

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI

CURSO DE SISTEMAS DE INFORMAÇÃO

VICTOR OLIVEIRA ALMEIDA

**DESENVOLVIMENTO DE UM SISTEMA INTELIGENTE DE INDEXAÇÃO DE
OFÍCIOS**

DIAMANTINA

2016

AGRADECIMENTOS

Agradeço primeiramente a Deus que me proporcionou realizar este sonho.

Aos meus pais, Laurismar e Glaucia, por toda confiança depositada em mim e sacrifícios feitos para que eu pudesse chegar até aqui. Aos meus irmãos, Farley e Thales, pelo apoio e força que me concederam e por estarem sempre comigo durante essa difícil caminhada.

Aos meus amigos, mestres e colegas que acreditaram, acompanharam e me auxiliaram durante todo o itinerário desta conquista.

Ao meu coorientador, Cristiano Grijó Pitangui, pela paciência, disposição e ensinamentos durante todo o curso e o desenvolvimento deste trabalho.

À Aline Pereira da Costa, minha colega durante todo o desenvolvimento desta monografia, por toda a contribuição para que este trabalho fosse desenvolvido.

Aos professores do curso de Sistemas de Informação que sempre estiveram dispostos a ajudar e contribuir na formação dos alunos, através de ensinamentos e apoio na busca do conhecimento.

Aos funcionários da UFVJM e, em especial, aos funcionários do curso de Sistemas de Informação, por estarem disponíveis nos momentos de necessidades, sempre realizando com excelência todas as atividades necessárias para proporcionar aos alunos as melhores condições de aprendizado.

RESUMO

Este trabalho surgiu de uma deficiência do Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM) em armazenar e recuperar seus ofícios. Tal departamento possui dificuldade no armazenamento e organização destes ofícios o que inviabiliza o acesso aos documentos e, conseqüentemente, dispende muito tempo na localização e recuperação da informação. Diante disso, foi desenvolvido um Sistema Inteligente de Indexação Automática utilizando-se técnicas da biblioteconomia através de ferramentas computacionais que visam automatizar o processo de indexação de novos ofícios, otimizar a recuperação e democratizar o acesso à informação. O banco de dados do Sistema foi construído baseado em um vocabulário controlado: tesouro. O Tesouro é um tipo de vocabulário controlado, mais complexo, que foi elaborado a partir de conceitos selecionados e termos relacionados presentes nos ofícios que permeiam o espaço de 2011 a 2014, totalizando 239 ofícios. O Tesouro trabalha com ambigüidade dos termos, sinonímia, relações hierárquicas e associativas. Para a alimentação do Tesouro com os novos documentos que surgirão propõe-se uma alimentação semiautomática, que fará a análise de relevância de novos termos contidos nos novos ofícios através de um algoritmo. 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 trabalho, sugere-se que os demais departamentos da UFVJM utilizem o mesmo sistema para a organização dos documentos, ganhando agilidade nos processos e satisfação do usuário final na localização do que procura.

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

ABSTRACT

This work came from a deficiency at Computer Department of the Federal University of Vales of the Jequitinhonha and Mucuri (UFVJM) to store and recover its crafts. This department has difficulty in the storage and organization of these crafts, thus prevents access to documents and, therefore, time is spent in the location and retrieval of information. Being aware of the problem found, was developed an Intelligent Automatic Indexing System using Techniques of the Library Science through computational tools aimed at automating the indexing process of new crafts, optimize recovery and democratize access to information. The data base was built based on a Controlled Vocabulary: thesaurus. The Thesaurus is a type of Controlled Vocabulary, more complex, which was developed from the selected and related concepts terms located in crafts which belongs to the period of 2011 until 2014, totaling 239 crafts. The Thesaurus works with ambiguity of the terms, synonymy, hierarchical and associative relations. The feeding process of the Thesaurus in new crafts insertions is proposed through a semiautomatic algorithm that will make relevancy analyses of terms found in new crafts. The system will be available in the Internet to be accessed from anyone and anywhere. Being positive the implementation of the Intelligent Automatic Indexing System it is suggested that other UFVJM's departments use the same system to organize their crafts as well, causing the system users to gain agility and satisfaction.

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

LISTA DE FIGURAS

Figura 2.1- Construção do Vocabulário Controlado	18
Figura 4.1 - Fluxograma do processo de indexação	33
Figura 4.2 - Exemplo de ofício	34
Figura 4.3 - Fluxograma do processo de alimentação	36
Figura 4.4 - Tela de alimentação do Tesouro	37
Figura 4.5 - Fluxograma do processo de recuperação	39
Figura 4.6 - Tela de busca de ofícios	40
Figura 5.1 - Tela de login	41
Figura 5.2 - Tela de cadastro	42
Figura 5.3- Tela de edição de dados	42
Figura 5.4 - Tela de instruções no anexo de ofícios	43
Figura 5.5 - Tela de anexo de ofícios	44
Figura 5.6 - Tela de gerenciamento de ofícios	44

SUMÁRIO

1	INTRODUÇÃO	8
1.1	Definição do problema	9
1.2	Objetivos gerais	10
1.3	Objetivos específicos	10
1.4	Organização do trabalho	10
2	INDEXAÇÃO.....	12
2.1	Ciência da informação e Biblioteconomia.....	12
2.2	História da indexação	13
2.3	Inteligência Artificial.....	14
2.4	Linguagem Documentária	15
2.5	Vocabulário Controlado: Tesauro.....	16
2.6	Métodos de indexação	19
2.7	Indexação quantitativa (métodos estatísticos)	20
2.8	Indexação qualitativa (métodos linguísticos)	21
2.9	Política de indexação	22
2.10	Exaustividade e especificidade	23
3	AUTOMATIZAÇÃO DA INDEXAÇÃO.....	25
3.1	Seleção de termos descritores	25
3.2	Vantagens e desvantagens da automação	26
4	PROPOSTA DO SISTEMA	28
4.1	Indexação e seleção	29
4.2	Algoritmo simplificado do sistema.....	29
4.3	Indexação de um ofício.....	30
4.4	Alimentação do Tesauro	31
4.5	Recuperação de um ofício	31
4.6	Processo de indexação	32
4.7	Exemplo de indexação	34
4.8	Processo de alimentação do Tesauro	35
4.9	Exemplo de alimentação do Tesauro	37
4.10	Processo de recuperação	38
4.11	Exemplo de recuperação de ofício.....	40
5	TELAS DO SISTEMA	41
6	RESULTADOS	45
	APÊNDICE A (TESAURO)	48
	APÊNDICE B (ARTEFATOS)	56

1 INTRODUÇÃO

Diante do conceito de informação, faz-se necessário delimitar o contexto de sua aplicação para utilizá-la de maneira mais adequada. 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. Essas informações aliadas às novas tecnologias assumiram um papel importante no compartilhamento do conhecimento. Trata-se de grandes volumes de informação e conhecimento, que são utilizados e compartilhados por inúmeras pessoas ao mesmo tempo.

Em uma instituição pública, grande parte das informações geradas são correspondências oficiais denominadas ofícios. Entende-se por ofício uma correspondência oficial, enviada normalmente a funcionários ou autoridades públicas. É o tipo mais comum de correspondência oficial expedida por órgãos públicos.

Os ofícios recebidos e expedidos no Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM) discorrem sobre solicitações, deferimentos, aquisições e diversas outras ações e recomendações importantes para o andamento das rotinas e processos desenvolvidos. Para que este trabalho fosse desenvolvido, foi necessário lidar com o controle do significado das palavras e escolher os melhores termos para a construção do Vocabulário Controlado. Este processo é tecnicamente chamado de indexação que, 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”.

A indexação possui como finalidade principal traduzir a Linguagem Natural utilizada pelos usuários para uma linguagem documental padronizada, aumentando a eficiência na entrada de dados no sistema e a recuperação da informação. Espera-se que com a conclusão do trabalho, o usuário final tenha acesso aos ofícios do Departamento de Computação da UFVJM e que consiga localizar o que necessita com agilidade.

Definição do problema

Identificou-se que o Departamento de Computação da UFVJM não possui um local aonde seja possível pesquisar os ofícios já expedidos. Quando há uma determinada demanda em localizar um ofício, recorre-se a e-mails enviados, a alguma pasta arquivada no computador ou fisicamente, processo tal que dispende muito tempo na busca. Surgiu então a necessidade de procurar uma solução para este problema: arquivar os ofícios de uma forma que seja fácil, rápida e eficaz, possibilitando a recuperação ágil das informações.

A indexação como um processo de análise documentária vem sendo bastante aprofundada pela Biblioteconomia devido ao crescimento exacerbado de informações em todo o mundo, sendo auxiliada pelas Ciências da Computação para buscar as melhores formas de automação, eficácia e rapidez na representação de conteúdos, utilizando de ferramentas computacionais e algoritmos apropriados.

A quantidade e o crescimento contínuo de ofícios dentro do Departamento de Computação, devido ao tempo e crescimento da UFVJM, permitiu a identificação da necessidade de uma indexação eficaz desses documentos de ofício. A indexação é realizada com um tratamento temático da informação, de modo que, as informações contidas nos ofícios são analisadas e, posteriormente, representadas através de termos (símbolos) que são capazes de descrever tudo o que aquele conteúdo representa, sendo assim um objeto fácil para a recuperação por parte dos usuários que desejam ter acesso àquele documento, preferencialmente, por um meio eletrônico.

Portanto, o problema consiste no desenvolvimento de uma ferramenta WEB para automatizar o processo de indexação de ofícios expedidos pelo departamento de computação (DECOM), já que todo documento de ofício que a UFVJM expede ou recebe poderá ser objeto de recuperação de algum usuário, podendo ser esses usuários: professores, técnicos administrativos, pró-reitores, alunos, usuários externos, entre outros. Sendo assim, se faz necessário estudo a fim de aprimorar o armazenamento, organização e recuperação dos ofícios expedidos, bem como ferramentas computacionais para a automatização de todo o trâmite do processo de indexação.

Objetivos gerais

Este trabalho visa utilizar técnicas das Ciências da Informação, dando ênfase às técnicas de indexação utilizadas na Biblioteconomia, para solucionar o problema identificado no Departamento de Computação. Através deste trabalho busca-se desenvolver algoritmos para indexação e recuperação dos ofícios do departamento de computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM). Em seguida, busca-se implementá-lo em um sistema *online*, utilizando das técnicas estudadas de análise documentária.

Objetivos específicos

- Estudar técnicas utilizadas na Biblioteconomia para indexação de documentos, tais como: livros, ofícios, artigos, entre outros;
- Selecionar as técnicas que se adequam melhor ao problema proposto e aos recursos computacionais disponíveis;
- Propor algoritmos próprios que se adequem ao problema específico e a solução, por meio das técnicas selecionadas;
- Implementar os algoritmos em uma ferramenta computacional para automatizar o processo de indexação e recuperação dos ofícios;

Organização do trabalho

A organização do texto deste trabalho baseia-se em capítulos assim divididos:

Capítulo 1 – Introdução: apresenta uma visão geral do trabalho, bem como os problemas que impulsionaram o desenvolvimento dessa pesquisa; o objetivo geral e os objetivos específicos.

Capítulo 2 – Indexação: apresenta a história da indexação e o que é o processo de indexação; Inteligência Artificial, Biblioteconomia e Ciência da Informação; alguns dos

métodos de indexação existentes; Conceitos básicos da biblioteconomia e ciência da informação para auxiliar no entendimento do Sistema Inteligente de Indexação Automática proposto pelo capítulo 4.

Capítulo 3 – Automação da indexação: Apresenta uma visão geral da automação do processo de indexação; o algoritmo simplificado e a indexação de um ofício; algumas formas de se selecionar termos descritores de forma automática e/ou semiautomática; as vantagens e desvantagens da automatização do processo de indexação.

Capítulo 4 – Proposta do sistema: Apresenta as ferramentas computacionais utilizadas para o desenvolvimento do Sistema Inteligente de Indexação Automática; formas de indexação e seleção escolhidas; o passo a passo dos processos de indexação, alimentação semiautomática do Tesouro e recuperação dos ofícios;

Capítulo 5 – Telas do sistema: Apresenta algumas telas do Sistema Inteligente de Indexação Automática.

Capítulo 6 – Resultados: Apresenta uma síntese do trabalho.

2 INDEXAÇÃO

A informação é um item, atualmente, essencial para todas as organizações, sendo alvo de profundo estudo por profissionais da Ciência da Informação. Segundo Rousseau e Couture (1998), a informação deve ser tratada, numa empresa, como um recurso tão importante quanto os recursos humanos, materiais ou financeiros. Ressalta-se ainda o fato que a informação se torna acessível por meio de Sistemas de Informação.

Com o objetivo de solucionar o problema identificado no Departamento de Computação, foram propostos três algoritmos. Para entender os algoritmos e o sistema desenvolvido é necessário entender minimamente as áreas e conceitos que tangem toda a pesquisa e a implementação da solução do problema por meio do Sistema Inteligente de Indexação Automática. Portanto, neste tópico, serão apontadas algumas das técnicas e conceitos utilizados na área da Ciência da Informação, mais especificamente na Biblioteconomia, para o desenvolvimento de soluções de problemas de recuperação da informação.

Ciência da informação e Biblioteconomia

A Ciência da Informação (CI) é um campo que vem crescendo constantemente, pois é a área preocupada com a análise, coleta, classificação, manipulação, armazenamento, recuperação e disseminação da informação. Dado a grande massa de informação existente na atualidade e o potencial poder que essas informações carregam intrinsecamente, os desafios em tornar essas informações em objetos fáceis de serem recuperados por parte de algum usuário se tornam cada vez mais desafiadores, complexos e indispensáveis. A CI tem sido muito visada por instituições de todos os níveis que almejam ter controle sob as informações, pois essa ciência, com ênfase na Biblioteconomia, estuda justamente técnicas ligadas à gestão de informações, conseguindo, assim, disponibiliza-las sempre que necessário de forma rápida e eficaz.

Tefko Saracevic (1996, p. 2), expõe que:

Existem características gerais que constituem a razão da existência e da evolução da Ciência da informação. Primeira, a CI é, por natureza, interdisciplinar, embora suas relações com outras disciplinas estejam mudando. A evolução interdisciplinar está longe de ser completada. Segunda, a CI está inexoravelmente ligada à tecnologia da informação. O imperativo tecnológico determina a CI, como ocorre também em outros campos. Em sentido amplo, o imperativo tecnológico está impondo a transformação da sociedade moderna em sociedade da informação, era da informação.

A Biblioteconomia é considerada uma área do conhecimento que envolve organismos, operações técnicas e princípios que direcionam os documentos a uma gestão eficaz, contribuindo com diversos segmentos sociais (SHERA, 1980).

A Biblioteconomia é uma das áreas de aplicação das Ciências da Informação que consiste no estudo de aspectos representativos, sistemáticos e do uso e disseminação da informação. É a área que trata de análise, planejamento, implementação e organização da informação em bancos de dados, Sistemas de Informação, bibliotecas, catálogos, entre outros.

Dada as duas definições, Biblioteconomia e Ciência da Informação, é notório que há uma relação indissociável entre as duas ciências, e ambas possuem diferenças pouco significativas no contexto do trabalho desenvolvido, sendo assim, mencionadas de forma intercambiável por essa monografia.

História da indexação

A Ciência da Informação teve sua origem, notoriamente, na revolução científica e técnica que se seguiu à segunda guerra mundial, assim como outros inúmeros campos interdisciplinares, como a Ciência da Computação, Pesquisa Operacional, etc. Como exemplo dessa assertiva, Segundo Tefko Saracevic (1996), ocorreu algo comprobatório nos Estados Unidos, quando o Congresso e outras agências governamentais aprovaram, durante os anos 50 e 60, inúmeros programas estratégicos que financiaram os esforços em larga escala para controlar a explosão informacional ocorrida naquela época, poucos anos após a segunda guerra mundial. Empresas privadas uniram-se a eles. Eventualmente, esses programas e esforços foram responsáveis pelo desenvolvimento da moderna indústria da informação e das concepções que a direcionam.

O surgimento da explosão informacional nos anos 50 e 60 (Tefko Saracevic, 1996) fez

com que alavancasse uma demanda na organização eficaz dessa gama de conhecimento. No entanto, o marco da automatização do processo de indexação surgiu, em evidência, quando houve um aumento considerável da produção científica e do desenvolvimento de novas tecnologias e a inviabilidade da organização manual de tantos conteúdos, em meados do século XX, levando ao aparecimento de discussões relacionadas à automatização dos processos de tratamento dos documentos (NARUKAWA, 2009). A automatização dos processos de tratamento desses documentos não era apenas para facilitar as tarefas e tornar possível a organização de tantos conteúdos, buscava-se também a garantia de resultados mais eficazes, conciliando eficácia e velocidade. Nesse sentido, há esforços para o desenvolvimento de softwares de indexação.

Inteligência Artificial

A Inteligência é um conceito muito abrangente e se estende por diversas definições. Segundo Richardson, “A inteligência é um conceito fundamental da psicologia moderna que todos utilizam; entretanto, quase ninguém consegue defini-la de modo definitivo ou pelo menos amplamente convincente” (1999 *apud* DALGALARRONDO, 2008, p. 277).

O homem é um ser naturalmente inteligente, pois possui capacidade única de raciocínio. A capacidade de raciocínio permite que o homem perceba, compreenda e manipule um mundo maior e mais complexo que ele próprio (GOMES, 2010). A compreensão e sintetização do próprio processo de raciocínio é estudada por uma das áreas da computação, denominada por Inteligência Artificial, que busca simular o processo de forma computacional.

A Inteligência pode ser definida como:

O conjunto das habilidades cognitivas dos indivíduos, a resultante, o vetor final dos diferentes processos intelectivos. Refere-se à capacidade de identificar e resolver problemas novos, de reconhecer adequadamente as situações vivenciais cambiantes e encontrar soluções, as mais satisfatórias possíveis para si e para o ambiente. (DALGALARRONDO, 2008, p. 277).

A Inteligência Artificial pode ser definida por um processo que busca simular o processo de aprendizagem do homem ou o processo de atuação do homem em algum campo específico. Segundo Kurzweil, uma das linhas de pensamento da Inteligência Artificial é “a arte de criar máquinas que executam funções que exigem inteligência quando executadas por

peessoas” (1990 *apud* GOMES, 2010, p. 235).

O trabalho foca no desenvolvimento de algoritmos que buscam imitar os profissionais da Biblioteconomia no momento da indexação e da recuperação da informação, utilizando das técnicas e objetos empregados pelos bibliotecários.

Linguagem Documentária

O aumento do acesso à informação devido às tecnologias de informação e a incapacidade humana em absorver todos estes materiais de conhecimento de uma só vez fez com que houvesse uma necessidade incontestável na “gestão” eficaz da informação, objetivando organizar o saber para, principalmente, possibilitar a sua recuperação. Por esta razão, se justifica a utilização das Linguagens Documentárias (LD) como uma forma de estruturar conhecimentos e tornar a disseminação do conhecimento algo essencial.

Para acessarmos um determinado conhecimento, faz-se necessário conhecer uma informação precisa que permitirá o retorno àquele conteúdo, como um ciclo. Sendo assim, as Linguagens Documentárias nos permitem estruturar certos conhecimentos, de acordo com diferentes segmentos sociais, transformando-lhes em informações uteis para que possam representar o conhecimento e, assim, possibilitar a sua recuperação. De acordo com Cintra, “(...) compete às Linguagens Documentárias transformar estoques de conhecimento em informações adequadas aos diferentes segmentos sociais” (1994 *apud* BELUCHE, 2008, p. 116).

As Linguagens Documentárias podem então ser definidas como “instrumentos de comunicação documentária, pois, realizam a tradução do conteúdo dos documentos e das perguntas dos usuários para unidades aptas a integrar sistemas de recuperação da informação” (CINTRA, 1994, p. 24). Dessa forma, pode surgir a pergunta: como se organiza a informação para que o conhecimento fique disponível? A resposta é: pelas Linguagens Documentárias, pois é uma das condições para plena recuperação da informação, representando o conhecimento e promovendo interação entre usuário e conteúdo através da transformação de informação de textos feitos com Linguagem Natural e/ou especializada para uma Linguagem Documentária controlada.

As Linguagens Documentárias restringem o acesso à dinamicidade dos textos,

transformando essa dinamicidade em versões representativas econômicas do conteúdo. As Linguagens Documentárias são capazes de reduzir a diversidade e ambiguidade da terminologia e resolver a problemática referente à utilização de diferentes palavras por diferentes autores. São também capazes de cobrir conceitos de uma determinada ciência, visando a uniformização do uso de linguagens especializadas e, impedindo, portanto, que existam Linguagens Documentárias gerais.

Vocabulário Controlado: Tesouro

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.

Em outras palavras, o controle do vocabulário nada mais é que o produto derivado da Linguagem Documentária. A Linguagem Documentária serve como um esqueleto para que o Vocabulário Controlado possa existir, pois, como dito no tópico 2.6, ela é a responsável pela transformação de informação de textos feitos com Linguagem Natural e/ou especializada para um Vocabulário Controlado.

Segundo a norma ANSI-NISO Z39. 19-2003, existem quatro funções distintas exercidas pelo controle dos vocabulários, sendo essas:

- 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;
- A recuperação, que serve como auxílio na busca e na recuperação de documentos.

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, 2011, 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 Tesouros (utilizado em todo o trabalho), os esquemas de classificação bibliográfica e as listas de cabeçalhos de assunto.

Um Vocabulário Controlado, à exemplo do Tesouro, é organizado da seguinte forma (ALVARES, 2014, p. 1):

- **Descritor:** Termo escolhido para representar um conceito no Tesouro 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.
- **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.

A estrutura de um Vocabulário Controlado contém inúmeras relações lógicas entre os termos, o que proporciona um acervo de termos altamente eficaz na representação de um conteúdo, pois descrevem ou visam descrever, com seus termos descritores e suas relações lógicas, exatamente a informação necessária para representar o objeto alvo da indexação. Como exemplo do poder dessas relações fornecidas pelo Vocabulário Controlado pode-se citar um ofício expedido por algum órgão de meteorologia, onde se solicita o nível pluviométrico de alguma região. Na indexação deste documento de ofício, o profissional e/ou sistema indexador poderia indexar este ofício pelo termo chuva, desde que o Vocabulário Controlado esteja estruturado explicitando essa relação lógica existente entre esses dois termos.

Dado o exemplo, observa-se que a utilização do Vocabulário Controlado expõe uma estratégia de indexação uma vez que se utiliza do controle lógico dos termos, ajudando a comunidade, em especial, o usuário leigo, do sistema de indexação, a encontrar e recuperar o

conhecimento transmitido pelos documentos.

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).

É, portanto, “uma Linguagem Documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos” (SALES, 2008, p. 103), que representa uma determinada área do conhecimento. O Vocabulário Controlado é desenvolvido por um profissional da Biblioteconomia, através das atividades:

- **Desenvolvimento da Linguagem Documentária:** consiste na extração dos termos de indexação dos documentos de uma determinada área por meio de técnicas desenvolvidas e utilizadas na Biblioteconomia;
- **Desenvolvimento do Vocabulário Controlado (Tesouro):** Consiste no desenvolvimento das relações lógicas existentes entre os termos por meio de técnicas desenvolvidas e utilizadas na Biblioteconomia.

Na Figura 2.1 é apresentado o fluxograma para o desenvolvimento de um Vocabulário Controlado.

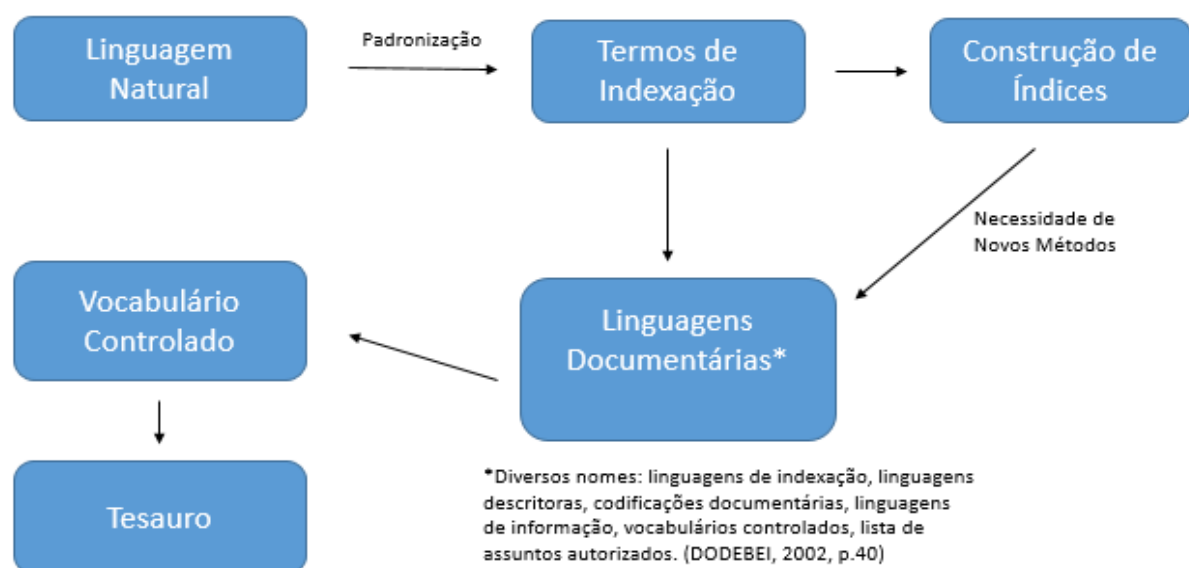


Figura 2.1- Construção do Vocabulário Controlado

Fonte: (DODEBEI, 2002)

Métodos de indexação

A indexação pode ser entendida como uma tradução de um documento qualquer em termos documentários, que são conhecidos por descritores, cabeçalhos de assuntos ou termos-chaves, que visam expressar todo o conteúdo pertencente ao documento; ou seja, é a tradução, no sentido literal da palavra, de termos em conteúdos armazenados em algum tipo de estoque de informações.

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”.

Para Cintra *et al* (2002, p. 39), a indexação é algo um pouco mais complexo e detalhado:

A operação de tradução de textos em LN [Linguagem Natural] para uma LD [Linguagem Documentária] denomina-se indexação. [...] As várias frases do processo analítico apresentam uma complexidade considerável, pois não se trata de adquirir os documentos e armazená-los numa ordem lógica. A documentação é memória, seleção de ideias, reagrupamento de noções e de conceitos, síntese de dados. Trata-se de triar, de avaliar, de analisar, de ‘traduzir’, de encontrar respostas para necessidades específicas.

Olhando sob o prisma da autora citada acima, a indexação configura-se em um processo bem mais abrangente do que a definição “NBR 12676” o coloca, visto a complexidade que existe inerente ao processo de indexação executado na prática.

As indexações de documentos demandam um alto nível de concentração e habilidade por parte do indexador. A etapa em que se definem os termos que virão a representar o conteúdo de um documento deve ser corretamente avaliada, pois deste processo dependerá a recuperação da informação, por parte do usuário final. Todavia, toda indexação é única e dependente de vários fatores, podendo ser citado como fator crucial um determinado público, haja vista que documentos idênticos, em algumas etapas da indexação, são suscetíveis a diferentes resultados, dependendo do seu público alvo. À exemplo podemos citar a análise conceitual, que se define por analisar do que se trata o documento, de uma forma mais geral. Segundo Lancaster (2004, p. 9):

A análise conceitual, em primeiro lugar, implica decidir do que trata um documento – isto é, qual seu assunto. [...] Esta afirmativa sobre análise conceitual está bastante simplificada. A indexação de assuntos é normalmente feita visando a atender às necessidades de determinada clientela – os usuários de um centro de informação ou de uma publicação específica. Uma indexação de assuntos eficiente implica que se

tome uma decisão não somente quanto ao que é tratado num documento, mas também porque ele se reveste de provável interesse para determinado grupo de usuários. Em outras palavras, não há um conjunto 'correto' de termos de indexação para documento algum. A mesma publicação será indexada de modo diferente, se os diferentes grupos de usuários estiverem interessados no documento por diferentes razões.

Visto algumas definições de indexação podemos concluir que não é um processo simples, é algo além de um simples algoritmo de indexação, que depende de inúmeros fatores, tais como: público alvo, nível de concentração do indexador, nível de conhecimento do indexador sobre o tema a ser indexado, entre outros.

No campo dos estudos organizacionais existem inúmeras abordagens metodológicas, podendo ser tanto de caráter quantitativo quanto qualitativo. No próximo tópico será falado sobre esses dois caracteres de indexação. O primeiro a ser explicado se utiliza de uma ideia numérica a qual é levado em consideração a quantificação de algo, podendo ser termos, frases, palavras compostas, entre outros, para realizar o processo de indexação. O segundo tem uma ideia mais complexa, levando em consideração fatores linguísticos de uma determinada língua. Ambos são modelos de indexação, que visam identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos.

Indexação quantitativa (métodos estatísticos)

A abordagem quantitativa utiliza a quantificação dos dados de um determinado documento como o cerne de todo o processo de indexação, sendo os dados coletados e analisados estatisticamente a fim de se identificar os termos que melhor o representam. Segundo Oliveira (2001), o método quantitativo:

Significa quantificar opiniões, dados, nas formas de coleta de informações, assim como também com o emprego de recursos e técnicas estatísticas desde a mais simples, como percentagem, média, moda, mediana e desvio padrão, até as de uso mais complexo, como coeficiente de correlação, análise de regressão, etc. [...].

A abordagem quantitativa, utilizada para o desenvolvimento de todo o trabalho, foca meramente em números. Essa abordagem se utiliza de técnicas estatísticas aplicadas a dados numéricos extraídos mediante a quantificação das palavras dentro de um contexto. As técnicas aplicadas permitem inferir possíveis termos representativos para a indexação dos documentos propostos (ofícios).

Existem duas maneiras de se realizar a indexação quantitativa. A primeira é a indexação quantitativa feita por atribuição, isto é, os termos representativos de um determinado documento são extraídos mediante a presença de um Vocabulário Controlado. Em outras palavras, os termos são selecionados e extraídos do documento de acordo com os dados numéricos obtidos por meio de cálculos estatísticos e, em seguida, são adequados aos termos presentes em um Vocabulário Controlado.

A segunda é a indexação quantitativa feita por extração, que se utiliza de termos empregados pelo expedidor do documento proposto (neste caso, um ofício), sem o suporte de um Vocabulário Controlado. Isto é, os termos indexados ao documento são extraídos do próprio documento (Linguagem Natural) sem se adequarem a um padrão pré-definido estabelecido pelo Vocabulário Controlado.

Indexação qualitativa (métodos linguísticos)

No início dos anos sessenta, surgiram estudos que associassem técnicas de processamento da Linguagem Natural (estudo e análise dos aspectos linguísticos de um texto mediante utilização de programas informáticos) e a automatização da indexação (Gil Leiva, 1999).

A indexação qualitativa tem um foco diferente da quantitativa, pois ela busca levar em consideração aspectos linguísticos dos objetos alvo de indexação, tais como: morfologia, sintaxe, semântica; sendo, portanto, um processo mais fácil de ser executado por seres humanos, já que os computadores têm um funcionamento totalmente numérico, impossibilitando ou, ao menos, dificultando a automação desse tipo de indexação.

Este tipo de indexação se divide basicamente em três categorias de análises, sendo elas:

- Analisador morfológico: Este analisador cuida do tratamento das palavras de forma individual, analisando e classificando as palavras às suas classes gramaticais;
- Analisador sintático: Este analisador se dedica ao tratamento das orações, ou seja, analisar as palavras e as conexões lógicas dentro de um contexto;
- Analisador semântico: Este analisador objetiva o tratamento de significado das

palavras em diferentes orações, ou seja, em diferentes contextos.

Os progressos proporcionados pelo estudo da abordagem de indexação qualitativa têm contribuído com diversas categorias de *softwares*, tais como: recuperação de informação, extração de informação, classificação, indexação e resumos de documentos (Gil Leiva, 1999).

Política de indexação

Quando se fala em Indexação e Vocabulário Controlado é inexorável que se fale sobre a política de indexação, pois ela que define “parâmetros” do processo de indexação e, conseqüentemente, reflete na recuperação da informação, uma vez que estes dois processos estão estritamente relacionados por meio de uma relação de causa e efeito.

À exemplo dos “parâmetros” mencionados acima, pode-se citar a especificidade de um processo de indexação, que é simplesmente o quão específico esses termos precisam ser para, assim, representar ou não um determinado conteúdo. Segundo Leiva e Fujita (2012), a indexação possui várias características que causam efeitos na recuperação da informação, ou seja, existe uma série de regras definidas pela política de indexação, que precisam ser seguidas para se obter a proposta do sistema de indexação implantado.

Os profissionais que realizam o tratamento dos documentos alvos do sistema de indexação (normalmente um bibliotecário), que realizam o desenvolvimento do Sistema de Informação que vai gerir o processo de indexação, que vão gerir os sistemas de recuperação da informação e que vão usufruir do sistema, são as pessoas interessadas pela política de indexação (LEIVA e FUJITA, 2012).

De acordo com Fujita (1999), a política de indexação:

Não deve ser vista como uma lista de procedimentos a serem seguidos, e sim um conjunto de decisões que esclarecem os interesses e os objetivos de um sistema de informação. A política decide não só sobre a consistência dos procedimentos de indexação em relação aos efeitos que se necessita obter na recuperação mas, principalmente, sobre a delimitação de cobertura temática em níveis qualitativos e quantitativos tendo em vista os domínios de assuntos e as demandas dos usuários. Isso nos leva a pensar na indexação sobre o ponto de vista gerencial e estratégico no contexto de unidades de informação haja vista ter efeitos na entrada e na saída de informações do sistema que a indicam como “[...]a parte mais importante porque condiciona o resultado de uma estratégia de busca, produzindo uma correspondência precisa com assuntos pesquisados em índices (FUJITA, 1999a, p. 17).

Para que essa correspondência, mencionada por Fujita (1999a), aconteça, no entanto, é necessário, imprescindivelmente, que se utilize de uma política de indexação, pois ela

norteará princípios e critérios que servirão como um guia nas decisões que estão diretamente ligadas à otimização dos serviços e racionalização dos processos.

Exaustividade e especificidade

Quando se há o intuito de analisar a eficácia de um determinado método é preciso que se utilize de métricas. Desta forma, o presente trabalho usa métricas para configurar o sistema e alcançar a eficácia desejada na indexação dos documentos trabalhados.

Uma das métricas utilizadas neste trabalho foi a exaustividade que é, segundo Milena Polsinelli (2009), o número de termos atribuídos como descritores do assunto tratado no documento. Ou seja, até que ponto os assuntos abordados no documento são relevantes para a indexação e, então, traduzidas pela Linguagem Documentária para indexá-los. Quanto mais exaustiva for a indexação, mais termos ela vai empregar. Um alto índice de exaustividade empregada na política de indexação é indicado para sistemas de indexação cujo público alvo é incerto. À exemplo, poderíamos citar uma biblioteca universitária, que é um local composto por um acervo considerável de livros de diversas áreas, e conseqüentemente, tem um público variado.

A outra métrica utilizada foi a especificidade da indexação, que determina o quão específico aquele termo precisa ser para indexá-lo a um determinado documento, ou seja, até que ponto a Linguagem Documentária utilizada pela instituição indexadora permite relacionar o documento a outros termos. Um índice baixo de especificidade definido na política de indexação da instituição poderá resultar em uma alta revocação na recuperação da informação, ou seja, em uma grande quantidade de documentos anexados (POLSINELLI, 2009).

Segundo Milena Polsinelli (2009):

A especificidade está relacionada ao nível de abrangência que a biblioteca e a Linguagem Documentária permitem especificar os conceitos identificados no documento. Exemplo: um livro cujo assunto seja especificamente sobre “tilápias” será indexado sob o assunto “peixes”. Essa situação é característica de bibliotecas que optam por uma baixa especificidade nos assuntos que, por sua vez, trará como resultados na recuperação uma alta revocação (POLSINELLI, 2009, p. 6).

Sendo assim, pode-se concluir que quanto mais elevado for o índice de exaustão em

uma política de indexação, mais termos serão indexados aos documentos; em contrapartida, menos precisos eles serão, uma vez que podem desviar do assunto realmente tratado.

Esses princípios de indexação são definidos pela política de indexação, e devem ser seguidos rigorosamente pelo bibliotecário ou pelo Sistema de Informação responsável no momento da análise de assuntos feita na indexação (POLSINELLI, 2009).

Existe uma forma de avaliação chamada de “Avaliação extrínseca mediante a recuperação”, proposta por Narukawa (2009) para comparar dois tipos de indexação derivados do mesmo sistema. Essa avaliação consiste em “interrogar duas bases de dados que contém os mesmos campos e idênticos conteúdos salvo os campos que armazenam a indexação” (NARUKAWA, 2009, p. 110). Dessa forma, pode-se definir o quão exaustivo e preciso o sistema de indexação é, através dos cálculos propostos por Narukawa (2009), conforme segue abaixo:

$$EXAUSTIVIDADE = \frac{N^{\circ} \text{ DE DOCUMENTOS RELEVANTES RECUPERADOS}}{N^{\circ} \text{ DE DOCUMENTOS RELEVANTES NA COLEÇÃO}} \quad (1)$$

$$PRECISÃO = \frac{N^{\circ} \text{ DE DOCUMENTOS RELEVANTES RECUPERADOS}}{N^{\circ} \text{ TOTAL DE DOCUMENTOS RECUPERADOS}} \quad (2)$$

3 AUTOMATIZAÇÃO DA INDEXAÇÃO

A análise automática de textos é, atualmente, vista como um campo de pesquisa essencial para as Ciências da Informação e tem sido alvo de discussão por profissionais da área há mais de trinta anos. Visto a possibilidade e, portanto, a busca em fornecer acesso à literatura técnico-científica mais rapidamente, por meio de computadores no processamento de dados e informações, este campo de pesquisa vem se tornando cada vez mais procurado (ROBREDO, 2005).

Com o objetivo de facilitar o processo de indexação devido à grande quantidade de documentos oficiais no departamento de computação, pensou-se no desenvolvimento de um sistema de indexação eletrônico cujo objetivo é imitar o processo de indexação realizado manualmente por um profissional da área. Segundo Robredo a indexação automática consiste em “[...] qualquer procedimento que permita identificar e selecionar os termos que representem o conteúdo dos documentos, sem a intervenção direta do indexador” (2005 *apud* NARUKAWA, 2009, p. 101).

No entanto, existe um problema intrínseco ao processo de indexação via computadores: as máquinas trabalham em base binária, ou seja, são desenvolvidas através de princípios numéricos. Esta assertiva acaba dificultando a indexação realizada por computadores, uma vez que os computadores não trabalham com linguísticas diretamente, tornando o processo de análise morfológica, sintática e semântica extremamente complexo de se implementar através de algoritmos computacionais. Considerando a posição de Lancaster (2004) e Moreiro Gonzáles (2004) que condiz:

Com a proposta dos *softwares* de indexação semi-automáticos na medida em que defendem o uso de software para indexar documentos com a ressalva de que é necessário ter em mente as limitações em seu processo, as quais só a análise apurada de um ser humano pode solucionar, sendo claro a necessidade de avaliação final dos termos de indexação por um profissional. (Lancaster e Moreiro Gonzáles, 2004, p. 29).

Seleção de termos descritores

Na indexação automática e/ou semiautomática existem diferentes formas de se

selecionar possíveis termos descritores do conteúdo do documento. Duas dessas formas são bastante conhecidas. A primeira forma é a extração dos termos mediante a retirada pura – sem a utilização de qualquer Linguagem Documentária para impor regras ou padronizações. A segunda é uma extração dos termos mediante a utilização de uma Linguagem Documentária para impor regras e padronizar a indexação. De acordo com Lancaster (2004), as duas formas são definidas por:

- Extração automática, que consiste em extrair as palavras do texto sem adequá-las a uma Linguagem Documentária, para serem usadas como representantes do conteúdo tratado pelo documento. Ou seja, se utiliza da própria Linguagem Natural para extrair termos representativos do documento a ser indexado;
- Existe também a seleção de termos descritores por atribuição automática que consiste em representar o conteúdo tratado pelo documento mediante termos selecionados de uma Linguagem Documentária. Segundo Lancaster (2004), é considerado mais difícil quando se pretende usar um computador para executar a extração, sendo assim necessário o desenvolvimento, para cada termo a ser atribuído, de um ‘perfil’ de palavras ou expressões que ocorram frequentemente.

De acordo com Gil Leiva, existem três conceitos diferentes de indexação automática:

- os programas realizam o processo de armazenamento dos termos de indexação obtido por um profissional, chamado de indexação assistida por computador;
- os programas realizam a análise dos documentos de modo automático e se necessário os termos são validados por um profissional, denominado indexação semiautomática;
- Os programas realizam o processo de análise dos documentos e não ocorre validação por profissionais ou, a indexação automática, propriamente dita (1999, p. 57; 2008, p. 320 *apud* NARAKAWA, 2009, p. 101).

Vantagens e desvantagens da automação

A descrição documental tradicional estabelecida basicamente na catalogação e classificação dos documentos não são mais suficientes. De acordo com João Sequeira (1998, p. 2), “As tarefas de resumo e indexação ganharam no seio da comunidade arquivística e bibliotecária importância crescente e ocupam ainda hoje nas bibliotecas e arquivos, um conjunto muito vasto de técnicos e consomem enormes recursos financeiros”. Portanto,

sistemas automatizados de indexação que fazem os acessos às informações serem rápidos e viáveis se transformaram em uma necessidade, e não mais uma opção.

Outro benefício gerado pelos sistemas de indexação automatizados é a disponibilidade que eles proporcionam aos consumidores de informação. De acordo com João Sequeira (1998), a complexidade dos processos manuais de indexação e resumo, observados principalmente pelo quesito mão de obra qualificada e, conseqüentemente, bem remunerada, faz com que os produtos documentais agreguem valor e se torne caros o suficiente para o não acesso por parte da generalidade dos consumidores de informação.

Ao reduzir os critérios que, naturalmente, são impostos por indexadores humanos, há uma tendência à padronização do processo de indexação, podendo ser uma vantagem, uma vez que os sistemas de indexação serão mais homogêneos e coerentes; conseqüentemente, mais fáceis de recuperar informação. Em contrapartida, serão suscetíveis a um maior número de termos descritores ineficazes para representação do conteúdo, pois o computador pode simular um processo humano, mas não ter autonomia para tomada de decisões em situações cruciais, onde apenas um ser humano teria o poder de análise e decisão.

Por fim, a velocidade da evolução em que ramos do conhecimento evoluem faz com que, cada vez mais, se torne mais difícil a atualização permanente dos temas por parte dos indexadores humanos - quesito considerado essencial para o processo de indexação, sendo sistemas automáticos uma ótima solução ao crescimento e disseminação de novas informações.

4 PROPOSTA DO SISTEMA

Diante dos problemas encontrados no Departamento de Computação da UFVJM, tais como a falta de organização dos arquivos, falta de padronização no armazenamento e, conseqüentemente, difícil recuperação desses arquivos, foi proposto um Sistema Inteligente de Indexação Automática, que possibilita o armazenamento de forma organizada, padronizada e digital, facilitando a recuperação dos arquivos. O Sistema Inteligente de Indexação Automática 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 metodologia aplicada a este trabalho foi uma pesquisa documental embasada nos arquivos do Departamento de Computação da UFVJM para dar suporte ao Sistema de Indexação Automática. Para o sistema proposto foi utilizada a linguagem de programação JAVA Standard Edition, na versão 1.7.0_10. Também foi utilizada a especificação TOMCAT, na versão 7.0, para JAVA Enterprise Edition, bem como o banco de dados MYSQL para armazenamento dos dados referentes aos atributos do sistema. Além da linguagem de programação e do Sistema Gerenciador de Banco de Dados, foram utilizadas bibliotecas disponibilizadas gratuitamente na internet, que auxiliaram em todo o processo de *upload* e análise dos arquivos de extensão PDF. As bibliotecas utilizadas foram:

- pdfbox-app-1.8.10;
- commons-fileupload-1.3.1;
- commons-io-2.4.

Todas as bibliotecas utilizadas são desenvolvidas pela fundação Apache, sendo, portanto, encontradas no endereço oficial da fundação.

Para que este trabalho atingisse os objetivos de automação na indexação, foi necessária a simulação de um processo de indexação manual passo-a-passo e a discussão para definir a melhor forma de “imita-lo” via algoritmos. Sendo assim, chegou-se à conclusão que a melhor forma seria através de uma base de dados previamente desenvolvida por um profissional da área da indexação, denominado por Vocabulário Controlado, na qual são selecionados todos os termos que, à priori, possivelmente representassem os assuntos tratados pelo departamento.

Para maiores detalhes técnicos do Sistema Inteligente de Indexação Automática, está

disponibilizado no Apêndice B alguns dos artefatos utilizados no desenvolvimento do sistema, tais como: Pseudocódigo, Modelo lógico do banco de dados, Diagrama de classes, Diagrama de casos de uso, Diagrama de sequência, Análise de requisitos.

Indexação e seleção

Dadas as definições e perspectivas de autores conceituados na área da Ciência da informação (Capítulo 2), a proposta de indexação automática para este trabalho abrange parte da indexação assistida por computador, definida acima por Gil Leiva (1999) e complementado por João Sequeira (1998) como sendo um processo que possui duas fases principais:

- Uma fase de pré-indexação automática, que consiste na análise de um documento por parte de um Sistema de Informação, e na conexão do documento aos descritores, que são normalmente extraídos de uma lista de autoridade ou de um Tesouro;
- Fase de comunicação entre o sistema e um profissional da área de indexação, que validará todos os termos associados na pré-indexação automática realizada pelo sistema.

O Sistema Inteligente de Indexação Automática utiliza também da seleção de termos descritores por atribuição automática citada por Lancaster (2004), mediante um Vocabulário Controlado (Tesouro) desenvolvido por uma bibliotecária e alimentado por usuários indexadores do Sistema Inteligente de Indexação Automática.

Algoritmo simplificado do sistema

Inicialmente, foi feita a construção de um Vocabulário Controlado (Apêndice A), extraíndo todos os termos relevantes dos 239 ofícios, compreendidos entre o período de 2011 e 2014. Esse vocabulário deu suporte a toda implementação do Sistema Inteligente de Indexação Automática de ofícios do Departamento de Computação 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 serão utilizados para representar (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, gestor, representante, avaliadores, professores, professor substituto, professor avaliador, professor adjunto, professor assistente, professor temporário, professor palestrante, e professor visitante. O Tesouro é alimentado no decorrer das interações dos usuários indexadores para com o sistema, por meio da pré-seleção dos termos e definição final por parte dos usuários.

Indexação de um ofício

A indexação de um ofício ocorre quando um novo ofício é inserido no Sistema Inteligente de Indexação Automática. O sistema recupera todos os termos presentes neste ofício e todo o conjunto de termos representativos do Tesouro (presente no banco de dados). Os termos do ofício que acabam de ser inseridos no sistema são comparados aos termos do Tesouro existente. Logo, todos os termos em comum, pertencentes aos dois conjuntos, são selecionados para a criação de um novo Tesouro.

Depois de feita a interseção e criação do novo Tesouro – menor e direcionado à indexação em questão, serão escolhidos 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. 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 feito um ranking com todas as pertinências e 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, com os pesos individuais de cada indexação sendo proporcional ao número de repetições do termo no ofício a ser indexado. O peso de cada

indexação é definido pela fórmula abaixo.

$$PESO = \frac{CONTAGEM\ DO\ TERMO\ INDIVIDUAL\ NO\ OFÍCIO}{NÚMERO\ TOTAL\ DE\ TERMOS\ NO\ OFÍCIO} * 100 \quad (1)$$

Alimentação do Tesouro

A possível alimentação do Tesouro se inicia com um pré-processamento dos termos que constitui o conteúdo do ofício. Inicialmente, o sistema segrega os termos potencialmente habilitados a serem inseridos no Tesouro, determinados através de uma simples comparação e, então, eliminação de todos os termos já existentes no Tesouro. Ao segregar esses termos, são desconsiderados quaisquer termos compostos existentes no conteúdo do ofício, já que a quantidade de termos compostos extraídos pelo algoritmo seria excessivamente grande se comparado aos termos capazes de representar o conteúdo, tornando a análise por parte do usuário indexador inviável. Após a retirada dos termos existentes no Tesouro, os termos são comparados a uma lista de palavras consideradas vazias no que concerne à representatividade de conteúdos no âmbito da língua portuguesa. As palavras pertencentes à lista de palavras vazias são extraídas de algumas classes gramaticais do português, tais como: artigos, pronomes, preposições, conjunções, interjeições, entre outras. Por fim, as palavras são mostradas ao usuário e, posteriormente, selecionadas, facultativamente, para a alimentação do Tesouro. Os termos selecionados pelo usuário indexador são inseridos ao Tesouro e indexados com peso conforme a fórmula abaixo.

$$PESO\ ALIMENTAÇÃO = \frac{SOMA\ DE\ TODOS\ OS\ PESOS\ NA\ BASE\ DE\ DADOS}{NÚMERO\ DE\ PESOS\ SOMADOS} \quad (2)$$

Recuperação de um ofício

Para a recuperação de um ofício, o usuário precisará entrar com dados do expedidor,

data do ofício e termos chaves. O número de termos chaves desejado será escolhido pelo usuário arbitrariamente (num total de no máximo cinco). Em seguida, com a ajuda do sistema, que irá propor opções de termos chaves baseadas no Tesouro, e também possíveis expedidores baseados nos usuários já cadastrados no sistema, o usuário preencherá os dados. Após a inserção destes dados, o sistema realizará uma busca ao banco de dados, recuperando todos os ofícios que se “ligam” (casam) a todos os dados fornecidos pelo usuário. Ademais, o sistema ordenará o resultado de acordo com o peso de cada “ligação” individual entre ofício e termo, que é definido pelo peso individual de indexação (explicado anteriormente).

Processo de indexação

O processo de indexação se inicia através do *upload* do ofício por parte do usuário indexador. Em caso de êxito nesta atividade do processo, verifica-se o conteúdo presente no arquivo de extensão PDF inserido pelo usuário e inicia a leitura do ofício. Em um eventual erro por parte do Sistema Inteligente de Indexação Automática ou por quaisquer equívocos realizados pelo usuário indexador, retorna-se ao estado inicial do processo de indexação.

Após a inserção e conferência de padrões exigidos pelo Sistema Inteligente de Indexação Automática é iniciado outras etapas deste processo. O arquivo inserido é, então, delimitado em conteúdos úteis para o processo de indexação e comparados individualmente a todo o Tesouro desenvolvido para dar base ao processo de indexação. Os três primeiros termos (ordenados pelo índice de frequência de cada termo) são selecionados e, então, investigados, novamente no Tesouro, em busca de todos os termos relacionados.

Por fim, será gerada uma lista de representativos. Essa lista de termos é indexada ao ofício inserido pelo usuário indexador. Na Figura 4.1 é apresentado o fluxograma do processo de indexação.

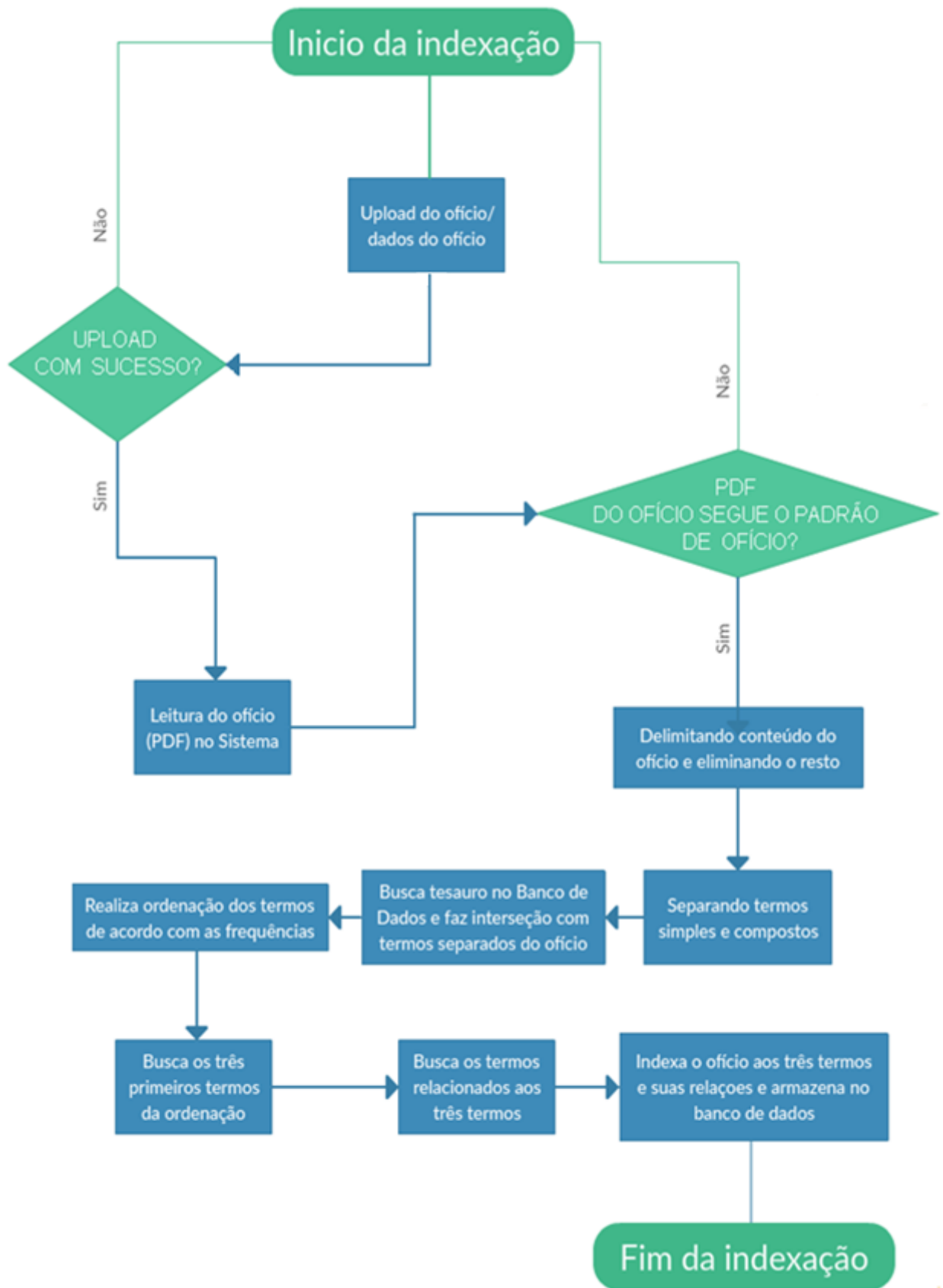


Figura 4.1 - Fluxograma do processo de indexação

Fonte: Desenvolvido pelo proponente

Exemplo de indexação

Para exemplificar, foi feita a inserção de um ofício no sistema - conforme a Figura 4.2, e em seguida, automaticamente, foi capturado todos os termos do ofício que fazem interseção com o conjunto representativo de termos.



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,

Figura 4.2 - Exemplo de ofício

Fonte: Ofício extraído do Departamento de Computação

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 de ocorrências relativamente alto se comparado aos outros termos, tornando-o um termo fraco para a indexação.

Em seguida, o sistema busca no Tesauro 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 no Tesauro, sendo o termo “renovação de garantia” ligado ao termo “impressoras multifuncionais”, “decom” e “ufvjm” ligados a “departamento de computação”, e “gabinetes” ligado a “cômodos”.

Processo de alimentação do Tesauro

O Fluxograma do processo de alimentação do Tesauro se inicia com a conferência da realização do processo anterior (Processo de indexação). Em um eventual erro por parte do Sistema Inteligente de Indexação Automática ou por parte do usuário indexador, por quaisquer equívocos cometidos, é retornado ao estado inicial do fluxograma.

Caso o processo anterior à Alimentação do Tesauro tenha ocorrido com êxito, as atividades de alimentação se iniciam. A primeira atividade deste processo é a delimitação do conteúdo (termos simples) cujos termos não possuem, à priori, nenhuma representatividade de acordo com o Tesauro. Em seguida, este conteúdo é filtrado por uma lista de palavras (pertencentes a algumas classes gramaticais do português) consideradas vazias no quesito representatividade de conteúdo.

Por fim, ocorre uma avaliação dos possíveis termos representativos por parte do usuário indexador. Caso o usuário indexador considere algum dos termos como um termo de forte representatividade para o ofício indexado, este termo é inserido no Tesauro e, posteriormente, também indexado ao ofício. Na figura 4.3 é apresentado o fluxograma do processo de alimentação.

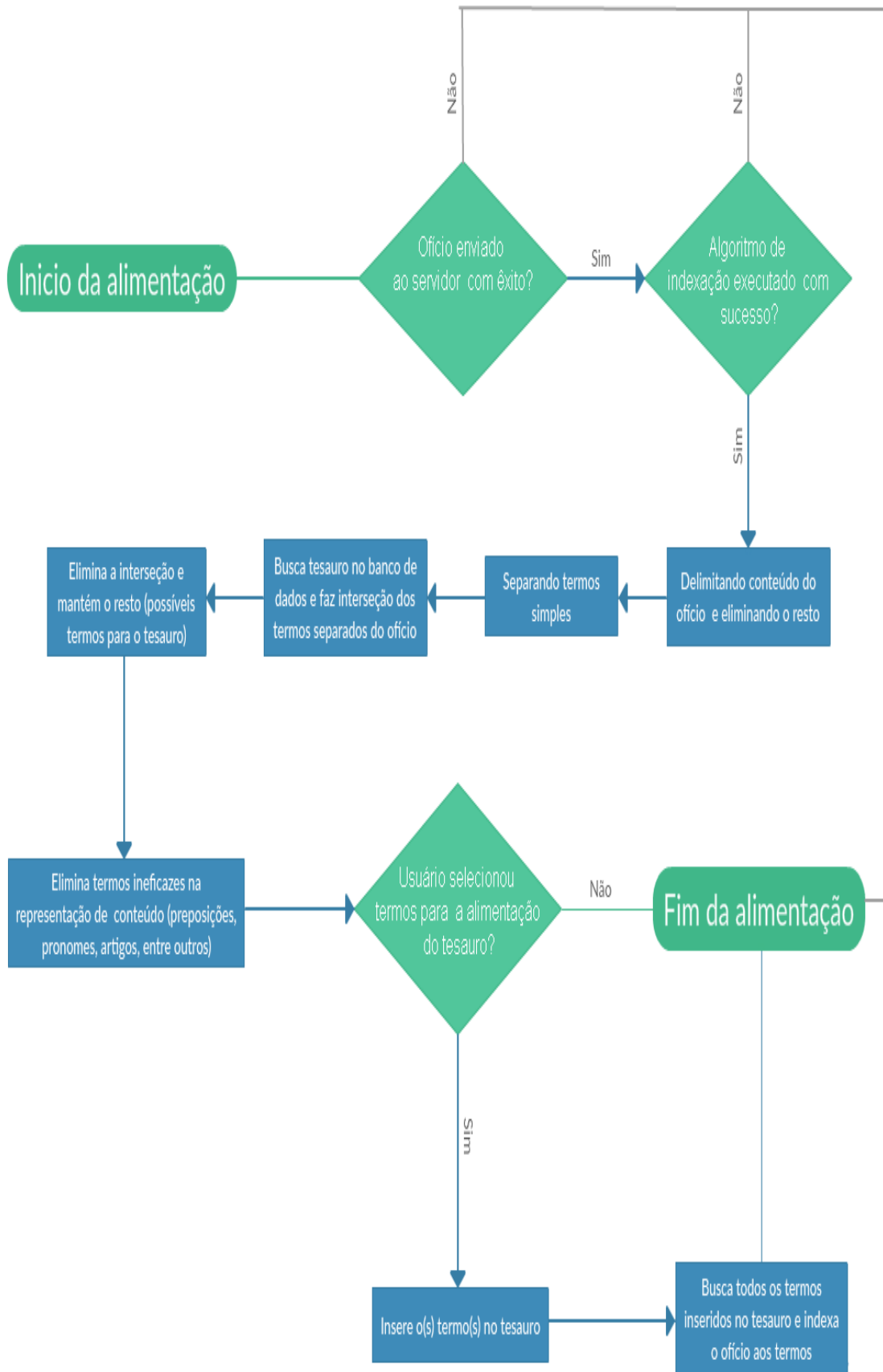


Figura 4.3 - Fluxograma do processo de alimentação

Fonte: Desenvolvido pelo proponente

Exemplo de alimentação do Tesouro

Feita a indexação convencional, apresentada no Capítulo 4.8, ocorre o processo de alimentação do Tesouro, que consiste na recuperação de todos os possíveis termos simples do ofício, eliminando todos os termos já existentes no Tesouro. Após obter o resto dos termos pertencentes ao ofício, são eliminadas todas as palavras consideradas vazias no quesito representatividade de conteúdo. Neste exemplo, restaram-se os termos: “aquisição”, “objetivo”, “adequar”, “dependências”, “referente”, “solicitar”, “deverão”, “adquiridas”.

Esses termos são fornecidos ao usuário, conforme Figura 4.4, como possíveis opções na representação do conteúdo do seu ofício. Após a seleção (colaboração facultativa por parte dos usuários indexadores), são inseridos no Tesouro todos os termos selecionados, pressupondo que há uma capacidade de representação intrínseca ao termo, identificada pelo usuário indexador. Dessa forma, o termo se torna, a partir de então, apto a representar qualquer outro ofício.

Possível termo	Alimentar tesouro
aquisição	<input type="checkbox"/>
objetivo	<input type="checkbox"/>
adequar	<input type="checkbox"/>
dependências	<input type="checkbox"/>
referente	<input type="checkbox"/>
solicitar	<input type="checkbox"/>
deverão	<input type="checkbox"/>
adquiridas	<input type="checkbox"/>

Concluir

Figura 4.4 - Tela de alimentação do Tesouro

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Por fim, o ofício é indexado aos termos selecionados com peso igual à média ponderada do peso de todas as indexações inseridas na base de dados. Este peso foi definido para buscar um equilíbrio na recuperação da informação, já que não há quaisquer dados

prévios dos novos termos, impossibilitando a realização da indexação comum.

Processo de recuperação

O processo de recuperação se inicia através do preenchimento do formulário pelo usuário com informações dos ofícios e possíveis termos indexados ao ofício. Caso o formulário tenha sido preenchido incorretamente, é retornado ao estado inicial. Caso contrário, são iniciadas as atividades de recuperação dos dados.

Mediante a quantidade de informações disponibilizadas pelo usuário, o sistema busca na base de dados todos os ofícios relacionados àqueles dados e retorna essa lista ao usuário buscador do ofício. Na Figura 4.5 é apresentado o fluxograma do processo de recuperação.

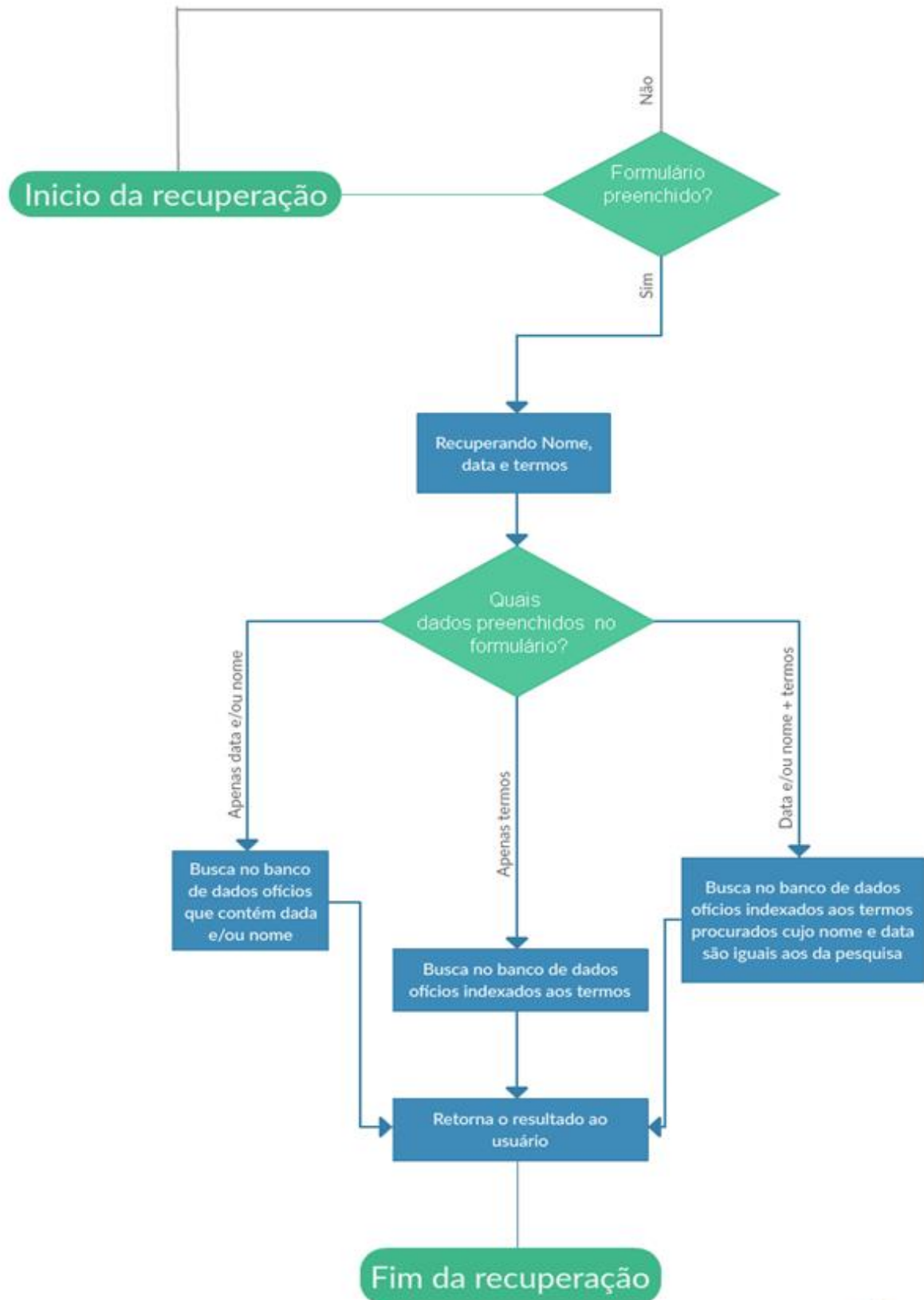


Figura 4.5 - Fluxograma do processo de recuperação

Fonte: Desenvolvido pelo proponente

Exemplo de recuperação de ofício

Após a indexação do ofício aos três termos ranqueados, seus afins e todos os termos selecionados pelo usuário indexador, 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, de maneira convencional ou colaborativa, e os dados de expedidor e data, individuais de cada ofício. Na Figura 4.6 é apresentada uma busca que retorna o ofício indexado.

UFVJM Home Anexar ofício **Buscar ofício**
Você está conectado(a) ao Sistema, Victor

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

Figura 4.6 - Tela de busca de ofícios

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Nesse exemplo, foi inserida a data do ofício, expedidor e alguns dos termos indexados definidos pelo algoritmo do sistema quando o ofício foi adicionado, que neste caso foram: “cômodos” e “computação”. Como é apresentado na Figura 4.6, à medida que é digitado 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.

5 TELAS DO SISTEMA

Conforme apresentado na Figura 5.1, o sistema tem como página principal uma tela para *login* que permite o acesso ao sistema.

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.

Login:
 Senha:

[Me cadastrar no Sistema agora](#)

Figura 5.1 - Tela de *login*

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Conforme apresentado na Figura 5.2, o sistema possui uma tela para cadastro do usuário. O sistema só permite a indexação de um ofício por parte de um usuário cadastrado.



UFVJM DE OFÍCIOS DA UFVJM

UFVJM Home Anexar ofício Buscar ofício Me cadastrar no Sistema

Cadastrando usuário

Nº de matricula/registro na UFVJM

Nome completo

Senha

Confirmar senha

Email

Cadastrar

Figura 5.2 - Tela de cadastro

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Conforme apresentado na Figura 5.3, o sistema possui uma tela para edição dos dados pessoais do usuário, em caso de equívocos ao se cadastrar no sistema.



UFVJM DE OFÍCIOS DA UFVJM

UFVJM Home Anexar ofício Buscar ofício Você está conectado(a) ao Sistema, Victor

Editando meus dados

Nº de matricula/registro na UFVJM

Nome completo

Senha

Confirmar senha

Email

Alterar

Figura 5.3- Tela de edição de dados

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Conforme apresentado na Figura 5.4, o sistema possui uma tela para *anexar um novo ofício*. A tela designada para essa função contém instruções básicas de como o usuário indexador deverá proceder para anexar um novo ofício.

Instruções

- 1 - É fundamental que se siga às instruções para que outros usuários consigam recuperar o ofício quando necessário.
- 2 - O campo "destinado à/ao"(1) é utilizado para informar a quem o ofício se destina, podendo ser a uma pessoa ou departamento.
- 3 - O campo de "assunto"(2) e "data"(3) são os mesmos mostrados no ofício, como segue nas instruções visuais ao lado.
- 4 - As informações inseridas na indexação do ofício são meramente auxiliares para identificação do ofício quando buscado posteriormente por um usuário.
- 5 - O sistema se encarregará de ler o ofício anexado e indexa-lo aos termos que melhor o representará.

Observação: Caso seu arquivo de ofício esteja em formato .doc ou outros formatos, recomendamos o uso do [Sistema online de conversão](#)

1 Assunto: [Assunto do Ofício]

2 Destinado a/ao: [Destinatário do Ofício]

3 Data: [Data do Ofício]

Figura 5.4 - Tela de instruções no anexo de ofícios
 Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Conforme apresentado na Figura 5.5, o sistema possui três caixas de entrada na aba de *anexar um novo 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.

Anexar um ofício

Destinado à/ao

Expedido em

Assunto

Anexar ofício(PDF) No file selected.

1 - É fundamental que se siga às instruções por ser necessário.
 2 - O campo "destinado à/ao"(1) é utilizado para pessoa ou departamento.
 3 - O campo de "assunto"(2) e "data"(3) são os visuais ao lado.
 4 - As informações inseridas na indexação do ofício quando buscado posteriormente por um usuário.
 5 - O sistema se encarregará de ler o ofício anexado e indexá-lo aos termos que melhor o representará.

Observação: Caso seu arquivo de ofício esteja em formato .doc ou outros formatos, recomendamos o uso do Sistema online de conversão

Figura 5.5 - Tela de anexo de ofícios

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

Conforme apresentado na Figura 5.6, o sistema possui uma tela para gerenciamento dos ofícios indexados pelo usuário, dando-lhe prerrogativas para exclusão do ofício, caso tenha procedido de forma inadequada ao indexar o ofício.

SISTEMA INTELIGENTE DE INDEXAÇÃO DE OFÍCIOS DA UFVJM

UFVJM Home Anexar ofício Buscar ofício Você está conectado(a) ao Sistema, Victor

Editando meus dados

Expedido por	Destinado à/ao	Assunto do ofício	Data do ofício	Download	Excluir ofício
Victor Oliveira Almeida	TESTE	TESTE	2016-05-18	<input type="button" value="Baixar ofício"/>	<input type="button" value="Excluir ofício"/>

Copyright 2015 - by Victor Oliveira Almeida | Facebook

Figura 5.6 - Tela de gerenciamento de ofícios

Fonte: Figura extraída do Sistema Inteligente de Indexação Automática

6 RESULTADOS

A partir do estudo sugerido neste trabalho, elaborou-se um sistema de ofícios do DECOM, o Sistema Inteligente de Indexação Automática, através do uso de um Vocabulário Controlado/Tesouro (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 indexação. Os artefatos utilizados no desenvolvimento do Sistema Inteligente de Indexação Automática encontram-se no apêndice B.

O Vocabulário Controlado foi construído tendo como 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 a 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 para tornar possível a recuperação do ofício desejado pelo consulente. Pretende-se que o Sistema seja disponibilizado de maneira *online* para que qualquer pessoa tenha acesso, necessitando somente de acesso à internet.

Sendo positiva a implementação do trabalho, 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.

REFERÊNCIA

- ALVARES, Lilian. *Fontes de informação: tesouro*. 2014. Disponível em: <<http://magisterandre.blogspot.com.br/2014/03/tesouro.html>>. Último 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.
- 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.
- BELUCHE, Renato. *Terminologia e linguagem documentaria: Uma perspectiva histórica*. *Biblos: Revista do Instituto de Ciências Humanas e da Informação*, Rio Grande, n.22, p.115-121, 2008.
- BUCKLAND, Michael. *Information as a thing*: Journal of the American Society of Information Science. *ABI/INFORM Global*, Berkeley, v.42, n.05, p.351-360, 1991.
- CINTRA, Anna Maria Marques *et al.* *Para entender linguagens documentárias*. São Paulo: Polis; APB, 1994. 72 p.
- DALGALARRONDO, Paulo. *Psicopatologia e Semiologia dos Transtornos Mentais*. 2. ed. Porto Alegre: Artmed, 2008.
- DODEBEI, V. L. D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.
- FAVRE, Liliana. *UML and the Unified Process*. 1. ed. Buenos Aires: IRM Press, 2003.
- FUJITA, Mariângela Spotti Lopes. *et al.* *Política de indexação*. São Paulo, Marília: Cultura Acadêmica, 2009.
- GOMES, S. Dennis. *Inteligência Artificial: Conceitos e Aplicações*. *Revista Olhar Científico: Faculdades Associadas de Ariquemes, Rondônia*, v.01, n.02, p.234-244, 2010.
- KURZWEIL, Ray. *The Age of Spiritual Machines*. Massachusetts: The MIT Press, 1990.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. ed. Brasília: Briquet de Lemos, 2004.

LEIVA, Isidoro Gil. *La automatización de la indización de documentos*. 1. ed. Gijón: Trea, 1999.

LEIVA Isidoro; FUJITA, Mariângela. *Política de indexação*. 1. ed. Marília: Cultura Acadêmica Editora, 2012.

NARUKAWA, Cristina Miyuki; LEIVA, Isidoro Gil; FUJITA, M. S. Lopes. *Indexação automática de artigos periódicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de Odontologia*. Inf. & Soc.: Est., João Pessoa: Inf & Soc, v.19, n.02, p.99-118, 2009.

OLIVEIRA, S. L. de. *Tratado de metodologia científica: projetos de pesquisas, TGI, TCC, monografias, dissertações e teses*. 2. ed. São Paulo: Pioneira Thomson Learning, 2001.

RICHARDSON, K. *Compreender a inteligência*. Lisboa: Instituto Piaget, 1999.

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

RUSSO, Mariza. *Fundamentos da Biblioteconomia e Ciência da informação*. Rio de Janeiro: E-Papers. 2010.

SEBESTA, Robert W. *Conceitos de linguagens de programação*. 5. ed. Porto Alegre: Artmed editora, 2003

SEQUEIRA, João. *Comunicação ao 6º Congresso da Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas*. Aveiro: Centro Cultural e de congressos, 1998.

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.

SILBERSCHATZ, Abraham; KORTH, Henry; SUDARSHAN,S. *Sistema de Banco de Dados*. 5. ed. Rio de Janeiro: Editora CAMPUS, 2006

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.

1 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 FACSAB - [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 COMITÊ - [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

ARTEFATOS DO SISTEMA/ARTEFATOS DO SISTEMA INTELIGENTE DE INDEXAÇÃO AUTOMÁTICA DO DEPARTAMENTO DE COMPUTAÇÃO DA UFVJM

ARTEFATOS

- Pseudocódigo
 - Modelo lógico do banco de dados
 - Diagrama de classes
 - Diagrama de casos de uso
 - Diagrama de sequência
 - Análise de requisitos
-

1 PSEUDOCÓDIGO

```
anexarOfício()
```

```
{
```

```
    Ofício objetoOfício;
```

```
    //Criando um objeto chamado 'objetoOfício' através da classe 'Ofício'
```

```
    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!
```

```
{
```

```
    objetoOfício.preencherOfício();
```

```
    // Neste momento o objeto 'ofício' criado acima será definido por meio do  
    método 'preencherOfício' da classe 'Ofício' com os dados e o ofício 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 minúsculas 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.pegarTesouro();
```

//Neste momento, o tesouro desenvolvido no trabalho é recuperado do banco de dados (palavra a palavra e suas relações) por meio do método ‘pegarTesouro’ da classe ‘BancoDeDados’

```
coleçãoPalavrasInterseccionadas = coleçãoPalavra.interseccionar(coleçãoPalavra);
```

//Neste momento, a coleção de palavras 'coleçãoPalavrasInterseccionadas' , que contém o tesouro 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 criado uma coleção de palavras 'coleçãoPalavrasRelacionadas' pra receber todas as palavras relacionadas (relações estabelecidas no Tesouro) 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 conterá todas as três

palavras rankeadas de acordo com a frequência, somadas às suas palavras relacionadas. (relações definidas pelo Tesouro).

```
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 oficiosBusca;

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

Coleção oficiosExpedidorData.

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

```
oficiosExpedidorData.buscarOficiosPeloExpedidorXdata(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 'buscarOficiosPeloExpedidorXdata'; em seguida, armazena na coleção
criada 'oficiosExpedidorData'
```

Coleção oficiosPelosTermos;

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

```
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 'Ofício'

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

alimentarTesauro()
{
    Oficio objetoOficio = oficioIndexado;

    //Criando um objeto chamado 'objetoOficio' através da classe 'Ofício', e inserindo
    neste objeto o ofício que já foi indexado no pseudocódigo "anexarOficio()"

    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

    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 minúsculas e tirando qualquer tipo de pontuação
```

```
Coleção coleçãoPalavra = textoOficio.separarTextoPorPalavrasSimples();
```

```
//Começa retirar todas as possíveis palavras simples por meio do método  
'separarTextoPorPalavrasSimples()' 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 coleçãoPalavrasTesauro = BD.pegarTesauro();
```

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

```
Coleção possiveisTermos;
```

```
//Cria uma coleção que receberá possíveis termos a serem inseridos no Tesauro(serão  
selecionados pelo usuário)
```

```
possiveisTermos = eliminarTermosDoTesauro(coleçãoPalavra e  
coleçãoPalavrasTesauro) ;
```

//Neste momento, a coleção de palavras ‘possiveisTermos’ está recebendo os termos contidos em ‘coleçãoPalavra’ menos todos os termos contidos no Tesouro, ou seja, na coleção ‘coleçãoPalavrasTesouro’. Deixando apenas o resto, que são os possíveis termos a serem inseridos no Tesouro.

```
possiveisTermos.eliminarPalavrasVazias();
```

//A coleção ‘possiveisTermos’, que possui todas as palavras capazes de pertencer ao Tesouro, está eliminando todas as palavras consideradas inúteis para representação de conteúdo; ou seja, neste momento, acontece a eliminação de algumas classes gramaticais do português, consideradas incapazes de representar conteúdos, tais como: artigos, pronomes, preposições, conjunções, interjeições, entre outras.

```
Coleção TermosSelecionados = possiveisTermos.mostrarParaUsuario();
```

//Cria uma coleção chamada ‘TermosSelecionados’ que receberá todos os termos capazes de representar o conteúdo do ofício (selecionados pelo usuário). O método ‘mostrarParaUsuario()’ mostra todos os termos, selecionados pelo sistema, possíveis para a alimentação do Tesouro. Dados os termos, o usuário seleciona os que melhor representam o conteúdo (se existir), e a função retorna inserindo na coleção ‘TermosSelecionados’.

```
BD.inserirNoTesouro(TermosSelecionados);
```

// Começa a inserção, no Tesouro (base de dados), das palavras selecionadas pelo usuário.

```
oficio.indexar(TermosSelecionados);
```

//Indexa as palavras selecionadas pelo usuário ao ofício responsável pelo processo de alimentação

```
}
```

```
//FIM DO PROCESSO
```

2 MODELO LÓGICO DO BANCO DE DADOS

O modelo lógico de um banco de dados explica várias das características de uma base de dados e seus comportamentos, facilitando o entendimento do desenvolvedor na hora da implementação do sistema. Este modelo mostra graficamente como foi construído a base de dados para dar suporte aos processos de negócio, como esses dados estarão organizados e quais os relacionamentos que pretendemos estabelecer entre eles.

Na Figura 1 é apresentado o modelo lógico do Sistema Inteligente de Indexação Automática.

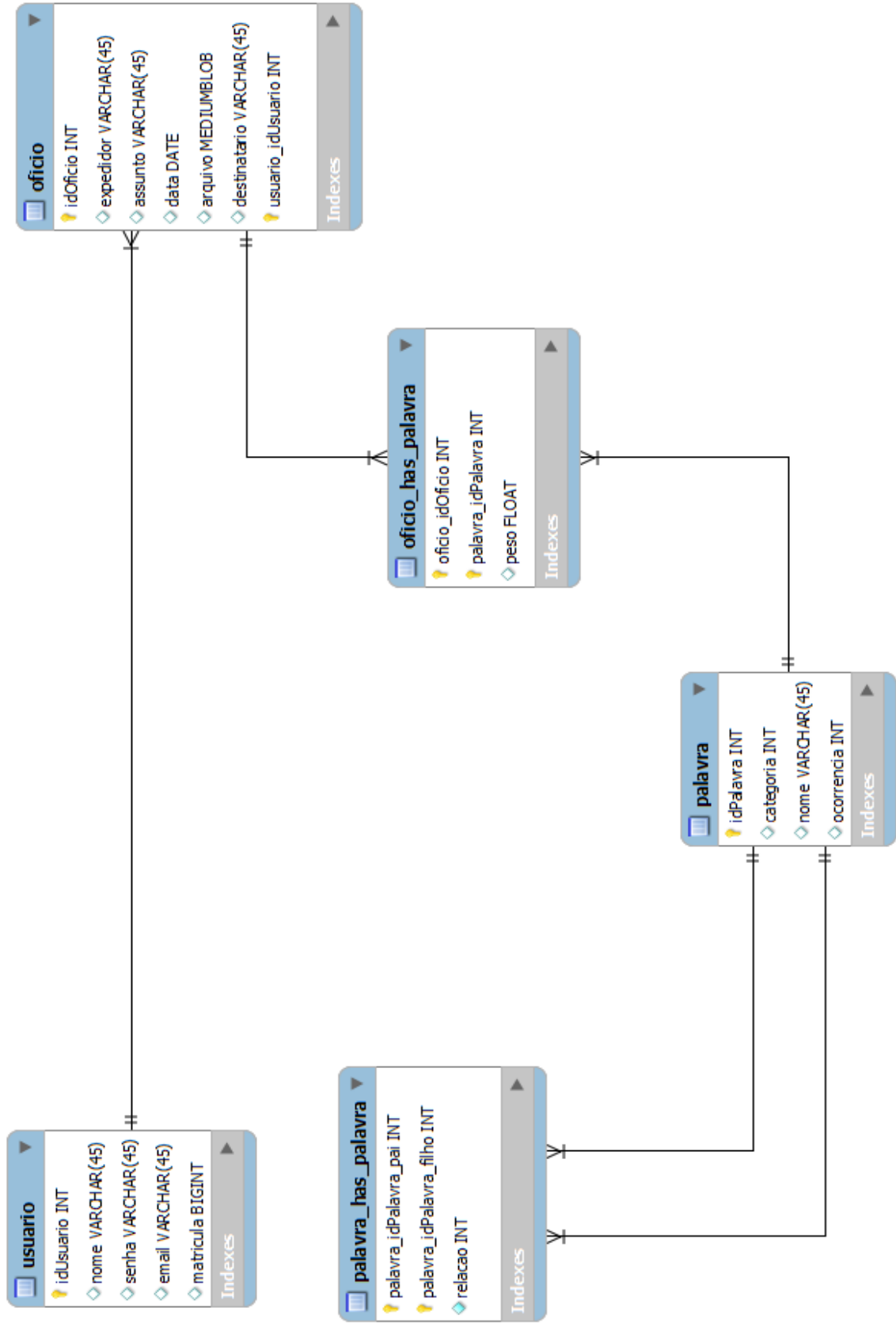


Figura 1 – Modelo lógico do Banco de dados

3 DIGRAMA DE CLASSES

O diagrama de classe é um dos diagramas mais utilizados para expressar as relações lógicas dentro de uma aplicação. Segundo Liliana Favre (2003), o diagrama de classe é de extrema importância na *Unified Modeling Language* (UML), pois ele é capaz de capturar toda a visão estática de um Sistema de Informação. O diagrama de classe modela classes de um mundo real de forma computacional e especifica a relação entre essas classes.

Na Figura 2 é apresentado o diagrama de classe do Sistema Inteligente de Indexação Automática.

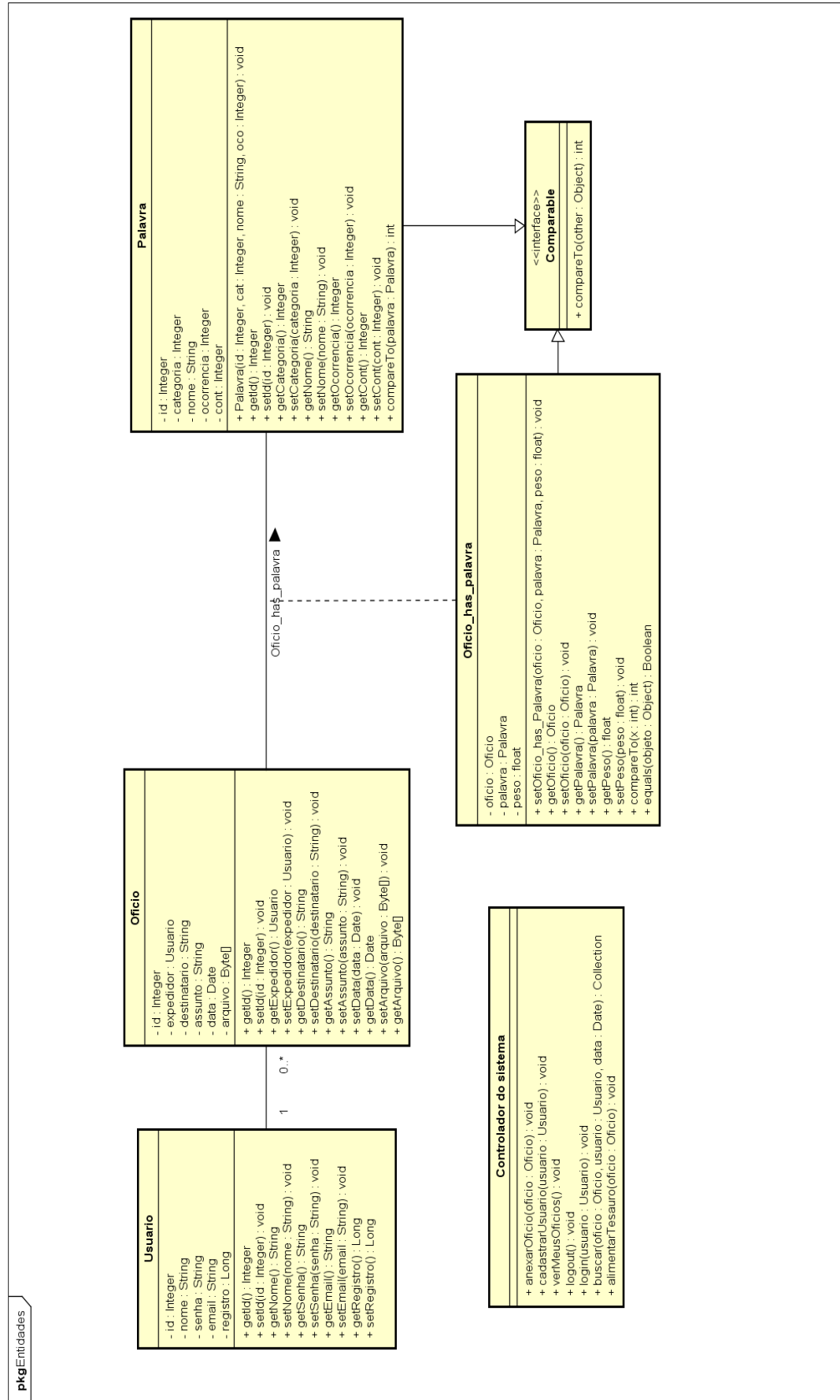


Figura 2 – Diagrama de classes

4 DIAGRAMA DE CASOS DE USO

O diagrama de casos de uso expressa a interação do usuário com o sistema, podendo ser elaborado de diversas maneiras. A conversação entre o sistema e o usuário acontece por meio de requisições emitidas pelo usuário ao sistema e respostas emitidas do sistema ao usuário.

Na Figura 3 é apresentado o diagrama de casos de uso do Sistema Inteligente de Indexação automática.



Figura 3 - Diagrama de casos de uso

5 DIAGRAMA DE SEQUÊNCIA

O diagrama de sequência é útil quando se deseja mapear as interações que acontecem entre os objetos e o cenário, realizadas através de operações ou métodos.

Na Figura 4 é apresentado o diagrama de sequência do Sistema Inteligente de Indexação automática.

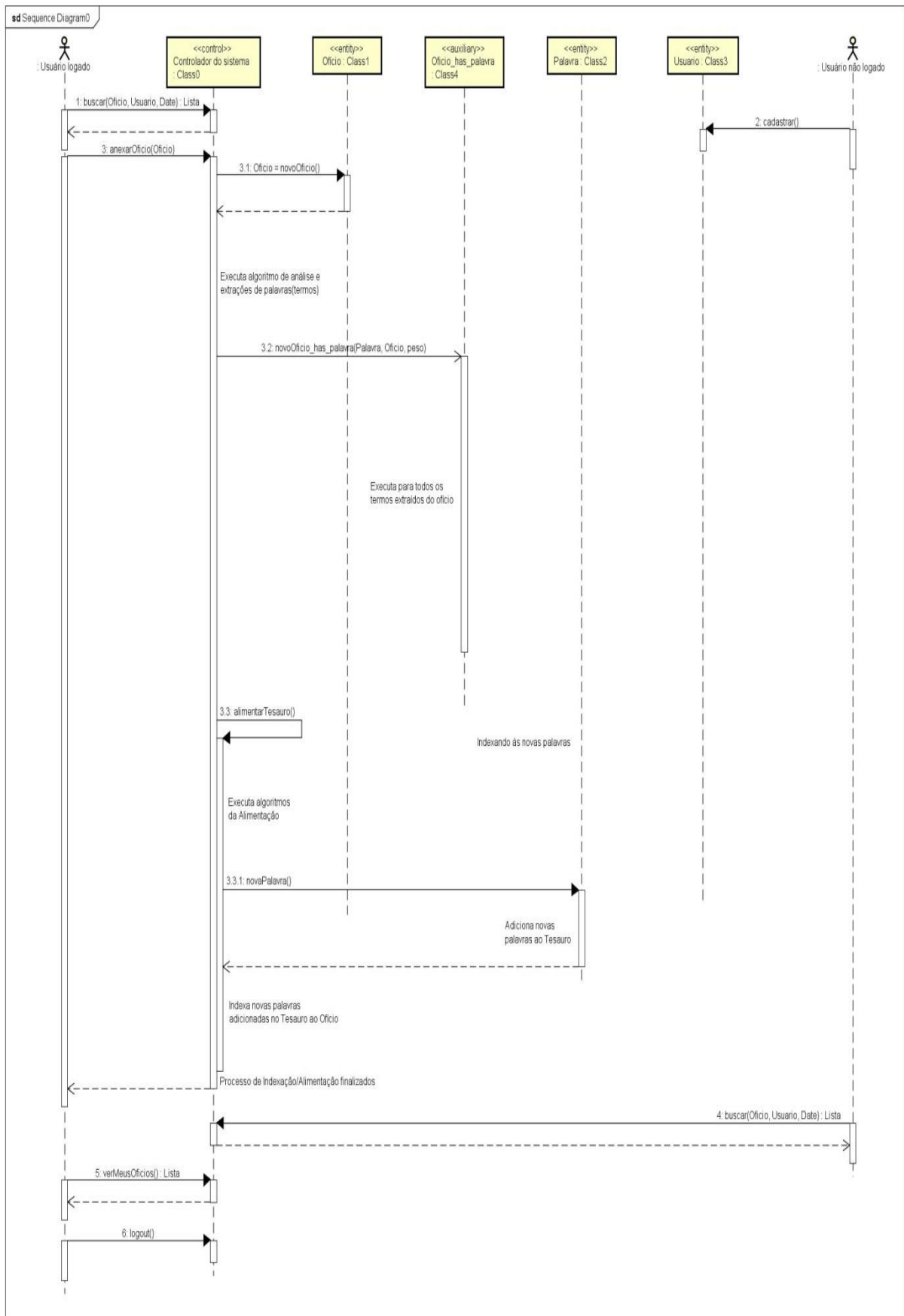


Figura 4 - Diagrama de sequência

6 ANÁLISE DE REQUISITOS

A análise de requisitos de um sistema engloba todas as tarefas predefinidas que serão executadas pelo sistema, bem como todas as configurações de ambiente para alcançar os objetivos definidos pelo escopo. A análise de requisito guia tudo o que se deve fazer no sistema sem especificar como será feito.

6.1 Requisitos funcionais

- O sistema deve permitir o cadastro dos funcionários aptos a confeccionar ofícios no departamento de computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri;
- O cadastro do funcionário deverá constar seu:
 - Número de registro;
 - Nome;
 - Senha;
 - Email.
- O sistema deve permitir a edição dos dados pessoais dos funcionários;
- O sistema deve possuir um *login* para que os funcionários acessem todas as funções disponíveis;
- O sistema deve permitir o gerenciamento dos ofícios indexados pelo expedidor do ofício;
- O sistema deve ser capaz de indexar um ofício de forma automática;
- O sistema deve ser capaz de alimentar os mecanismos de indexação propostos (Tesauro);
- O sistema deve permitir a recuperação dos ofícios por parte de qualquer usuário, logados e não logados;
- O sistema deve conter instruções de como prosseguir na realização das funcionalidades disponíveis.

6.2 Requisitos não funcionais

- O sistema deve ter boa usabilidade;
- O sistema deve ser desenvolvido com a linguagem de programação JAVA;
- O sistema deve ser desenvolvido com um Sistema Gerenciador de Banco de Dados gratuito.