

# **Sistema Inteligente de Indexação Automática de Ofícios do Departamento de Computação da UFVJM**

**Aline Pereira da Costa** (UFJF) - alinenovaera@gmail.com

**Victor Oliveira Almeida** (UFVJM) - victor.almeida.ti@gmail.com

**Cristiano Grijó Pitangui** (UFSJ) - pitangui.cristiano@gmail.com

**Alessandro Vivas Andrade** (UFVJM) - prof.alessandrovivas@gmail.com

**Luciana Pereira Assis** (UFVJM) - lupassis@gmail.com

## **Resumo:**

*Este projeto tem por objetivo sanar uma deficiência do Departamento de Computação da UFVJM – Universidade Federal dos Vales do Jequitinhonha e Mucuri. Tal Departamento não possui um local tanto físico quanto digital para armazenar seus ofícios. A falta de uma organização eficiente e padronizada destes ofícios inviabiliza o acesso aos documentos e dispende muito tempo na localização e recuperação da informação, podendo trazer grandes prejuízos para a instituição. Visando solucionar o problema de recuperação e organização das informações contidas nos ofícios, propõe-se a criação de um Sistema Inteligente de Indexação Automática utilizando um vocabulário controlado, tesauro, para que seja possível localizar com rapidez o ofício desejado. É importante que o sistema esteja em ambiente virtual para que o acesso seja democratizado e o processo de tratamento de novos ofícios seja automatizado. O projeto permeia o espaço de 2011 a 2014, totalizando 239 ofícios, sendo estes a base para a criação do tesauro. Sendo positiva a implementação do projeto, sugere-se que os demais departamentos da UFVJM utilizem o mesmo sistema para organização dos seus documentos, ganhando agilidade nos processos e satisfação do usuário final na localização do que procura.*

**Palavras-chave:** *Vocabulário Controlado 1. Tesauro 2. Sistema Inteligente 3. Indexação 4.*

**Área temática:** *Eixo 3 - Ecologia da Informação*

**Subárea temática:** *Preservação de acervos digitais*

## 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 da informação gerada 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 expedido por órgãos públicos. Seu destinatário pode ser outro órgão público, como particular.

Este trabalho tem o objetivo de, através do uso de vocabulário controlado, criar um Sistema Inteligente de Indexação Automática para localização dos ofícios do Departamento de Computação da UFVJM – Universidade Federal dos Vales do Jequitinhonha e Mucuri. Os ofícios recebidos e expedidos neste departamento discorrem sobre solicitações, deferimentos, aquisições e diversas outras ações e recomendações importantes para o andamento das rotinas e processos desenvolvidos no departamento.

Para que este projeto 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 projeto, o usuário final tenha acesso aos ofícios do Departamento de Computação da UFVJM e que consiga localizar o que necessita com agilidade na hora da busca.

Este trabalho se organiza como segue. A seção atual, a introdução, apresenta e justifica os objetivos do projeto. A segunda seção apresenta uma revisão da literatura e expõe os fundamentos teóricos nos quais este trabalho se baseia. A terceira seção descreve a metodologia utilizada na execução do trabalho. A quarta seção apresenta os resultados parciais já alcançados. Finalmente, a última seção, indica os próximos passos a serem seguidos visando a implantação completa do sistema.

### 1.1 Justificativa

Identificou-se que o Departamento de Computação da UFVJM não possui um local onde seja possível pesquisar os ofícios já expedidos. Quando há uma determinada demanda em localizar um ofício, recorre-se a e-mails enviados ou a alguma pasta arquivada no computador ou fisicamente - o 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 mais fácil e mais rápida sua recuperação.

Para sanar tal deficiência na busca pelos ofícios, propõe-se o desenvolvimento de um tesouro, tipo de vocabulário controlado. Vocabulário controlado é definido por Naves (2004, p. 10-11), “como um tipo de linguagem artificial. Esta é construída cuidadosamente, através

do processo minucioso da escolha dos termos a serem usados, o significado de cada um e as relações que apresentam.”. O tesouro servirá como base para uma indexação automática dos documentos e, assim, classificar os ofícios, de tal maneira que o usuário localize o que procura.

Após a construção do vocabulário controlado, será necessário indexar os ofícios utilizando os termos definidos no tesouro.

Segundo Lancaster (2004, p. 84) a prioridade da indexação precisa ser o usuário final, pois:

a indexação de assuntos é normalmente feita visando a atender às necessidades de determinada clientela, ou seja, é preciso que se tome uma decisão não somente quanto ao que é tratado no documento, mas que por ele se reveste de provável interesse para determinado grupo de usuários. (LANCASTER, 2004, p. 84).

A necessidade de informação do indivíduo é expressa no momento da busca. Este processo pela busca da informação é cíclico: digamos que um usuário está em um centro de informação (biblioteca, arquivo) e necessita de uma informação. Ele então irá procurar o especialista da informação, um bibliotecário, por exemplo, e diante do diálogo entre os dois, será feito um pedido, uma solicitação do usuário, para o bibliotecário do que ele deseja. Diante deste pedido, o especialista da informação prepara uma estratégia de busca, utilizando para isso os termos da indexação. A estratégia de busca será então confrontada em uma Base de Dados. Após a busca, alguns itens serão recuperados. O especialista da informação fará uma filtragem e apresentará ao solicitante somente as informações pertinentes ao seu pedido. Tal ciclo é representado na Figura 1.

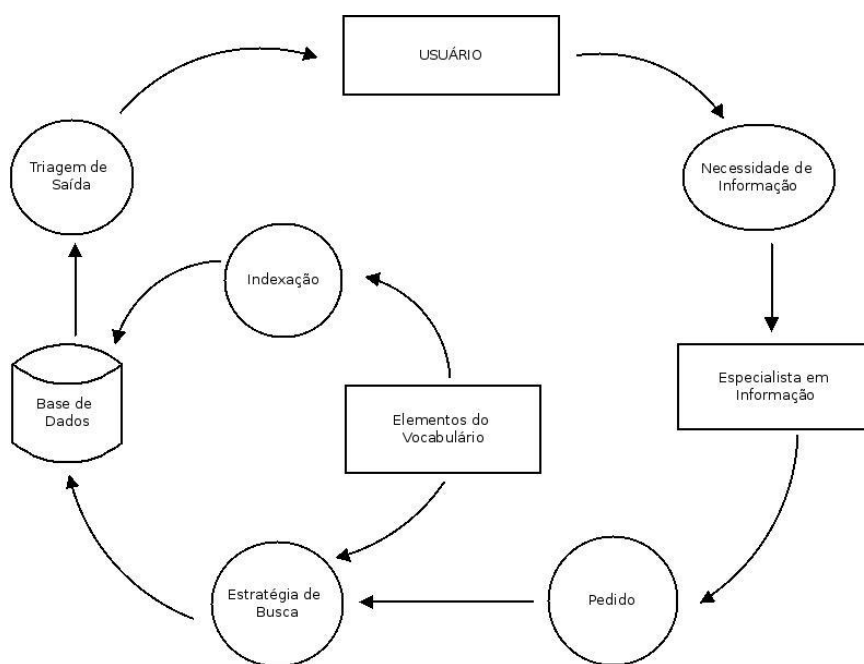


Figura 1 – Fatores que influem nos resultados de uma busca pela informação em uma Base de Dados

O grande desafio para a gestão documental e do conhecimento nas empresas é, corroborando com Araújo (2006, p. 62), estabelecer métodos que ofereçam ao usuário “documentos que sejam realmente relevantes para ele, que contribuam para satisfazer suas necessidades de pesquisa e, desse modo, a construção do conhecimento” de forma eficiente e rápida, diminuindo o gasto de recursos e tempo. Neste contexto, elaborar eficientes políticas de indexação e um sistema de recuperação são imprescindíveis para o sucesso da gestão da informação no meio organizacional.

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

### 1.2 Objetivo geral

Propor a criação de um Sistema Inteligente de Indexação Automática, através do uso de vocabulário controlado, para localização dos ofícios do Departamento de Computação da UFVJM.

### 1.3 Objetivos específicos

- Construir um tesauro como instrumento base para a indexação automática do sistema;
- Propor um método de ranqueamento dos ofícios, para que a busca retorne em primeiro lugar o que o usuário solicita.
- Criação de um Sistema Inteligente de Indexação Automática online para implementar o objetivo específico acima citado e armazenar/recuperar os ofícios.

## 2 Revisão de literatura

Para que este projeto fosse desenvolvido, foi necessário lidar com o controle do significado das palavras para uma futura recuperação dos termos para a construção do vocabulário controlado.

O vocabulário controlado é um instrumento que através de termos previamente determinados, isto é, uma linguagem padronizada, permite a representação e a recuperação de documentos. Segundo (SOUZA, 2010), os vocabulários controlados são:

utilizados desde 1950 principalmente nas agências dos governos dos Estados Unidos e dos países da Europa com o objetivo de gerenciar melhor a explosão do conhecimento registrado na literatura científica e tecnológica de áreas especializadas. (Souza, 2010).

Existem diferentes tipos de vocabulários controlados como mostra a Figura 2. São eles: lista, anel de sinônimo, taxonomia e tesauro. O direcionamento da régua mostra um aumento da complexidade das estruturas existentes. Essa complexidade é determinada pelas formas de relacionamentos que cada tipo de vocabulário possui.

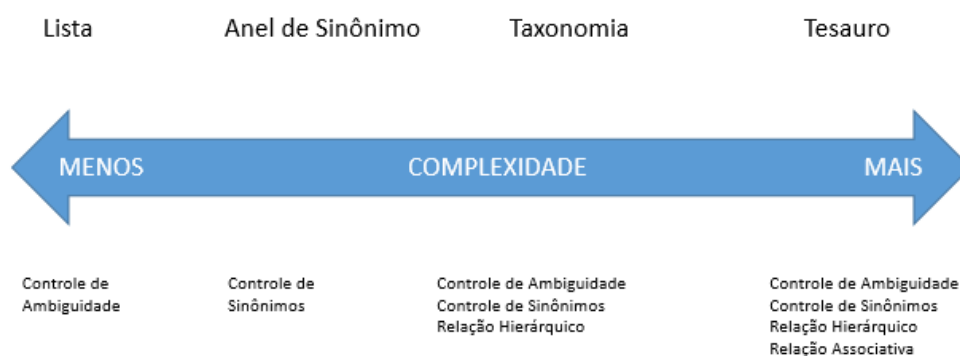


Figura 2 – Crescimento estrutural da complexidade entre tipos de Vocabulário Controlado

Observando-se a Figura 2, percebe-se que taxonomia e o tesouro são os tipos mais complexos, enquanto que a lista e o anel de sinônimos são mais simples. Enquanto o tesouro possui vários tipos de controles e relacionamentos em sua estrutura de termos, tais como ambiguidade, sinonímia, relações hierárquicas e associativas, a lista considerada simples, é uma lista de palavras em ordem alfabética ou em outra lógica evidente, sem nenhum outro tipo de controle ou relacionamento. Segundo a ANSI/NISO Z39.19 (2005), as listas são utilizadas para descrever aspectos do conteúdo de objetos ou entidades que tem um número limitado de possibilidades.

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

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

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

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

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

Lancaster (1993) sugere como método para construir vocabulário controlado:

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

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

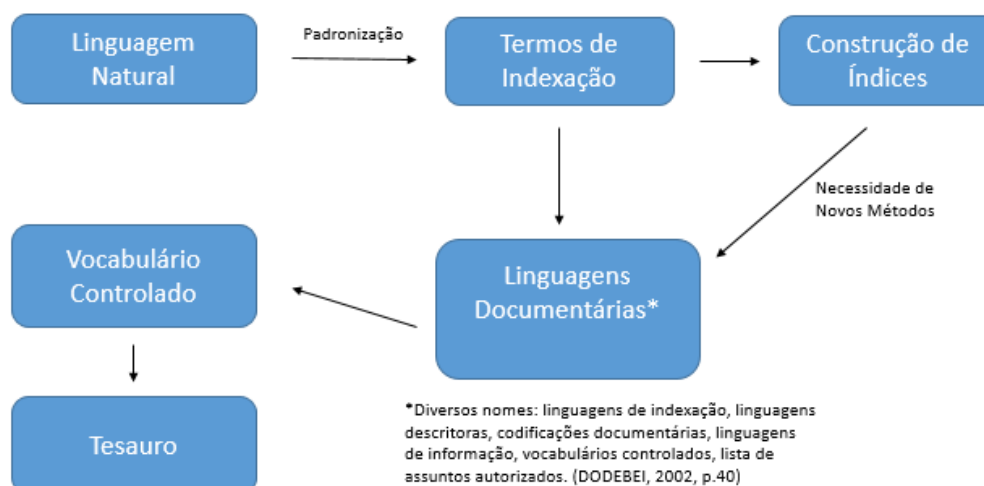


Figura 3 – Construção de vocabulário controlado

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

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

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

No entanto, segundo Lancaster (2004, p. 88-89), alguns fatores podem afetar a qualidade do processo de indexação e conseqüentemente a elaboração do vocabulário controlado. Estes fatores podem estar ligados ao: indexador, vocabulário, documento, processo e ambiente.

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

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

O processo de indexação também pode ser influenciado de acordo com as regras estabelecidas para a indexação, o nível de produtividade exigida para a tarefa, a exaustividade da indexação, definida por Lancaster (2003, p. 23) como sendo “[...] o emprego de termos em número suficiente para abranger o conteúdo temático do documento de modo bastante



completo”, e o tipo de indexação escolhida: por atribuição, onde os termos são selecionados a partir do vocabulário controlado ou por extração, que utiliza a linguagem natural empregada pelo autor do documento.

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

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

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

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

A análise do acesso de informações relacionadas a legislação e especificadamente a portarias em alguns SRIs de instituições públicas que utilizam vocabulários controlados na indexação de seus documentos mostraram alguns aspectos importantes que valem ser ressaltados: no site do INMETRO<sup>1</sup> por exemplo, foi observada a existência de várias bases de dados distintas em um único repositório e campos com possibilidades de buscas padronizadas que diminuem a polissemia dos termos e melhoram as chances de sucesso e precisão na recuperação da informação. Já no portal da UFJF<sup>2</sup> a categorização deste tipo de informação é realizada apenas por ano e/ou número da portaria, sem campos que orientem ou possibilitem uma busca avançada. Isto é um aspecto que limita o acesso ao documento, pois, o usuário precisa ter um conhecimento prévio para ter acesso à informação desejada. Na UFSC<sup>3</sup> a pesquisa por portarias pode ser feita através de várias possibilidades de categorização e critérios de pesquisa/classificação que podem refinar a procura, apresentando, assim, maior relevância dos resultados obtidos. A UFAL<sup>4</sup> também possui um sistema com campos padronizados para estabelecer as estratégias de busca por portarias, o que favorece a autonomia do usuário no processo de recuperação da informação.

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

<sup>1</sup> Instituto Nacional de Metrologia, Qualidade e Tecnologia – INMETRO. Disponível em: <[www.inmetro.gov.br/legislacao/](http://www.inmetro.gov.br/legislacao/)>.

<sup>2</sup> Universidade Federal de Juiz de Fora – UFJF. Disponível em: <<http://www.ufjf.br/portal/legislacao/portariasufjf/>>.

<sup>3</sup> Universidade Federal de Santa Catarina – UFSC. Disponível em: <<http://notes.ufsc.br/aplic/portaria.nsf/>>.

<sup>4</sup> Universidade Federal de Alagoas – UFAL. Disponível em: <<http://sistemas.ufal.br/portarias/>>.

## 3 Materiais e métodos

### 3.1 O Sistema Proposto

Diante dos problemas encontrados no Departamento de Computação da UFVJM, tais como a falta de organização dos ofícios, falta de padronização no armazenamento e, conseqüentemente, difícil recuperação desses ofícios, foi proposto um Sistema Inteligente de Indexação Automática, que possibilita o armazenamento de forma organizada, padronizada e digital, facilitando a recuperação dos ofícios. O Sistema de Indexação 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 projeto foi uma pesquisa documental embasada nos ofícios 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* e *JAVA Enterprise Edition*, bem como o banco de dados *MYSQL* para armazenamento dos dados referentes aos atributos do sistema.

### 3.2 O algoritmo simplificado do sistema

Inicialmente, foi feita a construção de um vocabulário controlado, extraindo 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.

#### 3.2.1 Indexação de um Ofício

Quando um novo ofício é inserido, 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.

Após feita a interseção e criação do novo Tesouro, serão escolhidos três termos (valor arbitrário e sujeito a mudança em estudos posteriores), mediante à 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.

### 3.2.2 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 desejados 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 definida pela combinação da pertinência do termo (explicado anteriormente como a frequência do termo dentre todos os ofícios cadastrados no sistema)

### 3.3 Exemplo de Indexação e Recuperação de um Ofício

Para exemplificar, foi feita a inserção de um ofício no sistema - conforme a Figura 4, e em seguida, automaticamente, foi capturado todos os termos do ofício que fazem intersecção com o conjunto representativo de termos.



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI  
DIAMANTINA - MINAS GERAIS

[www.ufvjm.edu.br](http://www.ufvjm.edu.br)



**Ofício nº 26 / 2012 / DECOM**

Diamantina, 04 de outubro de 2012.

*À Sua Senhoria, o Senhor  
Prof. Alexandre Ramos Fonseca  
Diretor do DTI / UFVJM*

**Assunto: Aquisição de impressoras multifuncionais para o Departamento de Computação.**

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,

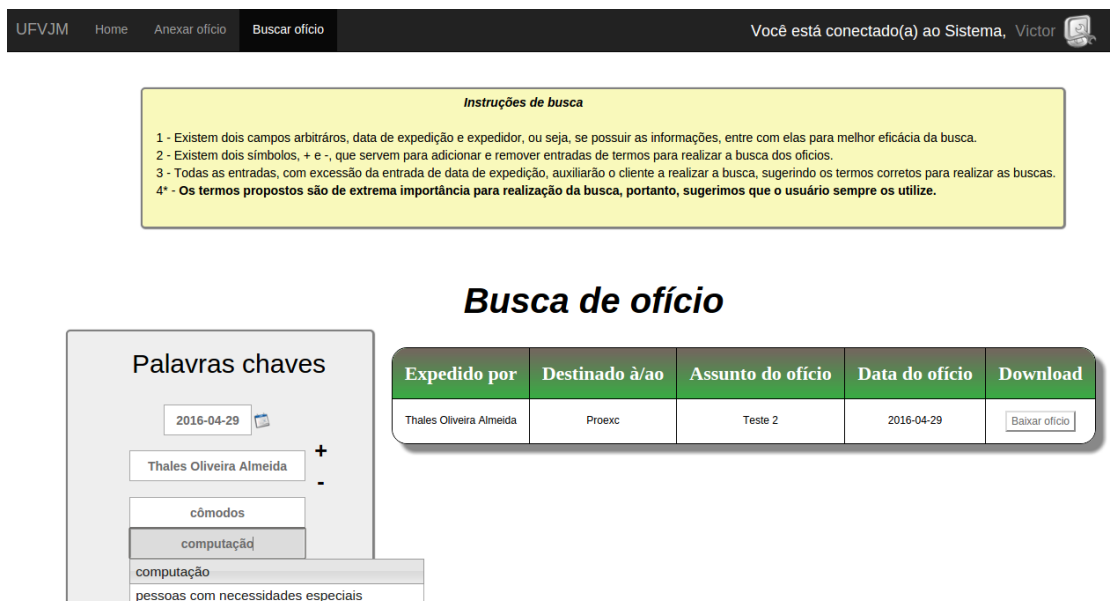
Figura 4 – Modelo de ofício

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

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

Em seguida, o sistema busca no Tesouro por todas as palavras interligadas aos três primeiros termos do *Ranking*. Neste exemplo, os termos interligados foram: renovação de garantia, decom, ufvmj e cômodos. Essas ligações e termos estão definidos em nosso Tesouro, sendo o termo renovação de garantia ligado ao termo impressoras multifuncionais, decom e ufvmj ligados à departamento de computação, e gabinetes ligado a cômodos.

Após a indexação do ofício aos três termos ranqueados e a todos os seus afins, qualquer usuário poderá recupera-lo facilmente. Para recuperar o ofício indexado basta inserir os dados relativos a esse ofício, que no caso seriam os termos indexados ao ofício, citados acima. A Figura 5 mostra uma busca que retornaria o ofício indexado.



**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  
computaçãd  
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 5 – Modelo da tela de busca do ofício

Nesse exemplo, foi inserido 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 mostra na Figura 5, 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.

## 3.4 Algumas telas do Sistema de Indexação Automática

Conforme apresenta a Figura 6, o sistema tem como página principal uma tela para login que permite o acesso ao sistema.



Figura 6 - Página principal do sistema de indexação

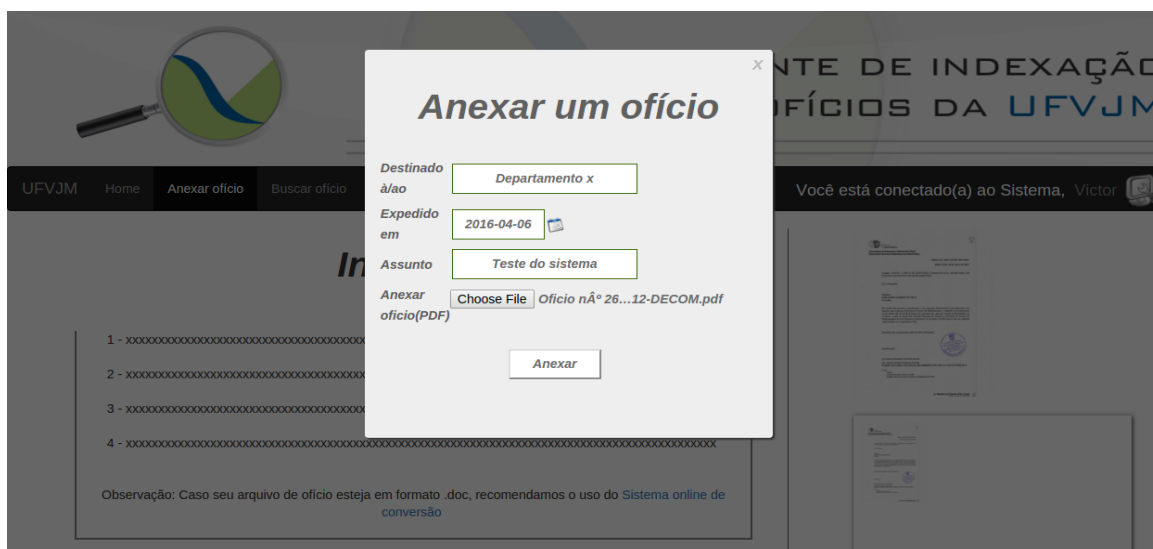


Figura 7 – Página de inserção de novo ofício no sistema

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

### 3.3 Pendências do Sistema

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

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

### 4 Resultados parciais/finais

Os resultados parciais deste projeto são: o Tesouro, que já está concluído e o Sistema Inteligente de Indexação Automática que está sendo desenvolvido. O desenvolvimento do Sistema precisa ser concluído, alimentado e implantado, para verificar sua utilidade e precisão na recuperação dos ofícios desejados.

### 5 Considerações parciais/finais

A partir do estudo sugerido neste projeto, elaborou-se um Sistema Inteligente de Indexação Automática (ainda em desenvolvimento), através do uso de um vocabulário controlado. O Sistema foi construído utilizando-se técnicas da biblioteconomia e de técnicas computacionais, sendo um sistema que utiliza tanto das análises qualitativas – feita inicialmente para a construção do vocabulário controlado por uma bibliotecária – quanto das análises quantitativas – utilizadas no algoritmo de auto alimentação do vocabulário controlado.

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

### 6 Referências

ARAÚJO, Vera Maria Araújo Pegozzi de. Documentação, terminologia e linguística: uma interface produtiva. Porto Alegre, 2006. Disponível em: <[www.ufrgs/termisul/biblioteca/dissertacoes/dissertacao\\_2006\\_ARAUJO.pdf](http://www.ufrgs/termisul/biblioteca/dissertacoes/dissertacao_2006_ARAUJO.pdf)>. Acesso em: 22 mar. 2016.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676*: Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <[www.unesp.br/revistadigital](http://www.unesp.br/revistadigital)>. Acesso em: 10 abr. 2016.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. *Fundamentos de metodologia científica*. 5. ed. São Paulo: 2003. 311p.

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

NAVES, Madalena M. L. Estudo de fatores interferentes no processo de análise de assunto. *Persp. Ci. inf.*, Belo Horizonte, v. 6, n. 2, p. 189 -203, jul./dez. 2001.

NAVES, Madalena M. L. Estudo de fatores interferentes no processo de análise de assunto. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 5, n. 2, p. 189-203, jul./dez. 2001.

OLIVEIRA, Carla Cristina Vieira de. **A interação dos usuários da UFMG com o catálogo online do sistema Pergamum**. Belo Horizonte, 2008. Dissertação (mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais – Escola de Ciência da Informação.

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

SALES, Rodrigo de; CAFÉ, Ligia.. Diferenças entre tesouros e ontologias. **Perspect. ciênc. inf.**, Belo Horizonte, v. 14, n. 1, p. 99-116, jan./abr. 2009. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362009000100008&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100008&lng=en&nrm=iso)>. Acesso em: 22 mar. 2016.

SOUZA, Geraldo Magela. Taxonomias: o que são e para que servem. *Websinder*. [S. l.], maio. 2010. Disponível em: <<http://webinsider.com.br/2010/05/11/taxonomias-o-que-sao-e-para-que-servem-1/>>. Acesso em: 20 de mar. 2016.