

APLICAÇÃO DE METODOLOGIA DE INDEXAÇÃO AO CONTEÚDO DO PORTAL DA CHAPADA

Emanuel de Jesus Lima¹; João Carlos Nunes Bittencourt², Delmar Broglio Carvalho³,
Acácia Batista Dias⁴

1. Bolsista PROBIC/UEFS, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: emanuel.uefs@gmail.com
2. Bolsista PROBIC-UEFS, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: jhonny.bittencourt@gmail.com
3. Orientador, Departamento de Tecnologia, Universidade Estadual de Feira de Santana, e-mail: carvalho.db@ecompu.uefs.br
4. Co-orientadora, Departamento de Ciências Humanas e Filosofia, Universidade Estadual de Feira de Santana, e-mail: acaciabatista02@gmail.com

PALAVRAS-CHAVE: Indexação, Arquitetura da Informação, Meta Portal, Software Livre.

INTRODUÇÃO

O crescimento exponencial da Web tem criado muitos desafios, para os quais somente a criação de padrões, aliados a normatizações, serão capazes de prover uma relativa organização. A remoção de grande parte das dificuldades do desenvolvimento permitirá maior flexibilidade ao desenvolvedor e produtor Web, possibilitando assim que as pessoas possam realmente exercer a sua criatividade ao mesmo tempo em que forjam elos de comunicação (Polizelli & Ozaki, 2008); (Catarino & Baptista, 2007).

O projeto Portal da Chapada está sendo desenvolvido pela Universidade Estadual de Feira de Santana (UEFS) com financiamento da Fundação de Apoio à Pesquisa do Estado da Bahia (FAPESB), sob a perspectiva do incentivo à Popularização da Ciência e da Tecnologia, e consiste no desenvolvimento de um Sistema Web para armazenamento, recuperação e disponibilização de conteúdos digitais (imagens, vídeos, mapas, músicas, resenhas), referentes a região da Chapada Diamantina, voltados à sociedade, ciência, cultura, lazer, entre outros (Bittencourt *et. al.*, 2011). O projeto é pautado na presença de atores de diversas áreas do conhecimento que, a partir da utilização de um sistema de gestão de conteúdo possam criar uma base de dados capaz de disponibilizar informações a todos os usuários do Portal. Sob esta perspectiva, a região da Chapada Diamantina foi escolhida por ser uma região com forte atuação da UEFS, onde foi identificada a ausência de um processo formal de sistematização das informações produzidas, dentro e fora da Universidade, sobre a região, bem como a pouca inserção desse material em formato digital. A situação encontrada não é única e reflete o que vem acontecendo em diversos setores, devido às consequências dos grandes avanços tecnológicos e ao caos informacional do mundo digital (Dias, 2001).

Este trabalho relata os aspectos relacionados à aplicação de metodologias capazes de fornecer recursos para a indexação dos diferentes conteúdos inseridos no Portal da Chapada Diamantina, desenvolvendo um modelo dinâmico, estruturado sob uma interface intuitiva, associada à mecanismos que permitam o aumento do nível de interação do usuário com o sistema, favorecendo o acesso à informação de forma rápida, segura e confiável. Este processo, inserido na construção de uma base de dados concisa, combinada com uma metodologia de usabilidade de portais Web, envolve estudos em diversas áreas do conhecimento, favorecendo a interdisciplinaridade dos atores envolvidos no desenvolvimento e uso do Portal.

METODOLOGIA

Segundo De Oliveira *et. al.* (2011), para possibilitar a recuperação da informação faz-se necessário aplicar uma representação, através de atributos caracterizadores para os

documentos, a fim de identificá-los de maneira única por meio da descrição que lhe foi designada, inclusive através de símbolos que sejam conhecidos tanto pelos operadores quanto pelos usuários do sistema de interface. Este processo resulta em facilidades ao sistema de indexação.

A indexação, dentro de um sistemas de buscas, tem como finalidade preparar uma representação do conteúdo temático dos documentos, para sua posterior recuperação em uma base de dados (De Souza, 2009). Para cada objeto digital pode-se associar várias palavras representativas, também designadas de *tags* (etiquetas), ou ainda adicionar particularidades que permitem ao mesmo ser identificado como peça única entre um grande número de objetos disponibilizados em um meio digital.

Um dos principais elementos no processo de indexação são as palavras-chave, usadas, em muitos casos, como o principal método de pesquisa dos mecanismos de busca, por isso é essencial que receba a máxima atenção. Dessa forma, torna-se muito importante usar apenas palavras-chave que o ajudarão a identificar um determinado conteúdo e não outro (De Oliveira *et. al.*, 2011).

Outro item de destaque no processo de indexação é a categorização de conteúdo, uma vez que as ferramentas de busca podem utilizá-las como filtros, facilitando a gestão das pesquisas. É imprescindível determinar a categoria mais apropriada para determinado objeto digital. No entanto, é preferível não associar uma determinada categoria a um arquivo em detrimento de escolher uma que não corresponda ao perfil do mesmo.

O processo de agrupamento, estruturação e indexação de conteúdo, no projeto em desenvolvimento, é realizado por um sistema de gerenciamento de conteúdo (CMS). No caso, foi escolhido o Joomla! (North, 2008; JOOMLA!, 2011), o qual é uma plataforma para aplicações Web, distribuído na forma de Software Livre (SL) e código aberto (*opensource*) sob licença GNU/GPL (GNU, 2012). O Joomla! conta ainda com um recurso de categorização de todo o material que é inserido. O gerenciador de categorias do Joomla! permite a intitulação, descrição e a vinculação com outra categoria.

RESULTADOS E DISCUSSÃO

De acordo com a proposta de desenvolvimento do projeto, o Joomla!, conjuntamente com as ferramentas acessórias, demonstrou ser capaz de gerenciar o processo de estruturação e indexação do conteúdo no Portal (Figura 1). No caso, existem três campos de resultados para diferentes contextos (Novidades e Atualizações, Páginas mais Visitadas e *Tags*), onde estes campos foram definidos na etapa de planejamento do Portal e cada um dos itens associados é oriundo de um mecanismo interno de ordenação. No caso do primeiro campo, o mecanismo está associado ao instante de postagem do objeto digital. Já no segundo campo a sequência obtida é fruto da ordenação em função dos número de acessos aos *links*correspondentes. No terceiro campo o resultado é oriundo do volume de objetos digitais associados através das *tags* (etiquetas).

Novidades e Atualizações	Páginas mais Visitadas	Tags
<ol style="list-style-type: none"> 1. Como criar artigos no Joomla! 2. SECTUR INFORMA: PROJETO DE DIGITALIZAÇÃO 3. SECTUR INFORMA: Secretária Nacional de Cidadania 4. CASA AFRÂNIO PEIXOTO APRESENTA: Filme Cidadão Kane 5. SECTUR INFORMA: Exibição do Programa da TV Brasil 	<ol style="list-style-type: none"> 1. Equipe 2. Meta Portal 3. Sobre o Portal da Chapada 4. Um Pouco sobre a Chapada Diamantina 5. "Até onde justo for" : Arranjos e Conflitos pela Propriedade e Posse da Terra 	<p>Características Sócio-geográficas Chapada Diamantina Conflitos Meta Porta Objetivos Portal Corporativo Portal da Chapada Portal da Chapada Diamantina Portal da Chapada Diamantina, UEMS chapada diamantina conteudo desenvolvimento</p> <p>equipe espaço virtual monografia portal da chapada</p> <p>Powered by Easytagcloud v2.0</p>

Figura 1 - Resultados oriundos do processo de indexação dos conteúdos do Portal.

No Joomla! o processo de indexação de textos também pode ser feito partir da inserção das palavras-chave, denominadas de *meta-tags*, em campos especiais no momento da criação ou edição do objeto digital. Internamente o sistema utiliza mecanismos instruídos a recolher informações (palavras-chave) que estão contidas dentro dos campos apropriados, sendo enviadas para o banco de dados do mecanismo de busca. No banco de dados esta informação será usada, juntamente com outras para classificar a página e colocá-la em uma nuvem de *tags* (Figura 2) que reúne um conjunto, ordenado alfabeticamente, de palavras-chave utilizadas no Portal, e o volume de conteúdo que o site contém em cada *tag* é representado proporcionalmente ao tamanho da fonte.



Figura 2 - Nuvem de TAGs.

A nuvem de *tags* é uma representação visual gerada pelo volume de acessos ao conteúdo indexado no Portal. Desta forma são disponibilizadas *tags*, as quais na nuvem correspondem aos links que redirecionam à coleções de itens relacionados às palavras das *tags*, proporcionando melhorias na navegabilidade do Portal e relacionando o conteúdo de maior interesse ao leitor.



Figura 3 - Nuvem de TAGsacionada.

Além dos mecanismos do processo de indexação demonstrados anteriormente, foi realizado um teste de verificação de busca de conteúdos digitais. Neste teste, foi avaliado se no momento da busca o sistema seria capaz de exibir conteúdos significativos referentes ao que

foi buscado. Para isso foram utilizados doze artigos relacionados a um tema específico, sendo que em oito destes objetos foram aplicadas técnicas de indexação aqui descritas e em quatro não. O resultado ocorreu de acordo com o esperado, onde os conteúdos indexados foram exibidos no resultado e aqueles não indexados foram deixados de fora. Em um segundo momento, esses conteúdos foram indexados e o mesmo procedimento de busca foi realizado, tendo como resultado a recuperação de todos os conteúdos.

CONSIDERAÇÕES FINAIS

A partir dos resultados apresentados, o processo de indexação do conteúdo do Portal da Chapada Diamantina é uma etapa que exige elaboração mais detalhada, uma vez que o mesmo conta com uma base de dados com um grande número de conteúdos digitais (monografias, teses, resenhas, fotos, mapas), abordando áreas como a biologia, geografia, história entre outros. Dessa forma, é recorrente que o usuário do Portal depare-se, no momento da leitura, ou em determinada busca, com sugestões de leitura ou pesquisas relacionadas ao tema em questão. Neste caso, o processo de indexação leva em consideração elementos como a escolha de palavras-chave e a categorização as quais determinam o conteúdo a ser exibido.

De modo geral, o processo de estruturação e indexação foi desenvolvido através das técnicas apresentadas no presente trabalho e os resultados apresentados asseguram que os objetivos foram alcançados com sucesso. Atualmente o Portal conta com uma base de dados concisa e indexada disponibilizando conteúdo rico e correlato para seus leitores. A versão do Portal da Chapada Diamantina atualizada e com conteúdo indexado encontra-se disponível em: <http://portaldachapada.uefs.br/novoportal>.

REFERÊNCIAS

- BITTENCOURT, J.C.N.; CARVALHO, D.B.; DIAS, A.B. 2011. META PORTAL: uso das tecnologias de informação e comunicação na construção de um ambiente colaborativo. In: XV Seminário de Iniciação Científica da UEFS, Feira de Santana, p.587-589.
- CATARINO, M.E.; BAPTISTA, A.A. 2007. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *DataGramaZero - Revista de Ciência da Informação* 8(3):a04.
- DE OLIVEIRA, A. M.; FERREIRA, H. T.; CASIMIRO, V. A.; MARQUES DOS SANTOS, C. A. C. 2011. SearchEngineOptimization - SEO: A contribuição do Bibliotecário na otimização de WebSites para os mecanismos de busca. *Perspectivas em Gestão & Conhecimento*, João Pessoa, v. 1, Número Especial, p. 137-159.
- DE SOUZA, M.S. 2009 [online]. *Indexação e Recuperação da Informação na Internet*. Homepage: http://sigplanet.sytes.net/nova_plataforma/monografias/7431.pdf.
- DIAS, C. A. 2001. Portal corporativo: conceitos e características. *Ciência da Informação* 30(1): 50–60.
- GNU. 2012 [online]. *A licença pública Geral do GNU*. Homepage: <http://www.gnu.org/licences/>.
- JOOMLA!. 2008 [online]. *O que é Joomla!?* Homepage: <http://www.joomla.com.br/o-que-e-joomla.html>.
- NORTH, B. 2008. *Joomla! Guia do Operador*. São Paulo, AltaBooks, 384p.
- POLIZELLI, D.L; OZAKI, A.M. (Org). 2008. *Sociedade da Informação: os desafios da era da colaboração e da gestão do conhecimento*. São Paulo, Saraiva, 300p.