporcionando melhor indicação do assunto de que trata o documento.

Para Fujita (2003), a estratégia de indexação adequada será aquela que produzir a correspondência mais precisa dos conceitos relacionados ao conteúdo do documento com os termos consultados pelos usuários. Neste sentido, caberá à organização, na qual o sistema de informação está inserido, decidir se a atribuição dos descritores será realizada de maneira mais seletiva ou se considerará um nível mais exaustivo.

Lima e Alvares (2012) afirmam que a qualidade da recuperação da informação depende dos procedimentos utilizados no processo de organização e representação da informação. A qualidade da indexação está relacionada os princípios de exaustividade e especificidade, que influenciam diretamente os indicadores de recuperabilidade dos SRIs, a revocação e a precisão.

A Figura 5 ilustra a relação entre as métricas de indexação e de avaliação da recuperação de informação:

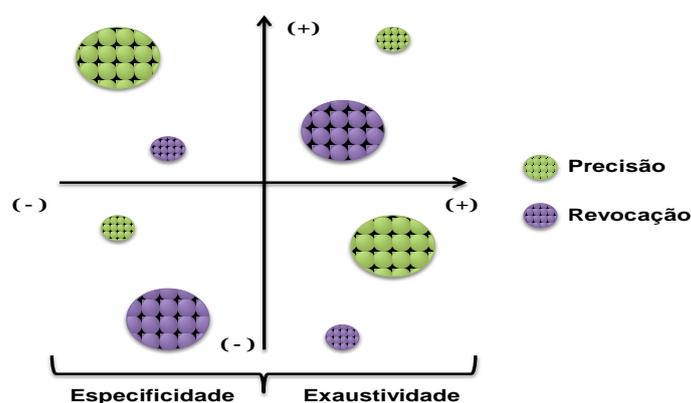


Figura 5 – Métricas de avaliação da recuperação de informação

Fonte: Elaboração própria

Lancaster (2004) enfatiza que, conforme aumenta-se o número de itens no resultado da busca, diminui-se a precisão do conjunto de itens recuperados, e o contrário também é verdadeiro, quanto mais se refinam as estratégias de busca, menor será o número de itens recuperados, aumentando a precisão da recuperação, ou seja, estas medidas são inversamente proporcionais.

A especificidade orienta à identificação dos assuntos de um documento da forma mais seletiva possível. Lancaster (2004) afirma que esta é a diretriz

mais importante da indexação, em que um item deve ser indexado sob o termo mais específico que o represente completamente.

A exaustividade orienta que os assuntos devem ser indexados na sua totalidade em função das necessidades informacionais dos usuários e do sistema de informação. Relaciona-se a capacidade do sistema indexar o documento em profundidade, ou seja, além do assunto principal são indexados também os assuntos secundários. Segundo Lancaster (2004), esta medida aumenta a extensão da representação e também aumenta a recuperabilidade do item, isto é, quanto mais elevado for o número de pontos de acesso, maiores são as chances do usuário encontrar o documento desejado. No entanto, sabe-se que este princípio deve ser utilizado com cautela, uma vez que pode afetar a precisão da recuperação.

As métricas revocação e precisão são utilizadas para avaliar os resultados apresentados pelos SRI, qualificando a informação recuperada como útil ou inútil, de acordo com as necessidades do usuário. Lancaster (2004) define a revocação (*recall*) como a capacidade do sistema de informação recuperar documentos úteis; e precisão, como a capacidade de evitar documentos inúteis.

Segundo o autor, a intenção do usuário ao fazer uma busca numa base de dados é encontrar os documentos que sejam úteis para satisfazer sua necessidade de informação e evitar a recuperação de itens inúteis.

Quanto ao método de execução, a indexação pode ser realizada por pessoas (indexação manual) ou por softwares computacionais (indexação automática). As pesquisas no campo da Ciência da Informação avaliam a eficiência dos métodos de indexação manual e automática.

A indexação manual, realizada por profissionais da informação, é um processo que, embora realizado de maneira profissional, não está isento de subjetividade, e por isso pode variar de um indexador para outro, ou até pelo mesmo indexador em momentos diferentes.

A indexação automática apresenta menor grau de subjetividade, uma vez que a subjetividade pela qual o indexador interpõe seus próprios valores, influenciando na análise do documento e na definição dos termos descritores, pode ser praticamente eliminada nos processos automatizados, quando um software fará todo o serviço de maneira padronizada seguindo parâmetros pré-estabelecidos.

A indexação automática tem sido reconhecida como o método mais adequado para o tratamento do grande volume de informações produzidas no ambiente digital. A indexação manual, em contrapartida, tornou-se dispendiosa e passível de erros de interpretação e no emprego de vocabulário fora do padrão estabelecido. Na visão de Lancaster (2004), os processos algorítmicos podem ser usados durante a indexação da informação em substituição ao processamento intelectual humano.

No cenário atual, onde é premente a necessidade dos documentos estarem disponíveis à comunidade de usuários tão logo sejam criados, a produtividade que os softwares de indexação automática prevaleça em relação à indexação manual.

2.4.2 Indexação automática

Hjørland (2008) define indexação automática como aquela realizada por procedimentos algorítmicos. O algoritmo pode tratar uma base de dados contendo documentos como texto completo, registros bibliográficos e materiais não textuais, como imagens ou música.

Segundo Robredo e Cunha (2003), a automatização da indexação consiste no emprego de algoritmos computacionais que realizam extração dos termos significativos do conteúdo de um documento, visando estabelecer rotinas que reduzam a subjetividade na representação da informação. É influenciada também por outras características do objeto informacional, que Feitosa (2006), identificou como: tamanho do registro, separador entre as palavras, tratamento para os números, caracteres individuais, entre outros.

Gil Leiva (2008) afirma que, com relação à execução da indexação automática, existem três diferentes métodos:

- Indexação assistida por computador durante o armazenamento;
- Indexação semi-automática ou assistida por computador;
- Indexação automática.

No primeiro métodos, o software de indexação apoia o trabalho intelectual de atribuir termos ao documento por meio de interfaces, onde são apresentados os termos e seus termos relacionados. No segundo, o software de indexação extrai do documento os termos de indexação, mas é necessário que o profissional, na função de indexador, valide estes termos refinando a lista de descritores propostos pelo software e fazendo os ajustes e as complementações necessárias. No terceiro método, os softwares analisam, extraem e atribuem os termos de indexação ao documento, sem nenhuma intervenção humana.

Lancaster (2004) subdivide a indexação automática em dois tipos:

- **indexação por extração automática:** consiste em retirar do documento as palavras que serão utilizadas para representar o seu conteúdo. Para os documentos digitais, Araújo Jr. (2007) sugere que o software será parametrizado para selecionar e extrair as palavras e/ou expressões que mais ocorrem no documento.
- **indexação por atribuição automática:** para cada termo que será indexado há um “perfil” de palavras ou expressões para vinculação, a partir de um vocabulário controlado.

Segundo afirma Lancaster (2004):

“Um vocabulário controlado é essencialmente uma lista de termos autorizados. Em geral, o indexador somente pode atribuir a um documento os termos que constem da lista adotada pela instituição para a qual trabalha. Comumente, no entanto, o vocabulário

controlado é mais do que uma mera lista. Inclui, em geral, uma forma de estrutura semântica.” (LANCASTER 2004, p. 19)

A Figura 6 exemplifica o fluxo de processamento de um software de indexação automática, cada fase executa um tipo de tratamento sob o conjunto de termos selecionados no conteúdo do documento.

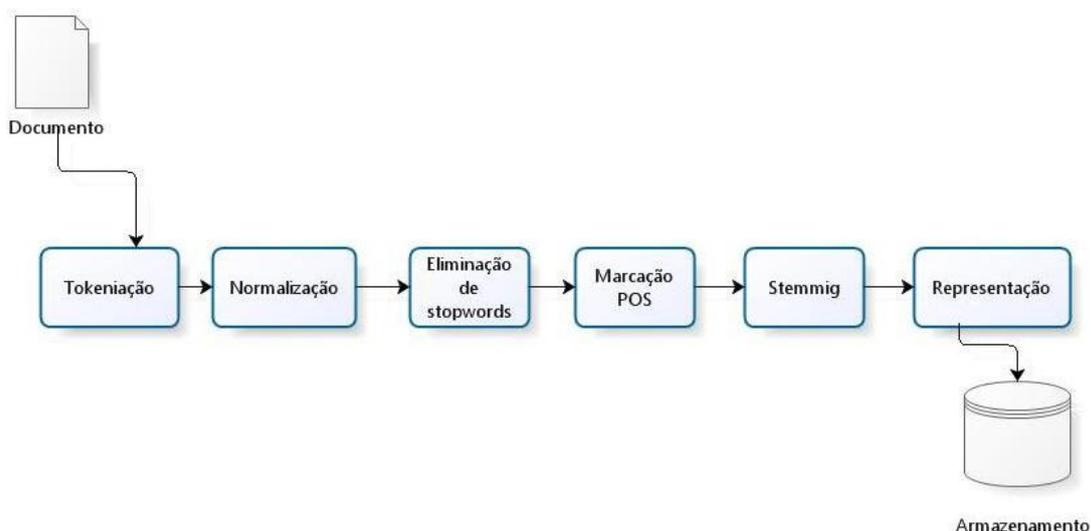


Figura 6 - Diagrama representativo do processo de indexação automática

Fonte: Adaptado de Baeza-Yates e Ribeiro-Neto (2011)

Na primeira fase, denominada tokenização, é realizada a decomposição do texto em um conjunto de fragmentos. Câmara Jr. (2007) define a tokenização como a separação do texto em palavras, por meio do reconhecimento do texto entre caracteres de marcação, que podem ser: hífen, parêntese, apóstrofo, ponto final e até mesmo um espaço em branco.

Na sequência, os *tokens* passam pelo processo de normalização para garantir, por exemplo, que todos estejam em caixa baixa.

Em seguida, as palavras com baixo valor informacional, denominadas *stopwords*, são subtraídas do conjunto de *tokens*.

Depois ocorre a marcação POS (parte de discurso) quando é identificada a classe gramatical dos *tokens*, e com essa classificação o software pode manter, por exemplo, apenas os termos substantivos.

A fase de *stemming* realiza a eliminação dos prefixos e sufixos dos *tokens*. *Stemming* tem significado específico no caso do tratamento da linguagem, designa capacidade do software determinar a raiz morfológica de uma palavra flexionada, ou derivada, reduzindo as palavras a um radical comum. Seu emprego para a língua inglesa é trivial, com apenas problemas ocasionais, tais como "dries" é o singular da terceira pessoa do presente do verbo "dry", mas torna-se complexo sua aplicação na morfologia, ortografia e codificação de caracteres na língua portuguesa, porque a possibilidade de flexões verbais é maior.

Após a execução de todas as fases, os termos selecionados do documento estão mais compactos e serão armazenado no banco de dados como índices para recuperação de informação.

Em sua pesquisa, Vieira (1988) identificou os métodos empregados pelos algoritmos de indexação automática:

- **Método de frequência ou análise estatística**

A contagem automática dos termos é realizada pela ocorrência e/ou coocorrência da palavra, palavras truncadas ou em raízes de palavras, e a frequência. Pode ser estabelecida pela:

a) ocorrência total da palavra no documento – a palavra é contada todas as vezes que aparece, fazendo-se o somatório das vezes em que ocorre;

b) ocorrência única da palavra no documento — conta-se somente uma vez a palavra, independentemente do número de vezes que ela aparece;

c) ocorrência da palavra na coleção — a contagem é realizada somando-se seu aparecimento na coleção.

A maior utilização desse método é para realização da seleção automática de descritores.

- **Método de atribuição de pesos**

O método de atribuição de peso aos descritores é baseado na frequência de cada descritor como uma forma de torná-los mais precisos sem, no entanto, diminuir sua capacidade de revocação. O peso pode ser atribuído por:

a) frequência total ou frequência única — a palavra recebe o mesmo valor do número de sua frequência;

b) fonte — se a palavra se encontrar em um documento reconhecido como relevante, receberá um peso maior do que outra existente em um documento menos relevante;

c) por fonte e usuário — o usuário é quem julgará se documento recuperado é relevante ou não.

d) frequência na coleção.

A indexação automática com pesos trabalha com a especificidade da palavra, sendo que as palavras de baixa frequência são as mais específicas e recebem maior peso que as palavras de alta frequência, mais exaustivas.

- **Método probabilístico**

O método probabilístico aplica a frequência de coocorrência em palavras truncadas automaticamente. As palavras truncadas são extraídas através de um critério estatístico de distribuição binomial, denominado distribuição de Poisson. Aquelas palavras cuja frequência de distribuição for descrita pela função de Poisson serão não significativas.

- **Método matemático**

O método matemático é baseado na identificação da frequência de coocorrência em pares de palavras, frases ou grupos de palavras significativas em documentos, através de algoritmo que as ordena alfabeticamente construindo um índice das várias combinações entre as palavras.

- **Análise de cluster**

A técnica se baseia no reconhecimento automático, em um grupo de documentos dos subgrupos de assunto que mais se assemelham entre si. É aplicada aos termos simples, pares de termos isolados em correlação matricial, termo a termo.

- **Método de associação entre palavras**

Utiliza a frequência de ocorrência e de coocorrência de palavras ou pares de palavras para identificar o conteúdo dos documentos. As palavras isoladas e as que se coassocia são identificadas em sentenças. Se as coassociações das mesmas palavras coocorrerem com determinada frequência, então serão consideradas "descritores associados".

Identificou-se que entre os fatores que mais favorecem a utilização da indexação automática nos SRIs, estão:

- fatores tecnológicos relativos a capacidade de processamento dos computadores, evolução dos softwares de indexação, o advento da web;
- menor subjetividade, uma vez que a subjetividade é inerente à indexação humana, que pode variar de um indexador para outro;
- eficiência operacional possibilitando o tratamento de grandes volumes de informação com custo e tempo inferior aos despendidos na indexação humana;
- exaustividade da indexação.

Segundo Cardoso Filho e Santos (2012), a indexação agrega valor à informação, organizando-a e tornado-a acessível. Sendo assim, a representação do conteúdo dos documentos assume uma dimensão estratégica, uma vez que o documento pode ser indexado de acordo com o público ao qual se destina, demandando a definição de estratégias e regras que direcionem a execução do processo de indexação.

A literatura pesquisada direciona para a implantação de políticas de indexação como o instrumento para estabelecer diretrizes estratégicas e corporativas de representação da informação pela indexação.

2.3.3 Indexação automática aplicada: Projeto TCDF

A Ciência da Informação aborda os aspectos teóricos e conceituais sobre o desenvolvimento da indexação automática, mas poucos são os estudos divulgados sobre a aplicação deste processo na representação da informação digital.

Relata-se a seguir o trabalho realizado no projeto de parceria para transferência de conhecimento entre o Centro de Apoio ao Desenvolvimento Tecnológico (CDT), órgão pertencente à Universidade de Brasília (UnB), e o Tribunal de Contas do Distrito Federal (TCDF) ocorrido no ano de 2012. (DOCUMENTO TÉCNICO TCDF, 2012)

O objetivo do projeto era realizar a indexação automática de 500 mil documentos do tipo: decisões, acórdãos, relatórios, votos e pareceres pertencentes à Base de Dados de Jurisprudência.

No que tange ao planejamento, o projeto foi composto das etapas ilustradas na Figura 7.

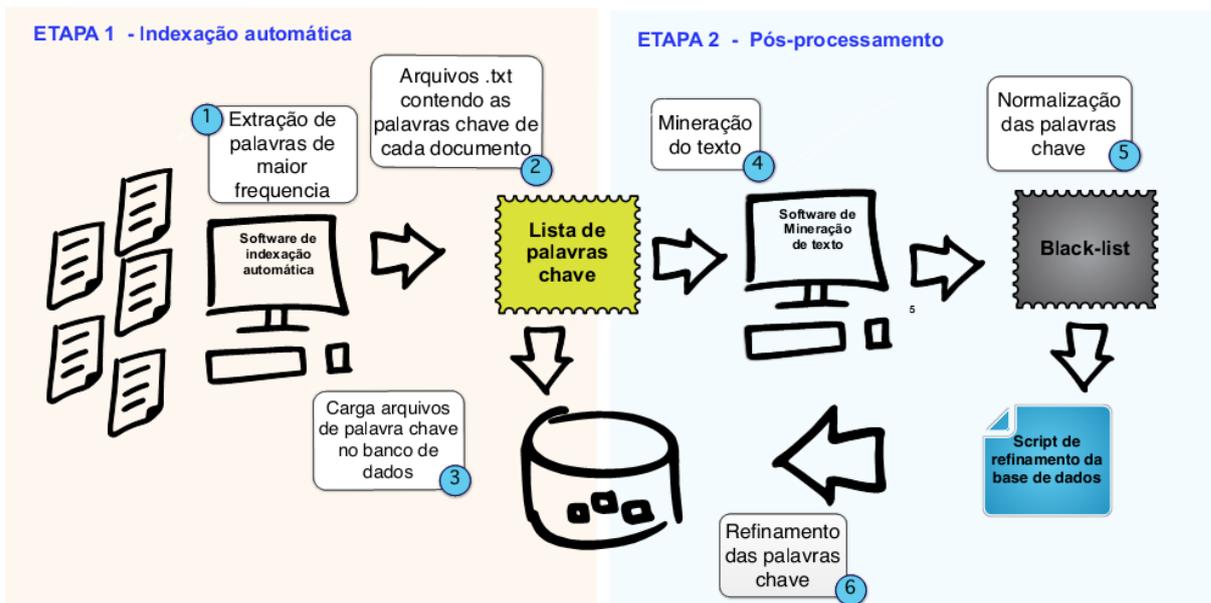


Figura 7 - Etapas do processo de indexação automática do TCDF

Fonte – Elaboração própria

Na primeira etapa, foi utilizado um software para o processamento da indexação automática, que extraiu as palavras de maior frequência no conteúdo dos documentos. Este software foi parametrizado para:

- ignorar palavras com menos de três letras e as entre aspas (“a” , ‘a’);
- ignorar caracteres especiais como §, “, ‘, \$, #;
- eliminar as palavras não significativas para o contexto, comparando-as com as palavras contidas na *black-list*. Esta lista é composta por: pronomes, preposições, artigos, conjunções, números romanos e cardinais, flexões verbais;
- obter os termos que iniciam com letra maiúscula e/ou os que são totalmente em letra maiúscula. Normalmente estes termos são substantivos que aparecem no início de frases ou nos títulos, sendo significativos para representar o conteúdo dos documentos.

Os produtos resultantes desta etapa foram os arquivos no formato .txt, denominados arquivos de índice, que continham os termos significativos de cada documento que foram extraídos pelo software de indexação.

Na sequência foi realizada a carga do conteúdo destes arquivos de índice no banco de dados. A carga foi realizada utilizando os próprios recursos de importação disponível no software de banco de dados, sem a necessidade de criação de outros programas.

Após a carga no banco de dados, foi realizado o pós-processamento, com o objetivo de garantir a normalização dos termos de indexação. Esta etapa compreendeu as seguintes atividades:

- execução do software de mineração de texto para análise da frequência dos termos e identificação dos termos: repetidos no mesmo documento, escritos somente em maiúsculo (que não fossem siglas) e demais termos não significativos para representação dos documentos, por exemplo: verbos no gerúndio, particípio passado, advérbios, números romanos etc.;
- criação da *black-list* com outros termos que deveriam ser excluídos dos arquivos de índice e atualização da *black-list* do software de indexação.

A nova *black-list* foi processada por meio de um *script*, executado no banco de dados para atualização dos arquivos de índices, concluindo a etapa do pós-processamento.

A Tabela 3 mostra exemplos de termos que foram normalizados durante a etapa de pós-processamento.

Antes do pós processamento	Depois do pós processamento
Tcdf	TCDF
Ordem Jud.	Ordem Judicial
Pres	Presidente
PROCESSO	Processo
Processos	Processo
CONTRATO	Contrato

Tabela 3 - Exemplo da lista de termos alterados no pós-processamento

Fonte – Documento Técnico TCDF (2012)

A estratégia do pós-processamento favorece a atualização contínua dos termos de indexação. Estes ajustes podem ser realizados de maneira global, considerando todos os termos de indexação, ou de maneira localizada, no limite de uma determinada massa documental. Novas *black-lists* poderão surgir junto com novos requisitos de conversão de termos.

Araújo Jr. (2007) afirma que a utilização de ferramentas de mineração de textos para compor listas de palavras chaves e enriquecer a linguagem documentária e favorecer a precisão na recuperação da informação.

Este projeto representa um importante referencial de aplicação da indexação automática para grandes volumes de documentos digitais, e evidencia a necessidade de políticas corporativas que direcionem o tratamento do conteúdo dos documentos, desde o início do ciclo da informação.

2.5 Sistemas de Recuperação da Informação

Os processos de organização e representação preparam os sistemas de recuperação de informação (SRI) para atender as demandas informacionais dos usuários. Assim, para a Ciência da Informação a recuperação da informação têm dois aspectos fundamentais: a indexação do conteúdo dos documentos e a busca por documentos úteis em uma coleção.

Cada vez mais os objetos informacionais migram dos ambientes físicos, no formato convencional, para o ambiente virtual, no formato digital, acessíveis por meio das redes de computadores e por sistemas informatizados que possibilitam a localização dos documentos. Lancaster e Warner (1993) *apud* Souza (2006) atribuem aos SRI a função de mediadores no processo de troca de informação por meio das interfaces entre os documentos, em suporte impresso ou não, e a comunidade de usuários.

Segundo Taylor e Joudrey (2009), no âmbito da organização da informação, o termo sistema de informação é análogo a sistema de recuperação da informação e executa funções como: armazenamento

organizado dos dados, recuperação por meio de consultas e interface de apresentação.

Baeza-Yates e Ribeiro-Neto (2011) ressaltam as três funções básicas dos SRI: representação da necessidade de informação; identificação no conteúdo dos documentos e a comparação, que seleciona os documentos relevantes baseado nas suas representações.

Segundo Baeza-Yates e Ribeiro-Neto (2011) as pesquisas acerca dos SRI dividem-se em:

- **Avaliação centrada no sistema:** estuda os aspectos tecnológicos incluindo algoritmos de recuperação, criação de índices, desempenho do processamento de busca, *design* de interfaces;
- **Avaliação centrada no usuário:** estuda o comportamento de busca do usuário, as necessidades de informações e as influências das abordagens cognitivas na recuperação.

Segundo Baeza-Yates e Ribeiro-Neto (2011), do ponto de vista tecnológico, o primeiro passo para criação de um sistema da informação é reunir uma coleção de documentos, que podem ser criados internamente na organização ou coletadas na web, esta coleção de documentos é então armazenada em banco de dados, usualmente chamado de repositório central. Para que os documentos possam ser recuperados devem ser indexados, o que envolve a criação de estruturas de dados associadas à parte textual dos documentos. Estando os documentos indexados, o processo de recuperação pode ser realizado. O usuário especifica uma consulta que corresponde à sua necessidade de informação, essa consulta é então processada em comparação com os índices, de modo a recuperar um subconjunto de todos os documentos que correspondam aos termos informados pelo usuário na consulta. Essa resposta é apresentada em forma de ranking, que tem o propósito de elencar os documentos que são provavelmente considerados relevantes para o usuário.

A Figura 8 representa uma visão geral dos sistemas de recuperação de informação.

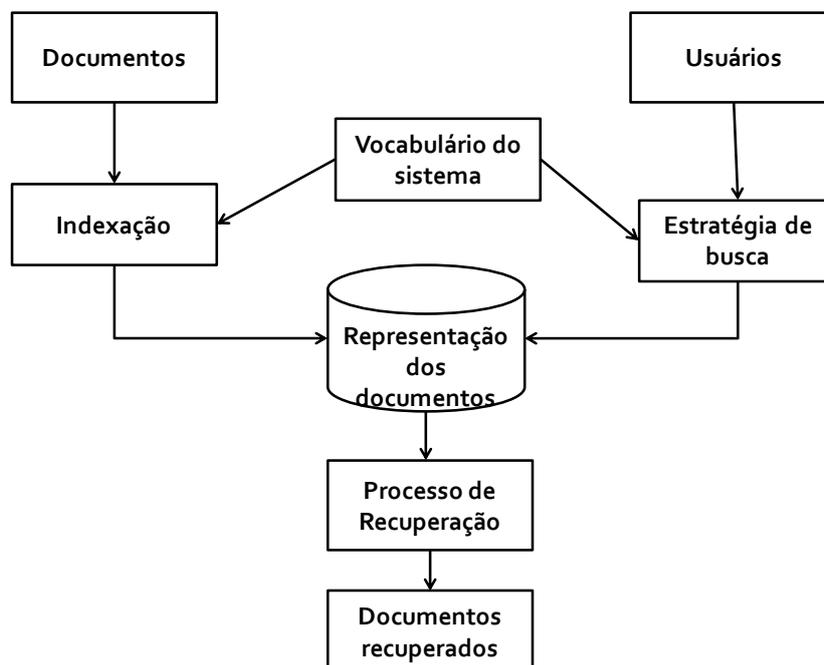


Figura 8 - Fluxo do sistema recuperação da informação

Fonte: Araújo Jr. (2007)

Segundo Chowdhury (2010), para o desenvolvimento de sistemas de recuperação de informação eficientes é preciso conhecer o contexto organizacional, as comunidades de usuários e suas necessidades de informação. Quanto às necessidades de informação, elas podem variar conforme o ambiente, o período de tempo e de pessoa para pessoa, o que significa que a mesma informação pode ser consultada de maneiras diferentes dentro de uma mesma organização.

Vários modelos de SRI foram propostos, dentre modelos clássicos, Baeza-Yates e Ribeiro-Neto (2011) citam o modelo booleano, o vetorial e o probabilístico. Estes modelos implementam a busca por documentos relevantes para a consulta, considerando os termos pelos quais os documentos foram indexados.

- **Modelo Booleano:** para cada consulta são recuperados os documentos que atendem à expressão lógica especificada pelo

usuário, que é formada pelo conjunto de termos ligados por operadores booleanos como 'e', 'não ou 'ou.

- **Modelo Vetorial:** os documentos são compostos por vetores contendo os termos de indexação e seus respectivos pesos para o documento. A similaridade entre os documentos e a consulta feita pelo usuário do SRI é estabelecida por meio dos pesos dos termos. Geralmente os pesos são atribuídos com base na frequência dos termos nos documentos.
- **Modelo Probabilístico:** calcula a probabilidade de cada documento ser relevante para a consulta. O modelo é aprimorado com base no *feedback* do usuário informando quais documentos recuperados são relevantes para sua consulta.

Kuramoto (1995) faz referência aos elementos que devem ser empregados por um sistema de recuperação de informação:

- uma interface de consulta para inserção da expressão de busca e visualização do resultado;
- a utilização de modelo de RI;
- estrutura conceitual da base de dados;
- os termos de indexação, os léxicos, os tesouros etc.

Os usuários dos SRI consideram eficientes aquelas ferramentas que apresentam os resultados mais significativos à sua consulta. A relevância é uma característica dos SRIs atuais, que além do conjunto de operações lógicas, procuram relacionar o contexto da pesquisa ao perfil do usuário e ao tipo de informação que se está consultando. Diante deste cenário, surgem novas tendências para SRI visando tornar as pesquisas cada vez mais eficazes: (FIORAVANTE, 2012)

- **Integração:** baseada na premissa que o valor de uma busca está no apoio e na facilidade de estabelecer conexões significativas entre

diferentes fontes de conhecimento, independentemente do seu formato. Assim, motores de busca começam a integrar diferentes formatos em uma mesma base de informações, possibilitando em uma única busca relacionar diferentes formas, sejam páginas web ou documentos de Word, mas também áudios, imagens, livros, filmes, etc.

- **Visualização por agrupamentos:** o agrupamento surge como uma nova forma de visualização dos resultados alcançados através das buscas, agrupando os resultados comuns em diretórios por assunto, ou em verdadeiros mapas de conhecimento, formados de acordo com a semelhança entre os conteúdos. Tal visualização é extremamente significativa, uma vez que oferece ao usuário a possibilidade de contextualizar mais facilmente sua pesquisa.
- **Linguagem Natural:** as buscas são baseadas em questões formuladas em linguagem natural e o pressuposto é que os usuários, ao realizarem suas pesquisas, querem como resultado uma resposta e não simplesmente referências sobre o assunto. Utilizam algoritmos que compreendem a pergunta formulada, trazendo a resposta de forma objetiva.
- **Segmentação:** consiste em segmentar as buscas por temas, perfil ou necessidades de informação específicas. Torna-se extremamente relevante quando o usuário tem conhecimento do tipo de informação que busca.
- **Personalização:** utiliza algoritmos que buscam compreender o que o usuário está pesquisando, cruzam as informações referentes ao contexto da pesquisa, por meio da análise da navegação dos links acessados. Com tais informações, o sistema tenta tornar as buscas mais efetivas.

O desafio maior dos SRIs está em possibilitar o acesso e uso da informação pelas diferentes comunidades de usuários, conforme as suas

necessidades, os diferentes comportamentos de busca, as influências do ambiente e pelo conhecimento prévio que o usuário traz consigo, seu aparato cognitivo.

Para Baeza-Yates e Ribeiro-Neto (2011) os índices são o centro dos modernos SRI, uma vez que sua atividade central é combinar uma consulta aos termos descritores para que satisfaçam a consulta do usuário.

2.6 Gestão de documentos digitais

Segundo Tarapanoff (2006), para estar acessível a informação deve ser organizada e gerenciada. A gestão da informação orienta as organizações para gerenciar os documentos que dispõem, integrando e coordenando a heterogeneidade das fontes de informação existentes.

A gestão da informação é definida por Wilson (1997) *apud* Tarapanoff (2006), como o conjunto de ações que permitem a aplicação de princípios administrativos para aquisição, organização, controle, disseminação e uso da informação nas organizações no tempo exato e necessário à tomada de decisão, proporcionando ganhos de eficiência operacional.

Tarapanoff (2006) esclarece que a importância estratégica da informação foi reconhecida depois que as organizações mudaram seu foco da gestão de dados para gestão dos objetos informacionais.

Choo (2006) afirma que cabe a gestão da informação planejar e criar sistemas, serviços, processos e recursos de informação, o que envolve um ciclo contínuo, análogo ao fluxo informacional, que deve compreender:

- a identificação das necessidades de informação;
- a aquisição da informação;
- a organização e armazenamento da informação;
- o desenvolvimento de produtos e serviços de informação;
- a distribuição da informação; e
- o uso da informação.

As TICs são intensamente empregadas nestes processos possibilitando o alinhamento do conhecimento contido nos documentos, bases de dados, páginas web, emails e aplicativos empresariais ao conhecimento pessoal dos indivíduos, o que Choo (2006) entende respectivamente como conhecimentos explícitos e tácitos.

McGee e Prusak (1994) reconhecem a informação como um recurso de valor estratégico e afirmam que a competitividade é influenciada pela capacidade das organizações em recuperar, tratar, interpretar e utilizar a informação de forma eficaz. O valor da informação está relacionado com a utilidade que ela terá para os usuários, conforme suas necessidades, circunstâncias e aplicabilidade.

Nas organizações, a informação é insumo para o desenvolvimento de atividades estratégicas, táticas e operacionais e, por essa razão é considerada um ativo tangível. Contribuindo, segundo afirmam Rossetti e Morales (2007), para aumento da eficiência operacional, tornando as organizações competitivas frente aos mercados em que atuam.

A gestão de documentos surge como faceta da gestão da informação, trata especificamente da informação registrada nos diferentes suportes, tanto nos acervos convencionais como nos acervos digitais e estabelece as normas e procedimentos técnicos que regulam o ciclo de vida dos documentos, da sua criação até o descarte.

No Dicionário Brasileiro de Terminologia Arquivística, gestão de documentos é definida como: “procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento de documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento.” (DICIONARIO, 2005).

Segundo Herrera (1993) *apud* Calderon *et al* (2004) estão entre os procedimentos da gestão de documentos:

- estudo da tipologia documental;
- normalização de documentos;

- informatização de processos de tramitação documental;
- coordenação e colaboração entre organismos produtores e arquivo;
- plano de classificação;
- plano de descarte;
- incorporação de fundos e manutenção da informação atualizada;
- elaboração de índices e de linguagens documentárias.

Concluí-se que a gestão da informação atua na identificação das demandas de necessidades de informação, no mapeamento do fluxo informacional e na definição de normas e padrões de tratamento da informação, visando o compartilhamento e o uso de informação de forma eficiente. Por sua vez, a gestão de documentos, como parte da gestão da informação, realiza o processo técnico que permeia o todo o ciclo de vida dos documentos.

2.7 Elementos da Política de indexação

A política de indexação fornece diretrizes administrativas que explicitam os critérios técnicos e operacionais a fim de garantir a qualidade na representação de conteúdos informacionais, contribuindo para o bom desempenho dos SRIs durante as consultas dos usuários.

Lousada *et al* (2011) afirmam que essas políticas assumem um caráter estratégico frente à criação do conhecimento, possibilitando identificar seus fluxos e administrá-los de modo a extrair maior valor da informação.

Segundo Fujita (2012), a política de indexação é um tema pouco explorado na literatura da Ciência da Informação e Biblioteconomia, embora seja reconhecida a importância da indexação nos resultados dos SRIs.

Para Fujita (2012), a política de indexação em um sistema de informação é determinada:

“pela seleção dos tipos de documentos a serem indexados, procedimentos de análise e representação de assuntos, aspectos

qualitativos da indexação como a precisão, especificidade, exaustividade e revocação, instrumentos de controle de vocabulário tais como linguagens documentárias ou opção de trabalhar com linguagem natural, além da avaliação da indexação pela consistência e pela recuperação.” (FUJITA 2012, p.23)

Outro aspecto de destaque é abordado por Fujita (2012) é que a política de indexação não deve ser considerada como uma lista de procedimentos a serem seguidos, e sim como uma filosofia que reflete os interesses e objetivos de uma unidade de informação.

Carneiro (1985) afirma que o objetivo de estabelecer uma política de indexação é assegurar a localização precisa dos documentos pelos usuários, a racionalização dos processos e otimização dos serviços.

Segundo Fujita (2012) existem duas visões complementares de política de indexação: uma visão mais ampla, que considera o contexto da gestão de sistemas de recuperação da informação em relação às atividades de indexação, e a outra, relacionada aos procedimentos e elementos de indexação circunscritos ao manual de indexação, sua operacionalização e avaliação.

Frente a estes aspectos, a política de indexação deverá estabelecer os princípios e critérios que proporcionam à otimização dos serviços e racionalização dos processos de indexação, para que os usuários tenham acesso a qualquer documento ou informação, além das fronteiras departamentais de onde foi produzido.

Para Carneiro (1985), a política de indexação deve estar associada ao contexto administrativo da organização e deve considerar os seguintes requisitos:

- conhecer a organização na qual a política de indexação está inserida, contemplando os objetivos da organização, suas atividades, as áreas de interesse, os tipos de documentos pertinentes a cada área, o tipo de atividades desenvolvidas;

- identificação dos usuários do sistema de informação, quem são os usuários, quais suas características e qual o seu perfil;
- recursos humanos, materiais e financeiros delimitadores dos esforços e investimentos no desenvolvimento da política.

Segundo Rubi (2012), estes aspectos também podem ser denominados como: contexto, destinatários e infraestrutura.

De acordo com Carneiro (1985), os elementos indispensáveis para construção de uma política de indexação são:

- **cobertura de assuntos:** quais são os assuntos centrais e periféricos, quais as necessidades que esse sistema precisa atender, como também qual tema deve ser abordado com ênfase e qual deve ser tratado apenas de modo superficial, permitindo identificar os assuntos que merecem destaque, levando em conta a extensão da cobertura que esses documentos podem ter em cada área de interesse, como também a qualidade desses documentos para usuários específicos.
- **seleção e aquisição de documentos fonte:** define de quais fontes serão originados os documentos;
- **processo de indexação:** conforme os níveis de especificidade, exaustividade, escolha da linguagem, capacidade de revocação e precisão do sistema;
- **estratégia de busca:** considera se a busca será auxiliada por um especialista da informação, também denominada de busca delegada, ou se o usuário procura diretamente no sistema, busca não delegada;
- **tempo de resposta do sistema:** O tempo utilizado entre o recebimento de uma busca de informação e o fornecimento da resposta à mesma, em atendimento as necessidades dos usuários;
- **forma de saída:** formato em que as informações são apresentadas ao usuário. Como resposta a uma busca, o sistema pode apresentar

número de acesso referente aos documentos, referências bibliográficas, resumos ou o texto completo dos documentos.

- **avaliação do sistema:** identificar se o usuário está satisfeito ou não com as respostas do sistema, se existe alguma falha operacional e alternativas de melhoria.

A Figura 10 ilustra os critérios estabelecidos por Carneiro (1985) que devem constituir uma política de indexação.



Figura 9 – Critérios para uma política de indexação

Fonte: Elaboração própria

Fujita e Rubi (2006) ressaltam a necessidade de que a política seja escrita de maneira clara e objetiva e esteja acessível aos profissionais da informação. Sobre esse aspecto, Guimarães (2000) também afirma que a política só poderá ter continuidade e aperfeiçoamento no decorrer dos anos se for devidamente registrada, de modo que se possa ter clareza (independentemente dos elementos humanos) do conjunto de decisões tomadas, suas razões e seu contexto.

Outras razões são apontadas por Chinelato Filho (1997) *apud* Lousada *et al* (2011), que justificam importância do manual e da política de indexação para os sistemas de informação:

- crescimento da organização;

- a necessidade de uniformização dos procedimentos;
- o registro dos procedimentos adotados para perpetuar a política de indexação independente da rotatividade dos profissionais da informação.

A respeito do manual da política de indexação, Fujita e Rubi (2006) justificam tratar-se da documentação oficial da unidade de informação, que deve abranger:

- Manual de operação ou procedimentos:** deve descrever a atividade de indexação, dar instruções sobre a realização dessa tarefa e proporcionar métodos que possibilitem sua execução de maneira uniforme;
- Manual de política:** deve descrever, de maneira geral as diretrizes a serem seguidas pelos analistas de informação;
- Manual de organização:** deve servir como um repositório das experiências acumuladas pelos analistas de informação e constituir-se numa fonte de consultas.

Bourdon e Landry (2007) *apud* Rubi (2012) indicam os requisitos essenciais para determinar uma política de indexação:

Aspectos estratégicos:

- Para quem a política é direcionada?
- Quais os limites de sua aplicação?
- A política deveria ser publicada?

Apresentação das ferramentas para indexação:

- Quais as diferentes linguagens e esquemas de classificação utilizados?
- Quais as características de cada linguagem?
- A indexação será pré ou pós-coordenada?

Aplicação das ferramentas para indexação:

- Como as ferramentas serão utilizadas na bibliografia nacional e no catálogo da biblioteca ou nas agências bibliográficas nacionais?
- Existirão níveis diferentes de catalogação?
- Devem ser levados em conta os projetos de indexação retrospectiva?
- Quais os tipos de documentos indexados e não indexados?
- O índice será bilíngue / multilíngue?
- Qual o uso das ferramentas para indexação na interface de busca?
- As ferramentas serão disponibilizadas para os usuários durante suas buscas?

A Figura 10 sintetiza os itens que devem constar na política de indexação.

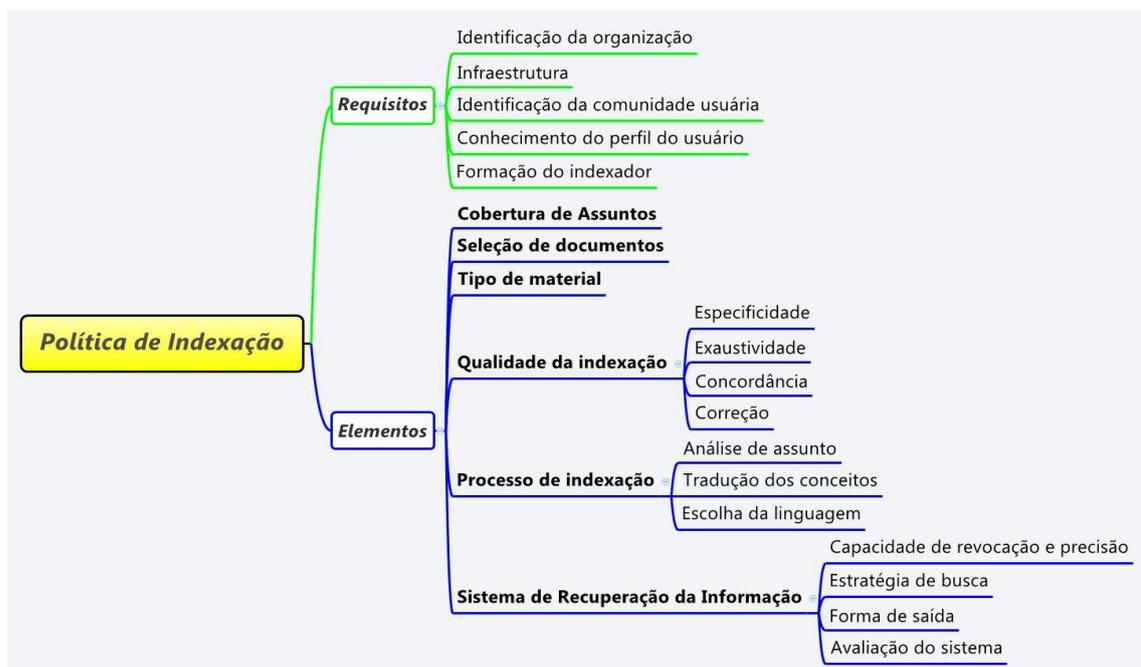


Figura 10 – Itens de uma política de indexação

Fonte: Adaptado de FUJITA (2012b)

2.8 O Acordo de Basileia II

O referencial teórico aborda também a análise do Acordo de Basileia II como temática do problema desta pesquisa, por sua relevância no contexto da instituição Banco do Brasil.

O sistema financeiro destaca-se como setor essencial ao desenvolvimento econômico. Entretanto, devido às peculiaridades de funcionamento das instituições que o constituem, existe a necessidade de regulamentações como forma de manter seus benefícios e evitar que possíveis problemas afetem o funcionamento da economia.

A regulamentação do Acordo de Basileia é aplicada a todo o sistema financeiro mundial. No Brasil o BACEN (Banco Central do Brasil), aliado aos órgãos internacionais, é o responsável por emitir as normatizações que devem ser seguidas pelas instituições financeiras que atuam no país.

Como consequência, os bancos enfrentam um novo desafio: estruturar os seus sistemas de informações operacionais, em conformidade com os requisitos do Acordo Basileia II, aprofundando ainda mais as características analíticas deste tipo de sistema e evoluindo o controle dos riscos de crédito, de mercado e operacional.

As regras estabelecidas pelo Comitê de Basileia¹, desde sua primeira edição, objetivaram a criação de um padrão internacional que os órgãos reguladores pudessem utilizar, no sentido de resguardar o mercado face aos riscos peculiares à indústria financeira.

Segundo Carvalho e Caldas (2006), em função das diversas crises econômicas que historicamente afetaram o mercado financeiro mundial, as

¹ O Comitê de Regulamentação Bancária e Práticas de Supervisão ou, simplesmente Comitê de Basileia, foi criado em um período de crise financeira, após forte volatilidade observada com a liberação das taxas de câmbio, situação que alcançou seu nível crítico em 1974. É sediado no Banco de Compensações Internacionais (BIS), em Basileia, na Suíça. e por esse motivo o comitê passou a ser denominado Comitê de Basileia.

principais preocupações do Comitê de Basileia são os fatores de risco presentes nas operações financeiras.

Um dos marcos do Comitê ocorreu em julho de 1988, com a celebração do Acordo de Basileia, que definiu mecanismos para mensuração do risco de crédito e estabeleceu a exigência de capital mínimo para suportar os riscos. Os principais objetivos do Acordo foram reforçar a solidez e estabilidade do sistema bancário internacional e minimizar as desigualdades competitivas entre os bancos internacionalmente ativos, em função das diferentes regras de exigência de capital mínimo pelos agentes reguladores nacionais.

No Brasil, o BACEN, sob a coordenação do Conselho Monetário Nacional (CMN), é o responsável pela supervisão bancária e pela definição quanto à forma de implantação das propostas apresentadas pelo Comitê. Sua comunicação com o mercado financeiro ocorre por meio das diversas Circulares BACEN e Resoluções do CMN.

Evolução do Acordo de Basileia

Basileia I

Os principais objetivos desse Acordo foram reforçar a solidez e estabilidade do sistema bancário internacional e minimizar as desigualdades competitivas entre os bancos internacionalmente ativos, em função das diferentes regras de exigência de capital mínimo pelos agentes reguladores nacionais.

Desse Acordo, ficaram definidos três conceitos básicos:

- **Capital Regulatório** – montante de capital próprio alocado para a cobertura de riscos, considerando os parâmetros definidos pelo órgão regulador;
- **Fatores de Ponderação de Risco dos Ativos** – a exposição a Risco de Crédito dos ativos é ponderada por diferentes pesos estabelecidos, considerando, principalmente, o perfil do tomador; e

- **Índice Mínimo de Capital para Cobertura do Risco de Crédito** (Índice de Basileia ou Razão BIS) – quociente entre o capital regulatório e os ativos ponderados pelo risco. Se o valor apurado for igual ou superior a 8%, o nível de capital do banco está adequado para a cobertura de Risco de Crédito.

Basileia II

Como evolução natural do cenário bancário mundial, e com o crescimento e integração dos diversos mercados financeiros, surgiu a necessidade de melhoria nas regras de exigência de capital. Assim, em junho de 2004, o Comitê de Basileia, divulgou o Novo Acordo de Capital, conhecido como Basileia II e apresentou, em linhas gerais, os seguintes objetivos:

- promover a estabilidade financeira;
- fortalecer a estrutura de capital das instituições;
- favorecer a adoção das melhores práticas de gestão de riscos; e
- estimular maior transparência e disciplina de mercado.

A proposta do Acordo de Basileia II apresenta um enfoque mais flexível para exigência de capital e mais abrangente com relação ao fortalecimento da supervisão bancária e ao estímulo para maior transparência na divulgação das informações ao mercado. Sua estrutura está baseada em três grandes premissas, que foram chamadas de pilares:

- **Pilar I** – fortalecimento da estrutura de capitais das instituições, com estímulo à utilização de modelos internos para mensurar riscos, introdução da exigência de capital para cobertura do risco operacional e aprimoramento da discussão sobre o risco do crédito;
- **Pilar II** – reafirma e fortalece os princípios de supervisão bancária, considerando que os supervisores devem avaliar e monitorar as estratégias dos bancos, além de cobrar que operem acima das exigências de capital mínimo e intervir buscando ações rápidas, em

caso de desenquadramento aos limites mínimos. Atribui responsabilidades à alta administração dos bancos pela estratégia de exposição aos riscos e pela compatibilização dos níveis de capital e enfatiza que os bancos devem possuir volume de capital adequado para suportar todos os riscos envolvidos nos negócios.

- **Pilar III** – redução da assimetria de informação e favorecimento da disciplina de mercado, por meio da divulgação de um conjunto mínimo de informações com vistas a aumentar a transparência das instituições quanto as suas práticas de gestão de riscos. A transparência possibilitará reconhecer e habilitar uma instituição financeira em uma abordagem de mensuração de capital específica e procura potencializar o poder de avaliação e atuação dos participantes do mercado.



Figura 11 - Acordo de Basileia II

Fonte: Relatório de Gestão de Riscos do Banco do Brasil (Brasil, 2012)

A importância estratégica dessas medidas foi de grande relevância para o mercado financeiro mundial, mas revelaram-se insuficientes para prevenir novas crises, como a verificada no final de 2008 e cujos reflexos são sentidos até o presente momento.

Na busca incessante de estabilidade no mercado financeiro mundial, o Comitê, nesse período de crise, promoveu uma série de reuniões, culminando com uma nova proposta que já foi batizada pelo mercado financeiro como Acordo de Basileia III.

Basileia III

Em 2010, o Comitê de Basileia para Supervisão Bancária divulgou as propostas que pretendem aumentar a qualidade do capital das instituições financeiras e promover a constituição de reservas adicionais de capital para serem usados em momentos de eventuais crises financeiras. Além dessas medidas, deve ocorrer a introdução de padrões mínimos de controle de liquidez. Essas propostas estão previstas para serem implantadas de forma gradual a partir de janeiro de 2013, até janeiro de 2019.

2.9.1 O Banco do Brasil e o Acordo de Basileia II

O Banco do Brasil tem buscado aprimorar suas práticas e seus instrumentos de gestão e controle de riscos alinhados com o Acordo de Basileia II, adotando modelos internos para os riscos de mercado, crédito e operacional.

A adoção das recomendações do Acordo de Basileia II eleva o patamar de confiança da instituição financeira frente o mercado mundial, propiciando maior visibilidade e favorecendo a conquista de novos clientes, inclusive internacionalmente.

Neste sentido, o modelo de governança de riscos utilizado no Banco do Brasil envolve a estrutura de comitês e subcomitês, com a participação das diversas áreas da instituição. As decisões relacionadas à gestão de riscos são tomadas em colegiado, de acordo com as diretrizes da instituição. A Diretoria de Gestão de Riscos (DIRIS) é a área responsável pela coordenação das ações para atendimento dos requisitos de Basileia II.

Ao estimular a abertura de informações, Pilar III, o Acordo de Basileia II procura ampliar o poder de avaliação e atuação dos participantes do mercado.

Sendo assim, cresce o papel dos reguladores no sentido de acessar e avaliar as posturas dos bancos diante de suas exposições ao risco.

O Pilar III trata do processo de comunicação com os clientes internos e externos envolvendo diversos documentos e sistemas de informação para gerenciamento de riscos.

Para atender a este Pilar o Banco do Brasil estabeleceu processos de comunicação da informação com clientes internos e externos. Os aspectos considerados na seleção e divulgação das informações são: as melhores práticas, a legislação bancária, as necessidades dos usuários, os interesses do Banco, a confidencialidade e a relevância da informação.

Os documentos são produzidos periodicamente para atender os clientes internos, resultam das análises realizadas pelos profissionais das áreas que evidenciam os riscos. Como exemplos destes documentos citam-se: (BRASIL, 2012)

- relatórios contábeis de gestão do capital, análise de risco e retorno da carteira de crédito;
- notas técnicas;
- planos de ação de riscos e investimentos;
- políticas e estratégias de mitigação de riscos;
- manual de procedimentos.

O caráter sigiloso desta documentação não permite que os documentos sejam divulgados nesta dissertação, mas a temática informacional pode ser identificada na regulamentação do Acordo de Basileia II e também nas normas disponíveis no site do BACEN.

Os documentos citados são produzidos e atualizados frequentemente e possuem informações gerenciais qualitativas e quantitativas sobre a exposição a riscos do Banco do Brasil, dando suporte às decisões dos seus subcomitês, Comitê de Risco Global, Conselho Diretor e Conselho de Administração.

Os clientes internos acessam estas informações via da intranet corporativa, por meio da qual as informações contidas nos documentos passem do status de informação departamental para constituir a base do conhecimento da instituição, garantindo assim agilidade e fidedignidade nas informações utilizadas para apoiar a tomada de decisão e o atendimento aos órgãos reguladores.

No processo de comunicação para clientes externos as áreas operacionais da estrutura de gerenciamento de riscos do Banco do Brasil, produzem as informações destinadas a este público e encaminham para a Unidade de Relações com Investidores (URI). A URI, como prática de governança transparente, divulga essas informações para o mercado, permitindo aos investidores e partes interessadas acompanhar as ações de gerenciamento de risco e a evolução do risco de crédito, e comprovar a suficiência de capital do Banco para cobertura de todos os riscos assumidos.

As informações destinadas ao público externo são disponibilizadas no *web site* do Banco do Brasil na internet, onde são publicadas nos seguintes documentos:

- relatório de administração;
- relatório de análise de desempenho;
- notas explicativas às demonstrações

O Banco do Brasil tem investido em soluções de TI para suportar os processos referentes ao Acordo de Basileia II. Encontra-se em fase de implementação um sistema de informação específico para gestão dos documentos digitais referentes ao Acordo de Basileia II. O acesso a esses documentos é controlado por políticas de segurança e sigilo, em que a classificação de acesso é restrita as unidades estratégicas do Banco. O acesso aos documentos é controlado por *logs*, que possibilitam rastrear todas as intervenções realizadas no documento. O gerenciamento eletrônico desses documentos é realizado por meio de um sistema de informação que tem como

função a gestão, a criação, o armazenamento, o acesso e a disseminação das informações.

A comunidade de clientes externos compõe-se dos funcionários dos órgãos reguladores, BACEN, auditorias, clientes e investidores. No caso destes usuários, a informação está disponível pelo *web site* do Banco do Brasil na internet (www.bb.com.br).

Sob o aspecto de transparência da informação, o Banco do Brasil está subordinado também a Lei 12.527 - Lei de Acesso a Informação, que conforme previsto no inciso II do art. 1º aplica-se, entre outros, as sociedades de economia mista, categoria que inclui o Banco do Brasil. Esta Lei reflete os esforços do governo brasileiro em assegurar a gestão transparente e proporcionar o ampla acesso à informação.

Como foi visto no decorrer deste capítulo os documentos produzidos no Banco do Brasil para atender as demandas regulatórias do Acordo de Basileia II e da Lei de Acesso a Informação são armazenados em sistemas de informação automatizados e devem atender às necessidades de informação de usuários internos e externos. Atualmente esses documentos não são submetidos a processo de representação do conteúdo informacional, fato que motiva a proposição de padrões corporativos para sua representação.

Considera-se que a política de indexação é a estratégia mais adequada para representação destes documentos por contribuir com a localização da informação mais adequada para atender com tempestividade as necessidades dos usuários.

2.10 Conclusão do referencial teórico

A revisão de literatura contribuiu para fossem alcançados os objetivos específicos definidos para nesta pesquisa, identificando na literatura da Ciência da Informação os elementos teóricos e metodológicos fundamentais para o tratamento dos documentos digitais e para criação uma política de indexação.

Foi possível observar as inter-relações entre os processos de produção dos documentos digitais e os seus reflexos no tratamento para representação da informação. A indexação, como método de representação da informação, trata do conteúdo dos documentos digitais relacionados ao contexto específico de cada organização e as necessidades informacionais dos seus usuários.

No contexto delimitado, que compreende a documentação bancária produzida para atender as exigências do Acordo de Basileia II, identificou-se a necessidade de tornar disponíveis as informações que possibilitem avaliar as estratégias do Banco do Brasil diante da exposição aos riscos. Para atender a esta exigência é conveniente estabelecer uma política de indexação que defina os padrões corporativos para sua representação da informação.

Observou-se que os estudos sobre política de indexação estão mais relacionados ao contexto dos centros de informação como as bibliotecas, por exemplo, e pouco relacionados a aplicação no contexto das instituições financeiras, onde a recuperação da informação é um fator determinante da eficiência dos seus processos operacionais estratégicos.

3 Procedimentos metodológicos

Esta dissertação, sob o ponto de vista de sua natureza, é caracterizada como pesquisa aplicada que, conforme Silva e Menezes (2005), “objetiva gerar conhecimentos para aplicação práticas e dirigidas à solução de problemas específicos”.

A pesquisa baseia-se em estudo de caso, que segundo Gil (2010), possibilita investigar um fenômeno, de modo a obter conhecimentos amplos e detalhados acerca do mesmo.

Do ponto de vista da Ciência da Informação insere-se na linha de pesquisa de organização e representação da informação, com foco nas políticas de indexação automática.

3.1 Definição do escopo da pesquisa

O escopo da pesquisa refere-se à população dos documentos digitais, de natureza bancária regulatória, em formato textual, na língua portuguesa e associados ao Acordo de Basileia II - Pilar III, produzidos no âmbito do Banco do Brasil.

3.2 Passos metodológicos

A realização desta pesquisa obedeceu aos seguintes passos metodológicos:

- apresentar o referencial teórico que trata sobre política de indexação automática;
- elaborar o instrumento de coleta de dados para verificar como ocorre a indexação dos documentos digitais produzidos no Banco do Brasil para atender as exigências do Acordo de Basileia II – Pilar III;
- apresentar os principais requisitos para indexação dos documentos digitais apontados no Acordo de Basileia II;

- analisar os dados coletados e propor diretrizes para uma política de indexação automática para os documentos digitais do Banco do Brasil referentes ao Acordo de Basileia II.

3.3 Coleta de dados

A coleta de dados foi realizada junto aos profissionais das áreas responsáveis pela gestão da documentação que atende o processo de comunicação do Acordo de Basileia II.

Foram realizadas entrevistas informais e abertas cujo roteiro consistia em três perguntas. O roteiro é apresentado no Apêndice 01.

Realizou-se também coleta de informações nas normas do Acordo de Basileia II – Pilar III, para seleção da temática considerada relevante às demandas informacionais dos usuários.

4 Descrição e análise dos dados

Foram selecionados dois gestores para responder às entrevistas, e com a intenção de preservar suas identidades, serão identificados como Gestor A e Gestor B.

A partir das entrevistas realizadas, conforme o roteiro no Apêndice 01 foram obtidas as seguintes respostas:

Gestor A

Pergunta 1: A qual público se destinam os documentos produzidos na sua equipe?

Os documentos são utilizados pelos funcionários do Banco, que atuam na administração da exposição a riscos e também para os altos executivos na tomada de decisão e nos direcionamentos da empresa.

Pergunta 2: Quais os tipos de documentos são criados nas atividades realizadas na área?

Produzimos os relatórios contábeis e financeiros, notas técnicas, planos de ação, políticas e estratégias e manuais operacionais no formato de planilhas eletrônicas ou textos em vários formatos.

Pergunta 3: De que forma o conteúdo dos documentos é representado?

O conteúdo é representado pelo título informado pelo autor durante a criação do documento.

Pergunta 4: Na sua equipe existem funcionários capacitados para atuar no tratamento do conteúdo dos documentos?

Não temos conhecimento de outras maneiras para representar o assunto do documento que não seja pelo título.

Gestor B

Pergunta 1: A qual público se destinam os documentos produzidos na sua equipe?

Clientes e investidores que acompanham o desempenho do Banco.

Pergunta 2: Quais os tipos de documentos são criados nas atividades realizadas na área?

O principal documento disponibilizado para o público externo é o relatório de gestão de riscos, mas também são produzidos os relatórios de análise e desempenho do Banco e as notas explicativas.

Pergunta 3: De que forma o conteúdo dos documentos é representado?

Não consideramos necessário representar o conteúdo porque entendemos que os nossos usuários sempre procuram pelo último documento publicado. Eles precisam da informação atualizada.

Pergunta 4: Na sua equipe existem funcionários capacitados para atuar no tratamento do conteúdo dos documentos?

Teríamos que capacitar os funcionários para realizar esse tipo de atividade.

A partir das respostas obtidas constatou-se que: os gestores responsáveis desconhecem os procedimentos para representação da informação abordados no âmbito da Ciência da Informação; não são aplicados processos de indexação automática nos documentos digitais e conseqüentemente não há uma política que oriente a representação do conteúdo dos documentos.

Foi possível observar também que a gestão desta documentação é descentralizada, cada departamento tem seu próprio sistema de informação, o que dificulta a identificação dos tipos de documentos e sua recuperação interdepartamental.

A contextualização da pesquisa no âmbito do Banco do Brasil permitiu verificar a existência de uma grande variedade de documentos digitais que poderiam ser organizados e representados, segundo a Ciência da Informação, e assim atender adequadamente as necessidades dos usuários aumentando alcance e a precisão da recuperação de informação nos seus sistemas de informação.

Diante do volume de documentos digitais produzidos, percebe-se a tendência de crescimento das dificuldades na recuperação de informação, o que poderá afetar o desempenho dos processos atividades e causar riscos na tomada de decisões por falta de informações adequadas, no momento que são demandadas.

Poucas são as pesquisas que tem esclarecido o processo de indexação automática aplicado às organizações empresariais. Os investimentos nas TICs deveriam convergir a gestão da informação, incorporando a utilização dos softwares de indexação e das linguagens documentárias. O Banco do Brasil é um exemplo de uma organização deficiente desta visão de convergência, uma vez que as estratégias para tratamento da documentação digital ocupam-se mais com as especificações tecnológicas de hardware e software, do que com os processos de organização e representação da informação contida nos documentos digitais.

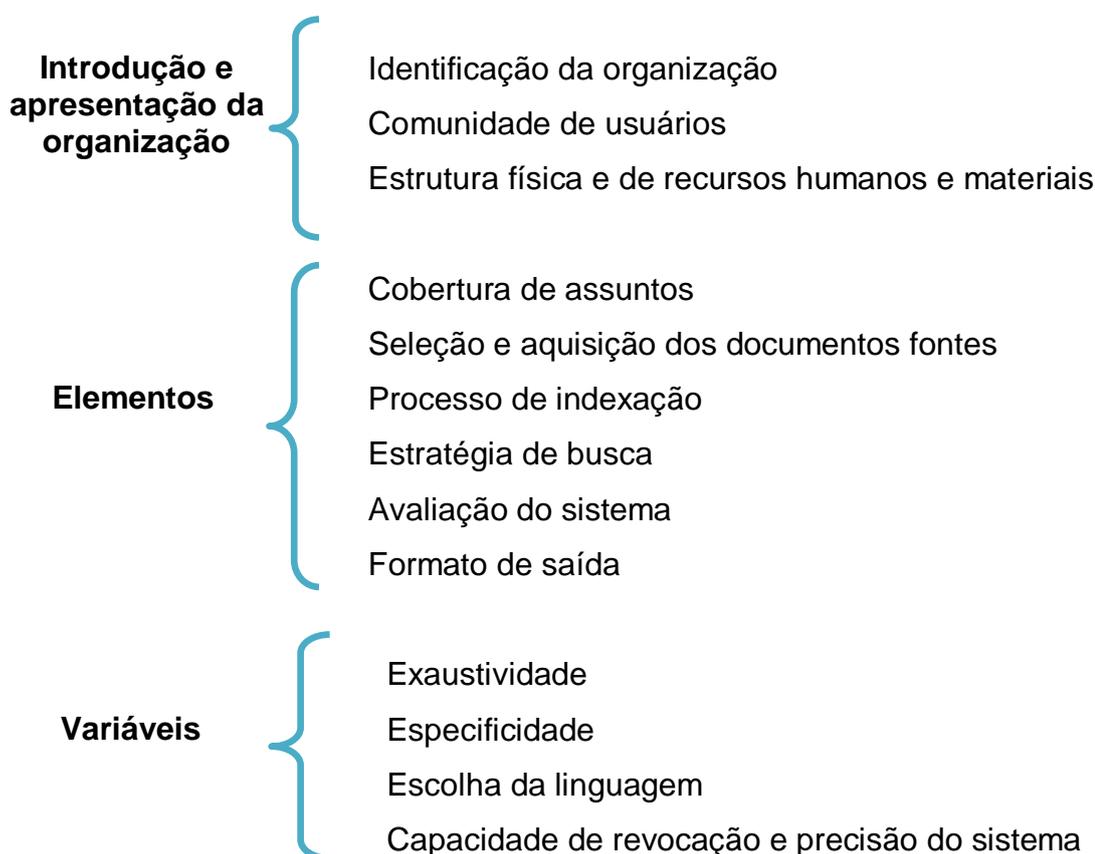
A partir da revisão de literatura, análise da regulamentação do Acordo de Basileia II e a análise de dados, foi elaborada uma proposta de diretrizes para uma política de indexação aplicável aos documentos que formam o escopo da pesquisa, constando das seguintes etapas:

- 1- identificação na literatura da Ciência da Informação quais são os elementos compõem uma política de indexação;
- 2- avaliação dos requisitos da documentação do Acordo de Basileia II – Pilar II que se aplicam ao contexto informacional delimitado;
- 3- verificação de como ocorre a indexação de documentos digitais produzidos no Banco do Brasil para atender as exigências do Acordo de Basileia II – Pilar III;
- 4- elaboração das diretrizes para a proposta de política de indexação, adequada para o escopo delimitado na pesquisa.

5 Proposta de política de indexação para documentos digitais no Banco do Brasil

Este capítulo apresenta os requisitos que devem ser considerados na elaboração de uma política de indexação orientada à estratégia corporativa de representação do conteúdo dos documentos digitais produzidos no Banco do Brasil para atender ao Acordo de Basileia II - Pilar III, que trata das informações sobre transparência bancária.

A elaboração desta proposta de política de indexação fundamentou-se nos trabalhos de Carneiro (1985) e Fujita (2012, 2012b), que propõe a seguinte estrutura:



Dentre os elementos citados pelos autores algumas adaptações foram realizadas para melhor adequação ao perfil da instituição Banco do Brasil. No que se refere à Introdução e apresentação da organização, serão tratados no item ambiente informacional. O elemento seleção e escolha dos documentos

constará no item perfil dos documentos, referindo-se a documentação digital que é escopo desta pesquisa. Foi desconsiderado o elemento formato de saída, porque em se tratando de informação digital, o mais provável é que os resultados sejam apresentados na forma de lista por meio de uma interface web do sistema de informação.

Apresenta-se a seguir o conjunto final das diretrizes que compõe a proposta de política de indexação, abrangendo os elementos mais significativos para o contexto informacional da instituição.

5.1 Diretrizes para política de indexação

As diretrizes expostas a seguir estão aderentes com o perfil do Banco do Brasil e às necessidades de informação dos seus usuários.



Figura 12 – Elementos que compõe as diretrizes da política de indexação

Fonte: Elaboração própria

Perfil do ambiente informacional

Esta diretriz prevê a implantação do Comitê de Gestor de Informações, motivada pela perspectiva de centralização dos esforços para o tratamento e gestão da informação, com amplitude corporativa para unificar os processos de atendimento às necessidades de informação das Unidades de Negócio.

O organograma na Figura 13 apresenta o Comitê como unidade autônoma, independente das Unidades de Negócio ou de Tecnologia, que tem como responsabilidade a coordenação das ações de tratamento da informação e discussões, com as demais unidades da instituição, para reconhecimento das necessidades de informação.

As Unidades Gestoras do Negócio detêm o conhecimento acerca do vocabulário que representa o seu nicho de negócios. No caso do Acordo de Basileia II, essas informações são de responsabilidade da Unidade de Gestão de Riscos.

A Unidade Gestora de Tecnologia será responsável pelo fornecimento dos recursos tecnológicos que irão apoiar o processo de indexação automática.

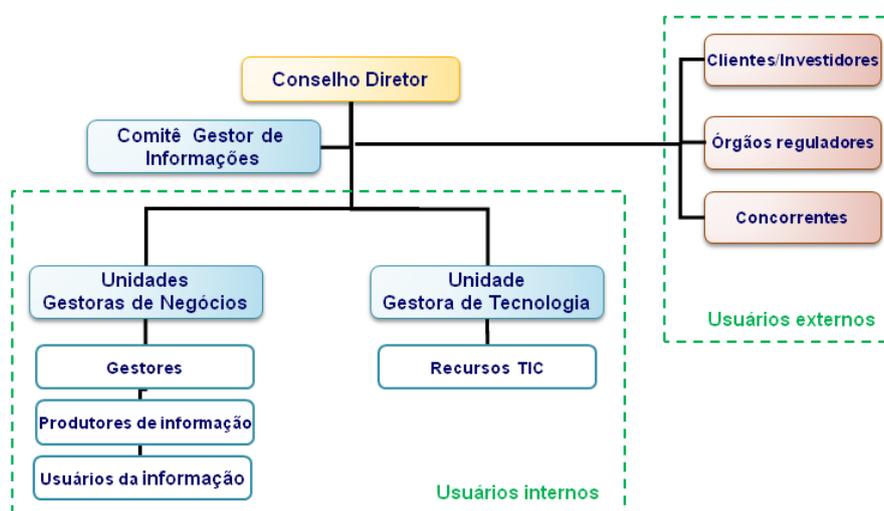


Figura 13 - Estrutura organizacional do Comitê de Gestor de Informações

Fonte – Elaboração própria

As atividades atribuídas ao Comitê de Gestor de Informações são:

- avaliar os processos que criam documentos digitais de conteúdo textual;
- identificar e definir os termos que devem compor o Tesouro, a partir da análise dos índices extraídos no processo de indexação e/ou por sugestões dos usuários;
- homologar, em conjunto, com as Unidades Gestoras do Negócio, os termos que serão incluídos no Tesouro;
- gerenciar e fornecer suporte ao Tesouro;
- definir, avaliar e manter os parâmetros de configuração do software de indexação e do software de mineração de textos;
- capacitar os funcionários do Banco para utilização dos termos do Tesouro, durante a criação de documentos e/ou na realização das estratégias de busca;
- divulgar a política de indexação nos canais de comunicação da instituição.

A equipe designada para formar o Comitê deve ser composta por especialistas com os perfis de: analista da informação, analista negócios, bibliotecários e arquivistas, capacitados para o tratamento da documentação e identificação das necessidades de informação das Unidades de Negócio.

A gestão da política de indexação é uma atividade colaborativa em que participam funcionários do Comitê de Gestor de Informações, das Unidades Gestoras do Negócio e da Unidade Gestora de Tecnologia. Estes participantes desempenham basicamente as seguintes funções:

- **analista de informação** – responsável pela gestão da política de indexação, manutenção da linguagem documentária e levantamento das necessidades de informação.

- **especialista no assunto** – apoia o analista de informação na identificação dos termos relevantes ao domínio de conhecimento da sua área de negócio.

Comunidade de Usuários

A referida documentação atende às necessidades de dois grupos de usuários:

- **Usuário interno** – funcionários da instituição que atuam na criação da documentação, e na utilização para apoiar os processos de negócio da instituição;
- **Usuário externo** – formado por órgãos reguladores nacionais ou internacionais, investidores, concorrentes e clientes do Banco que utilizam as informações para atendimento das suas necessidade de informações específicas ou para apoiar seus processos decisórios.

A Tabela seguinte sintetiza os principais papéis e funções na política de indexação.

Funções	Analista de Informação	Especialista no Assunto	Usuário Interno	Usuário Externo
Levantamento dos conceitos e palavras de temática relevante				
Gestão da linguagem documentária (Tesauro)				
Parametrização do software de indexação automática				
Elaboração e gestão do manual técnico da política de indexação				
Inclusão de documentos no sistema de informação				
Realização de estratégias de busca				

Tabela 4 - Política de indexação - papéis e responsabilidades

Fonte - Elaboração própria

Perfil dos documentos

A informação bancária, especificamente derivada do Acordo de Basileia II – Pilar III, é produzida em documentos digitais no formato textual.

Quanto aos tipos de documentos produzidos constam: relatórios financeiros, relatórios contábeis, relatórios de gestão do capital, relatórios de análise de risco, relatórios de retorno de crédito, notas técnicas, planos de ação de riscos e investimentos, políticas e estratégias de mitigação de riscos, manuais de procedimentos.

Estratégia de Busca

As estratégias de busca são elaboradas por usuários especialistas na temática do Acordo de Basileia II, sem a necessidade de intermediários para recuperação de informação.

Os usuários realizam suas buscas por meio das interfaces que possibilitam a interação com o sistema de informação. As respostas às buscas são apresentadas na forma de um ranking, que consiste na classificação dos documentos recuperados por ordem decrescente de relevância. Esses sistemas executam as operações de: análise da questão de busca, conversão da questão para a linguagem documentária, execução da estratégia de busca no banco de dados do sistema de informação, apresentação do ranking com os resultados que atendem a estratégia de busca.

Avaliação do sistema

O ambiente tecnológico, os processos e a política de indexação serão periodicamente avaliados pelo Comitê Gestor de Informações, a partir dos critérios de seleção e divulgação de informações demandados pelo Acordo de Basileia II. A avaliação visa aprimorar as diretrizes, melhorar a representação do conteúdo dos documentos e maximizar a utilização do Tesouro pelos usuários, tanto na criação de novos documentos quanto nas suas estratégias de busca.

Os usuários também podem contribuir enviando comentários, críticas e sugestões por meio do canal de comunicação com o Comitê.

Recursos de infraestrutura e financeiros

Os recursos de infraestrutura de TIC devem ser definidos e disponibilizados por acordo entre o Comitê Gestor de Informações e a Unidade Gestora de Tecnologia.

Dentre esses investimentos estão os softwares para indexação automática, para gerenciamento eletrônico dos documentos (banco de dados textual), para mineração textual, os hardwares (servidores) para suportar os softwares e os dispositivos para armazenamento dos documentos e dos índices, e o link de rede de telecomunicações que favoreça o tempo de resposta aos usuários.

A estimativa da necessidade desses recursos será calculada considerando variáveis como: quantidade de documentos criados e indexados por dia, número de transações de consulta e recuperação de documentos, tamanho dos documentos, quantidade de usuários.

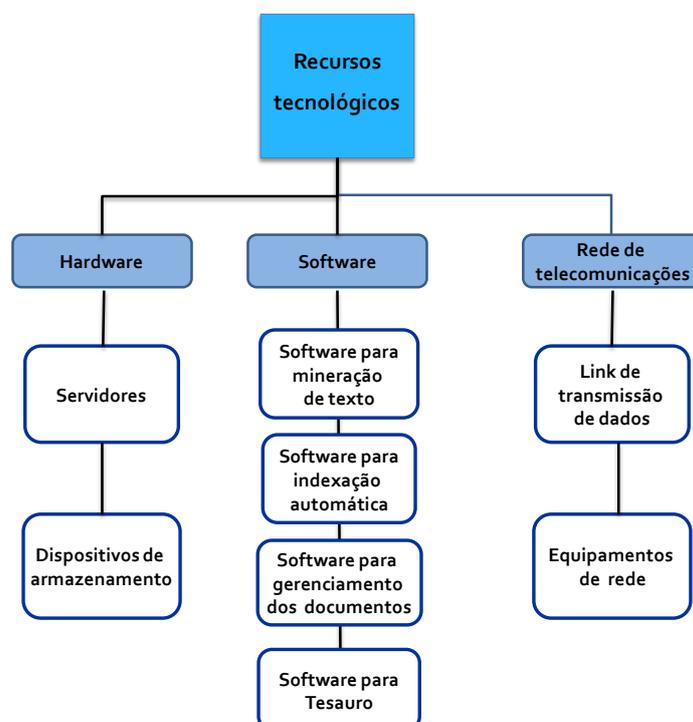


Figura 14 - Recursos TIC para suportar o processo de indexação automática

Fonte – Elaboração própria

Os recursos financeiros para aquisição dos recursos de TIC serão disponibilizados conforme o cronograma da área de Orçamento. A previsão orçamentária deverá ser apresentada com antecedência, normalmente, durante o primeiro semestre do ano anterior à execução do orçamento.

Cobertura de assuntos

A cobertura dos termos extraídos dos documentos será seletiva, direcionada à especificidade temática da legislação bancária estabelecida pelo Acordo de Basileia II compreenderá termos da Língua Portuguesa, em linguagem natural.

Propõe-se uma listagem de termos levantados no Acordo de Basileia II. Não é intenção nesta pesquisa realizar um levantamento exaustivo, mas sim um recorte representativo na temática do assunto.

Inadimplência - impossibilidade de completar uma transferência de fundos ou de valores mobiliários em conformidade com os termos acordados por razões que não são técnicas ou temporárias.

Pilar - princípios orientadores do Acordo de Basileia II.

Risco de crédito - risco decorrente da possibilidade de que a contraparte não honre a entrega de papéis ou fundos pactuados.

Risco de mercado - O risco de perdas nas posições de balanço e extra-balanço e que surge dos movimentos nos preços do mercado.

Risco operacional - risco de haver erro humano ou falha de equipamentos, programas de informática ou sistema de telecomunicações imprescindíveis ao funcionamento de determinado sistema.

Processo de indexação automática

O processo de indexação proposto é centrado no domínio, uma vez que considera a documentação digital de natureza regulatória, referente ao Acordo de Basileia II.

O Tesouro é a linguagem documentária utilizada para organização dos termos de indexação e durante a realização das estratégia de busca ao conteúdo dos documentos indexados.

O software de indexação automática executa a estratégia operacional da política de indexação. Sua configuração deverá considerar as seguintes premissas:

- indexação por extração dos termos na linguagem natural dos documentos;
- extração dos termos substantivos no singular;
- comparar termos com a lista de *stopwords* que contém as palavras não significativas como: artigos, preposições, numerais, interjeições, advérbios, numerais romanos e cardinais, caracteres especiais (‘, ”, #, \$ etc.);
- deverá reconhecer os substantivos da língua portuguesa pertencentes ao Tesouro da instituição, permitindo o controle de sinônimos.

Após a indexação automática ocorre a etapa denominada de pós-processamento, onde o arquivo de índices será tratado pelo software de mineração de textos para refinar os termos extraídos dos documentos durante a indexação. Nesta etapa serão identificados os termos que ainda precisam ser normalizados, os termos repetidos, os termos não significativos para o domínio do assunto (que serão acrescentados a lista de *stopwords*) e os termos de relevância que deverão ser incluídos no Tesouro.

A Figura 15 ilustra o processo de indexação baseado nas diretrizes da política de indexação.

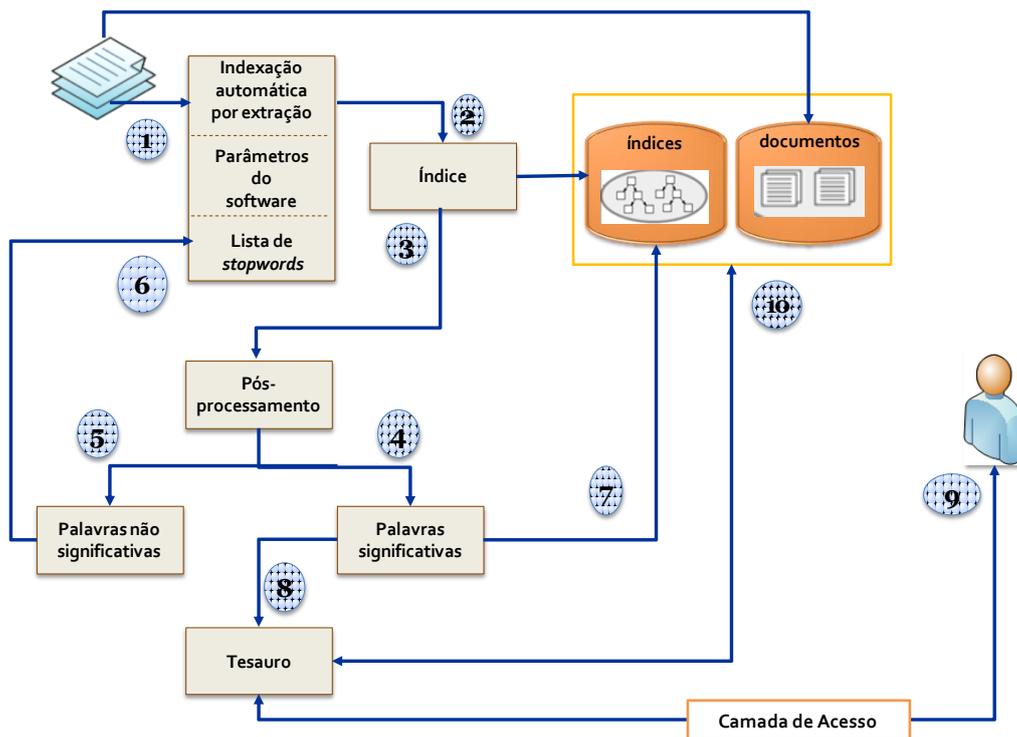


Figura 15 - Processo de indexação

Fonte – Elaboração própria

5.2 Conclusão

O capítulo tratou das diretrizes para uma proposta de política de indexação automática, selecionadas a partir da revisão de literatura e adequadas para o perfil informacional do Banco do Brasil.

É um instrumento passível de ser utilizado nos processos internos do Banco do Brasil, mas que demanda um esforço da administração da empresa em aderir a ideia e patrocinar sua implantação. Além do que, serão necessários investimentos na capacitação dos funcionários e na aquisição de recursos de infraestrutura tecnológica.

6 Considerações finais

No atual cenário de produção e busca de informações são necessários estudos que abordem a organização e representação da informação dos documentos digitais. Esta pesquisa se propôs, por meio dos seus objetivos geral e específicos, definir diretrizes para uma política de indexação automática para documentação digital de natureza bancária regulatória.

Para atingir o primeiro objetivo, foi realizada pesquisa na literatura da Ciência da Informação, onde foram identificados os métodos para tratamento da informação, especificamente relacionados com a indexação automática e os elementos que devem ser considerados para compor uma política de indexação automática.

A indexação possibilita a mediação entre o documento e os usuários do sistema de informação. Este processo evoluiu das etapas realizadas manualmente pelos profissionais da informação, para o processo realizado automaticamente pelos algoritmos dos softwares computacionais. Os benefícios agregados pelos softwares de indexação automática na representação da informação dizem respeito à maior agilidade na leitura do texto e seleção dos termos descritores, aumento da coerência e na redução da subjetividade.

O segundo objetivo, consistiu em avaliar como ocorre a representação da informação dos documentos digitais no Banco do Brasil, referentes ao Acordo de Basileia II – Pilar III. A escolha dessa documentação ocorreu em razão da sua relevância estratégica, da visibilidade no contexto institucional e da necessidade de disseminação das informações entre os usuários internos e externos.

Verificou-se a ausência de métodos de tratamento formalizado do conteúdo dos documentos, fato que reforçou a necessidade de definição de procedimentos operacionais corporativos voltados à indexação automática, como também a necessidade de capacitação dos funcionários para desempenhar tais atividades. A avaliação dos documentos não considerou a

documentação produzida anteriormente, para atender às versões anteriores do Acordo, uma vez que o tema transparência bancária foi introduzido a partir do Acordo de Basileia II.

O terceiro objetivo foi a identificação no Acordo de Basileia II dos elementos essenciais que orientam a definição de diretrizes para uma política de indexação automática. Entende-se, dentro deste contexto informacional, que a política de indexação não pode ser considerada como um elemento estático e definitivo, mas sim dinâmico, que evolui em conjunto com as necessidades de informação dos usuários, contribuindo para que os processos decisórios e operacionais sejam assistidos por informações de qualidade obtidas em fluxos informacionais efetivos.

Espera-se que esta pesquisa possa cooperar, no âmbito da Ciência da Informação, com novas reflexões e análises sobre a indexação automática na representação do conteúdo dos documentos digitais.

Em relação aos trabalhos futuros sugere-se o desenvolvimento de pesquisas que explorem os aspectos semânticos da linguagem bancária para construção de linguagens documentárias específicas e estudos da aplicação do processo de indexação automática para documentação digital de organizações públicas e privadas.

7 Referências bibliográficas

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 8, n. 15, p. 18-40, 2003. Disponível em:

<<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p18>>. Acessado em: set./2012.

ARAUJO JR., Rogério Henrique. *Precisão no Processo de Busca e Recuperação da Informação*. Brasília: Thesaurus, 2007, 176 p.

BACEN, Banco Central do Brasil. *Recomendações Basileia*. Disponível em: <<http://www.bcb.gov.br/?recbasileia>> Acesso em: jan./2013.

BORKO, Harold. *Information science: what is it?* American Documentation, Washington, v. 19, n. 1, p. 3-5, Jan. 1968. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.5090190103/abstract>>. Acesso em: mar./2011.

BANCO DO BRASIL. *Ambiente de GED*. Disponível em: <<https://intranet.bb.com.br>> . Acesso em jun./2012.

BANCO DO BRASIL. *Relações com Investidores, Relatório de Gestão de Riscos*. Disponível em:

<<http://www.bb.com.br/portalbb/page22,136,3604,0,0,1,8.bb?codigoNoticia=28847&codigoRet=5494&bread=1&codigoNoticia=28847&codigoMenu=208>>.

Acesso em mar./2013.

BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. *Modern information retrieval*. Second editon, Pearson, 2011.

BAPTISTA, Dulce. Maria.; ARAÚJO JR., Rogério Henrique de; CARLAN, Eliana. O escopo da análise de informação. In: ROBREDO, J. e BRÄSCHER, M. (Orgs.) *Passeios no Bosque da Informação: estudos sobre representação e organização do informação e do conhecimento*. Brasília DF: IBICIT, 2010, p. 335. Edição eletrônica. Disponível em: <<http://www2.senado.gov.br/bdsf/handle/id/189812>>. Acesso em: mar./2011

BRASCHER, Marisa; CAFÉ, Lígia. Organização da informação ou organização do conhecimento? In: Encontro Nacional De Pesquisa Em Ciência Da Informação, 9, 2008, São Paulo. *Anais*. Disponível em: <<http://www.ancib.org.br/media/dissertacao/1835.pdf>>. Acesso em: mai. /2012.

CALDERON, Wilmara R.; CORNELSEL, Julce M.; PAVEZI, Neiva; LOPES, Maria A.. O processo de gestão documental e da informação arquivística no ambiente universitário. *Ciência da Informação*, Brasília, v. 33, n. 3, p. 97-104, set./dez. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n3/a11v33n3.pdf>> Acesso em: mai./2013.

CÂMARA JR. Auto Tavares. Indexação automática de acórdãos por meio de processamento de linguagem natural. *VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação*. GT 2 – Organização e Representação do Conhecimento Comunicação oral. Salvador, 2007. Disponível em: <http://www.enancib.ppgci.ufba.br/prog_gt2.htm> Acesso em mai./2013

CARDOSO FILHO, Jair. Cunha.; SANTOS, Márcia. Mazo. Processo e temas selecionados In: ALVARES, Lilian. (Org.). *Organização da Informação e do Conhecimento: conceitos, subsídios interdisciplinares e aplicações*. São Paulo: B4 Editores, 2012. 248 p. Capítulo 4, p. 185-223.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 2005. Disponível em: <www.brapci.ufpr.br>. Acesso em: ago./2012.

CARVALHO, Demerval Bicalho; CALDAS, Marcelo Petroni. Basileia II: a abordagem para acompanhamento de risco operacional nas instituições financeiras. *Resenha BM&F Brasil*, São Paulo, n. 169, p. 76-84, jul./set. 2006, Disponível em:<www.febraban.org.br> Acesso em: ago./2012.

CHOO, Chun Wei. A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. Tradução Eliana Rocha. 2ª Ed., São Paulo: Editora Senac, São Paulo, 2006.

CHOWDHURY, G.G. *Introduction to modern information retrieval*. 3rd Ed. 2010, Neal-Schuman, New York.

CONARQ - *Glossário*: Versão 5.0 - Agosto de 2009. Disponível em: <https://d99edb8a-a-fa05922d-sites.googlegroups.com/a/arquivista.org/arquivista/arquivos/Arquivista.org-CONARQ-CTDE-Gloss%C3%A1rio_vers%C3%A3o_5.0.pdf?>. Acesso em: mai./2013.

CURRÁS, Emília. *Ontologias, taxonomia e tesouros em teoria de sistemas e sistemática*. Tradução Jaime Robredo. Brasília: Thesaurus, 2010.

DIAS, Eduardo W., Contexto digital e tratamento da informação. *DataGramaZero – Revista de Ciência da Informação* – v.2, n.5, out/01. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: set./2012.

DICIONARIO Brasileiro de Terminologia Arquivística. Rio de Janeiro: Arquivo Nacional, 2005.

DAHLBERG, Ingetraut. *Knowledge organization: a new science? Knowledge Organization*. v. 33, n. 4, p. 11-19, 2006.

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

FIORAVANTE, Felipe. Tendências emergentes em mecanismos de busca. Disponível em: <<http://www.terraforum.com.br/biblioteca/Documents/libdoc00000175v002Tendencias%20emergentes%20em%20mecanismos%20de%20bus.pdf>>. Acesso em: mar./2012

FLORIDI, Luciano. On defining library and information science as applied philosophy of information. *Social epistemology*, v. 16, n. 1, p. 37-49, 2002.

Disponível em:< <http://www.tandfonline.com/doi/abs/10.1080/02691720210132789>>.

Acesso em: jul./2010

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v.1, n.1, p. 60-90, jul./dez. 2003. Disponível em:

<http://www.sbu.unicamp.br/seer/ojs/index.php/sbu_rci/article/viewFile/267/167>. Acesso em: mar./2012.

FUJITA, Mariângela Spotti Lopes, RUBI, Milena Polsinelli. O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância do bibliotecário. *Perspectiva Ciência da Informação*, Belo Horizonte, v.11, p. 48-66, jan./abr. 2006. Disponível em: <www.scielo.br/pdf/pci/v11n1/v11n1a05.pdf> Acesso em: out./2011.

FUJITA, Mariângela Spotti Lopes. A política de indexação para representação e recuperação da informação In: GIL Leiva, Isidoro e FUJITA, Mariângela Spotti Lopes (editores). *Política de Indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012, p. 32-105.

FUJITA, Mariângela Spotti Lopes. Elaboração e avaliação da política de indexação na formação inicial do indexador. In: GIL Leiva, Isidoro e FUJITA, Mariângela Spotti Lopes (editores). *Política de Indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012b, p. 32-105.

GALVÃO, Maria Cristina Barbosa. A análise, a síntese, a representação da informação e a gestão do conhecimento nas empresas. In: RODRIGUES, G. M.; LOPES, I. L. (orgs.). *Organização e representação do conhecimento na perspectiva da Ciência da Informação*. Brasília: Thesaurus, 2003. v. 2, p. 230-239.

GIL, Antonio Carlos. *Como elaborar projetos de pesquisa*. 5. ed. São Paulo: Atlas, 2010.

GIL LEIVA, Isidoro. *Manual de indización: teoría y práctica*. Gijón: Trea, 2008.

GIL LEIVA, Isidoro. Aspectos conceituais da Indexação. In: GIL Leiva, Isidoro e FUJITA, Mariângela Spotti Lopes (editores). *Política de Indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012, p. 32-105.

GONZÁLEZ, José A. Moreira. Palavra, Termo, Conceito: das linguagens documentária até os vocabulários semânticos para a Web. In: FREITAS, Lídia S. de; MARCONDES, Carlos Henrique; RODRIGUES, Ana Célia (Orgs.) *Documento: gênese e contextos de uso*. Niterói: EdUFF, 2010.

GUIMARÃES, José Augusto Chaves. Políticas de análisis y representación de contenido para la gestión del conocimiento en las organizaciones. *Scire*, Zaragoza, v. 6, n. 2, p. 48-58, jul./dic. 2000. Disponível em: <<http://ibersid.eu/ojs/index.php/scire/article/view/1133>> Acesso em: dez./2012.

GUIMARÃES, José Augusto Chaves. Análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais In: RODRIGUES, G. M. & LOPES, I. L. (org) *Organização e Representação do Conhecimento na Perspectiva da Ciência da Informação*. Brasília: Thesaurus, 2003.

GUIMARÃES, José Augusto Chaves; SALES, Rodrigo, Análise documental: concepções do universo acadêmico brasileiro em Ciência da Informação. *DataGramaZero – Revista de Ciência da Informação*, Rio de Janeiro,. v. 11, n. 1, fev/2010, Artigo 02. Disponível em:<http://www.dgz.org.br/fev10/Art_02.htm>. Acesso em: nov./2012.

HJØRLAND, Birger. Automatic Indexing. In: *Lifeboat for Knowledge Organization*, 2008. Disponível em: <http://www.db.dk/bh/lifeboat_ko/CONCEPTS/automatic_indexing.htm>. Acesso em: dez./2010.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. *Ciência da Informação*, Brasília, v. 25, n. 2, 1996. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/435>> Acesso em: jun./2012.

LARA, Marilda Lopes G., Documento e significação na trajetória da epistemologia da Ciência da Informação. In: FREITAS, Lídia S. de; MARCONDES, Carlos Henrique; RODRIGUES, Ana Célia (Orgs.) *Documento: gênese e contextos de uso*. Niterói: EdUFF, 2010.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. 2. ed. Brasília: Briquet de Lemos, 2004.

LE COADIC, Yves-François. *A ciência da informação*. 2ª Ed., Brasília, DF: Briquet de Lemos / Livros 2004.

LIMA, José Leonardo Oliveira; ALVARES, Lilian. Organização e representação da informação e do conhecimento. In: ALVARES, Lilian. (Org.). *Organização da Informação e do Conhecimento: conceitos, subsídios interdisciplinares e aplicações*. São Paulo: B4 Editores, 2012. 248 p. Capítulo 1, p. 21-48.

LOUSADA, Mariana; LOPES, Elaine Cristina; FUJITA, Mariângela Spotti Lopes; VALENTIM, Lígia Pomim. Políticas de indexação no âmbito da gestão do conhecimento organizacional. *Informação & Sociedade*, João Pessoa, v.21, n.1, p.191-202, jan./abr.2011. Disponível em: <<http://www.brapci.ufpr.br>> Acessado em: 24/mar/2013.

McGEE, James; PRUSAK, Laurence. *Gerenciamento estratégico da informação*. Rio de Janeiro: Campus, 1994.

MIRANDA, Antônio; SIMEÃO, Elmira. A conceituação de massa documental e o ciclo de interação entre a tecnologia e o registro de conhecimento. *Data Gram Zero, Revista de Ciência da Informação*, v.3, n.4 ago/02. Disponível em: <http://www.dgz.org.br/ago02/Art_03.htm>. Acesso em: set/2012.

NAVES, Madalena Martins Lopes. Princípios e Técnicas de indexação, com vistas à **recuperação** da informação. Universidade Federal de Minas Gerais: Belo Horizonte, 2004.

NOVELINO, Maria Salet F., Instrumentos e metodologias de representação da informação. *Informação & Informação*, Londrina, v.1, n.2, p. 37-45, jul./dez. 1996. Disponível em: < <http://www.brapci.ufpr.br> > Acesso em: nov./2012.

RUBI, Milena Polsinelli, Política de indexação. In: GIL LEIVA, Isidoro; FUJITA, Mariângela. Spotti Lopes., (editores) *Política de Indexação*. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012.

ROBREDO, Jaime. Redes de informação e de gestão do conhecimento: modelagem e estrutura de informações In: TARAPANOFF, K. (org.) *Inteligência, informação e conhecimento em corporações*. Brasília: IBICT, UNESCO, 2006, p. 303-335.

ROBREDO, Jaime e CUNHA, Murilo Bastos da. *Documentação de hoje e de amanhã*. Brasília: Thesaurus, 2003.

ROSSETTI, Adroaldo, MORALES, Aran. B. O papel da tecnologia da informação na gestão do conhecimento, *Ciência da Informação*., Brasília, v. 36, p. 124-135, jan./abr. 2007. Disponível em:< <http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/795/644>> Acesso em:out./2013.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. *em Ciência da Informação*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: mar./2012.

SILVA, Edna Lúcia da; MENEZES, Estela Muszkat. *Metodologia da pesquisa e elaboração de dissertação*. 4 edição, Florianópolis: UFSC, 2005.

SILVA, Maria dos Remédios; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, Campinas, 16(2): 133-161, maio/ago., 2004. Disponível em: <<http://www.brapci.ufpr.br>>. Acessado em: dez./2010.

SOUZA, Joice Cleide Cardoso Ennes de. Avaliação de linguagem de indexação aplicada à informação jornalística: estudo de caso. Niterói, 2007. 156f. Dissertação Mestrado Ciência da Informação - Universidade Federal Fluminense. Instituto Brasileiro de Informação em Ciência e Tecnologia, 2007. Disponível em: <http://www.bdtd.ndc.uff.br/tde_busca/arquivo.php?codArquivo=2365>. Acesso em: mai./2013.

SOUZA, Renato Rocha. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspect. ciênc. inf.*, Belo Horizonte, v. 11, n. 2, ago./2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000200002&lng=en&nrm=iso>. Acesso em: set./2012.

TARAPANOFF, Kira. (org.) *Inteligência, informação e conhecimento em corporações*. Brasília: IBICT, UNESCO, 2006.

TAYLOR, Arlene G.; JOUDREY, Daniel N. *The organization of information*. 3 ed.; Greenwood Publishing Group, Inc. 2009.

VIEIRA, Simone Bastos. Indexação Automática e manual: revisão de literatura. *Ciência da Informação*, Brasília, 17(1): 43-57, jan./jun.1998. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1391/1017>> Acesso em: abr./2012.

APÊNDICE

Apêndice 01 - Instrumento de coleta de dados

Roteiro das entrevistas

Está entrevista é informal e se baseia nas seguintes perguntas:

- 1- A qual público se destinam os documentos produzidos na sua equipe?
- 2- Quais os tipos de documentos são criados nas atividades realizadas na área?
- 3- De que forma o conteúdo dos documentos é representado?
- 4- Na sua equipe existem funcionários capacitados para atuar no tratamento do conteúdo dos documentos?