

INDEXAÇÃO AUTOMÁTICA DE CONTEÚDOS NA WEB: análise de *sites* de museus

AUTOMATIC INDEXING IN CONTENT ON THE WEB: analysis of sites of museums

Cíntia Holanda*

Márcia Ivo Braz**

RESUMO

Na atualidade várias ferramentas foram criadas para a recuperação da informação no ambiente *Web*, bem como para a disponibilização de conteúdos. As características destas ferramentas influenciam no modelo de indexação utilizado e na forma que os assuntos serão recuperados. As tecnologias da informação e comunicação são cada vez mais utilizadas pelas instituições culturais, como os museus, objeto desse estudo. Dessa forma, serão identificadas algumas estratégias e procedimentos que os museus podem adotar ao disponibilizar conteúdos no ambiente *Web* para posterior recuperação no motor Google. A partir de um levantamento de literatura, este estudo explana conceitos da indexação manual, chegando à abordagem da indexação automática, apresentando caráter exploratório, cujo delineamento é definido com base em assuntos relacionados a fontes de informações na Internet, diretórios e motores de buscas. Após a revisão teórica foi possível identificar que não houve a devida preocupação em estruturar os conteúdos seguindo os critérios estabelecidos por mecanismos de busca, como o Google, para indexar páginas. Para se chegar a essa conclusão, analisou-se três itens fundamentais na constituição de uma página na *Web*: a *tag title* e as *meta tags description* e *keywords*. Os resultados podem possibilitar a remodelagem das *home pages*, de modo a melhorar a recuperação, promovendo o desenvolvimento dos métodos que o Google considera relevante no momento da indexação das páginas, bem como proporcionar estudos futuros baseados nas ferramentas de recuperação na *Web*.

Palavras-chave: Recuperação da informação. Indexação automática. Mecanismos de busca. Google. Museus.

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

ABSTRACT

At present several tools have been created for information retrieval in the Web environment as well as the availability of content. The characteristics of these tools influence in the model of indexing used and in the form that the affairs will be recovered. The information and communication technologies are increasingly used by cultural institutions such as museums, object of this study. Thus, some strategies will be identified and procedures that museums can adopt to make content available in the Web environment for later retrieval engine Google. From a survey of literature, this study explains the concepts of manual indexing, reaching approach to automatic indexing, presenting exploratory, whose design is set based on issues related to sources of information on the Internet, directories and search engines. After the theoretical review, we found that there was no proper concern to the contents structure following the criteria established by search engines like Google use to index pages. To reach this conclusion, we analyzed three key items in the constitution of a web page: the tag title and meta tags description and keywords. The results can enable the reshaping of home pages, so as to improve recovery, promoting the development the methods which the Google considers relevant in moment of indexing of pages, as well providing future studies based on recovery tools on the Web.

Keywords: Information retrieval. Automatic indexing. Search engines. Google. Museums.

1 INTRODUÇÃO

É evidente a influência da tecnologia nos diversos âmbitos da nossa sociedade, e também crescente a sua utilização em

processos que envolvam o gerenciamento da informação, como a indexação, armazenamento e recuperação da informação. Por volta do século XX, depois da segunda grande Guerra Mundial, com a chamada “explosão documental”, em razão da produção em massa de registros informacionais (BIBLIOTECA, 2007). As tecnologias de informação surgem, mesmo que rudimentarmente nesse âmbito, como tentativa de organizar os documentos contidos em sistemas analógicos. Dessa forma, a preocupação em desenvolver meios de organização da informação aliadas às tecnologias disponíveis é exponencial, visto que, por meio dessas, os usuários têm a capacidade de encontrar, identificar, selecionar e obter a informação desejada, elaborando meios que permitem o acesso, “principalmente em uma sociedade onde a explosão documental atingiu seu ápice e os usuários sozinhos não conseguem obter as informações desejadas” (MAIMONE; SILVEIRA, 2007, p. 55).

Com desenvolvimento da Internet, surgiu também a necessidade de interfaces que tornasse a “navegação” mais interativa, possibilitando que os conteúdos fossem mais facilmente compreendidos. Nessa perspectiva, no ano de 1991, Tim Berners-Lee apresenta ao mundo um estudo iniciado no ano de 1989, a *Web*, o que tornou a Internet mais dinâmica e, posteriormente, fazendo com o que se tornasse uma ferramenta popular considerada uma fonte valorosa de informação, disponibilizando textos, imagens, sons, fotos, vídeos, músicas, animação, multimídia, etc.

Com o advento da Internet e mais tarde com o surgimento da *Web*, a troca de informações atingiu proporções que não foram ainda alcançadas por outro meio de comunicação, uma vez que, se tornou um mecanismo de disseminação da informação, independentemente de suas localizações

geográficas. Além do mais, seu crescimento é exponencial, e, a cada dia, o número de páginas disponíveis aumenta, chegando a estimativa do surgimento de 320 milhões de endereços por semana, mas somente 20% desse conteúdo estará online dentro do prazo de um ano, o que demonstra a dinâmica existente em termos de crescimento e de mudanças na *Web* (BRAND; FARIA, 2006), refletindo consideravelmente nas formas de se buscar a informação no ambiente virtual.

Na atualidade, vários motores de busca são desenvolvidos de modo a facilitar a pesquisa na *Web* e cada característica destas ferramentas influencia diretamente no resultado final da procura pela informação. Nessa perspectiva a empresa Google foi fundada de acordo com o objetivo principal dos seus criadores, Larry Page e Sergey Brin, que é organizar as informações do mundo todo e torná-las acessíveis e úteis em caráter universal. Desde então, a empresa vem estudando e fornecendo abordagens para a recuperação mais precisa da informação, sendo o principal mecanismo de pesquisa no ambiente *Web*.

Assim, é evidente que as tecnologias da informação e comunicação são cada vez mais empregadas para diversas atividades e setores da sociedade, inclusive pela comunidade científica e pelas instituições culturais, como por exemplo, os museus, que são um exemplo do uso dessas tecnologias, pois desempenham papel fundamental na preservação da memória em diversos seguimentos sociais, culturais e econômicos. O uso das tecnologias por parte dos museus tenta repercutir para a sociedade projeções do acervo formado na instituição, empregando a *Web* como fonte valorosa de disponibilização de conteúdos.

Mesmo usando técnicas sofisticadas para indexar conteúdos e recuperando páginas que sejam igualmente importantes como

relevantes à busca, é possível o uso de alguns artifícios para facilitar a indexação automática das páginas que constituem um *site*, por parte de mecanismos como o Google. Mas será que os conteúdos disponibilizados pelos *sites* pertencentes aos museus são facilmente recuperados pelo sistema Google? Existem formas para que uma página seja recuperada com mais facilidade assim que solicitada a pesquisa?

Com o intuito de oferecer melhor recuperação da informação e preservação da memória no ambiente *Web*, este estudo teve como objeto de pesquisa o conteúdo e análise de *sites* de museus. Toda esta pesquisa foi conduzida permeando estudos ligados à recuperação da informação, desenvolvimento do ambiente *Web*, indexação manual e automática, bem como as ferramentas disponibilizadas pelo Google na indexação, organização e recuperação dos conteúdos.

A análise foi realizada do ponto de vista do usuário, que tem necessidade de recorrer frequentemente a *sites* de museus, procurando informações ou recursos para a realização de trabalhos acadêmicos ou profissionais. O objeto da análise foram os museus nacionais que, devido à sua natureza e às suas coleções, são mais solicitados e dispõem de maior número de recursos técnicos, financeiros e humanos para a concepção e desenvolvimento dos seus relativos *sites*.

Teve-se como hipótese que os museus estejam dando maior ênfase à Internet como meio seguro, fácil, econômico e de grande visibilidade para a divulgação das suas coleções, o que possibilita que este passasse a ser um dos mais importantes meios para difundir e promover as instituições museológicas, o seu acervo, ações e iniciativas culturais, através da disponibilização de um conjunto de informações e recursos.

Desse modo, o objetivo deste trabalho é identificar estratégias e procedimentos que os museus podem adotar ao disponibilizar conteúdos no ambiente *Web* para que sejam, posteriormente, recuperados de forma mais precisa quando buscados através do motor Google. É necessário, entretanto, levar em consideração que os museus, enquanto instituições de importante valor cultural, social e educativa, devem ter como função analisar, guardar e divulgar as diferentes tipologias de documentos, integrando o patrimônio cultural a sociedade e disponibilizando acesso independentemente da localização dos usuários.

2 RECUPERAÇÃO DA INFORMAÇÃO (RI)

Saracevic (1996) aponta que a expressão “recuperação da informação” foi definida por Calvin Mooers como sendo o englobamento dos aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação. Neste contexto, são estudadas técnicas de digitalização de objetos originados de fontes heterogêneas, técnicas de armazenamento, processos de busca, recuperação e apresentação de forma amigável das informações, sendo a indexação, ainda, a principal ferramenta para recuperação (CARDOSO, 2000).

O sistema de recuperação da informação deve responder às demandas dos usuários com tempo de respostas razoáveis, mas para que isso aconteça, é preciso que os documentos sejam submetidos a um tratamento prévio que permita “a extração dos descritores e sua estruturação com vistas a um acesso rápido às informações” (MAIA; SOUZA, 2010, p. 156). De acordo com Teixeira e Schiel (1997) a recuperação da informação compreende basicamente três etapas: indexar, armazenar e recuperar.

Na recuperação da informação, o principal desafio é compreender a necessidade de informação do usuário para atender de forma precisa e rápida, o que foi solicitado. Na RI a unidade básica de acesso à informação é a palavra (conhecida também como palavra-chave, descritor ou termo-chave), pois, a partir desta unidade, diversos modelos podem ser desenvolvidos “com o intuito de facilitar o acesso à informação e melhorar a precisão do resultado de uma consulta ou busca” (BORSATO et al, 2006, p. 1).

Barreto (1996) afirma que as medidas tradicionalmente aceitas na avaliação do rendimento operacional dos estoques informacionais são a *revocação*, que diz respeito à habilidade do estoque de informação em proporcionar itens relevantes para o receptor e a *precisão*, que é definida como a capacidade do estoque de informação em reter itens de informação não relevantes para o receptor. Em outras palavras, a revocação mede a proporção de documentos relevantes recuperados e a precisão mede quantos documentos relevantes foram recuperados

2.1 RECUPERAÇÃO DA INFORMAÇÃO NA INTERNET

Segundo Lancaster (2004), a recuperação da informação na Internet tem recursos informacionais diferentes dos utilizados para os registros bibliográficos. Todavia, há também algumas semelhanças: a rede mundial de computadores utiliza a URL, nome do *site*, quaisquer metadados incluídos nos *sites*, páginas de conteúdo (onde os termos que ali ocorrem podem ser considerados um tanto análogos aos termos de indexação de registros bibliográficos convencionais) e todo o texto encontrado no *site*. Já os registros bibliográficos utilizam título, número de classificação (às vezes), termos de indexação, texto de resumo (embora a quantidade de texto no *site* possa ser consideravelmente

maior). A diferença entre a recuperação na Internet e a realizada na maioria dos registros bibliográficos seria a existência de apontadores (hipertextos) para outros *sites*, nos quais os termos dos vínculos são também pesquisáveis. Neste sentido, um *site* não é uma unidade independente (como um registro bibliográfico comum), mas um nó de uma rede.

De acordo com Mercado (2002), a Internet é um dos elementos da globalização, além de ser um meio de comunicação fácil, barata e difundida, sendo uma ferramenta poderosa que reúne amplo número de informações que não fazem parte de livros, ou seja, é o ponto principal que permite acesso remoto às bases de dados e informações que seriam difíceis de encontrar. Entretanto, assim como a maioria das ferramentas, a Internet também é passível de problemas, e, nesse sentido, Mercado (2002, p. 193) aponta algumas dificuldades em relação à recuperação da informação na Internet:

a) **Informação e conhecimento utilizados como termos análogos** – dispomos de uma grande quantidade de dados e informações. Assim, conhecer é integrar a informação no nosso referencial, no nosso paradigma, apropriando-a, tornando-a significativa para nós. O conhecimento não é efêmero, o conhecimento se cria, se constrói;

b) **Facilidade de dispersão** – é comum que os usuários se percam no emaranhado de possibilidades de navegação, pois facilmente deixam-se arrastar para áreas de interesse pessoal, além de ser fácil perder tempo com informações pouco significativas, ficando na periferia dos assuntos, sem aprofundá-los. Conhecer acontece no momento de filtrar, selecionar, comparar, avaliar, sintetizar, contextualizar o que é mais relevante, significativo;

c) **Perde-se muito tempo na rede** – Dificuldade de se achar respostas corretas e confiáveis em pouco tempo, sem precisar passar horas abrindo todos os *sites* oferecidos pelo serviço de busca;

d) **Impaciência** – acontece principalmente pela necessidade de mudar de um endereço a outro constantemente. Isso pode levar a aprofundar pouco as possibilidades em cada página encontrada. É difícil avaliar rapidamente o valor de cada página, porque há muita semelhança estética na sua apresentação;

e) **Conciliar os diferentes tempos dos usuários** – alguns respondem imediatamente, outros, por fatores como habilidade ou familiaridade com as ferramentas, são mais lentos. Entretanto, a lentidão pode permitir um maior aprofundamento, e nesse sentido, na pesquisa individual esses ritmos diferentes podem ser respeitados.

3 INDEXAÇÃO: CONCEITOS FUNDAMENTAIS

O objetivo principal da indexação é assegurar a recuperação de qualquer documento ou informação no momento em que houver solicitação em um sistema de informações. Para Cintra (1983), a indexação é definida como a tradução de um documento em termos documentários, isto é, em descritores, cabeçalhos de assunto, termos-chave, que tem por função expressar o conteúdo do documento. A indexação assim é definida como uma “tradução lexical” das unidades lexicais da língua em que está escrito o documento, para unidades lexicais de uma linguagem documentária.

A indexação conduz ao registro dos conceitos contidos num documento de uma forma organizada e facilmente acessível, mediante a constituição de instrumentos de pesquisa documental, como índices e catálogos alfabéticos. A informação de um documento é

representada, então, por um conjunto de conceitos ou combinações de conceitos e a linguagem de indexação seria a transformação dos conceitos selecionados em termos ou símbolos para a representação da informação em um determinado sistema. Para Lancaster (2004), a indexação seria um processo subjetivo, sendo possível, portanto, que duas ou mais pessoas acabem divergindo sobre os termos mais adequados para a indexação de determinado documento, sendo também possível que existam pontos de vista divergentes em momentos diferentes.

Conforme Lima (2003), a indexação é o procedimento intelectual que envolve atividades cognitivas a compreensão e a composição da representação da informação, ou seja, a indexação necessita da utilização de uma estrutura de linguagem própria para definir a formação de categorias. Segundo Lancaster (2004), há dimensões da indexação: a *exaustividade*, a *seletiva* e a *especificidade*. A exaustividade é quando se atribui mais termo a indexação, a seletiva é quando menos termos são incluídos e a especificidade se refere ao conceito da segunda dimensão da indexação – quando um documento precisa ser indexado com o termo mais específico que o compreenda integralmente.

Através da indexação é possível se obter um aproveitamento melhor no processo de busca e recuperação da informação, uma vez que o elemento principal/fundamental estabelecido é a representação do conteúdo dos documentos (ARAÚJO JÚNIOR, 2007). Em síntese, indexar seria a substituição do texto de um documento por uma descrição do conteúdo tratado, com o objetivo de que se possa recuperar as informações contidas nessa unidade documental. Segundo Bruzina, Maculan e Lima (2007), existem pelo menos duas formas para se fazer a análise do conteúdo semântico de um documento: indexação manual e indexação automática.

A indexação manual ou controlada seria a seleção cuidadosa da terminologia empregada, realizada por um indexador que escolhe um conjunto de termos para representar o conteúdo de um documento. Na segunda forma de indexação, a chamada automática, cada palavra presente no texto tem a capacidade de estabelecer uma entrada no índice, sendo desempenhada por um sistema computacional.

A indexação automática pode ser dividida em indexação por extração automática e indexação por atribuição automática. A indexação automática por extração é aquela que, por métodos estatísticos, pode extrair palavras com certo grau de ocorrência em um documento. Este processo pode consistir em grande subsídio para um indexador identificar possíveis descritores no texto avaliado (indexação semi-automática), ou anunciar um termo a ser modificado para virar um descritor. A indexação automática por atribuição é a que mais se aproxima ao processo de indexação humana, pois são atribuídos outros conceitos aos termos escolhidos, na maioria das vezes, vinculados a um vocabulário controlado, o que oferece a termos repetidos detectados por um sistema, novos termos relacionados e paralelos, acrescentando novos valores à indexação automática realizada no primeiro momento (LANCASTER, 2004).

4 OS MÉTODOS DE INDEXAÇÃO NA WEB PELOS MECANISMOS DE BUSCA

Os buscadores não podem vasculhar toda a Internet sempre que um usuário realizar uma pesquisa, e desse modo, são criados bancos de dados para cada termo disponibilizado, recuperando as páginas que estão associadas às palavras utilizadas na busca. Periodicamente esses bancos de dados são renovados, eliminando os *sites* inválidos e atualizando as mudanças realizadas.

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

Robredo (2005) indaga sobre a indexação na *Web* e averigua que existem variadas formas de interpretação, ou até mesmo, aplicação, afirmando, ainda, que esse tipo de indexação levará a um entendimento diferenciado dos “processos de catalogação, classificação e indexação, com seus aspectos sintáticos, semânticos e taxonômicos” (ROBREDO, 2005, p. 314). Um ponto principal que pode ser apresentado é o fato que os mecanismos de busca não indexarem os *sites* em si, mas sim os conteúdos das páginas que os compõem. Isso significa que uma página de um determinado *site* pode estar entre os primeiros resultados de uma busca, como pode estar entre os últimos em outra consulta.

Campos e Bax (2000) apresentam os principais fatores analisados pelos mecanismos de busca no momento da indexação de uma página e que devem, portanto, ser considerados no momento de se disponibilizar conteúdos na *Web*. Segundo os autores, o documento deve estar estruturado através das palavras-chave, das *meta tags* e de conhecimentos de HTML. As *meta tags* são linhas de código HTML, ou “etiquetas”, que, entre outras coisas, descrevem o conteúdo de um *site* para os mecanismos de busca. É nelas que se é inserido as palavras-chaves. HTML (*HyperText Markup Language*) – é uma linguagem de marcação utilizada para produzir páginas na *Web*, para serem interpretados por navegadores. É o código usado para transformar qualquer documento comum em documento acessível e disponibilizável à rede mundial de computadores e navegável por meio das ferramentas de busca de conteúdos na Internet. A utilização do HTML faz com que os motores de busca entendam melhor as páginas dos *sites*, sendo capazes de indexar o conteúdo de forma mais eficaz. Os outros tipos de linguagem de programação, tais como Java Script, Flash e arquivos de imagens,

não são identificadas no momento da busca, por isso o HTML é mais indicado na constituição de um *site*, embora isso não exclua a possibilidade de utilização dos outros tipos de linguagens, basta que haja a combinação, tendo como base a estrutura HTML.

O grande problema dos mecanismos de busca com as outras linguagens de programação é que eles não conseguem lidar com os erros de programação, não alcançando entendimento com códigos que contenham falhas. Em outras palavras, uma página que não estiver estruturada com uma configuração precisa, será ignorada pelos motores de busca. Campos e Bax (2000) ressaltam ainda que os processos de indexação aplicados a páginas da *Web*, dependem do preparo prévio dos conteúdos a serem indexados e que essa tarefa deve ser realizada por especialistas.

Os *meta tags* são parâmetros que descrevem uma página para que os buscadores a recuperem. Alguns motores atribuem maior relevância às páginas que contêm os termos de busca nos *meta tags* de descrição ou nos *meta tags* de palavra-chave. Este método pressupõe que os autores utilizarão criteriosamente estes *tags*, o que pode não ocorrer. Alguns autores podem utilizar estes *meta tags* para chamar atenção sobre suas páginas, utilizando termos que não correspondem ao seu conteúdo (CENDÓN, 2001, p. 45). Muitos mecanismos de busca consideram as informações contidas nos *meta tags* tanto no momento que indexam uma determinada página, quanto quando calculam a relevância para efeito de classificação. O HTML admite que sejam especificados metadados em um conteúdo, ou seja, informações a respeito de um documento além do seu conteúdo, por meio do tag <META>. Esta *tag* tem a capacidade de ser empregada para adicionar pares nome/valor que descrevem características do documento, como por exemplo, o autor, uma lista de

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

palavras-chave, etc. Estas informações são invisíveis em um *browser* e, conseqüentemente ao visitante da página. De acordo com Campos e Bax (2000), os principais *meta tags* são *description* e *keywords*.

O *meta tag description* (descrição do conteúdo) é usado para adicionar uma descrição do conteúdo da página dentro do HTML. Esta descrição é utilizada pelos mecanismos de busca que suportam *meta tags* no momento em que eles apresentam a página como fruto de uma pesquisa. Caso a página não tenha o *meta tag description*, os mecanismos empregam as primeiras linhas de documento como uma descrição da página.

O *meta tag keywords* (palavras-chave) é utilizado para apontar as palavras-chave associadas à página. Essas palavras são separadas por vírgulas. Segundo Campos e Bax (2000), esse *meta tag* deve ser explorado para incluir todas as palavras-chave que tenham alguma relação com a página, mesmo as que não aparecem no seu corpo.

4.1 COMO O GOOGLE INDEXA?

O Google é um índice composto por dados extraídos da *Web*, que está continuamente sendo atualizado. Essas atualizações são feitas através do “Googlebot”, uma forma de rastreamento algorítmico: “os programas de computador determinam quais os *Web sites* que devem ser rastreados, com que frequência e quantas páginas de cada *Web site* devem ser analisados” (GOOGLE, 2010).

Após o “rastreo”, o Googlebot processa todas as páginas e as reúne em um gigantesco índice, organizado pelas palavras encontradas com as suas respectivas localizações. Além disso, as informações que compõem o conteúdo são processadas e incluídas nos principais atributos e *tags* de conteúdo, como *tags Title* e atributos ALT. O atributo ALT é

definido, para um conjunto de *tags*, como algo que tem a finalidade de permitir ao autor da página fornecer um texto equivalente para o objeto incluso, como por exemplo, a descrição de uma imagem.

O Googlebot, apesar da eficiência no rastreamento e processamento dos conteúdos inclusos nas páginas, não pode processar todos os tipos de conteúdos, como por exemplo, as páginas que contém arquivos de mídias elaboradas ou páginas dinâmicas.

Entretanto, o grande negócio do Google consiste em um mecanismo chamado de *PageRank*, uma fórmula matemática desenvolvida na Universidade de Stanford, nos Estados Unidos, que rastreia os *sites* em busca dos link gerados. O funcionamento ocorre da seguinte forma: o usuário realiza uma pesquisa na página principal do Google e as máquinas fazem uma busca no índice das páginas que correspondem e retornam os resultados que “parecem” ser os mais relevantes. Essa relevância é julgada por mais de duzentos fatores, mas é o *PageRank* que ministra toda a protuberância dos resultados. Ou seja, para cada link gerado de uma página, em outro *site* é adicionado um *PageRank* ao *site* “linkado”, mas nem todos os *links* são iguais, o sistema identifica *spams* e outras ameaças ao resultado da pesquisa (GOOGLE Central do Webmaster, 2010).

O *PageRank* seria o modo que o Google utiliza para ordenar a importância que cada página tem na Internet. Em outras palavras, seria a contagem de votos para cada espaço gerado no banco de páginas do *PageRank*. O Google utiliza também os denominados *spiders* ou *crawlers*, que indexam a informação coletada nos servidores centrais localizados no Googleplex em Mountain View, Califórnia. Os métodos utilizados pelo Google giram em torno da identificação exata de textos para encontrar páginas que sejam tanto importantes como relevantes para a busca, Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

recuperando apenas as páginas que incluem todos os termos utilizados na busca. Por exemplo, quando uma página é analisada, verifica-se o que as outras páginas que estão vinculadas à esta tem a dizer sobre ela, preferindo as páginas nas quais os termos de busca estejam próximos entre si. O Google também ignora palavras e caracteres comuns, conhecidos como palavras descartáveis, desconsiderando automaticamente termos como “*http*” e “.*com*”, assim como ignora também dígitos ou letras isoladas, pois raramente ajudam na busca e podem torná-la consideravelmente mais lenta. Segundo a empresa, refinar ou estreitar a pesquisa consiste no simples fato de se adicionar ou não palavras aos termos que já foram incorporados a busca. “Ao adicionar mais palavras, seus resultados conterão um subconjunto específico das páginas retornadas pela sua consulta original” (GOOGLE Central de Ajuda do Google, 2010). Tudo isso melhora o resultado da pesquisa que em média, dependendo dos fatores de conexão, dura algo em torno de 5 segundos (MONTEIRO, 2008, p. 119).

4.2 MÉTODOS DE RECUPERAÇÃO NO GOOGLE

O Google é um dos *sites* de busca mais utilizados em toda Internet, e sua apresentação chama a atenção por sua simplicidade e por dispor de várias ferramentas para que haja uma resposta mais clara da pesquisa solicitada. Alguns pontos devem ser esclarecidos com relação à busca que esse motor realiza:

- O Google pesquisa apenas páginas que correspondem exatamente ao texto escrito, desconsiderando acentos gráficos;
- Não são diferenciadas as letras maiúsculas e minúsculas: a pesquisa é feita independente da forma que está escrita, ou seja, considera-se todas as letras em minúsculo. Por exemplo, a expressão “Competência Informacional”, ser for

escrita “COMPETÊNCIA INFORMACIONAL”, “cOMPETÊNCIA INFORMACIONAL” ou “CoMpEtÊnCia InFoRmAclONaL” será considerada como “competencia informacional”.

O Google dispõe de vários recursos para que haja um refinamento na recuperação da informação, entre eles a busca pela frase exata, onde os termos são colocados entre aspas ou ligados por hífen entre as palavras. Com o buscador Google também é possível expandir uma pesquisa, incluindo ou retirando termos da recuperação. Também são ignoradas as *stop words* que são palavras comuns, preposições, artigos, advérbios (ex: *onde, o(s), a(s), como*) e outros dígitos e letras que diminuem a velocidade da pesquisa. **Stop words** (ou palavras de parada) são palavras que podem ser consideradas irrelevantes para o conjunto de resultados a ser exibido em uma busca realizada em um motor de busca. Se uma palavra comum for essencial para obter os resultados que se deseja, será possível incluí-la inserindo um sinal de "+" antes do termo escolhido. Além do mais, cada termo separadamente (se não estiver entre aspas) e ao se colocar o sinal de "+" na frente de uma palavra, será feita a pesquisa da forma que a palavra está escrita sem variações. Na pesquisa no Google, também é possível a busca de um ou outro termo; a exclusão de uma palavra; a busca por domínio e no título da página; a pesquisa por sinônimos ou palavras similares, por tipos específicos de arquivos e em um site característico, além de operações matemáticas.

5 ANÁLISE DOS SITES

Após apresentar e discutir questões teóricas, tais como: o surgimento da *Web* e da Internet, a utilização da *Web* como fonte de informação, como os mecanismos de busca indexam, assim como, a história e construção da empresa Google, será apresentada a *Biblionline*, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

análise realizada no buscador Google, onde foram avaliados *sites*, buscando analisar se estes adotaram estratégias que proporcionaram melhor recuperação por buscadores na *Web*.

5.1 PROCEDIMENTOS METODOLÓGICOS

A análise dos *sites* dos museus tem como objetivo identificar a melhor maneira para se disponibilizar conteúdos, considerando que possam ser indexados de modo mais eficaz e conseqüentemente, melhor recuperados pelo motor de busca Google. Sabe-se que os buscadores indexam isoladamente cada página que constitui os *sites*, entretanto, as *home pages* (Página de entrada ou página principal de um *site*. É nesta página que estão os links para as demais páginas do *site*.) de cada *site* foram analisadas como um todo.

Esta pesquisa se configura como de caráter exploratório, uma vez que estas “[...] tem como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores” (GIL, 2009, p. 27). Nesse sentido, o objetivo deste trabalho é analisar *sites* de museus e apontar meios para que sejam melhor indexados pelo motor de busca Google, dando-lhes maior visibilidade.

Para que seja alcançado o objetivo, foi realizado um levantamento teórico, tendo como base assuntos relacionados a fontes de informações na internet, diretórios e motores de buscas, indexação, indexação no ambiente *Web*, com a finalidade de se fazer uma análise sobre o tema em questão. Esse levantamento teórico possibilitou a enumeração de critérios e dos processos utilizados na indexação realizada pelo Google. Ainda de acordo com Gil (2009), o delineamento aplicado neste

trabalho é do tipo levantamento, uma vez em que não serão analisados todos os *sites* de museus recuperados pelo Google, mas, selecionada uma amostra como objeto de investigação.

Gil (2009, p. 90), define universo ou população, na pesquisa social, como sendo “um conjunto definido de elementos que possuem determinadas características” e Amostra como sendo “um subconjunto do universo ou da população, por meio do qual se estabelecem ou se estimam as características desse universo ou população” (GIL, 2009, p. 92). Dessa forma, o tipo de amostra que será utilizada neste trabalho é a por conglomerados, uma vez que “é a mais indicada em situações em que é bastante difícil a identificação de seus elementos” (GIL, 2009, p. 93), em outras palavras, esse tipo de amostra é o mais indicado para este trabalho, pelo fato da existência de inúmeros *sites* que compõem a Internet e por isso só será utilizada uma pequena parte desse universo.

A amostra foi designada da seguinte forma: buscou-se no Google o termo “*museu*” e foram encontrados aproximadamente 45.300.000 páginas (busca realizada no dia 29 de maio de 2011, às 15h35min.). Para limitar a amostra, foram considerados os cinco primeiros resultados, eliminando-se as páginas que tratavam de definições da palavra, como por exemplo, páginas da Wikipédia (a **Wikipédia** é uma enciclopédia multilíngue *online* livre colaborativa, ou seja, escrita internacionalmente por várias pessoas comuns de diversas regiões do mundo, todas elas voluntárias.), páginas de *blogs* (com domínios do tipo *blogspot*), páginas de faculdades e universidades e páginas do *Twitter* - (**Twitter** é uma rede social e servidor para *microblogging* que permite aos usuários enviar e receber atualizações pessoais de outros contatos (em textos de até 140 caracteres, conhecidos como “tweets”), por meio do website do serviço, por SMS e por

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

softwares específicos de gerenciamento), o que resultou em três *sites* para serem avaliados. Feito isso, os *sites* foram classificados de A a C, de acordo com a sequência do resultado, chegando-se a seguinte ordem:

- Site A – Museu Imperial (<http://museuimperial.gov.br/>);
- Site B – Museu Nacional UFRJ (<http://museunacional.ufrj.br/>);
- Site C - Museu de Arte Moderna de São Paulo (<http://mam.org.br/>).

Os *sites* foram analisados a partir dos campos da *tag title* e das *Meta Tags Description* e *Keywords*, visíveis apenas no “Código Fonte” de cada página. Para exibição do Código fonte (Apresenta as informações da página na codificação de HTML), foi necessário seguir os seguintes passos: um clique com o botão direito do *mouse* em uma área da página que não contivesse *links*, imagens, gráficos e etc.; e escolha o subitem “Código fonte”; ou ainda, através da barra de navegação, na opção “Exibir” no subitem “Código fonte”; ou através das teclas de atalho “Ctrl+U”. Seguindo qualquer um desses três passos, o código fonte será exibido no programa “bloco de notas”. O **Bloco de notas** é um editor de textos básico que pode ser utilizado para criar documentos simples. O uso mais comum do Bloco de notas é exibir ou editar arquivos de texto (.txt), mas muitos usuários o consideram uma ferramenta simples para criar páginas da *Web*.

Ao se identificar a *tag title* e as *meta tags description* e *keywords* foi feito um diagnóstico do conteúdo de cada uma, em relação às técnicas de otimização de busca, de modo que o conteúdo da *tag title* foi comparado com o título apresentado na *home page* do *site*; na *meta tag description* buscou-se perceber se o *site* está sendo descrito através de um breve resumo do seu conteúdo com clareza; assim, as palavras-chave

encontradas na *meta tag keywords* foram testadas no Google, de forma a saber em qual posição o *site* apareceu com a utilização de cada termo; já nas *home pages* que não possuíam palavras-chave na *meta tag keyword*, foram selecionadas alguns possíveis termos que poderiam ser usados na recuperação.

Os critérios usados para a escolha dos termos sugeridos foram básicos, utilizando-se de siglas, termos contidos no título da página, possíveis expressões do conteúdo e etc. Cada palavra-chave estava associada ao teor das páginas, de forma que cada *home page* fosse identificada claramente com o termo buscado.

Tanto as palavras contidas na *meta tag keywords* quanto às sugestões, foram

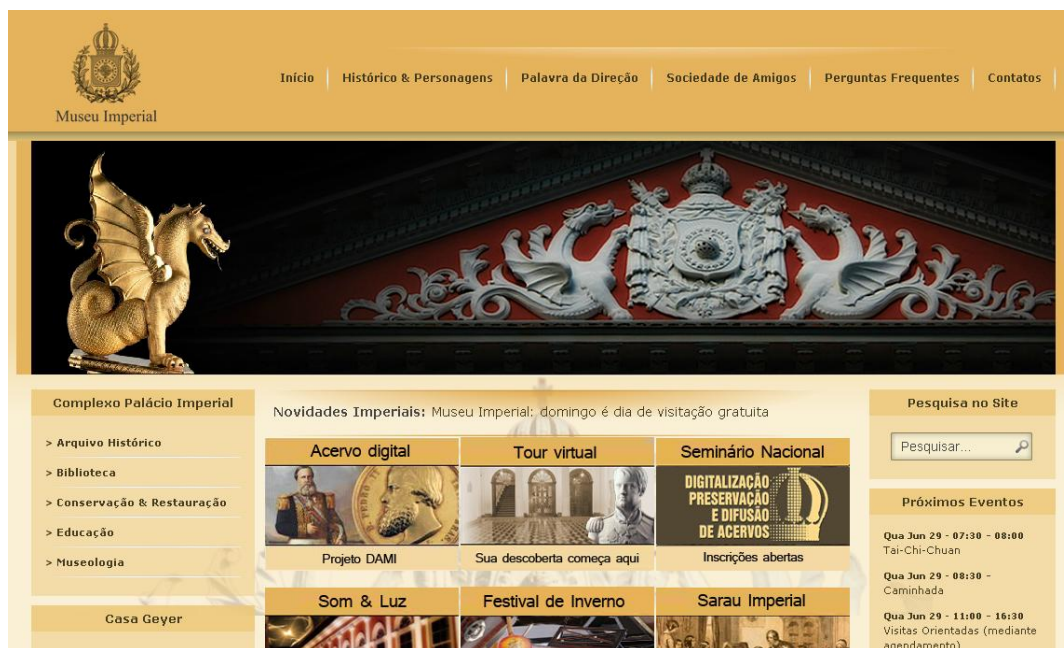
colocadas numa tabela com a sua respectiva colocação na recuperação no Google, considerando até os 50 primeiros resultados.

Por fim, foram relatadas as observações feitas de cada *site* através de um quadro comparativo, onde todos os critérios foram julgados de acordo com três classificações: Consta de forma adequada, Consta de forma inadequada e Não consta.

- **Site A – MUSEU IMPERIAL**

De acordo com o *site*, o Museu Imperial é um museu histórico-temático, localizado no centro histórico da cidade de Petrópolis, no estado do Rio de Janeiro, Brasil, instalado no antigo Palácio de Verão de Dom Pedro II.

Figura 1 - Imagem extraída da *home page* do museu imperial.



Fonte: <http://museuimperial.gov.br/>

➤ **Tag title**

O *site* utiliza o seguinte texto no título assim que a página é carregada: “Museu Imperial” e no Código Fonte também faz uso da mesma

expressão. Conforme já mencionamos, Campos e Bax (2000) afirmam que os mecanismos de busca consideram o título o local mais importante para determinar o assunto tratado, logo, o título considerado será o que estiver no código fonte.

➤ *Meta tag description*

A *meta tag description* não foi preenchida, ou seja, não foi utilizada a funcionalidade da *meta tag*, apenas foi colocado a expressão “Museu Imperial” (como pode ser observado na Figura 2). Essa *meta tag* é avaliada como importante, uma vez que, é usada pela maioria dos motores de busca para oferecer um pequeno conteúdo sobre a página.

Figura 2 - Código fonte - museu imperial.

```
<link href="favicon.ico" rel="shortcut icon" type="image/x-icon" />
<META NAME="KEYWORDS" CONTENT=" ">
<META NAME="TITLE" CONTENT="Museu Imperial">
<META NAME="DESCRIPTION" CONTENT="Museu Imperial">
<META NAME="LANGUAGE" CONTENT="pt">
<META NAME="DOCUMENTCOUNTRYCODE" CONTENT="br">
</head>
```

➤ *Meta tag keyword*

O site do Museu Imperial não utiliza a *meta tag keywords*, o que não é recomendável, uma vez que os motores de busca consideram as *meta tags* para indexar conteúdos de uma página. Segundo Engelmann (2006, p. 71), a prática da não utilização das *meta tags*, por intencionalidade ou descuido,

deve influenciar de forma negativa a otimização dos resultados de busca, por

dois motivos. Primeiro porque os *spiders* não terão essas *meta tags* como referência para fazer suas buscas e, segundo, porque a equipe que atualiza o site não tem referência dos termos que precisam receber ênfase na construção de textos, títulos e outras etapas já citadas, as quais são relevantes para otimização.

Foram feitos alguns testes no buscador Google, com palavras que poderiam ser consideradas “chave” para o site do Museu Imperial, (Quadro 1):

Quadro 1 - Palavras-chave relacionadas ao museu imperial.

PALAVRAS-CHAVE	POSIÇÃO DOS TERMOS TESTADOS COMO PALAVRAS-CHAVES SUGERIDAS
Museu	2º lugar
Museu imperial	1º lugar
Museu Petrópolis	1º lugar

• *Site B – MUSEU NACIONAL UFRJ*

De acordo com o site do Museu Nacional UFRJ, a instituição está vinculada ao Ministério da Educação, sendo a mais antiga instituição científica do Brasil e o maior museu de história natural e antropológica da

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

América Latina. Criado por D. João VI, em 06 de junho de 1818 e, inicialmente, sediado no Campo de Sant'Anna, serviu para atender aos interesses de promoção do progresso cultural e econômico no país. Originalmente denominado de Museu Real, foi incorporado à Universidade do Brasil em 1946.

Atualmente o Museu integra a estrutura acadêmica da Universidade Federal do Rio de Janeiro.

Figura 3 - Imagem extraída da *home page* do museu nacional UFRJ.



Fonte: <http://museunacional.ufrj.br/>

➤ **Tag title**

A *tag title* apresenta o mesmo título na *home page* do site do Museu Nacional UFRJ. Assim, percebe-se que o título apresenta relevância ao conteúdo da página.

➤ **Meta tags description e keywords**

As *meta tags description* e *keywords* não fazem parte do código fonte da página; sugerimos o uso das duas *meta tags*, uma vez que ambas são consideradas pelos motores de busca na hora da indexação. O Quadro 2 apresenta sugestões de termos que poderiam ser usados na composição da página e suas respectivas posições no buscador Google.

Quadro 2 -Palavras-chave relacionadas ao museu nacional UFRJ

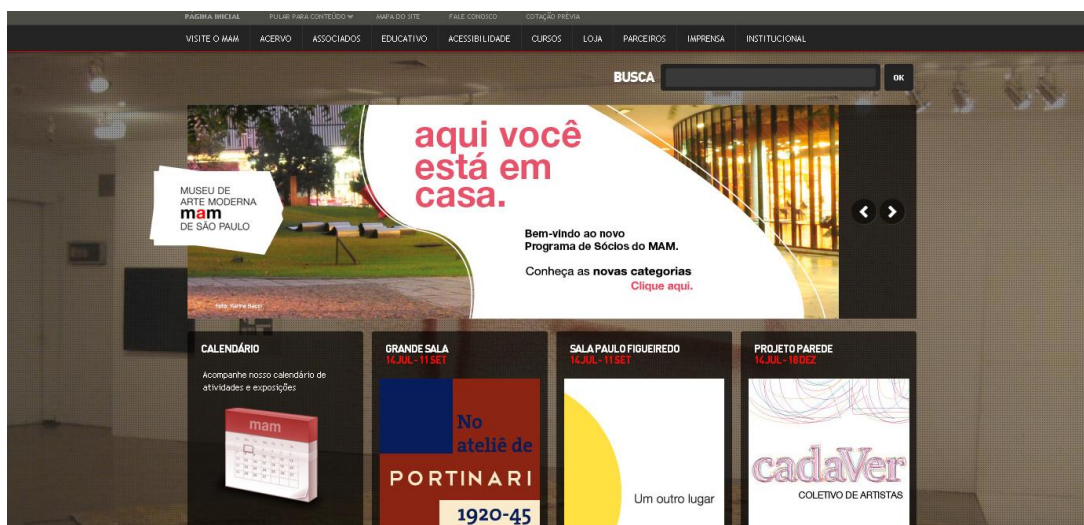
PALAVRAS-CHAVE	POSIÇÃO DOS TERMOS TESTADOS COMO PALAVRAS-CHAVES SUGERIDAS
Museu nacional	1º lugar
Museu UFRJ	1º lugar
Acervo museu	9º lugar

- **Site C – MUSEU DE ARTE MODERNA DE SÃO PAULO**

O Museu de Arte Moderna de São Paulo MAM, de acordo com o *site*, é uma sociedade civil de interesse público, sem fins lucrativos,

e é um dos mais importantes museus da América Latina. Fazem parte de sua coleção mais de 5 mil obras, reunindo mais de mil artistas entre os mais expressivos da arte moderna e contemporânea brasileira.

Figura 4 - Imagem extraída da *home page* do museu de arte moderna de São Paulo.



Fonte: <http://mam.org.br/>

➤ **Tag title**

O título apresentado na *tag title* e na *home page* são iguais, favorecendo a recuperação e gerando relevância ao conteúdo à *home page*.

➤ **Meta tags description e keywords**

Percebe-se que o *site* não apresenta descrição para explicar o conteúdo da página,

mas algumas informações poderiam se acrescentadas, tais como: a temática do museu, o tipo de acervo, há quanto tempo de atende à sociedade etc. A *home page* também não escolheu palavras para serem associadas *meta tag keywords*, dessa maneira, foram testadas algumas palavras que poderiam compor o código fonte (Quadro 3).

Quadro 3 - Palavras-chave relacionadas ao museu de arte moderna de São Paulo.

PALAVRAS-CHAVE	POSIÇÃO DOS TERMOS TESTADOS COMO PALAVRAS-CHAVES SUGERIDAS
Mam	1º lugar
Museu de arte moderna	1º lugar
Museu São Paulo	11º lugar

5.2 Análise dos dados

Após a análise das três *home pages* foi possível elaborar um quadro explicativo para apresentar, em síntese, os critérios utilizados pelo motor de busca Google considerados no

momento da indexação. Percebeu-se que a *tag title* e as *meta tags description e keywords* constam (de forma adequada ou inadequada) ou não na constituição das *home pages* (Quadro 4).

Quadro 4 - Quadro explicativo – análise dos sites.

<i>Site</i>	Crítérios	Tag title	Meta tags	
			<i>Description</i>	<i>Keywords</i>
A		Consta de forma adequada	Não consta	Não consta
B		Consta de forma adequada	Não consta	Não consta
C		Consta de forma adequada	Não consta	Não consta

O Quadro 4 mostra que a *tag title*, em todos os *sites* analisados, está composta de forma adequada. O alto índice de composição da *tag title*, possivelmente, se deve ao fato da relevância que a *tag* proporciona no momento da indexação feita pelo Google. Segundo Costa (2007) a *tag title* é de extrema importância para os motores de busca e que deve ser, preferencialmente, constituída de forma mais elucidativa possível.

A *meta tag description* não constou nos três *sites* estudados. Não se pode entender de fato a razão para essa totalidade, uma vez que, segundo Miygusku (2007), a *meta tag* descrição é a mais externa, o que delimita a definição do conteúdo da página. Um dos fatores que certamente contribuem para o uso negativo é a falta de conhecimento da funcionalidade existente das *meta tags*.

As palavras-chave também não constaram nos *sites* avaliados. Tal situação gera irrelevância para os motores de busca, no caso deste trabalho, para o motor de busca Google.

6 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi realizar um estudo sobre as estratégias que instituições como os museus podem utilizar ao disponibilizar conteúdos no ambiente *Web*, a fim de que sejam indexados pelo Google e, posteriormente, recuperados de forma mais precisa quando procurados. Assim, os *sites* analisados foram aqueles focados especificamente em museus nacionais.

Após a revisão teórica foi possível identificar que não houve a devida preocupação em estruturar os conteúdos seguindo os critérios estabelecidos por mecanismos de busca como o Google usa para indexar páginas. Para se chegar a essa conclusão, analisou-se três itens fundamentais na constituição de uma página na *Web*: a *tag title* e as *meta tags description* e *keywords*. A *tag title* apresentou um alto índice de relevância na categoria “Consta de forma adequada”. Possivelmente o indicador se deve ao fato de que no momento da criação da página considera-se que os usuários atribuirão um título ao conteúdo disponibilizado, ou seja, que durante a busca o título seja mais evidente a partir dos termos utilizados pelos consulentes.

Sugere-se, então, que os *sites* analisados sejam reestruturados, de modo que as configurações estabelecidas sejam adequadas e mais ‘visíveis’ pelo Google quando o conteúdo da página for cabível à pesquisa realizada. Para tanto, é necessário um reforço maior no que diz respeito à constituição do código HTML da página, para que seja favorável à recuperação.

Portanto, concluímos que o campo de estudo da indexação feita pelos motores de busca para posterior recuperação é ainda incipiente, e que novos estudos são necessários. Partindo dessa premissa, as reflexões aqui tratadas têm como finalidade ampliar a discussão sobre a necessidade de haver preocupação com o modo como as informações estão

sendo disponibilizadas no ambiente *Web*, além de sugerir novos estudos nessa perspectiva, a fim de que seja mais trabalhada a natureza da indexação automática. Assim, podemos apontar ainda o caráter interdisciplinar da Ciência da Informação, agregada ao fato das tecnologias estarem cada vez mais presentes em seus processos, o que torna evidente que a Ciência da Computação contribui muito para o desenvolvimento da área, e, portanto, deve ser um dos assuntos na pauta da Ciência da Informação na era da Sociedade da Informação.

REFERÊNCIAS

- ARAÚJO JÚNIOR, R. H. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007. 175 p.
- BARRETO, A. A. A eficiência técnica e econômica e a viabilidade de produtos e serviços de informação. **Ciência da Informação**, v. 25, n. 3, 1996. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/download/466/425>>. Acesso em: 05 jun. 2011.
- BIBLIOTECA Virtual do Governo do Estado de São Paulo. **História da biblioteca e do bibliotecário no mundo e no Brasil**. São Paulo: Biblioteca Virtual de São Paulo, 2007. Disponível em: <<http://www.bibliotecavirtual.sp.gov.br/especial/docs/200703hIstoriadabiblioteca.pdf>>. Acessado em: 30 abr. 2011.
- BORSATO, G. G. et al. Recuperação de informação em situações de urgência-emergência no atendimento pré-hospitalar. **Revista Eletrônica de Sistemas de Informação**, v. 5, n. 3, 2006. Disponível em: <<http://revistas.facecla.com.br/index.php/reinfo/article/view/549/434>>. Acesso em: 13 jun. 2011.
- BRAND, G. B; FARIA, T. G. GOGETIT: Um método de indexação *Web* baseado em redes peer-to-peer. In: WORKSHOP DE REDES PEER-TO-PEER, 2., 2006, Curitiba. **Anais eletrônicos...** Curitiba: BDBComp, 2006. Disponível em: <http://www.lbd.dcc.ufmg.br:8080/colecoes/wp2p/2006/st1_2.pdf>. Acesso em: 13 de maio 2011.
- BRUZINGA, G. S.; MACULA, B. C. M. S.; LIMA, G. A. B. O. Indexação automática e semântica: estudo da análise do conteúdo de teses e dissertações. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 8., 2007, Salvador. **Anais eletrônicos...** Salvador: ENANCIB, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--117.pdf>>. Acesso em: 14 jun. 2011.
- CAMPOS, F; BAX, M. P. Como os mecanismos de busca da web indexam páginas HTML. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 2000, Rio Grande do Sul. **Anais eletrônicos...** Rio Grande do Sul: FEBAB, 2000. Disponível em: <<http://www.bax.com.br/research/publications/ComoMaquinasBuscaIndexamPaginasWeb.pdf>>. Acesso em: 17 jun. 2011.
- CARDOSO, O. N. P. Recuperação de informação. **Inforcomp**, v. 2, n. 1, p. 33-38, 2000. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v2.1/art07.pdf>>. Acesso em: 13 jun. 2011.
- CASTRO, A. indexação manual versus indexação automática: estudo comparativo preliminar da eficiência de recuperação da informação na internet. **Usina de Letras**, 2001. Disponível em: <http://www.usinadeletras.com.br/exibelotexto.php?cod=172&cat=Teses_Monologos&vinda=S>. Acesso em: 14 jun. 2011.
- CENDÓN, B. V. Ferramentas de busca na *Web*. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 39-49, 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a06v30n1.pdf>>. Acesso em: 20 jun. 2011.
- CINTRA, A. M. M. Elementos de lingüística para estudos de indexação. **Ciência da Informação**, Brasília, v. 12, n. 1, p. 5-22, 1983. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/1526/1144>>. Acesso em: 13 jun. 2011.
- COSTA, C. J. **Desenvolvimento para web**. Lisboa, PT: Lusocredito, 2007. 660 p.
- Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

ENGELMANN, F. **Web marketing no webdesign: a influência do marketing de otimização de busca (mob) na comunicação digital.** 96 f. Trabalho de Conclusão de Curso (Curso de Comunicação Social – Habilitação Publicidade e Propaganda)–Instituto de Ciências Sociais Aplicadas, Centro Universitário Feevale, Novo Hamburgo, 2006. Disponível em: <<http://ged.feevale.br/bibvirtual/Monografia/MonografiaFelipeEngelmann.pdf>>. Acesso em: 29 jun. 2011.

GIL, A. C. **Métodos e técnicas de pesquisa social.** São Paulo: Atlas, 2009.

GOOGLE Central do Webmaster. Ajuda das Ferramentas do Google para Webmasters. Disponível em: <<http://www.google.com/support/Webmasters/bin/anSwear.py?answer=70897&hl=pt-BR#2>>. Acesso em: 13 maio 2010.

GOOGLE Informações corporativas. Visão geral da empresa. Disponível em: <<http://www.google.com.br/corporate/>>. Acesso em: 13 maio 2011.

LANCASTER, F. W. **Indexação e resumos: teoria e prática.** 2. ed. Brasília: Briquet de Lemos, 2004. 452 p.

LIMA, G. A. B. Interfaces entre ciência da informação e ciência cognitiva. **Ciência da Informação**, Brasília, v. 32, n. 1, p. 77-87, 2003.

MAIA, L. C.; SOUZA, R. R. Uso de sintagmas nominais na classificação automática de documentos eletrônicos. **Perspectivas em Ciência da Informação**, v. 15, n. 1, p. 154-172, jan./abr. 2010. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/875/717>>. Acesso em: 10 jun. 2011.

MAIMONE, G. D.; SILVEIRA, N. C. Cognição humana e os Paradigmas da Ciência da Informação. **Revista Eletrônica Informação e Cognição**, v. 6, n. 1, p. 55-67, 2007. Disponível em: <http://74.125.155.132/scholar?q=cache:wbSAYzFzQ4J:scholar.google.com/+explos%C3%A3o+documental&hl=pt-BR&as_sdt=2000>. Acesso em: 16 jun. 2011.

Biblionline, João Pessoa, v. 8, n. 1, p. 42-59, 2012.

MERCADO, L. P. L. A Internet como ambiente de pesquisa na escola. In: MERCADO, L. P. L. (Org.). **Novas tecnologias na educação: reflexões sobre a prática.** Maceió: EDUFAL, 2002. p. 191-204.

MIYAGUSKU, R. **Crie sites arrasadores.** São Paulo: Digerati Books, 2007. 128 p.

MONTEIRO, S. D. Os mecanismo de busca: a guisa de uma tipologia das múltiplas sintaxes. In: TOMAÉL, M. I. (Org.). **Fontes de informação na Internet.** Londrina: EDUEL, 2008. p. 97-122.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas.** 4. ed. Brasília: Edição do autor, 2005. 409 p.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectiva em Ciências da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/235/22>>. Acesso em: 13 jun. 2011.

TEIXEIRA, C. M. S; SCHIEL, U. A Internet e seu impacto nos processos de recuperação da informação. **Ciência da Informação**, Brasília, v. 26, n. 1, jan/abr. 1997. Disponível em: <http://www.scielo.br/scielo.php?pid=S010019651997000100009&script=sci_arttext&tlng=en>. Acesso em: 06 jun. 2011.

Dados sobre Autoria

*Graduada em Biblioteconomia pela Universidade Federal de Pernambuco. Mestranda em Ciência da Informação pela Universidade Federal de Pernambuco. Bolsista da FACEPE - Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco. E-mail: cmsholanda@gmail.com

**Mestranda do Programa de Pós-Graduação em Ciência da informação pela Universidade Federal de Pernambuco. Bacharel em Biblioteconomia pela mesma universidade. E-mail: marcia_ibraz@hotmail.com

Artigo enviado em dezembro de 2011 e aceito em janeiro de 2012.