



INDEXAÇÃO SEMIAUTOMÁTICA DE PUBLICAÇÕES ATRAVÉS DE TÉCNICAS DE MINERAÇÃO DE TEXTO

Área temática: Gestão do Conhecimento Organizacional

Nilton Freitas Junior
niltonfjunior@gmail.com

Geórgia Gomes
georgiargomes@gmail.com

Nícollas Cretton
nicollas_nc@hotmail.com

Valderedo Fontana
valderedo@gmail.com

Resumo: *Com o crescente uso das tecnologias da informação, percebe-se um aumento significativo no volume de publicações digitais. Técnicas de indexação semiautomática de documentos, que visam selecionar índices importantes em publicações, tornam-se imprescindíveis quando aplicados à Ciência da Informação, uma vez que auxilia de forma gerencial o trabalho dos profissionais da área, na manipulação desses grandes volumes de documentos. O objetivo deste artigo é apresentar um estudo de caso que descreve a construção de um protótipo de software para extração semiautomática para extração de informação e sua utilização na indexação de publicações digitais. Testes realizados ao final do estudo demonstram a viabilidade de sua utilização em ambiente real no auxílio à indexação semiautomática de publicações, mostrando-se como um recurso eficaz para o processo de extração da informação.*

Palavras-chaves: *Mineração de Texto, Indexação Semiautomática, Ciência da Informação, Recuperação da Informação, Gestão da Tecnologia da Informação.*

1. INTRODUÇÃO

Com avanço das tecnologias da informação e comunicação (TIC's), sobretudo de ferramentas que criam e disponibilizam informação em ambiente WEB, é possível perceber o crescimento da grande quantidade de publicações de forma digital, disponíveis aos usuários da rede mundial de computadores. Esses dados, muitas vezes, constituem-se informações necessárias à atividade produtiva humana.

Ao fazer uso das TIC's para o tratamento temático da informação, as mesmas possibilitam não só operacionalização de *softwares*, mas também o seu desenvolvimento, manutenção e gerenciamento de forma a otimizar a automação da atividade de indexação semiautomática das publicações existentes. Chaumier (1988, p.63), conceitua o termo indexação como:

Operação que escreve e caracteriza um documento, com o auxílio da representação dos conceitos nela contidos. Tal ação pode ser feita por um indexador humano, sendo nesse caso denominada na literatura indexação manual; por um programa de computador, sendo denominada indexação automática; ou ainda por um programa de computador e depois revista por um indexador humano, sendo denominada indexação semiautomática. (CHAUMIER,1988, p.63)

A Ciência da Informação, ao trabalhar a organização da informação, busca abranger em suas atividades, tanto os processos de tratamento descritivo, de ordem física, como os tratamentos temáticos, ou seja, os conteúdos dos documentos. Tais técnicas de indexação possibilitam identificar o conteúdo dos documentos por meio da busca e evidência dos principais assuntos de um documento (REZENDE, 2005, p.352).

Assim, com o crescimento da produção científica em ambiente digital, tornam-se necessários estudos que visam à indexação dessa grande quantidade de dados disponíveis ou armazenados em centros de informação, como por exemplo: bibliotecas, serviços de informação e sistemas, sejam feitos de modo a apoiar os profissionais a uma melhor gestão do conhecimento, pois facilita a busca pela informação (TURBAN e VOLONINO, 2011, p.67; REZENDE, 2005, p.337).

Neste contexto, o presente trabalho tem como objetivo propor a arquitetura de um protótipo que auxilia, de forma semiautomática, a indexação de documentos. Em cada documento processado, o protótipo trabalha com mineração de texto em seu conteúdo,

retornando as palavras com maior frequência (moda), que tendem a ser mais significantes, facilitando a busca de conhecimentos e a eleição dos índices.

Pinto (2001, p. 227), define a indexação semiautomática como a junção da indexação manual com a indexação automática. De início, o software realiza uma indexação automática dos documentos quantificando as ocorrências das palavras mais encontradas no texto. A seguir, o indexador humano trabalha com a lista dos tópicos propostos pelo sistema fazendo os ajustes e/ou complementações necessárias.

Em centros de documentação, como por exemplo bibliotecas, a indexação mostra-se como um processo documentário que deve ser realizado de modo que facilite a operacionalização dos profissionais ali presentes, sempre que existir a necessidade de buscar um documento de temática específica, atendendo assim as necessidades de recuperação da informação (CENDÓN, et al, 2005, p. 62, 63). O foco deste trabalho se dá sobre textos semi ou não estruturados, armazenados em formato PDF¹, pois esse tipo de formato digital de arquivos é mais comum em sua utilização neste contexto.

A Mineração de Textos ou Descoberta de Conhecimento em Textos surge, neste contexto, como uma abordagem à aquisição de informação útil a partir de textos. As Técnicas de Mineração de Textos conseguem manipular documentos em formato textual, objetivando a aquisição do conhecimento implícito presente nos textos. (ARANHA & PASSOS, 2006). Descobrir conhecimentos em formato textual, nesses grandes volumes de dados eletrônicos, é de extrema importância.

O processo de Descoberta de Conhecimento Textual é originado do *Knowledge Discovery in Databases* (KDD), que tem como objetivos extrair padrões relevantes em bases de dados estruturadas (FAYYAD et al., 1996). Inspirado pelo *data mining*, ou mineração de dados, que procura descobrir padrões emergentes de banco de dados estruturados, a mineração de textos pretende extrair conhecimentos úteis de dados não estruturados ou semiestruturados.

¹Portable Document Format, ou simplesmente PDF, é um padrão aberto para compartilhamento de documentos digitais, desenvolvido pela Adobe Systems e mantido pela International Standards Organization (ISO).

2. MINERAÇÃO DE TEXTOS

A Mineração de textos é um conjunto de técnicas e métodos usados para navegar, organizar, achar e descobrir informação inovadora em bases textuais. Pode ser vista como uma extensão da área de *Data Mining*, focada na análise de textos. Também é chamada de *Text Data Mining* ou *Knowledge Discovery in Texts* (KDT) (REZENDE, 2005, p.337, 338).

O KDT é uma subárea da Inteligência Computacional, responsável por extrair informações relevantes e ainda ocultas de documentos ou bases de dados não estruturadas. O KDT é composto das etapas de pré-processamento, com a finalidade de preparar, transformar, organizar e melhorar a qualidade do texto para a etapa seguinte, o processamento ou Mineração de Textos, que é o objetivo do KDT, onde as técnicas aplicadas variam de acordo com a finalidade, recuperação da informação, indexação, extração da informação, associação de documentos, sumarização, clusterização e classificação/categorização. Por fim, a etapa de pós-processamento avalia os resultados da mineração (BARION e LAGO, 2008).

A utilização dessas técnicas permite extrair informações de um conjunto de dados, resumir e conhecer melhor os documentos e realizar análises qualitativas ou quantitativas em documentos de texto, de modo que possamos usar esses dados em objetivos futuros.

3. METODOLOGIA

Para o desenvolvimento do protótipo foi utilizada a linguagem de programação PHP, que é de código livre, possuindo recursos adequados para a construção da aplicação, além de permitir que a aplicação seja utilizada tanto em plataforma *desktop*, quanto para plataforma web. PHP é acrônimo de *Hypertext Preprocessor* e, segundo Soares (2010, p.28):

É uma poderosa linguagem de programação open source, mundialmente utilizada, principalmente no ambiente WEB. [...] Uma das características mais marcantes no PHP é sua capacidade de se misturar ao HTML, tornando mais fácil a geração de páginas web dinâmicas. (SOARES, 2010, p.28)

O desenvolvimento foi feito sob o paradigma de orientação a objetos, técnica em que os sistemas devem ser construídos baseados em uma coleção de componentes reutilizáveis, conhecidos como objetos (SOARES, 2010, p.292). O protótipo conta também com o uso de banco de dados relacional. Como sistema gerenciador de banco de dados (SGBD) foi

utilizado o MySQL, que é amplamente difundido entre sistemas web (MySQL, 2014). A interface gráfica foi implementada através da linguagem HTML5, amparada na utilização do *framework Bootstrap*².

Para descrever o protótipo neste trabalho serão apresentadas imagens de telas capturadas durante sua utilização, demonstrando como a mineração de texto de uma publicação é realizada. A figura 01 apresenta a arquitetura do protótipo desenvolvido.

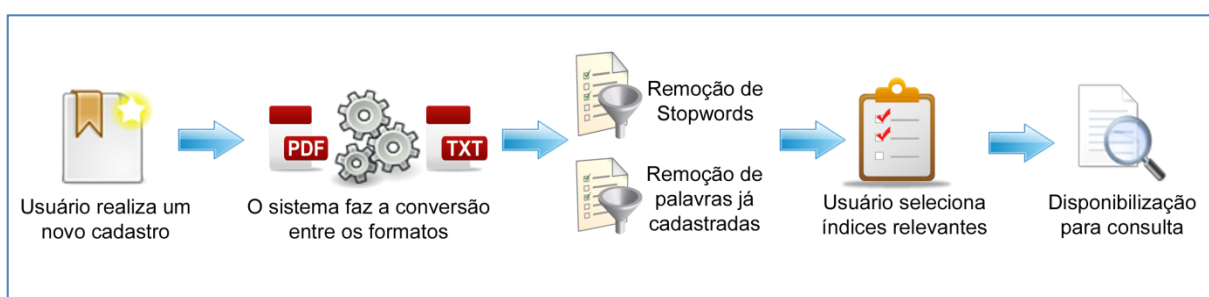


Figura 1 - Arquitetura do protótipo. Fonte: autoria própria³

A estratégia de programação utilizada para a mineração de texto foi a utilização de uma classe específica, escrita em linguagem PHP, para proceder a conversão de um arquivo PDF para um pacote de dados em texto (TXT). Uma vez realizada a conversão, o texto extraído é vetorizado, ou seja, as palavras são inseridas em uma estrutura de dados do tipo vetor, para posterior manipulação.

Em um texto, existem muitas palavras (*tokens*) que não representam nenhum valor semântico ao texto, estando presente apenas por conta da estrutura textual e compreensão geral do mesmo. Os *tokens* são classificados como *stopwords*⁴ e fazem parte de *stoplist*⁵ em um sistema de Mineração de Textos. (BASTOS, 2006, apud JUNIOR, 2007, p.46).

A utilização de vetores justifica-se em dois momentos da programação. Primeiro, para a extração das palavras mais frequentes do texto, o vetor é submetido a um algoritmo de ordenação, o que torna mais simples o cômputo das palavras repetidas dentro de um laço de repetição. A eliminação de *stopwords* também é beneficiada pelo uso de vetor, uma vez que a

² *Bootstrap* é um conjunto *open source* de padrões para os elementos HTML, CSS e Javascript mais utilizados, agilizando o desenvolvimento de interface (*front-end*) para páginas web, podendo ser também personalizado conforme as necessidades do desenvolvimento.

³ Imagens disponíveis em <http://www.iconspedia.com/>

⁴ *Stopwords* são palavras potencialmente irrelevantes para o conjunto de resultados a ser exibido na busca realizada em um texto. Exemplos: as, e, os, de, para, com, sem, foi.

⁵ Conjunto estruturado de *Stopwords*.

linguagem PHP conta com funções apropriadas de busca dentro desse tipo de estrutura de dados.

Sabe-se que uma *stoplist* bem elaborada, permite o tratamento do texto, de modo que elimina os termos irrelevantes, melhorando o resultado final da Mineração de Texto. Sabe-se também que 40 a 50% do total de palavras presentes em um texto serão removidos com uma *stoplist* quando se pretende trabalhar com a extração do conhecimento. (SILVA, 2007, apud JUNIOR, 2007, p.46).

O uso das linguagens PHP/MySQL permitirá que a aplicação desenvolvida seja facilmente disponibilizada, pois é factível que a grande maioria de servidores, atualmente, oferecem esta plataforma de serviço em inúmeros modelos de contratação de serviços de hospedagem web.

A modelagem dos dados utilizada para o protótipo foi construída observando-se o nível simples do esquema de metadados Dublin Core.

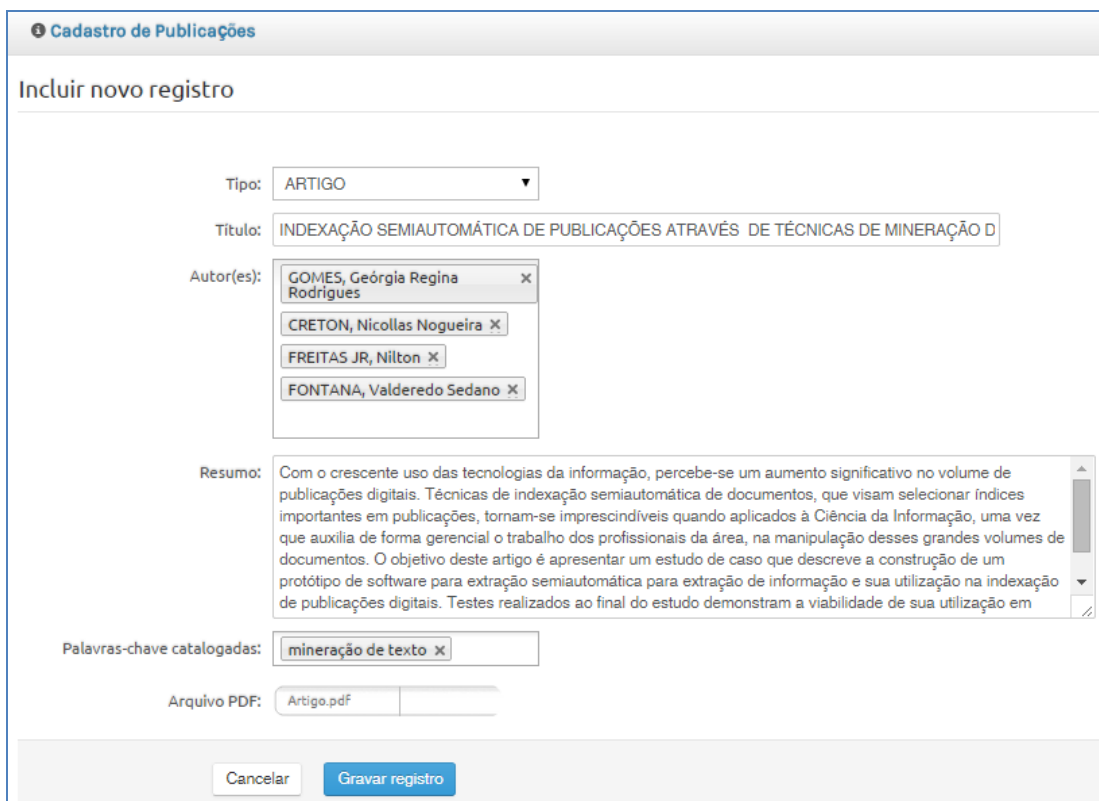
A Iniciativa de Metadados Dublin Core, ou "DCMI", é uma organização aberta apoiando a inovação em design de metadados e as melhores práticas em toda a ecologia de metadados. Atividades de DCMI incluem trabalhos sobre arquitetura e modelagem, discussões e trabalho colaborativo em Comunidades DCMI e DCMI globais, conferências, reuniões e oficinas, e os esforços educacionais para promover a aceitação generalizada de padrões de metadados e as melhores práticas (The Dublin Core® Meta data Initiative).

Por se tratar de um protótipo, a aplicação conta com apenas alguns elementos do nível simples DCMI. Entretanto, sua modelagem permite futuras implementações mais completas seguindo os padrões DCMI.

3. ESTUDO DE CASO

Segundo (Gomes, Moraes Filho, 2011), tratar documentos digitais não é uma tarefa fácil, pois existem vários fatores que devem ser levados em consideração, como por exemplo, a falta de estruturação que eles apresentam, detectar as palavras que são consideradas irrelevantes para a mineração, a quantidade de palavras relevantes que aparecem nos documentos. Portanto, ao se trabalhar com mineração de textos é necessário tratar os dados textuais para que a falta da limpeza dos dados não comprometa a eficácia do método.

O processo de utilização da aplicação inicia com o registro de uma publicação, onde o usuário preenche os dados de um formulário baseado no nível simples DCMI, como demonstrado na figura 02:



Cadastro de Publicações

Incluir novo registro

Tipo: ARTIGO

Título: INDEXAÇÃO SEMIAUTOMÁTICA DE PUBLICAÇÕES ATRAVÉS DE TÉCNICAS DE MINERAÇÃO D

Autor(es): GOMES, Geórgia Regina Rodrigues
CRETON, Nicollas Nogueira
FREITAS JR, Nilton
FONTANA, Valderedo Sedano

Resumo: Com o crescente uso das tecnologias da informação, percebe-se um aumento significativo no volume de publicações digitais. Técnicas de indexação semiautomática de documentos, que visam selecionar índices importantes em publicações, tornam-se imprescindíveis quando aplicados à Ciência da Informação, uma vez que auxilia de forma gerencial o trabalho dos profissionais da área, na manipulação desses grandes volumes de documentos. O objetivo deste artigo é apresentar um estudo de caso que descreve a construção de um protótipo de software para extração semiautomática para extração de informação e sua utilização na indexação de publicações digitais. Testes realizados ao final do estudo demonstram a viabilidade de sua utilização em

Palavras-chave catalogadas: mineração de texto

Arquivo PDF: Artigo.pdf

Cancelar Gravar registro

Figura 02: Registro de uma publicação. Fonte: autoria própria

O protótipo executa processos de mineração em um texto manipulável e informado por meio de quatro etapas para o processo de extração. A primeira etapa trabalha com a seleção e preparação do texto, seguindo para a segunda etapa, que consiste na retirada das *stopwords*, palavras previamente catalogadas, além da retirada de caracteres ortográficos. Este processo é feito de forma automática, sem depender da intervenção do usuário.

A terceira etapa será realizada após o usuário escolher a frequência mínima (moda) das palavras no texto, de forma que somente aquelas palavras que tiverem uma frequência maior ou igual à mínima definida sejam apresentadas para seleção de indexação (figura 03). Aqui se configura a característica do protótipo ser semiautomático, já que todo o processo depende da intervenção do usuário.

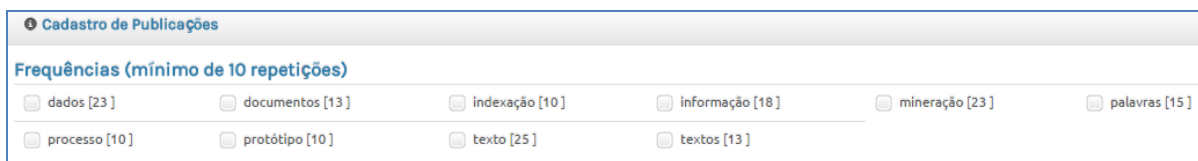


Figura 03: Frequência de palavras extraídas do texto. Fonte: autoria própria

Na quarta etapa, com a lista de palavras apresentadas ao usuário, este deverá selecionar dentre aquelas informadas quais julgar relevantes, de modo que seja possível inserir em uma base de dados e associa ao documento em questão. Como na etapa anterior, há a dependência do usuário para este processamento. A figura 04 apresenta um diagrama de atividades com as etapas aqui descritas.

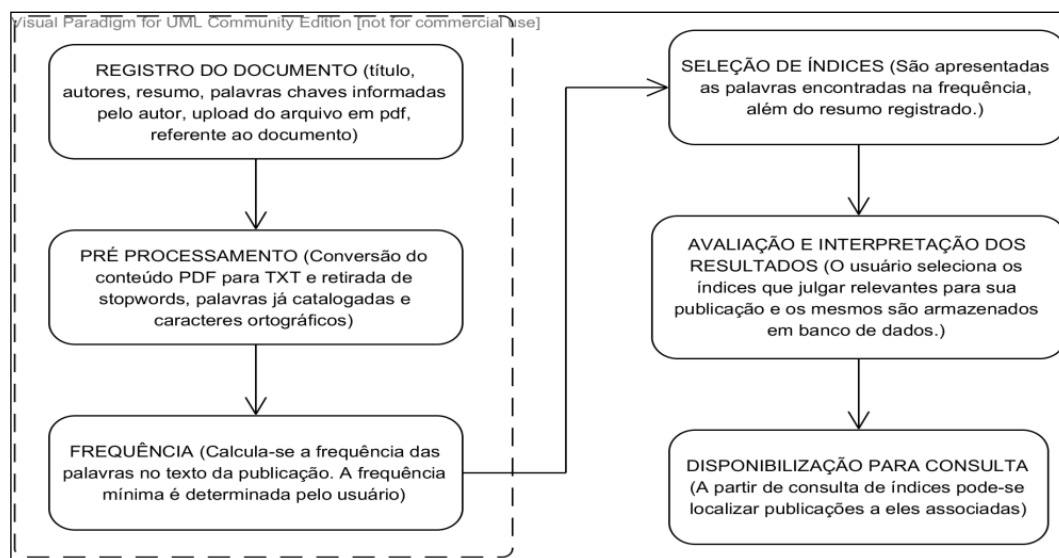


Figura 04: Etapas do processo de extração. Fonte: Autoria própria⁶

As etapas descritas representam bem o processo de mineração de texto. Em seu trabalho, Medeiros (2004) explica que:

O processo de Mineração de texto possui duas fases principais e sequentes: a extração de informação e a mineração de dados. A primeira fase tem papel bastante importante no processo de mineração, pois permite o tratamento da representação de dados e tem como objetivo extrair conceitos, estatísticas e palavras relevantes de um conjunto textual, visando proporcionar uma estrutura mínima. Na segunda fase são aplicadas diretrizes e algoritmos de mineração de dados destinados a gerar regras, classificações ou agrupamentos. (MEDEIROS, 2004, p.14)

⁶ Visual Paradigm Community Edition - FREE for non-commercial use only. Disponível em <<https://www.visual-paradigm.com/download/community.jsp>>

Pressupõe-se que uma alta frequência de uma mesma palavra no texto representa uma alta relevância da mesma, para o texto minerado. Mesmo assim, caberá ao usuário analisar essa relevância para eleger uma ou mais palavras como índices de sua publicação. Moraes e Ambrósio (2007) apresentam uma motivação do uso da frequência para a mineração de texto.

O cálculo de relevância de uma palavra em relação ao texto em que está inserido pode basear-se na frequência da mesma, na análise estrutural do documento ou na sua posição sintática de uma palavra. As análises baseadas em frequência costumam ser as mais utilizadas por serem mais simples, pois os outros tipos de análise muitas vezes necessitam de outras técnicas mais complexas, como processamento de linguagem natural. (MORAIS e AMBROSIO, 2007).

O processo de frequência não foi o único utilizado para extração de texto da publicação, uma vez que o resumo também irá fornecer possibilidades ao usuário de escolher outros índices os quais julgar da mesma forma, relevantes para sua publicação (figura 05).

SUGESTÕES					
<input type="checkbox"/> crescente	<input type="checkbox"/> uso	<input type="checkbox"/> tecnologias	<input type="checkbox"/> informação	<input type="checkbox"/> percebe-se	<input type="checkbox"/> aumento
<input type="checkbox"/> significativo	<input type="checkbox"/> volume	<input type="checkbox"/> publicações	<input type="checkbox"/> digitais	<input type="checkbox"/> técnicas	<input type="checkbox"/> indexação
<input type="checkbox"/> semiautomática	<input type="checkbox"/> documentos	<input type="checkbox"/> visam	<input type="checkbox"/> selecionar	<input type="checkbox"/> índices	<input type="checkbox"/> importantes
<input type="checkbox"/> tornam-se	<input type="checkbox"/> imprescindíveis	<input type="checkbox"/> aplicados	<input type="checkbox"/> ciência	<input type="checkbox"/> auxilia	<input type="checkbox"/> forma
<input type="checkbox"/> gerencial	<input type="checkbox"/> trabalho	<input type="checkbox"/> profissionais	<input type="checkbox"/> área	<input type="checkbox"/> manipulação	

Figura 05: Palavras extraídas do resumo da publicação. Fonte: autoria própria

Todo este processo pode ser repetido quantas vezes o usuário desejar, embora os resultados sejam sempre mais restritos. Isso acontece porque a programação verifica as palavras-chave já associadas à publicação. Dessa forma, cada vez que o usuário executar o procedimento de mineração de texto, menores resultados serão oferecidos como possibilidade para indexação da publicação.

4. RESULTADOS

Testes de usabilidade realizados com o protótipo mostraram-se consistentes, atingindo os objetivos propostos na construção da aplicação. Como esperado, alguns arquivos PDF apresentaram problemas em sua leitura, todos decorrentes de possíveis inconsistências em sua geração. Entretanto, a mineração de texto aplicada ao resumo da publicação mostrou-se eficaz em todos os testes realizados.

O uso da linguagem PHP para o processamento da mineração de texto mostrou-se adequado, principalmente, pela facilidade de integração entre os métodos de conversão do arquivo PDF e a construção da aplicação propriamente dita.

Os índices selecionados pelo usuário foram devidamente armazenados em tabelas SQL, com associação para com publicação a partir da qual foram selecionados. Um módulo de busca por esses índices também retorna as publicações associadas aos mesmos (figura 06).



The screenshot shows a search interface with two main sections. The left section, titled 'Pesquisa', contains a 'Palavras-chave:' field with two input boxes containing 'semiautomática' and 'mineração', each with a small 'X' icon to its right. Below the input boxes are 'Cancelar' and 'Pesquisar' buttons. The right section, also titled 'Pesquisa', displays the search results for the query 'INDEXAÇÃO SEMIAUTOMÁTICA DE PUBLICAÇÕES ATRAVÉS DE TÉCNICAS DE MINERAÇÃO DE TEXTO [artigo]'. The results list authors: GOMES, Geórgia Regina Rodrigues; FREITAS JR, Nilton; FONTANA, Valdeir de Sedano; and CRETON, Nicollas Nogueira. Below the list is a paragraph of text describing the research and the application's purpose.

Figura 06: Pesquisa por índice e resultado. Fonte: Autoria própria

Segundo Brandau e Monteiro (2005), quando feita de forma correta, a escolha de palavras-chave aumenta as chances de o artigo ser encontrado e, conseqüentemente, lido e também referenciado.

É interessante destacar que os índices armazenados também podem ser utilizados por publicações futuras, já que é oferecida essa possibilidade ao usuário no momento da inserção de um novo registro de publicação. O protótipo tende, com sua constante utilização, a se tornar um repositório de índices úteis para catalogação de publicações diversas.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou e descreveu o desenvolvimento do protótipo de uma aplicação semiautomática de extração de conhecimento e sua utilização na indexação de documentos digitais, juntamente com suas funcionalidades. Dessa forma, a aplicação auxilia na indexação de um determinado documento em formato PDF, retirando primeiramente do resumo, já tratado, cada palavra separada para que as mais adequadas possam ser eleitas como índices. Já do texto completo, é levada em consideração a frequência das palavras, facilitando e agilizando assim, o processo de seleção dos índices do documento.

Segundo Narukawa, Leiva e Fujita (2009), o constante avanço tecnológico e científico tem acarretado numa grande quantidade de produções científicas, gerando novos

conhecimentos e difundindo informações. Dessa forma, o tratamento de tais documentos se torna necessário para que seja possível ser feita a recuperação destas informações.

A Ciência da Informação é responsável por tais tratamentos, sendo estes temáticos e se baseando no conteúdo dos documentos. Com tal crescimento, surgiu a necessidade de ferramentas automáticas ou semiautomáticas de indexação para não só garantir resultados mais eficazes, mas também para facilitar as tarefas envolvidas (NARUKAWA, LEIVA e FUJITA, 2009).

Como trabalhos futuros, o desenvolvimento do protótipo pode continuar no sentido de associar mecanismos de representação semântica entre os índices armazenados a fim de oferecer maiores recursos gerenciamento de informação e resultados de pesquisas dentro da própria aplicação, de modo a contribuir com os avanços da área da ciência da informação.

REFERÊNCIAS

ARANHA, C.; PASSOS, E. .A Tecnologia de Mineração de Textos. RESI-Revista Eletrônica de Sistemas de Informação, N°2-2006 .Lab.ICA Elétrica PUC-Rio, 2006.

BARION, Eliana Cristina Nogueira; LAGO, Decio. Mineração de textos. Revista de Ciências Exatas e Tecnologia, v. 3, n. 3, p. 123-140, 2015.

BRANDAU, Ricardo; MONTEIRO, Rosangelaand BRAILE, Domingo M.. Importância do uso correto dos descritores nos artigos científicos. RevBrasCirCardiovasc [online]. 2005, vol.20, n.1, pp. VII-IX. ISSN 0102-7638. <http://dx.doi.org/10.1590/S0102-76382005000100004>.

CENDÓN, B.V, et al. Ciência da Informação e Biblioteconomia: novos conceitos e espaços de atuação. Belo Horizonte, MG: Editora UFMG, 2005.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

DAMAS, L. SQL, Structured query language.Rio de Janeiro, RJ: LTC, 2012.

DUBLIN CORE METADATA INITIATIVE. The Metadata Community — Supporting Innovation in Metadata Design, Implementation & Best Practices.Disponível em <<http://dublincore.org/>>. Acesso em 06/12/2014.

GOMES, G. R. R.; MORAES FILHO, R. O. CADWeb: categorização automática de documentos digitais. *Ciência da Informação*, v. 40, n. 1, p. 68–76, abr. 2011.

JUNIOR, J.R.C. Desenvolvimento de uma Metodologia para Mineração de Textos. Rio de Janeiro: PUC-Rio, 2007, 113 f.Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007

MEDEIROS, E. A. Técnica de aprendizagem de máquina para categorização de textos. Recife: Escola Politécnica de Pernambuco, 2004.

MORAIS, E.A.M, AMBROSIO, A.P. Mineração de Textos. Instituto de Informática: Universidade Federal de Goiás, 2007.

MySQL. Disponível em: <www.mysql.com/>. Acesso em: 20 nov 2014.

NARUKAWA, C. M.; LEIVA, I. G.; FUJITA, M. S. L. (2009). Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de Odontologia. *Informação & Sociedade: Estudos*, v. 19, n. 2, 2009.

PHP. Disponível em: <<http://php.net/>>. Acesso em: 20 nov 2014.

PINTO, V. B. Indexação documentária: uma forma de representação do conhecimento registrado. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 6, n. 2, p. 223 - 234, jul./dez. 2001.

REZENDE, S.O. Sistemas Inteligentes: fundamentos e aplicações. Barueri, SP: Manole, 2005.

SOARES, W. PHP 5: Conceitos, Programação e Integração com Banco de Dados. 6. ed. São Paulo, SP: Érica, 2010.

TERUEL, E.C. HTML5: Guia Prático. São Paulo, SP: Érica, 2011.

TURBAN, E, VOLONINO, L. Tecnologia da Informação para Gestão: em busca do melhor desempenho estratégico e operacional. 8.ed. Porto Alegre, RS: Bookman Editora, 2011.