

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI

Programa de Pós-Graduação em Educação

Aline Pereira da Costa

**SISTEMA DE INDEXAÇÃO AUTOMÁTICA DE OFÍCIOS DO DEPARTAMENTO DE
COMPUTAÇÃO DA UFVJM**

Diamantina

2016

Aline Pereira da Costa

**SISTEMA DE INDEXAÇÃO AUTOMÁTICA DE OFÍCIOS DO DEPARTAMENTO DE
COMPUTAÇÃO DA UFVJM**

Dissertação apresentada ao programa de Pós-Graduação em Educação da Universidade Federal dos Vales do Jequitinhonha e Mucuri, como requisito para obtenção do título de Mestre.

Orientador: Prof. Dr. Cristiano Grijó Pitangui

Diamantina

2016

Janete Cristiana Lucas – CRB6/2071

025.49
C837s

Costa, Aline Pereira da.
Sistema de indexação automática de ofícios do
Departamento de Computação da UFVJM / Aline
Pereira da Costa. – 2016.

69f. : il.

Orientador: Dr. Cristiano Grijó Pitangui.
Dissertação (Mestrado em Educação) –
Universidade Federal dos Vales do Jequitinhonha e
Mucuri, 2016. Programa de Pós-Graduação em
Educação.

1. Vocabulário controlado. 2. Tesouro. 3. Sistema.
4. Indexação. I. Pitangui, Cristiano Grijó. II.
Universidade Federal dos Vales do Jequitinhonha e
Mucuri. Programa de Pós-Graduação em Educação.
III. Título.

Aline Pereira da Costa

**SISTEMA DE INDEXAÇÃO AUTOMÁTICA DE OFÍCIOS DO DEPARTAMENTO DE
COMPUTAÇÃO DA UFVJM**

Dissertação apresentada ao programa de Pós-Graduação em Educação da Universidade Federal dos Vales do Jequitinhonha e Mucuri, como requisito para obtenção do título de Mestre.

Orientador: Prof. Dr. Cristiano Grijó Pitangui

Data de aprovação 30/09/2016.

Prof. Dr. Cristiano Grijó Pitangui
Dep. de Tecnologia e Eng. Civil, Computação e Humanidades/UFVJM

Profa. Dra. Luciana Pereira de Assis
Departamento de Computação/UFVJM

Prof. Dr. Alessandro Vivas Andrade
Departamento de Computação/UFVJM

Prof. MSc Rafael Santin
Departamento de Computação/UFVJM

**Diamantina
2016**

Dedico este título aos meus pais Eliezer e
Maria Bernardete.

AGRADECIMENTOS

Qualquer trabalho que nos dispomos a realizar, embora seja um objetivo pessoal, não poderia ser concluído sem o apoio daqueles que compartilham conosco a caminhada da vida.

Agradeço a Deus que sempre me guiou, orientou e iluminou. A Nossa Senhora Aparecida, meu porto seguro, em todos os momentos da minha vida.

Agradeço carinhosamente aos meus pais Eliezer e Maria Bernardete, que nunca pouparam esforços para que fosse possível eu estudar. Que nunca duvidaram do meu potencial e que sempre me incentivaram a subir os degraus da vida acadêmica. O amor que tenho por vocês vai além da vida.

Aos meus irmãos: Allan, Pingo e Ronan, pela eterna relação de afeto, carinho e respeito. Mesmo estando geograficamente longe, eu nunca me senti distante de vocês. As minhas cunhadas Lícia e Ingrid, por serem tão importante em minha vida, assim como meus irmãos.

Aos meus sobrinhos: Alexia, Mary Hellen, Allan Alexandre, Gabriel e Bernardo. A certeza de estar fazendo algo em que vocês possam se espelhar eu já fico satisfeita.

No âmbito acadêmico, meu agradecimento especial para o meu orientador Prof. Dr. Cristiano Grijó Pitanguí. Esse "louco" apostou em uma bibliotecária na linha de tecnologia. Certa vez ele falou: "Como se constrói um diamante? Pressão e tempo.", eu me sinto lapidada. Quanto conhecimento adquirido, quanto amadurecimento, quanto eu pude perceber que posso mais. As pessoas precisam de gente que apostem nelas. Obrigada por apostar em mim. Eternamente grata!

Não menos importante os Profs Drs Alessandro Vivas Andrade e Luciana Pereira Assis, que confiaram na minha capacidade de estar nesta linha e por serem indiretamente meus co-orientadores. Como foram ricos nossos seminários. Devo muito do que sou hoje a vocês. Obrigada!

Ao Prof MSc Rafael Santin, por aceitar participar da banca como membro externo, fica aqui meu muito obrigada e desde já antecipo meus agradecimentos pela contribuição no meu projeto.

Ao Victor Oliveira Almeida, estudante de graduação em Sistemas de Informação da UFVJM, por ajudar a fazer este sonho de projeto ser real. Sucesso na jornada meu caro.

Aos amigos bibliotecários, Eliana, Welerson e Allan, por terem me ensinado o sentido do trabalho em equipe. Agradeço ainda por terem assumido o meu trabalho e as minhas atribuições em todas as vezes que precisei me ausentar para aulas em Diamantina. Muito obrigada!

Aos assistentes e auxiliares das Bibliotecas ICV/UFJF e ICSA/UFJF: Aline Mello, Amélia, Alexandre, Henrique, Athos, Diogo, Talitely e Fausto, como eu gosto de vocês! Obrigada por sempre estarem na torcida pelo meu sucesso. Equipe nota mil!

Aos demais colegas e amigos que a UFJF – Campus Avançado Governador Valadares, sempre na certeza de estarmos tentando fazer o melhor para nossa comunidade. Juntos somos mais!

A minha amiga/irmã Alessandra Gomes Ferreira, presente da graduação UFMG. Pelo carinho constante comigo, pela ajuda na revisão do texto. Eu sempre te disse que conseguiria, mas com seu apoio certamente seria mais fácil. Obrigada minha irmã. Agora é a sua vez!

As minhas amigas Fernanda de BSB; Sheila e Juliana de SAO, provas reais que não existe distância para amizades verdadeiras. Obrigada meninas por existirem em minha vida!

Aos especiais amigos que Arinos me proporcionou. Elza, Gustavo, Luana, Mires, Kesiane, Diego e Vera. Aquele ano foi um ano difícil, longe de tudo e de todos e foi em vocês que eu me apoiei até que conseguisse voltar para mais próximo dos meus; isso não tem preço. Obrigada!

A prima Lourdinha e o primo Fernando meus referenciais acadêmicos. Obrigada pelo exemplo nesse caminho que optei seguir por ver em vocês a satisfação da vitória alcançada.

Ao primo/irmão/amigo Giovanni, você não imagina o quanto é importante na minha vida. Obrigada por sempre se preocupar comigo.

As primas Ge, Adriana e Sílvia, minhas irmãs mais velhas! Amo vocês!

A todos os meus tios e primos, pelo carinho sempre que temos o prazer de estarmos juntos.

Ao meu amigo Altamir, por me apoiar e incentivar na vida acadêmica. Não lembro de nenhum encontro nosso que este assunto não foi pauta. Obrigada!

Aos Parças: Viviane, Cristiane, Priscila, Greyd, João, Josué e Salvador, amigos do mestrado, que levarei para a vida toda. Na certeza de termos criados

laços enquanto passávamos os apertos acadêmicos juntos.

Ao grupo dos Travas: Willcker, Gustavo, Leandro, Clariana, Meire, Sandra, Boreu, Jader, Gton, Wanderson, Luciano e Estelita, por terem a capacidade de colocar cor em qualquer dia cinza. Mais de 20 anos de amizade não é para qualquer um.

Aos meus amigos bibliotecários: Cléia, Georgia, Josiane, Janett, Marizete, Raquel, Fred, Fabian, Daniele, Isabela, Rosimeire, por desde a graduação estarmos juntos e em contato, tirando dúvidas e reforçando cada vez mais a nossa amizade.

Aos amigos de Nova Era e Itabira que acompanharam de perto minha trajetória.

Ao meu amigo bibliotecário, Felipe Lopes Alves, por ser meu ombro nos momentos mais difíceis dessa vida acadêmica.

Ao amigo cientista da computação Haroldo Lage, por me tirar dúvidas técnicas ao longo do curso. Obrigada!

Agradeço finalmente a UFVJM, por me proporcionar novos rumos a minha vida. Agradeço também o período que me foi autorizado usufruir a Casa de Apoio no período de aulas.

Aos amigos e colegas de facebook, que muito acompanharam minha trajetória.

Caso tenha esquecido de alguém, fica aqui meu muito obrigada na certeza de sempre caminhar rumo ao conhecimento e tentar usufruir dele da melhor maneira possível convergindo para o meu próximo, para a sociedade.

“Se vi mais longe foi por estar de pé sobre ombros de gigantes”
Isaac Newton

RESUMO

Este projeto surgiu de uma deficiência do Departamento de Computação da UFVJM (Universidade Federal dos Vales do Jequitinhonha e Mucuri), em armazenar e recuperar seus ofícios. Tal Departamento possui dificuldades na organização e armazenamento eficiente destes ofícios o que inviabiliza o acesso aos documentos e dispende muito tempo na localização e recuperação da informação. Diante disso, foi desenvolvido um Sistema de Indexação Automática utilizando-se técnicas da biblioteconomia e técnicas computacionais que visa automatizar o processo de indexação de novos ofícios, otimizar a recuperação e democratizar o acesso a informação. O banco de dados do Sistema foi construído baseado em um vocabulário controlado: o tesouro. O tesouro é um tipo de vocabulário controlado, mais complexo, que trabalha com ambiguidade dos termos, sinonímia, relações hierárquicas e associativas e foi elaborado a partir de conceitos selecionados e seus termos relacionados presentes nos ofícios. O escopo inicial do projeto permeia o espaço de 2011 a 2014, totalizando 239 ofícios. Para a alimentação do Sistema com os novos documentos que surgirão propõe-se a auto alimentação do tesouro que fará a análise de relevância de novos termos nos novos ofícios através de um algoritmo em construção. O sistema estará em ambiente virtual, para que o acesso seja democratizado e o processo de tratamento de novos ofícios seja automatizado. Sendo positiva a implementação do projeto, sugere-se que os demais departamentos da UFVJM utilizem o mesmo sistema para organização dos seus documentos, ganhando agilidade nos processos e satisfação do usuário final na localização do que procura.

Palavras chave: Vocabulário Controlado 1. Tesouro 2. Sistema 3. Indexação 4.

ABSTRACT

This project arose from an actual deficiency of the UFVJM's Department of Computer Science (Universidade Federal dos Vales do Jequitinhonha e Mucuri) to store and retrieve archives. This institution has been having to put up with problems to store documents efficiently. This difficulty prevents access to documents and imposes a heavy burden to the department's staff, in terms of time to locate and retrieve information. Therefore, we developed a system to index information efficiently. To this end, we have used techniques borrowed from various areas within information science. Such techniques allowed us to automate the process of indexing new archives, optimize their recovery and, as a result, we have been able to democratize the access to information. The proposed database was built based on a controlled vocabulary: the thesaurus. The thesaurus is a controlled type of vocabulary, more complex, working with ambiguity of terms, synonymy, hierarchical relationships and associativity. It was prepared from selected concepts and terms related to these concepts in the archives. The initial scope of the project covers a time period from 2011 to 2014, totaling 239 offices. To feed the system with new documents yet to emerge we propose a self-feeding mechanism. This self-feeding process will lead to the relevant analysis of new terms in the new archives through an algorithm, currently under implementation. The system will be in a virtual environment, which ensures not only the democratic access to information, but also the automatic handling of new documents. In case this project obtains positive feedback, we shall suggest that other UFVJM departments use the same system for organizing documents; hence, decreasing their response time, and improving the experience of their end-users.

Keywords: Controlled Vocabulary 1. Thesaurus 2. System 3. Indexing 4.

LISTA DE ILUSTRAÇÕES

Figura 1 – Crescimento estrutural da complexidade entre tipos de Vocabulário Controlado.....	32
Figura 2 – Influência de fatores nos resultados de uma busca numa base de dados	38
Figura 3 – Construção de vocabulário controlado	39
Figura 4 – Abrangência da Lei de Acesso a Informação	43
Figura 5 - Página principal do Sistema de Ofícios do DECOM	45
Figura 6 – Indexação de ofício	46
Figura 7 – Página de inserção de novo ofício no Sistema	48
Figura 8 - Recuperação de um ofício	49
Figura 9 – Modelo de ofício	50
Figura 10 – Modelo da tela de busca do ofício.....	51

LISTA DE QUADROS

Quadro 1- Relacionamentos Conceituais.....	29
--	----

LISTA DE ABREVIATURAS E SIGLAS

CAT	Categoria
DECOM	Departamento de Computação
INMETRO	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
LD	Linguagem Documentária
LN	Linguagem Natural
NE	Nota Explicativa
SRI	Sistema de Recuperação da Informação
TA	Termo Associado
TE	Termo Específico
TG	Termo Genérico
TR	Termo Relacionado
UFAL	Universidade Federal de Alagoas
UFJF	Universidade Federal de Juiz de Fora
UFSC	Universidade Federal de Santa Catarina
UFVJM	Universidade Federal dos Vales do Jequitinhonha e Mucuri
UP	Usado por

SUMÁRIO

1 INTRODUÇÃO	20
1.1 Justificativa.....	22
1.2 Objetivos	23
1.2.1 <i>Objetivo geral</i>	23
1.2.2 <i>Objetivos específicos</i>	23
1.3 Apresentação da dissertação	23
2 FUNDAMENTOS TEÓRICOS E METODOLÓGICOS	25
2.1 Linguagem documentária.....	25
2.2 Vocabulário controlado: tesouro.....	27
2.3 Indexação.....	34
2.4 O uso do vocabulário controlado em sistemas de recuperação da informação	40
2.5 Lei de Acesso a Informação	42
3 SISTEMA DE OFÍCIOS DO DECOM	43
3.1 Metodologia	43
3.2 Indexação de um Ofício	45
4 TESTES COM O SISTEMA	48
4.1 Recuperação de um Ofício	48
4.2 Exemplo de Indexação e Recuperação de um Ofício.....	50
4.3 Pendências do Sistema de Ofícios do DECOM	52
4.4 Resultados parciais/finais	52
5 CONSIDERAÇÕES FINAIS	53
REFERÊNCIAS	54
APÊNDICE A	58
APÊNDICE B	66

1 INTRODUÇÃO

Diante do conceito de informação, corroborado com Capurro e Hjørland (2007), como registro e comunicação do conhecimento e seu forte impacto na sociedade atual, organizar a informação tornou-se um processo fundamental na medida que vêm crescendo o volume de informações e a dificuldade de se localizar o que se procura, seja em um ambiente físico ou digital. Por isso faz-se necessário, delimitar o contexto de sua aplicação para utilizá-la de maneira mais adequada.

Com as novas tecnologias da informação houve um aumento considerável na produção, registro e intercâmbio de informações em diversas áreas do conhecimento. Nasce então a preocupação na armazenagem, organização e recuperação das informações.

Neste sentido a proposta deste estudo trata da informação que se localiza em ambiente administrativo e organizacional. Rousseau e Couture (1998) consideram que a informação deve ser tratada, numa empresa, como um recurso tão importante quanto os recursos humanos, materiais ou financeiros.

A Biblioteconomia é considerada como uma área do conhecimento, na medida em que compreende um conjunto de organismos, operações técnicas e princípios que dão aos documentos a utilização máxima, em benefício da humanidade (SHERA, 1980).

É a disciplina que se preocupa com o fluxo da informação desde a sua origem até a sua utilização pelo usuário final. Trata do “diagnóstico, planejamento, implementação, organização, administração, disseminação e uso da informação em bibliotecas, centro de documentação, sistemas de informação, sites, entre outros” (HOLANDA e NASCIMENTO, 2010, p.02).

O gerenciamento de informações em ambientes organizacionais, de acordo com PIZARRO e DAVOCK (2008), se inicia desde a coleta de dados até a disponibilização e disseminação das informações tendo em vista sempre os interesses e necessidades do usuário e os objetivos da empresa. A gestão das informações deve ser estruturada e sistematizada em sistemas de fácil acessibilidade em um processo de tratamento da informação controlado e realizado por profissionais da informação.

Os profissionais da informação, como explica Rezende (2002), possuem a

função de mediadores entre as necessidades informacionais dos usuários e o diverso e crescente universo de informações disponíveis.

A Ciência da Computação é a disciplina que busca construir uma base científica para tópicos como projeto e programação de computadores, processamento de informação, soluções algorítmicas de problemas e o próprio processo algorítmico. Ela fornece a estrutura das aplicações computacionais atuais, bem como a base para a futura infraestrutura da computação (BROOKSHEAR, 2013).

A produção incessante de informações aliada às novas tecnologias assumiu um papel importante no compartilhamento do conhecimento. Trata-se de grandes volumes de informação e conhecimento acessíveis por meio de Sistemas de Recuperação de Informação (SRIs) que são utilizados e compartilhados por inúmeras pessoas ao mesmo tempo, como discuti Cardoso (2000), com o intuito de melhorar a busca e a apresentação de informações para responder às necessidades dos usuários.

Segundo Araújo (1995), os SRIs são sistemas que organizam, processam e recuperam as informações contidas em documentos neles registrados com a finalidade de maximizar o uso da informação em um ciclo contínuo de entrada-tratamento-saída de informações. Na entrada, os documentos, através da indexação, são representados de forma que possam ser armazenados no SRI durante a etapa de tratamento. Na saída, as informações são disponibilizadas e podem ser recuperadas pelo usuário.

A forma com que o documento é representado nos SRIs, de acordo com Araújo (1995), influi diretamente nas possibilidades do uso da informação. Para que a recuperação seja relevante e bem-sucedida é necessário que os dados sejam processados e armazenados com qualidade, ampliando as oportunidades do usuário final de satisfazer as suas necessidades informacionais.

Este trabalho tem o objetivo de através do uso de vocabulário controlado, criar um Sistema de Indexação Automática para localização dos ofícios do Departamento de Computação da UFVJM – Universidade Federal dos Vales do Jequitinhonha e Mucuri

Ressalta-se que este projeto é interdisciplinar envolvendo a Biblioteconomia e a Ciência da Computação ambas aplicadas no contexto de resolver o problema alicerçado neste projeto. Do lado da Biblioteconomia, a contribuição é feita pela

criação do Tesouro que irá solucionar o problema com os termos dos ofícios de modo que seja possível localizar o que se procura, além de assegurar uma boa indexação. Já no lado da Ciência da Computação, é implementado o Sistema propriamente dito, para que a busca seja concluída com sucesso.

1.1 Justificativa

O ofício “é uma correspondência oficial, enviada normalmente a funcionários ou autoridades públicas. É o tipo mais comum de correspondência oficial expedido por órgãos públicos”. Seu destinatário pode ser outro órgão público ou particular (PORTAL EDUCAÇÃO, 2013, p.01). O Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), não possui um local, onde seja possível pesquisar os ofícios já expedidos. Quando há uma demanda por determinado ofício, recorre-se a e-mails enviados ou a alguma pasta arquivada no computador ou, ainda, fisicamente - o que demanda muito tempo na busca.

A indexação, segundo a NBR 12676 é: “o ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação”.

Diante do cenário apresentado, observou-se a necessidade de desenvolver um sistema de indexação automática que represente e armazene os ofícios de forma que seja mais fácil e rápida sua recuperação.

Um vocabulário controlado é definido por Naves (2004, p. 10-11), “como um tipo de linguagem artificial. Esta é construída cuidadosamente, através do processo minucioso da escolha dos termos a serem usados, o significado de cada um e as relações que apresentam”. Para que este projeto fosse desenvolvido foi necessário desenvolver uma linguagem de indexação, escolher os melhores termos para a construção de um vocabulário controlado.

Para sanar tal deficiência na busca pelos ofícios, propõem-se desenvolver um vocabulário controlado. O tesouro, tipo de vocabulário controlado, servirá como base para uma indexação automática dos documentos e, assim, classificar os ofícios, de tal maneira que o usuário localize rapidamente o que procura.

Após a construção do vocabulário controlado, será necessário indexar os

ofícios utilizando os termos definidos no tesouro.

1.2 Objetivos

1.2.1 Objetivo geral

Propor a criação de um sistema de indexação automática, o Sistema de Ofícios do DECOM, através do uso de vocabulário controlado, para localização dos ofícios do Departamento de Computação da UFVJM.

1.2.2 Objetivos específicos

- Construir um tesouro como instrumento base para a indexação automática do Sistema;
- Propor um método de ranqueamento dos ofícios, para que a busca retorne em primeiro lugar o que o usuário solicita.
- Criação de um sistema de indexação automática online para armazenar/recuperar os ofícios.

1.3 Apresentação da dissertação

A organização do texto resultante desta pesquisa baseia-se em capítulos assim divididos:

Capítulo 1 – Introdução: apresenta os fatores que impulsionaram o desenvolvimento desta pesquisa; suas justificativas; o problema central e os objetivos pretendidos.

Capítulo 2 – Fundamentos teóricos e metodológicos: apresenta o referencial teórico em que o trabalho foi norteado e a metodologia utilizada para a construção do vocabulário controlado; a diferença entre linguagem natural e linguagem controlada; e explica a indexação.

Capítulo 3 – Sistema de Ofícios do DECOM: a metodologia do Sistema; o algoritmo simplificado e a indexação de um ofício.

Capítulo 4 – Testes com o Sistema: recuperação de um ofício; exemplo de indexação e recuperação de um ofício; algumas telas do Sistema; pendências do sistema; resultados parciais/finais.

Capítulo 5 – Considerações finais: Relata a elaboração do Sistema de Ofícios do DECOM, ainda em desenvolvimento; a construção do Tesouro; e a possibilidade de implementar o Sistema em outros departamentos da UFVJM.

2 FUNDAMENTOS TEÓRICOS E METODOLÓGICOS

Neste capítulo, apresentam-se fundamentos teóricos e metodológicos que orientam esta pesquisa. A revisão de literatura abordou os temas essenciais para a reflexão e discussão referente a melhor forma de construção do vocabulário controlado visando a recuperação da informação eficaz e a satisfação do usuário final.

2.1 Linguagem documentária

Vivemos na sociedade da informação e para que o conhecimento produzido “não se perca e possa ser compartilhado, ele é registrado num dado suporte: livro, imagem, foto, disco, etc” (CINTRA, 1994, p. 14).

A globalização, o advento da Internet e o aprimoramento das tecnologias de informação e comunicação provocaram mudanças na produção e disseminação do conhecimento. Houve um grande aumento no número de informações disponíveis a cada minuto, e que a cada dia cresce mais e em grande escala. Diante deste cenário de explosão informacional, como discute Cintra (1994), não há como um indivíduo conhecer tudo e por isso, tende a procurar uma forma de satisfazer as suas necessidades e objetivos procurando informações atuais, confiáveis e disponíveis.

De acordo com a autora, o homem possui uma “tendência natural de compreender, governar e modificar o mundo e, por isso, [...] busca, incansavelmente, encontrar uma ordem para as coisas, [...] uma estrutura capaz de explicar as coisas” (CINTRA, 1994, p. 19). Diante deste contexto, são criados os sistemas de organização do conhecimento com estruturas de representação do conhecimento voltadas para atender as necessidades informacionais dos usuários.

Segundo Hodge (2000) e Boccato (2009), esses sistemas possibilitam a organização do conhecimento, a administração e sua recuperação em ambientes informacionais impressos e eletrônicos. São “instrumentos comunicadores e mediadores da informação, [...] possibilitam a obtenção de resultados precisos e relevantes, que atendam às necessidades de buscas dos usuários” (BOCCATO, 2009, p. 43).

Visando a construção de instrumentos mais eficientes no tratamento da informação, armazenamento e disseminação do conhecimento, investiu-se na elaboração de linguagens documentárias. As Linguagens Documentárias (LDs) também denominadas como Vocabulário Controlado, podem ser definidas como um conjunto limitado de termos autorizados para uso na indexação e busca de documentos e são concebidas “como instrumentos de comunicação documentária, pois, realizam a tradução do conteúdo dos documentos e das perguntas dos usuários para unidades aptas a integrar Sistemas Recuperação da Informação”. (CINTRA, 1994, p. 24).

Conforme explica Sales (2007), as linguagens documentárias auxiliam as atividades de representação dos documentos, proporcionando uma representação padronizada do conteúdo informacional e uma recuperação eficiente.

Entende-se como Linguagem Natural (LN) a linguagem usada habitualmente na comunicação cotidiana, isto é, na fala e na escrita do discurso comum. Os estudos de Cintra (1994) e Sales (2007) demonstram que as LDs têm por finalidade padronizar a passagem da linguagem natural, termos utilizados em uma determinada língua, para a linguagem documentária, “termos e símbolos empregados para a identificação temática nos sistemas de informação”. (CAVALCANTI, 1978, p. 21).

De acordo com Dodebei (2002) tanto o uso da LD quanto o uso da LN apresentam vantagens e desvantagens na indexação e na recuperação da informação. Enquanto a linguagem documentária é precisa, rígida e inflexível, a linguagem natural é altamente expressiva, flexível, no entanto, potencialmente ambígua.

A função das linguagens documentárias, como afirmam Cavalcanti (1978) e Cintra (1994), vai além da simples seleção e tradução de um conjunto restrito de palavras e frases do conteúdo de documentos em linguagem natural, através de um conjunto de regras, símbolos e termos pré-estabelecidos, com o objetivo de indicar os assuntos pertinentes dos documentos; ela tem a função de compartilhar e disseminar a informação; de possibilitar que os diferentes segmentos sociais tenham acesso ao conhecimento produzido.

Na construção das linguagens documentárias, a luz de Gardin¹ (1968 citado por Cintra, 1994, p.25), deve-se observar a presença de três elementos importantes:

- O léxico – lista de descritores/termos;
- Uma rede paradigmática que organiza e traduz as relações essenciais entre os descritores e
- Uma rede sintagmática que expressa as relações de coordenação entre os descritores/termos de acordo com um contexto específico onde estão inseridos.

Dentre as linguagens documentárias podem-se destacar, como cita Sales (2007), os cabeçalhos de assunto, os glossários, os dicionários, os sistemas de classificação, as taxonomias, os tesauros e as ontologias.

2.2 Vocabulário controlado: tesauro

O vocabulário controlado, conforme Araújo (2006), é uma linguagem documentária que identifica e representa uma determinada área específica do conhecimento. O principal objetivo do controle do vocabulário é padronizar, através de normas e procedimentos, a indexação, visando aprimorar o resultado no processo de busca e recuperação da informação. Propicia a verbalização de conceitos e a tradução da linguagem natural para a linguagem documentária das informações contidas nos documentos de forma resumida e padronizada.

Portanto, o vocabulário controlado pode ser entendido como “uma lista de descritores/termos autorizados, visando à recuperação do conteúdo temático de um documento; ele estabelece a organização lógica dos conceitos de uma área do saber”. (ARAÚJO, 2006, p. 71). O controle do vocabulário tem por finalidade:

- Controlar sinônimos, optando por uma única forma padronizada, como remissiva de todas as outras;
- Diferenciar homógrafos;
- Reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si (LANCASTER, 2004, p. 19).

Corroborando com Araújo (2006), são considerados vocabulários controlados os tesauros, os esquemas de classificação bibliográfica e as listas de cabeçalhos de assunto.

¹ GARDIN, J. C. et al. L'automatisation dès recherches documentaires: um modèle général: LE SYNTOL. 2. ed. revue et augmentée. Paris: Gauthier Villars, 1968.

Diante da crescente quantidade de documentos especializados produzidos, o tesouro surgiu da necessidade de manipulação e recuperação da informação e da organização do conhecimento (MOREIRA e MOURA, 2006).

A palavra “tesouro”, de acordo com Motta (1987, p. 21), “provém do vocabulário latino “thesaurus”, o qual deriva do grego “thesaurós” e que foi usado durante vários séculos com o significado de ‘tesouro ou armazém/repositório de palavras’”.

O termo tesouro “foi utilizado pela primeira vez por Brunetto Latini (1220-1294) para designar uma enciclopédia sistematizada que ele chamou de 'Os livros do tesouro'” (MIKHAILOV² 1973 citado por MOTTA, 1987, p. 398). Em 1532, Thierry e R. Etienne publicaram um dicionário da língua latina com arranjo alfabético chamado 'Dicionário ou tesouro do idioma latino'. Em 1572, A. Etienne empregou o termo na obra "*Thesaurus language Graecae*", outro dicionário iniciado por seu pai R. Etienne (MOTTA, 1987). Um dos tesouros mais conhecido é o "*Thesaurus of English Words and Phrases*" de Roget publicado em 1852. Nele, tesouro é definido por “uma coleção arranjada, não em ordem alfabética como um dicionário, mas segundo as ideias que elas representam...”. (ROGET³, 1962, *apud* MOTTA, 1987, p. 22).

Tesouro pode ser entendido como um

sistema de vocabulário baseado em conceitos, incluindo termos preferidos (descritores), termos não preferidos (não descritores) e suas inter-relações, que se aplica a um determinado ramo do conhecimento e que se destina a controlar a terminologia utilizada para a indexação/recuperação de documentos. (MOTTA, 1987, p. 24).

O tesouro possui quatro finalidades importantes para a recuperação da informação: i) controlar os termos usados na indexação, através da tradução do assunto do documento da linguagem natural para a linguagem documentária; ii) assegurar o desenvolvimento de uma política de indexação padronizada; iii) limitar o número de termos necessários atribuídos aos documentos sem que haja a necessidade de incluir termos de conotação mais ampla ou termos associados; iv) servir como um auxiliar de busca em sistemas de recuperação da informação (IBICT, 1984).

² MIKHAILOV, A. I. Et al. Nociones generales acerca del tesouro. In: _____. *Fundamentos de la informática*. Moscú, La Habana, Nauka: Academia de Ciências de Cuba, Inst. Document. E Inf. Científica y Técnica, 1973, v. 2, p. 397-495.

³ ROGET, Peter M. *Roget's thesaurus of English words and phrases*. London: Longman, 1962.

A organização de um tesauro apresenta diversos campos e abreviaturas relacionados, conforme mostra Alvares (2014):

- **Descritor:** Termo escolhido para representar um conceito no Tesauro e que será utilizado na indexação e na recuperação de determinado assunto. Quando houver outros termos que representem o mesmo conceito, antes do termo descritor, constará a sigla USE.
- **Não-descritor:** Termo que, embora descreva o mesmo conceito que o descritor, não é autorizado na indexação, para evitar a proliferação de sinônimos. Antes de cada não-descritor, constará a sigla UP.
- **Nota explicativa (NE):** Fornece uma definição do termo ou uma orientação sobre como utilizá-lo em uma indexação.
- **Termo genérico (TG):** Indica que há relação hierárquica entre termos com relação gênero-espécie e que este descritor representa o termo com o conceito mais abrangente.
- **Termo específico (TE):** Indica os termos subordinados ao termo genérico na cadeia hierárquica.
- **Termo relacionado (TR):** Indica relação entre termos que não formam uma hierarquia (gênero-espécie), mas que são associados mentalmente, de forma automática. Servem para orientar o indexador quanto às possibilidades de encadeamento de descritores e para sugerir ao usuário formas de limitar ou expandir uma pesquisa.
- **Categoria (CAT):** Grande grupo ao qual pertence o descritor. (ALVARES, 2014, p.01).

A estrutura de um tesauro, concordando com Alvares (2014), é norteada pelas relações existentes entre os conceitos. Os conceitos, como explica Araújo (2011), podem ser entendidos como reproduções mentais de um objeto e suas características representados por um termo que denomina, aglutina e descreve essas reproduções e suas propriedades. Os relacionamentos entre conceitos podem ser classificados, de acordo com Gomes (1990) em: relacionamento lógico, ontológico e de efeito, apresentados no Quadro 1.

Quadro 1- Relacionamentos Conceituais

Tipo	Subdivisões	Exemplo
Relacionamento lógico: aquele que possui algumas características em comum	Genérico/específico: são conceitos que pertencem ao mesmo gênero.	Solo / Solo ácido
	Analítico: relação associativa entre dois conceitos quando um deles for uma característica do outro e ambos não fizerem parte da mesma hierarquia.	Arquivamento / Gestão de documentos

	<p>De oposição: são conceitos opostos quase sinônimos. Este relacionamento possui três espécies:</p> <ul style="list-style-type: none"> - Relacionamento de oposição contraditória; - Relacionamento de oposição contrária; - Relacionamento positivo / indiferente / negativo. 	<p>Oposição contraditória. Ex.: Numérico/Não numérico; Oposição contrária. Ex.: Amizade/Inimizade; Positivo/indiferente/negativo. Ex.: Muito valioso/Valioso/Pouco valioso.</p>
<p>Relacionamento ontológico: são relações indiretas entre conceitos. Ocorrem quando o conceito é visto também como um objeto individual.</p>	<p>Partitiva: que apresentam partes constituintes</p>	<p>Avião / Avião anfíbio</p>
	<p>Contiguidade no termo: é um relacionamento de sucessão.</p>	<p>Os três Planos Nacionais de Desenvolvimento (PND): PND / I PND / II PND / III PND</p>
	<p>Material-produto: mostra diferentes estágios na produção de bens, que vão desde a matéria prima até o produto final.</p>	<p>Filme/ Filme fotográfico / Fotografia</p>
<p>Relacionamento de efeito: possui três tipos.</p>	<p>Causalidade: se dá entre conceitos que designam a causa e seu efeito.</p>	<p>Acidente/ Ofuscamento Pista molhada</p>
	<p>Instrumentalidade: se dá entre o conceito relativo ao instrumento e sua ação.</p>	<p>Tabela de temporalidade / Eliminação</p>
	<p>De descendência: se dá entre conceitos que guardam entre si uma relação genealógica, ontogenética ou de estágios da substância.</p>	<p>Larva / Crisálida / Ovo</p>

Fonte: Quadro elaborado pela autora baseada em GOMES, Hagar Espanha. *Manual de elaboração de tesouros monolíngues*. Brasília: MEC/ Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990. p. 40-46.

Os tesouros, de acordo com Gomes (1990), demonstram com maior incidência os seguintes tipos de relações: genérico-específicas, associativas e de equivalência. Vargas (2010) explica que nas relações hierárquicas, os relacionamentos se dão em nível vertical através da subordinação e superordenação baseadas nas semelhanças entre os objetos. Destacam-se aqui as relações gênero/espécie, as partitivas e as exemplares. Nesse tipo de relação há um termo geral identificado pela sigla (TG) e o específico ou subordinado (TE). São exemplos de relações hierárquicas: Ave (TG) / Gavião (TE); Árvore (TG) / Folha (TE); Biblioteca (TG) / Biblioteca escolar (TE). Nos relacionamentos associativos, as relações se dão a nível vertical, por coordenação dos termos que representam

associações mentais, por exemplo, animal/defesa do meio ambiente. São representados pelas siglas (TR) termo relacionado ou (TA) termo associado. Nas relações de equivalência só um termo pode ser utilizado como descritor autorizado. Essas relações são utilizadas em termos sinônimos ou quase sinônimos. São representados pelas siglas USE ou UP (usado por). Exemplo: Letra de imprensa maiúscula / USE Caixa alta (VARGAS, 2010).

Quanto a sua tipologia, como expõe Gomes (1990), o tesauro pode ser classificado quanto à língua, a sua especificidade e ao assunto.

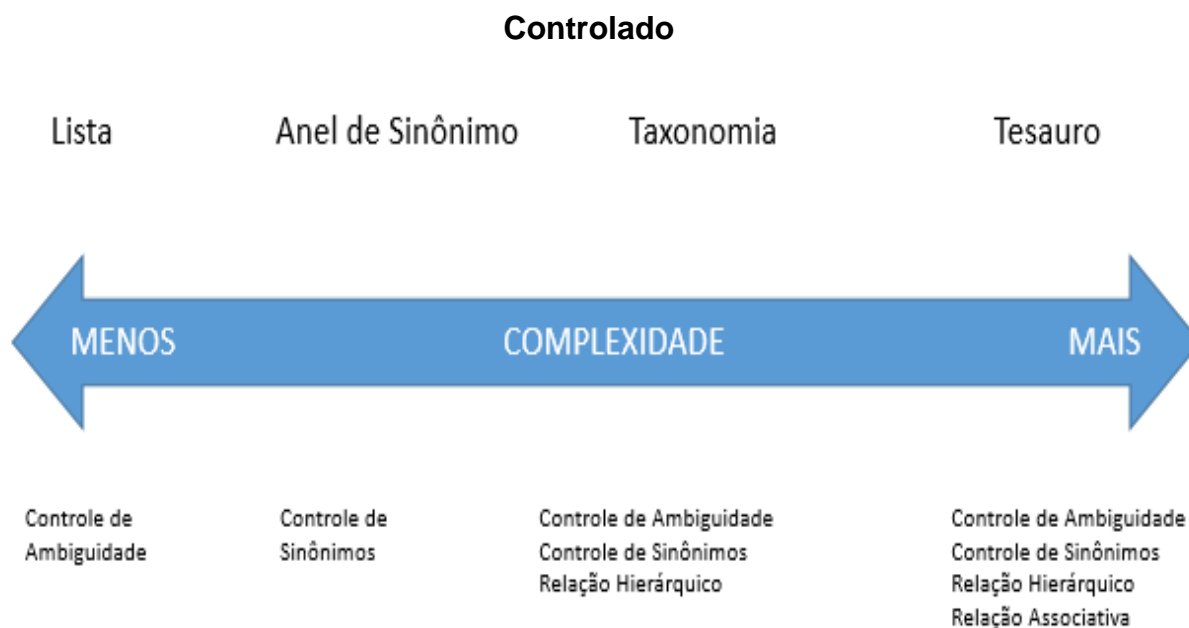
- Quanto à língua:
 - Monolíngues
 - Multilíngues
- Quanto à especificidade:
 - Macrotesauros: os conceitos são mais amplos, com um número menor de descritores e número de remissivas elevado.
 - Microtesauros: o nível de especificidade dos conceitos é maior e se referem a uma área determinada do conhecimento.
- Quanto ao assunto:
 - Voltados para uma missão ou problema (multidisciplinares).
 - Voltados para um determinado assunto (apenas uma disciplina científica).

A construção de um tesauro proporciona o controle total dos termos selecionados, minimizando os problemas de comunicação entre o indexador e o usuário; uma melhor assimilação dos conteúdos dos documentos pelos indexadores através das notas de escopo constantes no tesauro; aumento da efetividade na recuperação e relevância no processo de busca; possibilidade de identificação de conceitos relacionados e redução no tempo de consulta (DODEBEI, 2002). No entanto, a elaboração do tesauro pode trazer algumas desvantagens: aumento dos custos; possibilidade de inadequação dos conceitos utilizados; inadequação aos objetivos da base de dados e favorecimento de falsos resultados no caso de desatualização (DODEBEI, 2002).

Existem diferentes tipos de vocabulários controlados como mostra a Figura 1. São eles: lista, anel de sinônimo, taxonomia e tesauro. O direcionamento da régua mostra um aumento da complexidade das estruturas existentes. Essa complexidade

é determinada pelas formas de relacionamentos que cada tipo de vocabulário possui.

Figura 1 – Crescimento estrutural da complexidade entre tipos de Vocabulário



Fonte: National Information Standards Organization (NISO), 2005

Observando-se a Figura 1, percebe-se que taxonomia e o tesouro são os tipos mais complexos, enquanto que a lista e o anel de sinônimos são mais simples. Enquanto o tesouro possui vários tipos de controles e relacionamentos em sua estrutura de termos, tais como ambiguidade, sinonímia, relações hierárquicas e associativas, a lista considerada simples, é uma lista de palavras em ordem alfabética ou em outra lógica evidente, sem nenhum outro tipo de controle ou relacionamento. Segundo a ANSI/NISO Z39-19-2005, as listas são utilizadas para descrever aspectos do conteúdo de objetos ou entidades que tem um número limitado de possibilidades.

Este trabalho utiliza o tesouro, tipo de vocabulário controlado, devido à complexidade (ambiguidade dos termos, sinonímia, relações hierárquicas e associativas) dos ofícios.

Em 1990, a Norma Internacional de Elaboração de tesouros ANSI/NISO Z39-19-1993, publicada pela *National Information Standards Organization* (NISO), sediada nos Estados Unidos, define tesouro como:

um vocabulário controlado organizado em uma ordem conhecida na qual as relações de equivalência, de homografias, hierárquicas e associativas entre os termos são claramente exibidas e identificadas por meio de indicadores de relação padronizadas, que devem ser empregados de forma recíproca (ANSI/NISO Z39-19-1993, p.09).

É, portanto, como salienta Sales (2008, p. 103), “uma linguagem documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos” que representam uma determinada área do conhecimento.

Quanto à funcionalidade dos tesauros, a norma ANSI-NISO Z39. 19-2003 define quatro funções distintas: a tradução, meio pelo qual a linguagem natural dos autores, indexadores e usuários é traduzida para um vocabulário controlado usado na indexação e recuperação da informação. A consistência que promove a coerência na escolha dos termos no processo de indexação. A indicação de relacionamentos, que indica os relacionamentos semânticos, ou seja, de sentido e significado entre os termos. E a recuperação, que serve como auxílio na busca e na recuperação de documentos.

De acordo com Oliveira (2008) os SRIs foram criados para atender as necessidades de organização, gestão, acesso e recuperação da informação de forma que as demandas informacionais sejam sanadas com maior rapidez. Portanto, a qualidade da indexação e da construção do vocabulário controlado é de vital importância para a alimentação do sistema e para nortear as estratégias de busca.

Os SRIs, em instituições, têm como objetivos coletar, organizar, preservar e disponibilizar informações e a produção do conhecimento organizacional. Elaborar um Sistema em ambiente virtual proporciona uma maior visibilidade dos documentos recebidos e produzidos, e viabiliza o acesso em qualquer tempo e lugar de forma rápida e eficiente.

Essa disponibilidade de informações de fácil acesso em um único repositório contribui para que as necessidades informacionais do usuário final sejam atendidas prontamente, e a gestão documental se torne mais eficiente, segura e ágil. No entanto, para que o sistema possa recuperar informações relevantes relacionadas a necessidade informacional, é de extrema importância que os documentos sejam indexados de acordo com uma linguagem padronizada; um vocabulário controlado. O vocabulário controlado tem por finalidade categorizar, organizar e representar o conteúdo dos documentos, e se torna um instrumento imprescindível para ampliar a eficiência na recuperação da informação por Sistemas.

A análise do acesso de informações relacionadas a legislação e especificadamente a portarias em alguns SRIs de instituições públicas que utilizam vocabulários controlados na indexação de seus documentos mostraram alguns

aspectos importantes que valem ser ressaltados: no site do INMETRO⁴ por exemplo, foi observada a existência de várias bases de dados distintas em um único repositório e campos com possibilidades de buscas padronizadas que diminuem a polissemia dos termos e melhoram as chances de sucesso e precisão na recuperação da informação. Já no portal da UFJF⁵ a categorização deste tipo de informação é realizada apenas por ano e/ou número da portaria, sem campos que orientem ou possibilitem uma busca avançada. Isto é um aspecto que limita o acesso ao documento, pois, o usuário precisa ter um conhecimento prévio para ter acesso à informação desejada. Na UFSC⁶ a pesquisa por portarias pode ser feita através de várias possibilidades de categorização e critérios de pesquisa/classificação que podem refinar a procura, apresentando, assim, maior relevância dos resultados obtidos. A UFAL⁷ também possui um sistema com campos padronizados para estabelecer as estratégias de busca por portarias, o que favorece a autonomia do usuário no processo de recuperação da informação.

O vocabulário controlado, portanto, é uma ferramenta importantíssima para a representação da informação e padronização da linguagem a ser utilizada no momento da busca. Os termos definidos na política de indexação e utilizados por SRIs norteiam o usuário nas estratégias de busca, favorecendo a autonomia no acesso à informação, a ampliação da relevância, precisão dos resultados, e agilidade no processo.

Corroborando com Foskett (1973) e Rubi & Fujita (2003), o tesauro não é apenas uma lista de termos, mas a representação das reproduções mentais a eles associadas de forma estruturada e lógica de uma determinada área do conhecimento.

2.3 Indexação

De acordo com os estudos de Fujita e Leiva (2010), Rubi e Fujita (2003) e Rubi (2008), a indexação pode ser compreendida como um processo que identifica os conceitos mais pertinentes através de uma análise do assunto dos conteúdos dos

⁴ Instituto Nacional de Metrologia, Qualidade e Tecnologia – INMETRO. Disponível em: <www.inmetro.gov.br/legislacao/>.

⁵ Universidade Federal de Juiz de Fora – UFJF. Disponível em: <<http://www.ufjf.br/portal/legislacao/portariasufjf/>>.

⁶ Universidade Federal de Santa Catarina – UFSC. Disponível em: <<http://notes.ufsc.br/aplic/portaria.nsf/>>.

⁷ Universidade Federal de Alagoas – UFAL. Disponível em: <<http://sistemas.ufal.br/portarias/>>.

documentos, com a finalidade de determinar um conjunto de palavras chaves que são traduzidas para termos de uma linguagem documentária, armazenados em base de dados, com o objetivo de facilitar a intermediação entre documento e usuário e atender a demanda informacional.

Rubi e Fujita (2003) enfocam que a indexação é a parte mais importante na elaboração de um tesouro, pois, é durante esta fase que a linguagem natural é traduzida para uma linguagem documentária. Da eficiência deste processo estão condicionadas a busca, a relevância da recuperação de documentos e a satisfação dos usuários.

Segundo Lancaster (2004, p. 83), uma indexação de boa qualidade “permite que se recuperem itens de uma base de dados durante buscas para as quais sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis”, ou seja, uma boa indexação visa a recuperação de documentos relevantes que atendam a demanda do usuário. Por isso, um mesmo documento, como explica Lancaster (2004), pode apresentar um conjunto de termos de indexação diferente de acordo com as necessidades informacionais de cada grupo de usuários a qual a base de dados se destina.

O procedimento de indexação depende da política de indexação adotada onde se define a extensão da cobertura de assuntos, os níveis de representação do conteúdo, a capacidade de revocação e precisão na recuperação da informação, a escolha da linguagem a ser utilizada e o tipo de sistema de recuperação da informação de acordo com os interesses e objetivos delineados para a comunidade de usuários em potencial (ARAÚJO, 2006).

O conteúdo de um documento pode ser representado de forma específica ou exaustiva. Araújo (2006) explica que a indexação específica visa determinar os assuntos principais considerados mais relevantes e representativos. Já na indexação exaustiva, a representação do conteúdo documental é realizada de forma mais abrangente e completa, procurando representar todos os assuntos que ocorrem no documento. Em seus estudos, Lancaster (2004), considera uma indexação seletiva aquela que inclui até cinco termos/descriptores e a exaustiva mais de cinco. Vale ressaltar que quanto maior a exaustividade, isto é, quanto mais abrangente for a indexação do documento com um maior número de termos/descriptores representando os diversos assuntos, maior será o número de vezes que esse documento será recuperado.

A capacidade de revocação está relacionada ao número total de documentos recuperados pelo sistema, no entanto, nem todos os documentos são relevantes. A capacidade de precisão está relacionada ao número de documentos recuperados que atendam a demanda informacional (RUBI, 2009). Portanto, quanto mais abrangente e exaustiva for a indexação, maior a capacidade de revocação e acessibilidade do documento, porém, quanto mais específica for a indexação, maior será o nível de precisão e mais relevantes serão os documentos recuperados capazes de satisfazer as necessidades informacionais do usuário.

A política de indexação, conforme (RUBI, 2009), deve servir como subsídio e guia para a organização e representação do conhecimento em bases de dados, garantindo a personalização do sistema e a individualização da recepção da informação pelo usuário.

Podem-se destacar dois principais tipos de processo de indexação: a indexação manual e a indexação automática. A indexação manual é realizada por um ser humano e os termos determinados são estabelecidos através do julgamento subjetivo do indexador. Na indexação automática, a representação do conteúdo documental é realizada através de qualquer procedimento que identifique e selecione os termos sem a intervenção direta do ser humano (ALVARES, 2016).

Os autores divergem do número de etapas no processo de indexação, no entanto, segundo a Norma ISO 5963 (1985), pode-se destacar: a análise e compreensão do assunto; a identificação dos conceitos mais pertinentes e a tradução desses conceitos em termos de uma linguagem de indexação. Na etapa de análise de assunto, o profissional indexador faz a leitura e determina do que se trata o documento com o olhar atento para os prováveis interesses e necessidades de um determinado grupo de usuários (LANCASTER, 2004).

Segundo (LANCASTER, 2004) a prioridade da indexação precisa ser o usuário final, pois:

a indexação de assuntos é normalmente feita visando a atender às necessidades de determinada clientela, ou seja, é preciso que se tome uma decisão não somente quanto ao que é tratado no documento, mas que por ele se reveste de provável interesse para determinado grupo de usuários. (Lancaster, 2004 p. 82).

Alguns fatores, segundo Lancaster (2004, p. 88-89), podem afetar a qualidade da indexação e conseqüentemente a elaboração do vocabulário controlado. Estes

fatores podem estar ligados ao: indexador, vocabulário, documento, processo e ambiente.

A inexistência de conhecimento prévio sobre o assunto, a excessiva ou falta de experiência no ato de indexar, o nível de capacidade de concentração em processos minuciosos e cansativos e da capacidade de leitura, compreensão e síntese dos textos são fatores que podem influenciar negativamente a qualidade da indexação realizada pelo profissional indexador.

O tipo de sintaxe utilizada para a criação do vocabulário controlado, a falta de controle da ambiguidade ou imprecisão dos termos, a eficiência da estrutura de classificação dos termos escolhidos e a falta da disponibilidade de instrumentos auxiliares, por exemplo, dicionários e glossários especializados, afetam diretamente o processo de indexação.

O idioma, a extensão da obra, a tipologia, estrutura textual e a temática do documento influenciam no ato de indexar.

O processo de indexação também pode ser influenciado de acordo com as regras estabelecidas para a indexação, o nível de produtividade exigida para a tarefa, a exaustividade da indexação, definida por Lancaster (2003, p. 23) como sendo “[...] o emprego de termos em número suficiente para abranger o conteúdo temático do documento de modo bastante completo” e o tipo de indexação escolhida: por atribuição, onde os termos são selecionados a partir do vocabulário controlado ou por extração que utiliza a linguagem natural empregada pelo autor do documento.

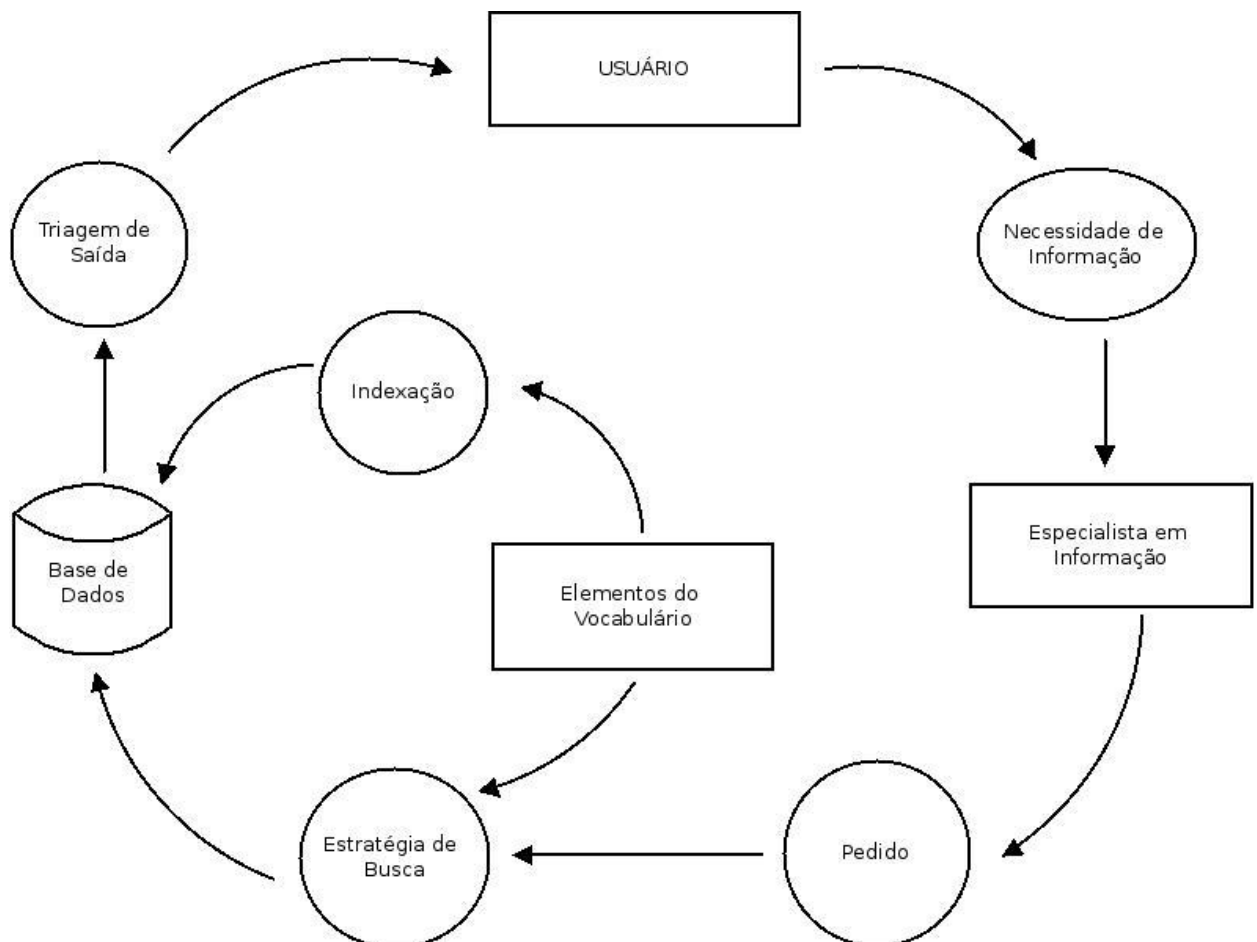
Ainda vale ressaltar que a qualidade da indexação pode ser afetada por fatores ambientais que agem diretamente no processo intelectual e na capacidade de concentração do profissional indexador: temperatura (frio/calor), iluminação, ruídos, entre outros.

Portanto, o crescimento do número de publicações, o tempo reduzido do profissional indexador para a análise e compreensão dos documentos, a subjetividade que possibilita várias interpretações válidas de acordo com o profissional que faz a leitura, o idioma do documento, o nível de conhecimento sobre o assunto por parte do indexador e fatores ambientais interferem na qualidade e coerência da indexação. (LANCASTER, 2004).

A necessidade de informação do indivíduo é expressa no momento da busca. Este processo pela busca da informação é cíclico: digamos que um usuário está em um centro de informação (biblioteca, arquivo) e necessita de uma informação. Ele

então irá procurar o especialista da informação, um bibliotecário, por exemplo, e diante do diálogo entre os dois, será feito um pedido, uma solicitação do usuário, para o bibliotecário do que ele deseja. Diante deste pedido, o especialista da informação prepara uma estratégia de busca, utilizando para isso os termos da indexação. A estratégia de busca será então confrontada em uma Base de Dados. Após a busca, alguns itens serão recuperados. O especialista da informação fará uma filtragem e apresentará ao solicitante somente as informações pertinentes ao seu pedido. Tal ciclo é representado na Figura 2.

Figura 2 – Influência de fatores nos resultados de uma busca numa base de dados



Fonte: LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. Brasília: Briquet de Lemos, 2004. p. 84

O grande desafio para a gestão documental e do conhecimento nas empresas é, corroborando com Araújo (2006, p. 62), estabelecer métodos que ofereçam ao usuário “documentos que sejam realmente relevantes para ele, que contribuam para satisfazer suas necessidades de pesquisa e, desse modo, a construção do conhecimento” de forma eficiente e rápida, diminuindo o gasto de recursos e tempo.

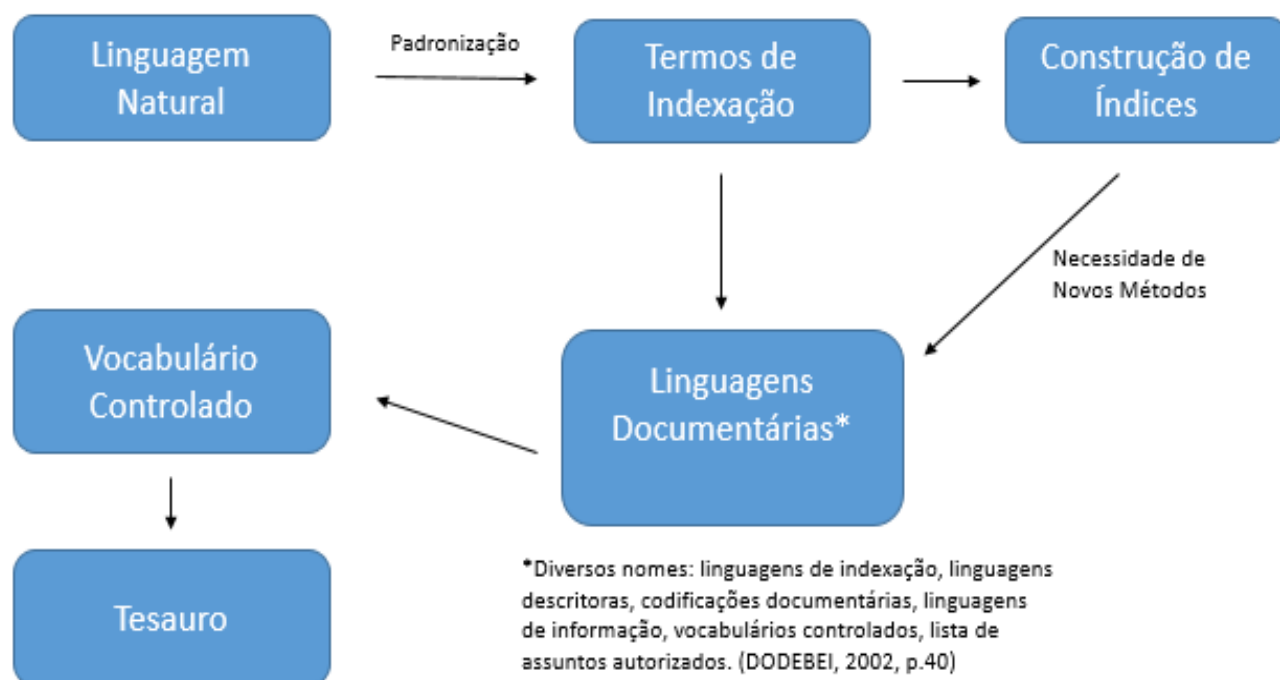
Neste contexto, elaborar eficientes políticas de indexação e um sistema de recuperação são imprescindíveis para o sucesso da gestão da informação no meio organizacional.

Lancaster (2004) sugere como método para construir vocabulário controlado:

- Gerar vocabulário empírico com base em um conjunto representativo de documentos;
- Modificar um vocabulário existente;
- Extrair o vocabulário de outro já existente;
- Reunir termos de diferentes fontes: especialistas, dicionários, glossários, índices, etc.

Como mostra a Figura 3, para a elaboração de um vocabulário controlado, é necessário que a linguagem natural do usuário seja traduzida para uma linguagem padronizada construída através de termos/descriptores que sintetizem o conteúdo do documento através do processo de indexação. A partir destes termos são elaborados índices que norteiam a construção das linguagens documentárias. A partir desta linguagem documentária obtém-se o vocabulário controlado e, conseqüentemente, o tesouro.

Figura 3 – Construção de vocabulário controlado



Fonte: DODEBEI, V. L. D. Tesouro: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

Em 2005, a norma ANSI/NISO Z39.19, estabeleceu diretrizes e convenções consideradas como “um conjunto de recomendações baseadas em técnicas e procedimentos preferenciais” utilizadas para o conteúdo, visualização, construção, ensaios, manutenção e gestão dos vocabulários controlados. A elaboração do vocabulário controlado deve ser efetivada observando três importantes princípios:

- Definição do escopo ou significado dos termos;
- Relação de equivalência de sinônimos e quase sinônimos;
- Distinção entre homógrafos.

Essas diretrizes visam nortear e padronizar a construção de vocabulários controlados para facilitar o processo de recuperação da informação.

Com o intuito de otimizar o acesso, a busca e a recuperação da informação, propõe-se neste trabalho um Sistema de recuperação de informação que utilize um vocabulário controlado elaborado através do conjunto representativo dos ofícios do Departamento de Computação da UFVJM, emitidos no período de 2011 a 2014.

2.4 O uso do vocabulário controlado em sistemas de recuperação da informação

Para Lancaster e Warner⁸ (1993 *apud* Souza, 2006), os SRIs são sistemas que funcionam como mediadores entre uma coleção de recursos de informação e uma comunidade de usuários.

Araújo (2006) e Jesus (2002) sintetizam bem as suas finalidades; os SRIs preocupam-se em representar, armazenar, organizar e fornecer acesso aos itens de informação, afim de garantir uma otimização no processo de busca e recuperação da informação.

Os SRIs organizam o conhecimento e dão acessibilidade através das atividades de: representação do conteúdo documental; armazenamento e gestão física dos itens informacionais, e a recuperação das informações, de forma a satisfazer aos interesses e necessidades informacionais dos usuários (SOUZA, 2006).

O aumento da produção do conhecimento, o desenvolvimento das tecnologias de comunicação e informação e o crescimento exponencial do número de

⁸ LANCASTER, F. W.; WARNER, A. J. **Information Retrieval Today**. [S. l.]: Information Resources Press, 1993.

informações disponíveis dificultam a recuperação de documentos relevantes que satisfaçam as buscas realizadas por usuários. No intuito de melhorar a eficiência no momento da recuperação da informação, como afirma Jesus (2002), são elaborados vocabulários controlados que funcionam como instrumentos que normalizam os termos determinados no processo de indexação e os utilizados pelo usuário nas estratégias de busca da informação.

Segundo Rowley (2002, p. 73), a recuperação envolve três etapas:

- a) Aceitação de uma consulta como uma representação da necessidade de informação, formulada pelo usuário;
- b) Execução de uma comparação da consulta com os registros existentes na base de dados, de forma a recuperar aquilo que atenda a necessidade do usuário;
- c) Produção como resultado a ser fornecido ao usuário, de um conjunto de registros recuperados e que foram identificados com base nessa comparação.

De acordo com Vieira (2011), para que o usuário tenha acesso a informação desejada, é importante que este consiga uma boa interação com o SRI que é explicitada no momento da expressão da necessidade do usuário ao sistema. Neste contexto, os SRIs exercem sua função de forma satisfatória quando recuperam informações relevantes, evitando o excesso de informações não úteis.

O vocabulário controlado, especificadamente o tesauro, neste caso, permite um controle terminológico da representação do conteúdo documental. Os termos selecionados na indexação que compõe o tesauro são também chamados de descritores. Esses descritores auxiliam o indexador na representação do conteúdo e o usuário na formulação de estratégias de busca, já que com a consulta ao tesauro ele poderá visualizar os termos de que ele necessita, e as formas de pesquisa possíveis para recuperar os documentos mais relevantes sobre o assunto (VIEIRA, 2011).

Em síntese, corroborando com Vieira (2011), a aplicação de um tesauro em SRIs é muito importante, pois, além de normalizar os termos de indexação na base de dados, evitando erros e reduzindo a subjetividade, proporciona ao usuário conhecer como a informação é organizada e quais os melhores termos que devem ser utilizados no momento da consulta. Isto aumenta a eficiência e o número de documentos relevantes que atendam a sua demanda informacional na recuperação da informação.

2.5 Lei de Acesso a Informação

Um dos propósitos deste trabalho é atender a Lei n 12.572/2011, que é a Lei de acesso a informação. Ela regulamenta o direito constitucional de acesso a informações públicas. A Lei entrou em vigor em 16 de maio de 2012, onde qualquer pessoa, física ou jurídica, sem necessidade de apresentar motivo, tem direito ao recebimento de informações públicas dos órgãos e entidades.

As informações de acordo com o artigo 4, inciso I, da Lei 12.572/2011, são dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, registrados em qualquer suporte ou formato.

O serviço de busca e fornecimento da informação é gratuito. O que poderá ser cobrado são os custos dos serviços e dos materiais utilizados na produção e envio dos documentos. Caso isso ocorra, o órgão deverá disponibilizar uma GRU - Guia de Recolhimento da União, para que o solicitante efetue o pagamento.

Quanto ao prazo de entrega, se a informação estiver disponível, deve ser entregue ao solicitando imediatamente; caso não seja possível o acesso imediato, o órgão ou entidade tem até 20 (vinte) dias para atender ao pedido, prorrogáveis por mais 10 (dez) dias, se houver justificativa expressa.

Principais aspectos da Lei de acesso à informação:

- Acesso é a regra, o sigilo a exceção.
- Requerente não precisa dizer por que e para que deseja a informação.
- Hipóteses de sigilo são limitadas e legalmente estabelecidas.
- Fornecimento gratuito de informação, salvo custo de reprodução.
- Divulgação proativa de informações de interesse coletivo e geral.
- Criação de procedimentos e prazos que facilitam o acesso à informação.

Figura 4 – Abrangência da Lei de Acesso a Informação

Todos os órgãos e entidades	Federais/Estaduais/Distritais/Municipais
Todos os Poderes	Executivo/Legislativo/Judiciário
Toda Administração Pública	Direta (órgãos públicos) / Indiretas (autarquias, fundações, empresas públicas, sociedades de economia mistas) / Demais entidades controladas direta ou indiretamente pela União, estados, Distrito Federal e/ou município
Entidades sem fins lucrativos	Aquelas que receberam recurso públicos para realização de ações de interesse público, diretamente do orçamento ou mediante subvenção social, contrato de gestão, termo de parceria, convênio, acordo, ajuste. Neste caso, a publicidade a que estão submetidas refere-se à parcela dos recursos recebidos e à sua destinação.

Fonte: Site: Acesso à Informação. Disponível em: <<http://www.acessoainformacao.gov.br>>. Acesso em: 26 maio 2016.

A abrangência da Lei de Acesso a Informação está descrita na Figura 4.

Toda e qualquer informação produzidas ou sob guarda do poder público são públicas e, portanto, acessíveis a todos os cidadãos. As ressalvas quanto a divulgação das informações pessoais onde o sigilo e legalmente estabelecido.

3 SISTEMA DE OFÍCIOS DO DECOM

Neste capítulo, apresenta a metodologia utilizada na pesquisa. Explica como é realizado a indexação do ofício, seu fluxograma e a inserção do mesmo.

3.1 Metodologia

Diante dos problemas encontrados no Departamento de Computação da UFVJM, tais como a dificuldade na organização dos ofícios, falta de padronização no

armazenamento e, conseqüentemente, difícil recuperação dos mesmos, foi proposto uma solução computacional para o armazenamento dos ofícios através de um sistema de indexação automática, o Sistema de Ofícios do DECOM, que possibilita o armazenamento de forma organizada, padronizada e digital, facilitando a recuperação dos ofícios. Este sistema foi uma estratégia encontrada com o propósito de sanar essas deficiências, e se utiliza de estudos aprofundados das Ciências da Informação junto às Ferramentas Computacionais que apoiam em todo o processo.

A base para a pesquisa foram os ofícios do Departamento de Computação da UFVJM. Para a elaboração do sistema utilizou-se a linguagem de programação *JAVA Standard Edition*⁹ e *JAVA Enterprise Edition*¹⁰, bem como o banco de dados *MYSQL*¹¹ para armazenamento dos dados referentes aos atributos do Sistema.

Inicialmente, foi feita a construção de um Vocabulário Controlado (Apêndice A), extraído todos os termos relevantes dos 239 ofícios, compreendidos entre o período de 2011 e 2014. Esse Vocabulário deu suporte a toda implementação do Sistema de Ofícios do DECOM da Universidade Federal dos Vales do Jequitinhonha e Mucuri. O conjunto de termos extraídos dos ofícios, chamado de Tesouro, foi inserido no banco de dados do Sistema, uma vez que ele define a união dos termos mais representativos de todos os ofícios analisados do Departamento de Computação, e todas as relações existentes entre estes termos, que são utilizados para representar e indexar os ofícios. As relações de cada termo são definidas no Tesouro, assim, cada termo pode possuir ou não termos afins. Por exemplo, o termo docente é definido no Tesouro possuindo os termos afins: mestres, doutores, gestores, representantes, avaliadores, professores, professor substituto, professor avaliador, professor adjunto, professor assistente, professor temporário, professor palestrante, e professor visitante.

⁹ JAVA SE - java version "1.7.0_101". Disponível em: <<https://www.java.com/>>.

¹⁰ JAVA EE - Como fazer o "download do Java EE"? O Java EE é apenas um grande PDF, uma especificação, detalhando quais especificações fazem parte deste. Para usar o software, é necessário fazer o download de uma **implementação** dessas especificações. Existem diversas dessas implementações. Já que esse software tem papel de **servir** sua aplicação para auxiliá-la com serviços de infraestrutura, esse software ganha o nome de **servidor de aplicação**. Neste projeto foi utilizado o TOMCAT 7.0, disponível em: <<http://tomcat.apache.org/>>.

¹¹ MYSQL DATABASE. Disponível em: <<https://www.mysql.com/>>.

A Figura 5 apresenta a tela inicial do sistema.

Figura 5 - Página principal do Sistema de Ofícios do DECOM

SISTEMA INTELIGENTE DE INDEXAÇÃO DE OFÍCIOS DA UFVJM

UFVJM Home Anexar ofício Buscar ofício Me cadastrar no Sistema

Etapas da indexação

Análise documental/conceitual

→ É conhecida como análise de assunto, ela remete o indexador ao assunto do livro, visando atender as necessidades de informação de usuários especializados através desses termos que irão representar o conteúdo do documento. Isso implica em tomada de decisão por parte do indexador tanto sobre a cobertura do assunto que ele está tratando no livro (abrangência do assunto) quanto a eficiência de representação (quais informações que o livro tem são de fato pertinentes/relevantes para o interesse do usuário).

Tradução

→ Traduz esses termos da linguagem natural (nossa linguagem) para a linguagem que está no documento (linguagem do bibliotecário). Por exemplo, temos um artigo que trate sobre o tema de chuva; O termo chuva é o tema central deste livro, mas o sistema da biblioteca pode não indexar este tema pela palavra chuva, podendo utilizar outro como 'pluviosidade' ou 'precipitação pluviométrica'. Ou seja, é uma coerência entre os termos que estão sendo tratados.

Me cadastrar no Sistema agora

Fonte: Imagem extraída do próprio sistema.

Para ter acesso à funcionalidade de indexação de um ofício no Sistema é necessário *login* e senha do usuário que irá operar o Sistema.

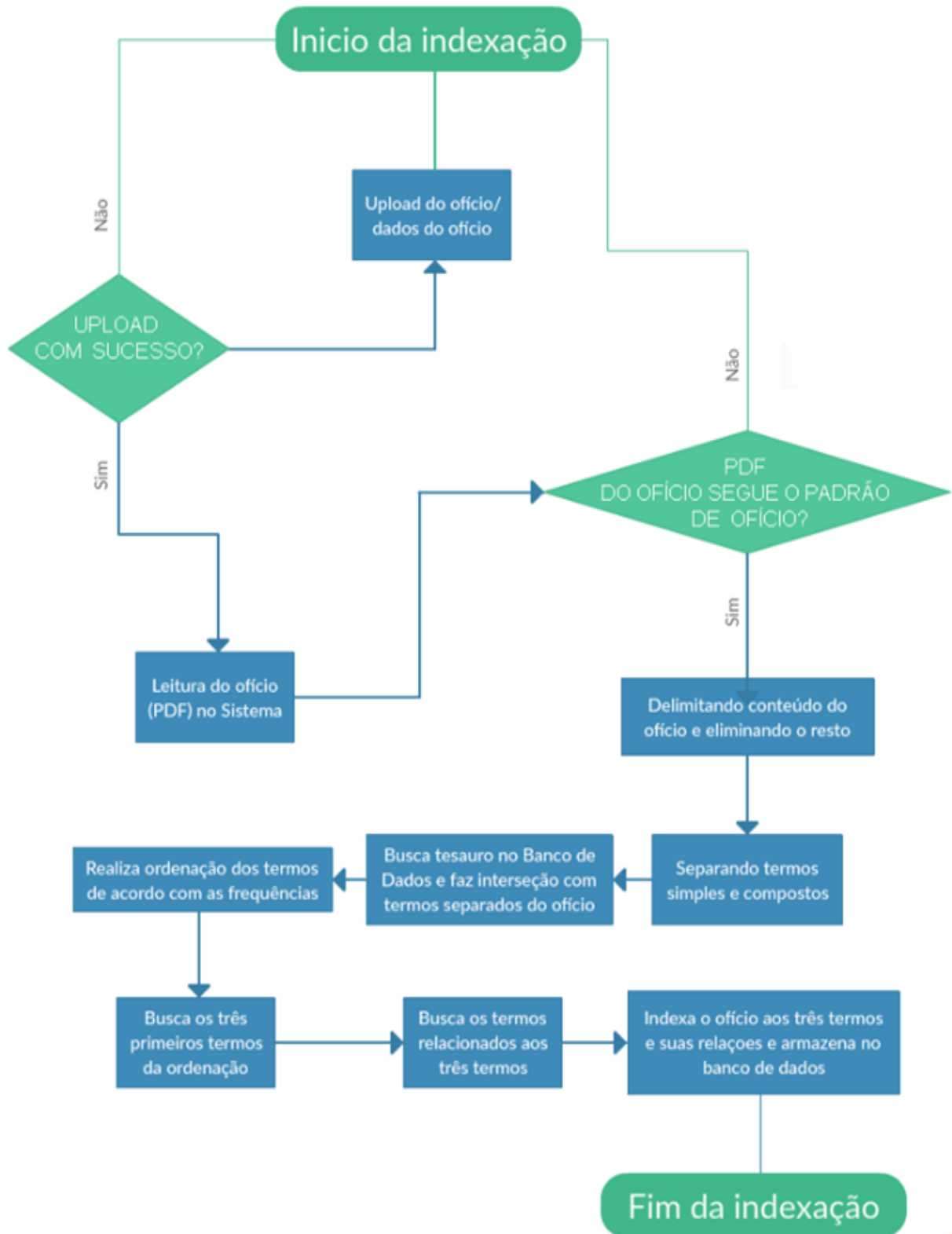
3.2 Indexação de um Ofício

Inicialmente todos os ofícios serão convertidos para o formato PDF. Para que esta conversão aconteça pode-se utilizar alguns programas gratuitos da Internet como o smallpdf¹², por exemplo.

Para visualizar este tipo de arquivo o usuário precisa ter instalado o Adobe Reader, que é um programa gratuito. A vantagem é que o ofício não poderá ser editado. Outra vantagem é que a visualização poderá ser tanto no Windows, Linux, Mac, ou em outro sistema operacional que exista um visualizador.

¹² Disponível em: <<https://smallpdf.com/pt>>.

Figura 6 – Indexação de ofício



Fonte: Desenvolvido pelo proponente

O processo de indexação do ofício está descrito na Figura 6. Quando um novo ofício é inserido, o Sistema de Ofícios do DECOM faz uma análise inicial se ele foi inserido com sucesso, caso não, ele deverá ser inserido novamente, se sim, o sistema executa o próximo passo.

Quando o ofício é inserido com sucesso, o sistema delimita todo o conteúdo do ofício e elimina o resto, separando termos simples e compostos. Feito isso, ele busca no Vocabulário Controlado (Apêndice A) os termos que farão a interseção dos termos separados do ofício. Será feita uma ordenação dos termos de acordo com as frequências individuais de cada termo, em todo o conjunto de ofícios que foram indexados pelo sistema até então, e a frequência no ofício que está sendo indexado; o número de frequência de cada termo - frequência do termo dado o conjunto de ofícios, consta em colchetes em frente ao termo no Apêndice A. O sistema irá recuperar os três primeiros termos da ordenação e irá buscar os termos relacionados no Vocabulário Controlado. Ele indexará aos três termos e suas relações e os armazenará no Banco de Dados. O fluxograma encontra-se na Figura 6. Optou-se por escolher três termos (valor arbitrário e sujeito a mudança em estudos posteriores), mediante a pertinência dos mesmos, que é definida por meio da contagem individual da ocorrência de cada termo em todos os ofícios indexados na Base de Dados do Sistema, e sua ocorrência no ofício a ser indexado. Quanto menor o número de ocorrências para um determinado termo pressupõe-se uma especificidade maior para aquele termo, já que ele está inserido em um número menor de ofícios, tornando-o mais representativo.

Posteriormente, é feita uma classificação ordenada com todas as pertinências e é selecionado os três primeiros termos do *ranking* – considerados os mais representativos do novo ofício a ser indexado – e, então, o ofício é indexado a estes três primeiros termos e aos seus termos interligados, pré-definidos no Tesouro.

Figura 7 – Página de inserção de novo ofício no Sistema

Fonte: Imagem extraída do próprio sistema.

Conforme apresenta a Figura 7, o Sistema possui três caixas de entrada na aba de *anexar ofício*: Destinatário, Data de Expedição, Assunto, e o botão para carregar ofício. A indexação de um ofício só poderá ser efetuada por algum usuário cadastrado no sistema.

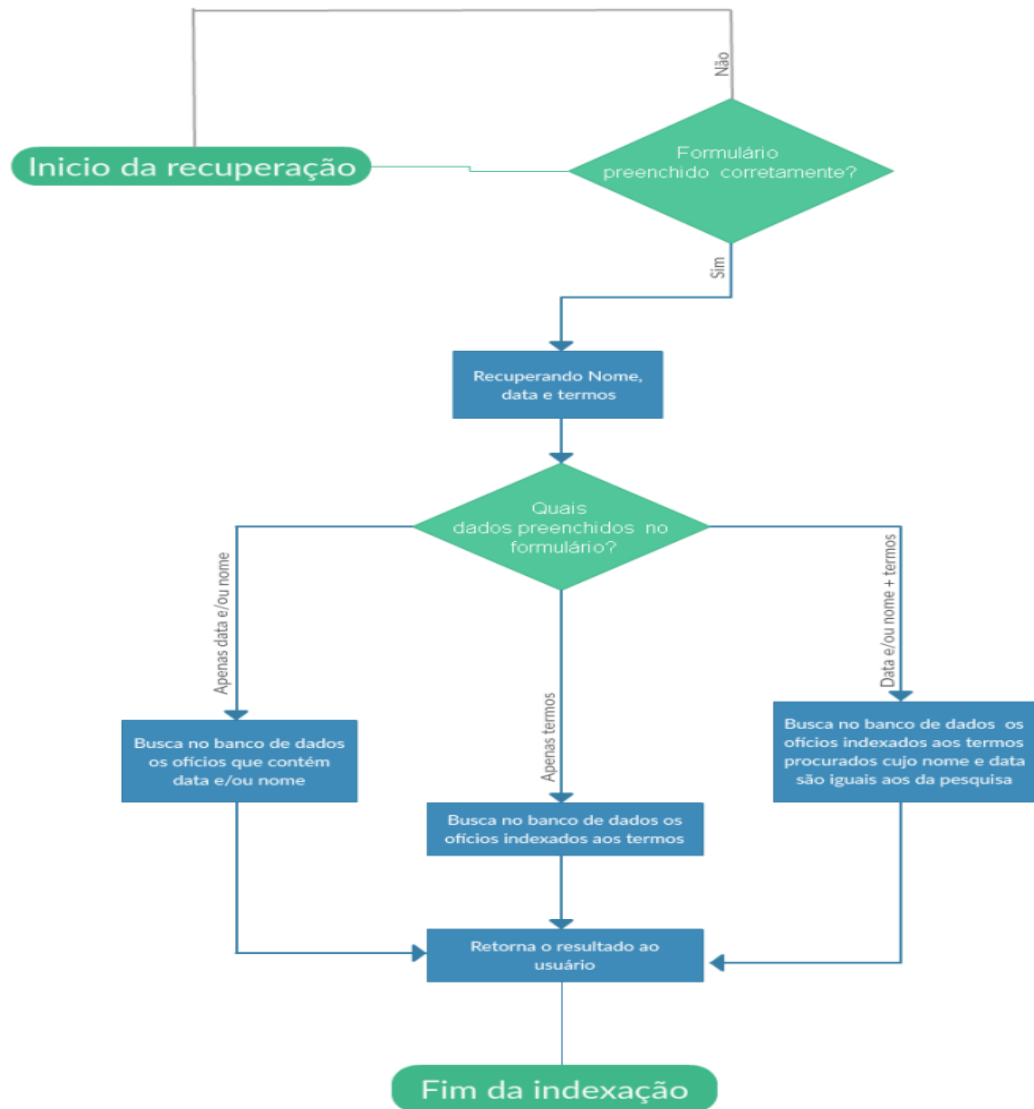
4 TESTES COM O SISTEMA

Este capítulo apresenta a recuperação do ofício, seu fluxograma e exemplifica como esta recuperação é feita. Relata as pendências do Sistema, os resultados parciais/finais.

4.1 Recuperação de um Ofício

Estando o ofício inserido no sistema de maneira correta, para que seja possível sua recuperação, será necessário seguir os passos conforme a Figura 8.

Figura 8 - Recuperação de um ofício



Fonte: Desenvolvido pelo proponente

Para a recuperação de um ofício, o usuário precisará preencher o formulário com pelo menos um dos seguintes dados: nome do expedidor, data do ofício e/ou termos chaves. O número de termos chaves desejados será escolhido pelo usuário arbitrariamente, num total de no máximo cinco, (valor sujeito a mudanças em estudos posteriores). Em seguida, o Sistema irá propor opções de termos chaves constantes no Vocabulário Controlado (Apêndice A), e também possíveis expedidores baseados nos usuários já cadastrados no Sistema e a provável data do ofício.

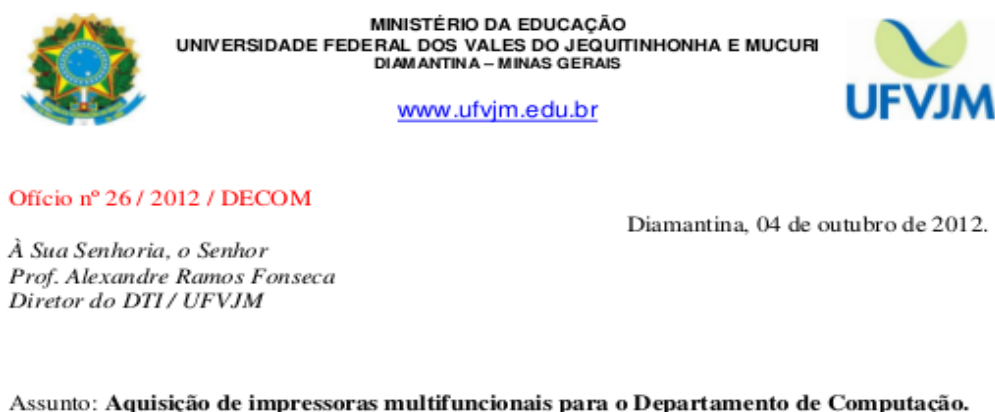
Após a inserção destes dados, o sistema realizará uma busca ao banco de dados, recuperando todos os ofícios que se associam a todos os dados fornecidos

pelo usuário. Ademais, o sistema ordenará o resultado de acordo com o peso de cada associação individual entre ofício e termo, que é definida pela combinação da pertinência do termo (explicado anteriormente como a frequência do termo dentre todos os ofícios cadastrados no Sistema e a frequência individual do termo no ofício a ser indexado pelo sistema).

4.2 Exemplo de Indexação e Recuperação de um Ofício

Para exemplificar, inseriu-se um ofício no sistema - conforme a Figura 9, e em seguida, automaticamente, foram capturados todos os termos do ofício que fazem intersecção com o conjunto representativo de termos.

Figura 9 – Modelo de ofício



Prezado Diretor,

Com o objetivo de adequar os gabinetes dos docentes e demais dependências referente ao Departamento de Computação, venho por meio deste, solicitar de Vsa à aquisição de 21 (vinte e uma) impressoras multifuncionais para o referido Departamento. Informo também que estas impressoras deverão ser adquiridas com recursos do PDTI.

Atenciosamente,

Fonte: Fornecido pelo Departamento de Computação da UFVJM.

Dentre os termos interseccionados pertencentes aos conjuntos ofício e Tesouro, estão: docentes, computação, gabinetes, e impressoras multifuncionais.

O sistema identificou quais dos termos mencionados acima possui menos ocorrências no conjunto total de ofícios armazenados no sistema, para então conseguir analisar a pertinência de cada termo e criar um *ranking*. Neste caso, os três primeiros termos do *ranking* foram: impressora multifuncional, computação e gabinetes; o sistema excluiu o termo docentes, pois este possui um número relativamente alto se comparado aos outros termos, tornando-o um termo fraco para a indexação.

Em seguida, o sistema busca no Vocabulário Controlado (Apêndice A) por todas as palavras interligadas aos três primeiros termos do *ranking*. Neste exemplo, os termos interligados foram: renovação de garantia, DECOM, UFVJM e cômodos. Essas ligações e termos estão definidos em nosso Tesouro, sendo o termo renovação de garantia ligado ao termo impressoras multifuncionais, DECOM e UFVJM ligados à Departamento de Computação, e gabinetes ligado a cômodos.

Após a indexação do ofício aos três termos ranqueados e a todos os seus afins, qualquer usuário poderá recupera-lo facilmente. Para recuperar o ofício indexado, basta inserir os dados relativos a esse ofício, que no caso seriam os termos indexados ao ofício, citados acima. A Figura 10 mostra uma busca que retornaria o ofício indexado.

Figura 10 – Modelo da tela de busca do ofício

Instruções de busca

- 1 - Existem dois campos arbitrários, data de expedição e expedidor, ou seja, se possuir as informações, entre com elas para melhor eficácia da busca.
- 2 - Existem dois símbolos, + e -, que servem para adicionar e remover entradas de termos para realizar a busca dos ofícios.
- 3 - Todas as entradas, com exceção da entrada de data de expedição, auxiliarão o cliente a realizar a busca, sugerindo os termos corretos para realizar as buscas.
- 4 - Os termos propostos são de extrema importância para realização da busca, portanto, sugerimos que o usuário sempre os utilize.

Busca de ofício

Palavras chaves

2016-04-29

Thales Oliveira Almeida +

cômodos -

computaçãq

computação

pessoas com necessidades especiais

Expedido por	Destinado à/ao	Assunto do ofício	Data do ofício	Download
Thales Oliveira Almeida	Proexc	Teste 2	2016-04-29	Baixar ofício

Fonte: Imagem extraída do próprio sistema.

Nesse exemplo, foi inserido a data do ofício, o expedidor e alguns dos termos indexados – definidos pelo algoritmo (Apêndice B) do sistema quando o ofício foi inserido– que neste caso foram: cômodos e computação. Como mostra na Figura 10, à medida que são digitados os termos desejados, o Sistema propõe termos para guiar o usuário ao acesso direto do Tesouro do Sistema, o que faz a busca se tornar eficaz.

4.3 Funcionalidades a serem implementadas no Sistema de Ofícios do DECOM

Na medida em que novos ofícios são inseridos, novos termos podem se tornar eficazes para representação dos ofícios do DECOM e, portanto, terão de ser inseridos no Vocabulário Controlado (Apêndice A) do sistema. Existirá a necessidade de mecanismos para prover a atualização constante do Vocabulário Controlado, devido ao aumento de ofícios e, conseqüentemente, o aumento de novos termos que possam vir a representar novos conteúdos. Assim, está-se desenvolvendo um algoritmo que propõe a auto alimentação do Tesouro ao longo do ciclo de vida do sistema, com a análise de relevância de novos termos em novos ofícios que advirão.

Acredita-se que com esta metodologia será possível sanar a deficiência do Departamento de Computação da UFVJM de maneira rápida e satisfatória.

4.4 Resultados parciais/finais

Os resultados parciais deste projeto são: o Vocabulário Controlado (Apêndice A), também chamado de Tesouro, que já está concluído e o sistema de indexação automática, o Sistema de Ofícios do DECOM, que está sendo desenvolvido. O desenvolvimento do Sistema precisa ser concluído, alimentado e implantado, para verificar sua utilidade e precisão na recuperação dos ofícios desejados.

5 CONSIDERAÇÕES FINAIS

A partir do estudo sugerido neste projeto, elaborou-se um sistema de indexação automática, o Sistema de Ofícios do DECOM (ainda em desenvolvimento), através do uso de um Vocabulário Controlado/tesauro (Apêndice A). O Sistema foi construído utilizando-se técnicas da biblioteconomia e técnicas computacionais, sendo um sistema que utiliza tanto das análises qualitativas – feita inicialmente para a construção do vocabulário controlado – quanto das análises quantitativas – utilizadas no algoritmo de auto alimentação do Vocabulário Controlado. O pseudocódigo do algoritmo do Sistema de Ofícios do DECOM encontra-se no apêndice B.

O Vocabulário Controlado foi construído tendo a base documental os 239 ofícios do Departamento de Computação da UFVJM, compreendidos entre o período de 2011 a 2014. Extraíu-se manualmente de cada ofício os melhores termos para a criação do Vocabulário Controlado. Para alcançar-se o objetivo do sistema, que é retornar o ofício desejado, foi realizada uma contagem dos termos, de quantas vezes o termo apareceu em cada ofício. Foi especificado que quanto mais vezes o termo aparecesse nos ofícios, menos relevância ele teria. Ou seja, este foi o critério utilizado (até o momento) para que seja possível retornar o ofício desejado pelo consulente. Pretende-se que o sistema seja disponibilizado de maneira *online* para que qualquer pessoa tenha acesso, onde quer que ela esteja, necessitando somente de acesso à Internet.

Sendo positiva a implementação do projeto, sugere-se que demais departamentos da UFVJM utilizem do mesmo sistema para organizar seus documentos. Espera-se, com isso, agilidade nos processos, bem como a satisfação do usuário final ao localizar o que procura.

REFERÊNCIAS

- ALVARES, Lillian. *Fontes de informação: tesouro*. 2014. Disponível em: <<http://magisterandre.blogspot.com.br/2014/03/tesouro.html>>. Último acesso em: 30 jan. 2016.
- ALVARES, Lillian. *4.2 Indexação*. [Material didático]. Disponível em: <<http://www.alvarestech.com/lillian/Fundamentos/conteudo.htm>>. Acesso em: 20 maio 2016.
- ANSI/NISO Z 39.19: 2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005. 176 p. Disponível em: <<http://www.niso.org/standards/resources/Z39-19-005.pdf?CFID=2436046&CFTOKEN=81298864>>. Acesso em: 30 jan. 2016.
- ARAÚJO, Ana Luiza Custódio de. *Vocabulário Controlado Básico da Rede RVBI: estudo sobre as relações conceituais na área de Direito Autoral e Direito do Consumidor*. Monografia (Graduação em Biblioteconomia) - Faculdade de Ciência da Informação, Universidade de Brasília (UnB), Brasília, 2011.
- ARAUJO, Vânia M. R. H. Sistemas de informação: nova abordagem teórico conceitual. *Ci. Inf.*, v. 24, n. 1, 1995.
- BOCCATO, Vera Regina Casari. *Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal*. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo 2009.
- BORGES, Graciane Silva Bruzina. *Indexação automática de documentos textuais: proposta de critérios essenciais*. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.
- BRASIL. Lei de acesso a informação, nº 12.527, de 18 de Novembro de 2011. Disponível em: <<http://www.acessoainformacao.gov.br/assuntos/conheca-seu-direito/a-lei-de-acesso-a-informacao>>. Acesso em: 27 maio 2016. Brasília, DF.
- BROOKSHEAR, J. Glenn. *Ciência da computação: uma visão abrangente*. 11.ed. São Paulo; Bookman, 2013.
- CAPURRO, Rafael; Hjørland, Birger. O conceito de informação. *Perspect. ciênc. inf.* v.12, n.1, p.148-207, jan./abr., 2007.
- CARDOSO, Olinda Nogueira Paes. Recuperação da informação. *INFOCOMP Journal of Computer Science*, v. 2, n. 1, nov. 2000. Disponível em: <<http://professores.dcc.ufla.br/ojs/index.php/INFOCOMP/article/view/46>>. Acesso em 26 out. 2015.
- CAVALCANTI, Cordelia R. *Indexação & tesouro: metodologia e técnicas*. Brasília:

Associação de Bibliotecários do Distrito Federal, 1978. 89 p.

CINTRA, Anna Maria Marques *et al.* *Para entender linguagens documentárias*. São Paulo: Polis; APB, 1994. 72p.

DODEBEI, V. L. D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

FUJITA, Mariângela Spotti Lopes; LEIVA, Isidoro Gil. As linguagens de indexação em bibliotecas nacionais, arquivos nacionais e sistemas de informação na América Latina. In SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 16 e SEMINÁRIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2. Rio de Janeiro, 2010. p. 1-18. Disponível em: < http://eprints.rclis.org/15137/1/Indexing_languages_FUJITA_GIL_LEIVA.pdf>. Último acesso em 30 maio 2016.

GOMES, Hagar Espanha. *Manual de elaboração de tesouros monolíngues*. Brasília: MEC/ Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia deCS na área de odontologia. *Inf. & Soc.:Est.*, João Pessoa, v.19, n.2, p. 99-118, maio/ago. 2009. Disponível em: <<http://webs.um.es/isgil/Evaluacion%20indizacion%20automatica%20SISA%20Odon%20Gil%20Leiva.pdf>>. Último acesso em: 30 maio 2016.

HOLANDA, C.; NASCIMENTO, A. Bibliotecário: gestor das unidades de informação. In: ENCONTRO REGIONAL DOS ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA E GESTÃO DA INFORMAÇÃO DA REGIÃO SUL, 12., 2010, Porto Alegre. *Anais Eletrônicos...* Porto Alegre: UFRGS, 2010. Disponível em: <http://rabci.org/rabci/node/127>. Acesso em: 08 jun. 2016.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. *Diretrizes para elaboração de tesouros monolíngues*. Brasília: IBICT, 1984. Disponível em: <[ISO 5963 Documentation. *Methods for examining documents, determining their subjects, and selecting indexing terms*. Geneve: International Organization for Standardization, 1985.](https://www.google.com.br/search?sclient=psy-ab&site=&source=hp&q=finalidades+do+tesouro&oq=finalidades+do+tesouro&gs_l=hp.3...4670.18596.0.20109.51.32.1.3.3.0.2472.2472.9-1.1.0...0...1c.1.64.psy-ab..46.5.2481.0Tr-9nL83lo&pbx=1&biw=1440&bih=732 &dpr=1&cad=cbv&sei=I4WwVuW2OsPYetGnilgP#>. Acesso em: 15 dez. 2015.</p>
</div>
<div data-bbox=)

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

NARUKAWA, Cristina Miyuki. *Estudo de vocabulário controlado na indexação automática: aplicação no processo de indexação do sistema de indexação semiautomática (SISA)*. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista —Júlio de Mesquita Filho (UNESP), Marília, 2011. Disponível em:<<https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/>

Dissertacoes/narukawa_cm_me_mar.pdf>.

fNARUKAWA, Cristina Miyuki; LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti Lopes.

NAVES, Madalena M. L. *Curso de indexação: princípios e técnicas de indexação, com vistas à recuperação da informação*. Belo Horizonte: UFMG, Biblioteca Universitária, 2004. [Material didático]. 23p.

PIZARRO, Daniella Câmara; DAVOK, Delsi Fries. O papel do bibliotecário na gestão da informação empresarial: uma pesquisa bibliográfica em periódicos nacionais de biblioteconomia e ciência da informação. *Revista ACB: Biblioteconomia em Santa Catarina*, Florianópolis, v.13, n.1, p.37-58, jan./jun., 2008.

PORTAL EDUCAÇÃO. *Redação comercial, modelo de ofício e circular*. 2013. Disponível em: <<http://www.portaleducacao.com.br/administracao/artigos/39303/redacao-comercial-modelo-de-oficio-e-circular>>. Último acesso em: 26 out. 2015.

REZEMDE, Yara. Informação para negócios: os novos agentes do conhecimento e a gestão do capital intelectual. *Ciência da Informação*, Brasília (DF), v. 31, n. 2, p. 120-128, maio/ago. 2002. Disponível em: <<http://www.ibict.br/cionline/>>. Acesso em: 8 jun. 2016.

ROUSSEAU, Jean Yves; COUTURE, Carol. *Os fundamentos da disciplina Arquivística*. Lisboa: Publicações Dom Quixote, 1998. 356 p.

ROWLEY, J. *A biblioteca eletrônica*. Brasília: Briquet de Lemos, 2002.

RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. *Perspect. cienc. inf.*, Belo Horizonte, v. 8, n. 1, p.66-77, jan./jun. 2003. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/375/193>>. Último acesso em: 20 maio 2016.

RUBI, Milena Polsinelli. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, MSL. (Org.) *et al. A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais* [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2008. cap. 4, p.81-93. Disponível em: <books.scielo.org/id/wcvbc/pdf/bocato-9788579830150-06.pdf>. Último acesso em: 31 maio 2016.

SALES, Rodrigo de. Suportes teóricos para pensar linguagens documentárias. *Revista Digital de Biblioteconomia & Ciência da Informação* v. 5, n. 1, p. 95-114, 2007. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000004583&dd1=4e8ca>>. Último acesso em: 03 nov. 2015.

SEVERINO, Antonio Joaquim. *Metodologia do trabalho científico*. São Paulo: Cortez, 1996.

SHERA, Jesse. H. Sobre biblioteconomia, documentação e ciência da informação. In: GOMES, Hagar Espanha (Org.). *Ciência da informação ou informática?* Rio de Janeiro: Calunga, 1980. 105 p.

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, p. 161 -173, mai./ago. 2006.

VARGAS, Dóris Fraga. *Estudo metodológico de elaboração de tesouros*. Monografia (Graduação em Biblioteconomia) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010

VIEIRA, Jessica Monique de Lira; SANTOS, Monick Trajano dos; LAPA, Remi Correia. Estudo da construção e aplicação do tesouro na recuperação da informação de teses e dissertações do programa de pós-graduação em desenvolvimento urbano. *Biblionline*, João Pessoa, v. 6, n. esp., p. 71-80, 2011. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/9841>>. Acesso em: 30 maio 2016.

APÊNDICE A

VOCABULÁRIO CONTROLADO/TESAURO DO DEPARTAMENTO DE COMPUTAÇÃO DA UFVJM

FONTES UTILIZADAS PARA A COLETA DOS TERMOS

Os termos foram selecionados através dos ofícios do Departamento de Computação da UFVJM entre o período de 2011 a 2014.

SIMBOLOGIA DO TESAURO

TG: Termo Geral

TR: Termo Relacionado

TE: Termo Específico

[número] – Quantas vezes o termo apareceu nos ofícios.

TESAURO

APRESENTAÇÃO SISTEMÁTICA

TG DOCENTES - [121]

TR MESTRES - [1]

TR DOUTORES - [8]

TR GESTOR - [10]

TR REPRESENTANTE - [7]

TR AVALIADORES - [2]

TR PROFESSORES - [127]

TE PROFESSOR SUBSTITUTO - [10]

TE PROFESSOR AVALIADOR - [0]

TE PROFESSOR ADJUNTO - [27]

TE PROFESSOR ASSISTENTE - [25]

TE PROFESSOR TEMPORÁRIO - [5]

TE PROFESSOR PALESTRANTE - [1]

TE PROFESSOR VISITANTE - [0]

TG DISCENTES - [4]

TR REPRESENTANTE - [0]

TR ALUNOS - [1]

TE ALUNO UNIVERSITÁRIO - [0]

TE ALUNO INTERCAMBISCA - [0]

TE USUÁRIO - [5]

TG TAE - [0]

TR TÉCNICOS ADMINISTRATIVOS EM EDUCAÇÃO - [70]

TE SERVIDORES PÚBLICOS - [20]

TE ANALISTAS - [4]

TE FUNCIONÁRIOS PÚBLICOS - [0]

TE ASSISTENTES EM ADMINISTRAÇÃO - [2]

TG CHEFES - [71]

TE VICE-CHEFE - [51]

TE CHEFIA - [32]

TE VICE-CHEFIA - [11]

TE PRESIDENTE - [9]

TE REITOR - [15]

TE PRO REITOR - [18]
 TE VICE-REITOR - [0]
 TE CHEFE INTERINO - [4]
 TE DIRETOR - [117]
 TE VICE DIRETORES - [1]
 TE SUPERINTENDENTE - [26]
 TE COORDENADOR - [9]
 TE VICE COORDENADOR - [1]
 TG PESSOA FÍSICA - [0]
 TG PESSOA JURÍDICA - [0]
 TG PORTADOR DE NECESSIDADES ESPECIAIS - [1]
 TR PESSOAS COM NECESSIDADES ESPECIAIS - [5]
 TR PNE - [4]
 TG TITULAÇÃO - [3]
 TR MESTRADO - [4]
 TR DOUTORADO - [19]
 TE RETRIBUIÇÃO POR TITULAÇÃO - [0]
 TG PLANO ANUAL DE CAPACITAÇÃO - [6]
 TG PROGRESSÃO - [0]
 TE PROGRESSÃO HORIZONTAL - [14]
 TE PROGRESSÃO VERTICAL - [0]
 TG AVALIAÇÃO - [9]
 TR PROVAS - [7]
 TR TESTES - [2]
 TR CONCURSO - [1]
 TE CONCURSO PÚBLICO - [81]
 TR ESTÁGIO - [0]
 TR ESTÁGIO PROBATÓRIO - [24]
 TR DESEMPENHO - [4]
 TR FICHA DE AVALIAÇÃO - [0]
 TR BANCA - [77]
 TE BANCA EXAMINADORA - [10]
 TE MEMBROS - [47]
 TR QUALIFICAÇÃO - [13]
 TE DOUTORAL - [3]
 TR SEMINÁRIO - [15]
 TR APROVAÇÃO - [3]
 TR NOTA - [2]
 TR ATIVIDADE - [0]
 TR APRESENTAÇÃO - [5]
 TR PESQUISA - [8]
 TR APROVEITAMENTO - [0]
 TR RESULTADO - [2]
 TR CORREÇÃO - [4]
 TR CERTIFICAÇÃO - [0]
 TG CONDUTA INADEQUADA - [1]
 TR ACUSAÇÕES - [1]
 TR DESRESPEITO - [1]
 TR CALÚNIA - [1]
 TR AGRESSÃO - [1]
 TR ASSÉDIO MORAL - [1]
 TR PROBLEMAS - [7]
 TR SUBSTITUIÇÃO - [14]
 TR TRANSFERÊNCIA - [7]
 TR REALOCAÇÃO - [0]

TR REMOÇÃO - [0]
 TR PENDÊNCIAS - [12]
 TG REUNIÃO - [14]
 TR REUNIÃO ORDINÁRIA - [0]
 TR CONGREGAÇÃO - [1]
 TR VOTAÇÃO - [1]
 TR APRECIÇÃO - [0]
 TR HOMOLOGAÇÃO - [3]
 TG UFVJM – [63]
 TE UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI - [1]
 TR PRO REITORIA – [12]
 TE PLANEJAMENTO E ORÇAMENTO - [2]
 TR DEPARTAMENTOS - [6]
 TR DEPARTAMENTO DE COMPUTAÇÃO – [189]
 TE DECOM – [125]
 TR FACET - [110]
 TE FACULDADE DE CIÊNCIAS EXATAS - [3]
 TR FACSAE - [0]
 TE FACULDADE DE CIÊNCIAS APLICADAS E EXATAS – [0]
 TR DEPAD - [0]
 TE DEPARTAMENTO DE ADMINISTRAÇÃO - [10]
 TR DTI - [8]
 TE DEPARTAMENTO DE TECNOLOGIA E INFORMAÇÃO - [12]
 TR DME - [1]
 TE DEPARTAMENTO DE MATEMÁTICA E ESTATÍSTICA – [2]
 TR DEPARTAMENTO DE QUÍMICA - [1]
 TR DIVISÃO DE PAGAMENTOS – [29]
 TE PAGAMENTOS – [43]
 TE ADICIONAL NOTURNO – [46]
 TE FREQUÊNCIA - [39]
 TR CONSELHO UNIVERSITÁRIO – [2]
 TR RECURSOS HUMANOS - [12]
 TG COMPUTAÇÃO - [19]
 TG INFORMÁTICA - [10]
 TR EQUIPAMENTOS - [16]
 TE NO-BREAKS - [0]
 TE COMPUTADORES - [5]
 TE MICROCOMPUTADORES - [3]
 TE VOIP - [1]
 TR INFORMÁTICA BÁSICA - [1]
 TE PROGRAMAS - [6]
 TE PROGRAMAÇÃO - [3]
 TE CONFIGURAÇÕES - [1]
 TE MANUTENÇÃO - [6]
 TE ATUALIZAÇÃO DE DADOS - [4]
 TE INTERNET - [4]
 TE SOFTWARE - [6]
 TE HARDWARE - [5]
 TE INTERFACE GRÁFICA - [6]
 TR SISTEMA OPERACIONAL - [5]
 TE WINDOWS - [1]
 TE LINUX - [0]
 TR PACOTE OFFICE - [0]
 TE WORD - [1]

- TE EXCEL - [1]
- TE POWERPOINT - [1]
- TE OUTLOOK - [1]
- TR DISCIPLINAS - [5]
 - TE LEITURA E PRODUÇÃO DE TEXTOS - [3]
 - TE INGLÊS - [2]
 - TE INGLÊS BÁSICO - [3]
 - TE INGLÊS INSTRUMENTAL - [4]
 - TE FÍSICA - [1]
 - TE LÓGICA - [1]
 - TE ALGORITMOS E ESTRUTURA DE DADOS - [9]
 - TE BANCO DE DADOS - [12]
 - TE SISTEMAS OPERACIONAIS - [2]
 - TE SISTEMAS DISTRIBUÍDOS - [7]
 - TE PESQUISA OPERACIONAL - [12]
 - TE ORGANIZAÇÃO E ESTRUTURA DE COMPUTADORES - [9]
 - TE SEGURANÇA E AUDITORIA DE SISTEMAS - [4]
 - TE ENGENHARIA DE SOFTWARE - [5]
 - TE REDES - [7]
 - TE REDES LOCAIS - [1]
 - TE REDES DE COMPUTADORES - [7]
 - TE REDES LÓGICAS - [2]
 - TE METODOLOGIA CIENTÍFICA - [6]
 - TE INTELIGÊNCIA ARTIFICIAL - [1]
 - TE TRABALHO COOPERATIVO APOIADO POR COMPUTADOR - [1]
 - TE INTERFACE HOMEM MÁQUINA - [1]
 - TE SISTEMAS DE APOIO A DECISÃO - [1]
 - TE TEORIA GERAL DE SISTEMAS - [1]
 - TE FUNDAMENTOS DE SISTEMAS DE INFORMAÇÃO - [1]
- TG CURSOS – [21]
 - TE CURSOS LIVRES – [6]
 - TE SISTEMAS DE INFORMAÇÃO – [52]
 - TE COMPUWAY - [1]
 - TE TURISMO - [1]
- TG PRÉDIO - [55]
 - TR INFRAESTRUTURA – [23]
 - TR CONSTRUÇÃO - [17]
 - TR CONSTRUTORA - [1]
 - TR OBRA - [23]
 - TR INSTALAÇÕES – [0]
 - TR PINTURA - [7]
 - TR PLANTAS - [1]
 - TR ARQUITETA - [18]
 - TR INSTALAÇÃO - [1]
- TG CÔMODOS - [0]
 - TR BANHEIRO - [14]
 - TE WC - [0]
 - TE TOALETE - [0]
 - TR COBERTURA - [1]
 - TR COPA - [21]
 - TR GABINETES - [20]
 - TR AUDITÓRIO - [7]
 - TR CORREDORES - [15]
 - TR SALA - [35]
 - TR LABORATÓRIO - [46]

- TR HALL - [2]
- TR ESCRITÓRIO - [7]
- TR COORDENAÇÃO - [23]
- TG PARTES DOS CÔMODOS - [0]
 - TR JANELAS - [9]
 - TR PORTAS - [9]
 - TR PAREDES - [23]
 - TR PISO - [7]
 - TR SACADAS - [11]
 - TR BANCADAS - [2]
 - TR PIAS - [11]
 - TR VASO SANITÁRIO - [1]
 - TR ELEVADOR - [4]
 - TR TELHADO - [12]
 - TR CABINES - [16]
 - TR TELAS DE PROTEÇÃO - [2]
 - TR ISOLAMENTO ACÚSTICO - [8]
 - TR REDE ELÉTRICA - [4]
 - TR REDE LÓGICA - [2]
 - TR QUADRO DE CHAVES - [1]
- TG MOBILIÁRIO - [0]
 - TR RACK - [6]
 - TR MESA - [1]
 - TR MESA CENTRAL - [1]
 - TR CADEIRAS - [1]
 - TR ASSENTOS - [1]
 - TR ARMÁRIOS - [1]
 - TR GAVETAS - [1]
 - TR POLTRONAS - [7]
 - TR ESQUADRIAS - [9]
 - TR QUADRO BRANCO - [1]
- TG ADEQUAÇÕES - [9]
 - TR ADAPTAÇÕES - [1]
 - TR CONCERTO - [2]
 - TR ENGENHEIRO ELETRICISTA - [8]
 - TR ENGENHEIRO CIVIL - [1]
 - TR INSTALAÇÃO ELÉTRICA - [1]
 - TR PINTURA - [1]
- TG APARELHOS ELETRÔNICOS - [0]
 - TR TOMADAS - [8]
 - TR PONTOS ELÉTRICOS - [0]
 - TR RENOVAÇÃO DE GARANTIA - [5]
 - TE AR CONDICIONADO - [5]
 - TE MICROONDAS - [1]
 - TE TELEFONE - [0]
 - TE PROJETOR MULTIMÍDIA - [0]
 - TE IMPRESSORA - [4]
 - TE IMPRESSORA MULTIFUNCIONAL - [6]
- TG UTENSÍLIOS - [5]
 - TE BANDEJA DE AÇO - [4]
 - TE COPOS DE VIDRO - [4]
 - TE FILTRO DE ÁGUA - [1]
 - TE GARRAFA TÉRMICA - [4]
 - TE JARRA DE VIDRO - [4]
 - TE FACAS - [4]

- TE GARFOS - [4]
- TE COLHERES - [4]
- TG LAYOUT - [5]
 - TE PAISAGISMO - [1]
- TG REFEIÇÕES - [0]
 - TR ALIMENTAÇÃO - [0]
 - TE CAFÉ DA MANHÃ - [0]
 - TE ALMOÇO - [0]
 - TE LANCHE - [13]
 - TE JANTAR - [0]
- TG HOSPEDAGEM - [16]
 - TE APARTAMENTO - [0]
 - TE HOTEL - [4]
 - TE POUSADA - [9]
 - TE CASA DE APOIO - [0]
 - TE CASA - [0]
- TG TRANSPORTE - [20]
 - TR MOTORISTA - [1]
 - TR DIÁRIAS - [30]
 - TR VIAGEM - [0]
 - TR RECURSO - [0]
 - TR DINHEIRO - [0]
 - TR PASSAGEM - [15]
 - TE AVIÃO - [3]
 - TE CARRO - [0]
 - TE ÔNIBUS - [1]
 - TE TAXI - [0]
- TG LOCALIDADES GEOGRÁFICAS - [0]
 - TE BELO HORIZONTE - [6]
 - TE BRASÍLIA - [2]
 - TE FOZ DO IGUAÇU - [1]
 - TE TEÓFILO OTONI - [2]
 - TE LAVRAS - [24]
- TG DOCUMENTAÇÃO - [0]
 - TR DOCUMENTOS - [7]
 - TE ATAS - [5]
 - TE OFÍCIOS - [17]
 - TE RELATÓRIOS - [7]
 - TE PLANILHAS - [4]
 - TE DIRETRIZ - [0]
 - TE LEI - [6]
 - TE PORTARIAS - [33]
 - TE EDITAL - [38]
 - TE COMUNICAÇÃO INTERNA - [0]
 - TE ARTIGOS - [0]
 - TE CONTRATO - [10]
 - TE PREGÃO - [1]
 - TE EMPENHO - [0]
 - TE CONVITE - [4]
 - TE MEMORANDO - [5]
 - TE LICITAÇÃO - [8]
- TG RECESSO - [2]
 - TR FÉRIAS - [69]
 - TR FERIADOS - [0]
 - TE REVEILLON - [0]

TE CARNAVAL - [0]
 TE SEXTA FEIRA DA PAIXÃO - [0]
 TE PAIXÃO DE CRISTO - [0]
 TE TIRADENTES - [0]
 TE DIA DO TRABALHO - [0]
 TE DIA DO TRABALHADOR - [0]
 TE CORPUS CHRISTI - [0]
 TE INDEPENDÊNCIA DO BRASIL - [0]
 TE SETE DE SETEMBRO - [0]
 TE 7 DE SETEMBRO - [0]
 TE NOSSA SENHORA APARECIDA - [0]
 TE DIA DAS CRIANÇAS - [0]
 TE FINADOS - [0]
 TE PROCLAMAÇÃO DA REPÚBLICA - [0]
 TE NATAL - [1]
 TE FINAL DE ANO - [2]

Termos que podem estar ligados a mais de um termo.

TG RENOVAÇÃO - [46]
 TG SOLICITO - [2]
 TG SOLICITAÇÃO - [13]
 TG INCLUSÃO - [7]
 TG ALTERAÇÃO - [2]
 TG MUDANÇA - [21]
 TG INTERINA - [1]
 TG INDICAÇÃO - [5]
 TG COMITÉ - [1]
 TG DIVULGAÇÃO - [8]
 TG REALIZAÇÃO - [4]
 TG DEVOLUÇÃO - [5]
 TG PARECER CONCLUSIVO - [7]
 TG PLANO DIRETOR - [6]
 TG EMPRÉSTIMO - [0]
 TG DESIGNAÇÃO - [0]
 TG PATRIMÔNIO - [0]
 TG RECONHECIMENTO - [0]
 TG RETIFICAÇÃO - [0]
 TG LIFE/UFVJM - [0]
 TG INFORME - [0]
 TG ABERTURA - [0]
 TG FECHAMENTO - [0]
 TG DEFERIDO - [0]
 TG INDEFERIDO - [0]
 TE AQUISIÇÃO - [8]
 TE ENCAMINHAMENTO - [15]
 TE DEMANDA - [5]
 TE AUTORIZAÇÃO - [0]
 TR RENOVAÇÃO - [0]
 TR REPROGRAMAÇÃO - [34]
 TR PRORROGAÇÃO - [15]
 TR CANCELAMENTO - [9]
 TR AFASTAMENTO - [21]
 TR AFASTADOS - [11]
 TR RETORNO - [1]

TR CORREÇÃO - [4]
TR COMPOSIÇÃO – [13]
TR EDUCAMUNDO – [1]
TR AUDITORIA [7]

APÊNDICE B

PSEUDOCÓDIGO DO SISTEMA DE OFÍCIOS DO DECOM

```
anexarOficio()
{
    OficioobjetoOficio;

    //Criando um objeto chamado 'objetoOficio' através da classe 'Oficio'

    IteradorDeFormulario itr;

    //Cria um objeto para auxiliar no upload do formulário com nome itr, através da classe
    'IteradorDeFormulario'

    enquanto( existir itr ) //enquanto ainda existir campos do formulario, faça!
    {
        objetoOficio.preencherOficio();

        // Neste momento o objeto 'oficio' criado acima será definido por meio do
        método 'preencherOficio' da classe 'Oficio' com os dados e o oficio fornecido pelo
        usuário através do formulário
    }

    PDFManager manipuladorPDF;

    //Cria objeto com nome 'manipuladorPDF' através da classe 'PDFManager' para auxiliar na
    leitura do PDF fornecido pelo usuário

    Texto textoOficio = manipuladorPDF.lerOficio(objetoOficio);

    // Começa ler o arquivo PDF de ofício, fornecido pelo usuário no formulário (que foi carregado
    acima no objeto 'oficio' por meio do método 'preencherOficio'), e adiciona este texto do 'oficio'
    no objeto 'textoOficio' criado pela classe 'Texto'

    textoOficio.delimitar();

    //Delimita o textoOficio por meio do método 'delimitar' da classe 'Texto', pegando apenas o
    assunto e o conteúdo do texto e eliminando o resto

    textoOficio.tratar();
    // Começa tratar o objeto 'textoOficio' por meio do método 'tratar' da classe 'Texto',
    padronizando com letras minusculas e tirando qualquer tipo de pontuação

    Coleção coleçãoPalavra = textoOficio.separarTextoPorPalavras;
```

//Começa retirar todas as possíveis palavras, simples e compostas (com até cinco palavras) por meio do método 'separarTextoPorPalavras' e insere palavra a palavra, individualmente, na 'coleçãoPalavra'.

BancoDeDados BD;

//Cria um objeto por meio da classe 'BancoDeDados' com nome 'BD'; objeto 'BD' auxilia no processo de recuperação no sistema gerenciador de banco de dados 'MYSQL'

Coleção coleçãoPalavrasInterseccionadas = BD.pegarTesauro();

//Neste momento, o tesauro desenvolvido no projeto é recuperado do banco de dados (palavra a palavra e suas relações) por meio do método 'pegarTesauro' da classe 'BancoDeDados'

coleçãoPalavrasInterseccionadas = coleçãoPalavra.interseccionar(coleçãoPalavra);

//Neste momento, a coleção de palavras 'coleçãoPalavrasInterseccionadas' , que contém o tesauro completo, está sendo comparada a todas as palavras da coleção 'coleçãoPalavra', por meio do método 'interseccionar' da classe 'Coleção', com parâmetro recebendo um objeto de 'Coleção' (objeto a ser comparado) e eliminando todas as palavras que não fazem intersecção entre as coleções.

coleçãoPalavrasInterseccionadas.ordenar();

// Ordena as palavras interseccionadas de acordo com a frequência dela no texto e a frequência no conjunto de todos os ofícios (busca os três primeiros objetos e elimina as palavras rankeadas acima da terceira posição), através do método 'ordenar' da classe 'coleçãoPalavrasInterseccionadas'

Coleção coleçãoPalavrasRelacionadas = palavrasInterseccionadas.pegarRelações();

// Neste momento, é criada uma coleção de palavras 'coleçãoPalavrasRelacionadas' pra receber todas as palavras relacionadas (relações estabelecidas no Tesauro) dos três primeiros objetos selecionados devido à ordenação pela frequência; isso é feito por meio do método 'pegarRelações' da classe 'Coleção'.

Coleção coleçãoPalavrasIndexação = coleçãoPalavrasInterseccionadas +
coleçãoPalavrasRelacionadas;

//Neste momento, está sendo criada a última 'Coleção' que conterà todas as três palavrasrankeadas de acordo com a frequência, somadas às suas palavras relacionadas. (relações definidas pelo Tesauro).

oficio.indexar('coleçãoPalavrasIndexação');

// Está indexando todas as palavras da coleção 'coleçãoPalavrasIndexação' ao ofício, de fato. Isso é feito por meio do método 'indexar' da classe 'Ofício', recebendo no parâmetro a coleção 'coleçãoPalavrasIndexação', que contém todas as palavras selecionadas para ser indexadas.

```
}
```

```
//FIM DO PROCESSO DE INDEXAÇÃO
```

```
recuperarOficio()
```

```
{
```

```
    TermosFormulário termosFormulário;
```

```
    //Cria um objeto para auxiliar no upload do formulário com nome 'termosFormulário', através da classe 'TermosFormulário'
```

```
    IteradorDeFormulario itr;
```

```
    //Cria um objeto para auxiliar no upload do formulário com nome itr, através da classe 'IteradorDeFormulario'
```

```
    Coleção coleçãoPalavrasBusca;
```

```
    //Inicializando uma coleção de palavras chamada 'coleçãoPalavrasBusca' que persistirá objetos de uma classe 'Palavra'
```

```
    enquanto(existir itr ) //enquanto ainda existir campos do formulario de busca, faça!
```

```
{
```

```
        termosFormulário.preencherColeçãoPalavrasBusca();
```

```
        // Neste momento o objeto 'termosFormulário' criado acima será definido com os dados e o ofício (arquivo com extensão PDF) fornecido no formulário, por meio do método 'preencherColeçãoPalavrasBusca' definido na classe 'TermosFormulário'
```

```
}
```

```
    Coleção ofíciosBusca;
```

```
    //Inicializa uma coleção de ofícios, com objetos de uma classe 'Ofício'
```

```
    Coleção ofíciosExpedidorData.
```

```
    //Inicializa uma coleção de ofícios, com objetos de uma classe 'Ofício'
```

```
    ofíciosExpedidorData.buscarOfíciosPeloExpedidorXdata(TermosFormulário);
```

```
    //Busca os ofícios cruzando ofícios que tem como característica a data e o expedidor passados no formulário e armazenados no objeto 'TermosFormulário', por meio do método 'buscarOfíciosPeloExpedidorXdata'; em seguida, armazena na coleção criada 'ofíciosExpedidorData'
```

```
    Coleção ofíciosPelosTermos;
```

```
    //Inicializa uma coleção de ofícios, com objetos de uma classe 'Ofício'
```



```
oficiosPelosTermos.buscaOficiosPelosTermos());
```

```
//Busca os ofícios cruzando ofícios que tem como característica os termos passados no formulário e armazenados no objeto 'TermosFormulário', por meio do método 'buscaOficiosPelosTermos'; em seguida, armazena na coleção criada 'oficiosPelosTermos'
```

```
Coleção oficiosResultado
```

```
//Inicializa uma coleção de ofícios, com objetos de uma classe 'Oficio'
```

```
oficiosResultado.intersecao(oficiosPelosTermos e oficiosExpedidorData);
```

```
//Faz a interseção das coleções 'oficiosTermos' e 'oficiosExpedidorData', por meio do método 'intersecao' da coleção, passando dois parâmetros, as coleções 'oficiosPelosTermos' e 'oficiosExpedidorData'; em seguida, armazena na coleção criada 'oficiosResultado'
```

```
}
```

```
//FIM DO PROCESSO
```