

MODELO PARA DIVULGAÇÃO E GESTÃO DE OBJETOS ARQUITETÔNICOS

Christiano Benvindo dos Santos¹
Renata Maria Abrantes Baracho²

RESUMO:

A presente pesquisa faz um panorama geral sobre gestão da informação de objetos arquitetônicos em diferentes órgãos. Em alguns órgãos existe muita informação sobre um determinado objeto arquitetônico, mas sempre fragmentada e sem normalização, tornando necessária a busca em diferentes caminhos e lugares. Por outro lado, alguns objetos possuem pouca ou nenhuma informação divulgada e de fácil acesso. Dentro desse contexto a pesquisa responde a perguntas de como obter informações referentes a objetos arquitetônicos de forma completa e com inferências. Inferências, por exemplo, de localizar todas as obras de um determinado autor em uma determinada região ou época. Apresenta um modelo para a representação, organização e obtenção de informações acerca de objetos arquitetônicos edificados e a inserção em base de dados para gerenciamento. Para tal foi desenvolvido um protótipo de sistema de informação que gera as páginas referentes à informação dos objetos arquitetônicos de forma automática. As informações inseridas são em formato de textos e imagens. As informações são georeferenciadas com geração de mapas temáticos para localização dos objetos. Os documentos são indexados e a recuperação considera a combinação e ponderação dos termos utilizados.

Palavras-chave: Gestão da informação. Indexação automática. Recuperação da informação.

ABSTRACT

This research provides an overview of management information about architectural objects through different local, state and federal agencies responsible for the information. In some of these local, state and federal agencies there is a lot of information about a particular architectural object, but that information is always fragmented and without standardization, making it necessary to search through different ways and places. On the other hand, some objects have little or no information disseminated and accessible. Within this context the research answers questions of how to obtain information related to architectural objects comprehensively and inferences. These inferences can be, for example, find all the works of an author in a particular region or period. This research presents a model for the representation, organization and obtaining information about built architectural objects and insertion into the database to assist the management process. We developed a prototype information system that generates pages for the information of architectural objects automatically. The information entered are in form of texts and images. The information is georeferenced with thematic maps for location of objects. The documents are indexed and recovery considers the combination and weighting of terms used.

¹ Mestrando em Ciência da Informação pela Escola de Ciência da Informação. Universidade Federal de Minas Gerais. E-mail: christianobsantos@eci.ufmg.br

² Professora do Departamento de Teoria e Gestão da Informação. Escola de Ciência da Informação. Universidade Federal de Minas Gerais. E-mail: renatabaracho@eci.ufmg.br

INTRODUÇÃO

Os problemas de recuperação e gerenciamento da informação suscitam estudos em diversas áreas. Na atualidade as organizações geram e consomem informação a uma velocidade nunca vista antes, e o que pode parecer em um primeiro momento como algo a ser louvado pode se transformar em um problema. Afinal como gerenciar a informação que deve ser utilizada por gestores em áreas distantes, como controlar essa informação de forma descentralizada e como permitir o acesso através de ferramentas de buscas que permitam ao usuário encontrar a informação que lhe interessa dentre tantas outras.

Atualmente tem-se um crescente interesse pelas informações referentes aos objetos arquitetônicos que apresentam de forma fragmentada, descentralização e sem padronização, comprometendo a recuperação e ampla disseminação.

Para conseguir uma boa recuperação é necessário focar na representação da informação. Essa pesquisa visa à elaboração de um modelo para a recuperação e o gerenciamento de informação a ser disponibilizada ao público e aos gestores acerca de objetos arquitetônicos edificados. Para desenvolvimento dessa pesquisa foi realizada coleta de dados em diferentes órgãos responsáveis, aquisição de imagens proprietárias e considerados padrões de metadados para representação da informação. Para atingir o objetivo de se alcançar o maior número de pessoas possível desenvolveu-se um sistema de recuperação de informação que utiliza a web, utilizando conceitos de georeferenciamento, de forma a potencializar o de alcance.

DESCRIÇÃO DO PROBLEMA

Os objetos arquitetônicos edificados (OAE), tais como casas, igrejas, palácios, palacetes, construções militares, mausoléus, dentre outros, vem com o passar do tempo, recebendo mais atenção. Seja por sua história, arte ou pelo que representa para a comunidade local, é visível que o número destes objetos que recebem proteção em âmbito federal, estadual e municipal vem aumentando.

Quando tais objetos são tombados as informações ficam nos órgãos responsáveis pelo tombamento, o qual pode ser realizado em uma das esferas federal, estadual, municipal ou em todas, o que pulveriza ainda mais os locais onde podem ser armazenadas, localizadas e recuperadas as informações para pesquisa.

Existem também diversos OAE que não são tombados, mas que podem possuir alto grau de relevância, tanto quanto qualquer outro nessa condição e, portanto não estão cadastrados em nenhum órgão. Com isso ampliamos ainda mais o número de tais objetos que representam interesse à população. Eles podem estar em várias cidades de qualquer estado do país. Supondo que em cada instância possa haver uma iniciativa para documentar tais objetos teremos no mínimo uma centena de lugares para visitar em busca de informação.

É comum em certas localidades não existir toda a infraestrutura necessária para manter, por exemplo, um bom programa de gestão e divulgação em funcionamento. Por vezes é comum vermos documentações deixarem de ser disponibilizadas não devido ao conteúdo sensível da mesma, mas por falta de ferramentas e treinamentos necessários. Vemos Brasil afora iniciativas muito interessantes que visam projetar o patrimônio para outras localidades de forma a levar maior visibilidade à sua região. Como são iniciativas dispersas o fato de não haver um ponto centralizador ainda causa alguma dificuldade ao usuário.

Quanto aos sistemas utilizados para disseminação das informações apresentam uma estrutura rígida de busca, onde por várias vezes o usuário pode apenas baixar um arquivo pdf. Apesar da qualidade do conteúdo esse método não facilita o gerenciamento e torna mais complexa a busca do usuário devido ao fato do conteúdo dos arquivos não serem indexados e assim não serem considerados na busca.

Existem ainda sites temáticos que visam fornecer a maior quantidade de informação possível acerca de um local determinado, no entanto são sites complexos para atualização e, portanto necessitam de mão de obra especializada para atualização do mesmo.

De tudo isso emerge o problema sobre como gerenciar e recuperar a informação acerca dos objetos arquitetônicos edificados presentes nas mais diversas localidades. É necessário criar uma metodologia para levantamento da informação, sua representação e inserção em uma base e sua indexação de forma simples sem a necessidade de grandes conhecimentos em tecnologia por parte do usuário bem como a recuperação de forma simples e eficaz com o intuito de agregar valor à mesma.

GESTÃO DA INFORMAÇÃO

Defini-se Gestão da Informação como a aplicação de princípios administrativos à aquisição, organização controle, disseminação e uso da informação para a operacionalização efetiva de organizações de todos os tipos (WILSON, 1997 on-line)

Segundo Davenport (1998, p.170) trata-se de um conjunto estruturado de atividades que incluem o modo como as empresas obtêm, distribuem e usam a informação e o conhecimento.

Segundo Lima-Marques (2011, p.241): No contexto organizacional, a geração e troca de informações e conhecimentos tornaram-se decisivas, e a capacidade de gestão de recursos informacionais passou a ser um diferencial (...).

METADADOS

Vários são os padrões de metadados existentes, sendo que para cada tarefa haverá um padrão que melhor irá se adequar a tarefa. Alvarenga (2001, on-line) diz que os metadados servem como chaves de acesso a serem utilizadas pelos internautas. Segundo a NISO existem três tipos principais sendo: Metadados descritivos, estruturais e

administrativos. Os mais comuns são os descritivos os quais representam um documento para fim de identificação e recuperação. Os estruturais servem para o agrupamento de objetos compostos e os administrativos ajudam no gerenciamento, por exemplo, identificando o responsável pelo cadastro, datas de inserção e alteração dentre outras coisas.

Os metadados que são mais lembrados são os descritivos que são utilizados para a recuperação de informação, mas os metadados podem ser utilizados para organização de recursos eletrônicos, interoperabilidade e integração com sistemas legados, prover identificação digital e fornecer suporte ao arquivamento e preservação.

SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO

Para Baeza e Yates (2013 - p. 1) A Recuperação da Informação trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse.

De acordo com Souza (2006, on-line) um SRI opera na organização e viabilização do acesso aos itens de informação e desempenha tarefas como: Representação da informação, armazenamento e recuperação de informação de forma a satisfazer as necessidades do usuário.

Para Lancaster e Warner (1993, p.1), Os SRIs são a interface entre uma coleção de recursos de informação, em meio impresso ou não, e uma população de usuários;

Vale ressaltar conforme Fernalda (2003, p.15) que existe diferença entre sistemas gerenciadores de banco de dados e sistemas de Recuperação da informação. O primeiro segundo o autor tem como foco a recuperação precisa dos dados baseando-se na expressão de busca. Já na recuperação de informação essa precisão não é o foco. Isso deve-se ao fato dos objetos tratados, enquanto que em um os dados estão em uma estrutura bem definida no outro os objetos não possuem uma estrutura rígida de forma ou linguagem.

INDEXAÇÃO AUTOMÁTICA

Um dos pontos de pesquisa em recuperação da informação é como realizar a indexação de um texto de forma automática. Existem diversas abordagens para o problema mas todas com o objetivo de representar um documento possibilitando recuperá-lo ao comparar a consulta do usuário com a representação do mesmo e em alguns casos possibilitar a ordenação de forma decrescente de acordo com o grau relevância. De forma geral alguns passos são comuns, entre diversos autores, para a indexação automática sendo eles: o processamento do texto, remoção das palavras com pouco significado (stop words) como artigos e preposições, cálculo da representatividade e a criação do índice.

A estrutura mais comum e utilizada é o índice invertido. Segundo Navarro (2013, p.339) Um índice invertido é um mecanismo orientado para a indexação de uma coleção de texto a fim de acelerar a tarefa de busca. Ou seja um índice invertido deve possuir pelo menos uma parte com o vocabulário indexado e outra com as ocorrências dos termos do vocabulário por documento. Desta forma é possível localizar em quais documentos estão presentes cada termo. Em formas mais complexas é necessário também saber em qual posição cada palavra está localizada no texto para permitir buscas por frases por exemplo.

Além do que já foi dito, também é necessário atribuir pesos aos termos para aferir o quanto cada um é importante na indexação de um determinado documento em uma coleção.

Essa a pesquisa utiliza o conceito de indexação automática e apresenta do protótipo onde a ponderação dos termos será realizada utilizando-se a técnica *Term Frequency - Inverse Document Frequency* (TF-IDF), descrita na próxima seção.

PONDERAÇÃO DE TERMOS TF-IDF

A ponderação de termos não é algo que possa ser considerado trivial, no entanto essa tarefa é necessária, é preciso identificar o peso de um termo em um documento de forma a possibilitar o ranking dos documentos entregando ao usuário uma lista ordenada de forma decrescente em relação ao grau de importância do documento frente à busca realizada pelo mesmo.

Existem algumas formas para realizar essa tarefa. Entre as quais temos uma bem conhecida e muito utilizada que é a TF-IDF (*Term Frequency - Inverse Document Frequency*).

TermFrequency(TF) indica a frequência de ocorrência de determinado termo em um documento. Parte-se do princípio que se um termo é repetido diversas vezes em um documento ele será um bom indexador para o mesmo.

InverseDocumentFrequency(IDF) indica a especificidade de determinado termo, ou seja, se determinado termo aparece em todos os documentos o mesmo não será um bom descritor pois uma vez que o mesmo está tão disseminado uma busca pelo mesmo devolveria como resultado praticamente toda a base, agora se determinado termo aparece em poucos documentos de uma coleção ele será considerado um bom descritor uma vez que sua utilização tende a melhorar a precisão da consulta. O idf é representado por $\log N/n$ onde N é o número de documentos na base e n o número de documentos onde o termo aparece.

Para um melhor resultado ainda seria necessário mais um levar em conta o tamanho do vetor gerado para cada documento. Documentos maiores com mais termos levariam vantagem sobre outros menores sem significar contudo que seriam mais adequados a resolução do problema do usuário. Conforme pode ser visto em (SALTON; BUCKLEY, 1988, p.518) foi realizada a normalização da fórmula de forma a que o tamanho do vetor não interfira no resultado.

CORREÇÃO LÉXICA DOS TERMOS BUSCADOS

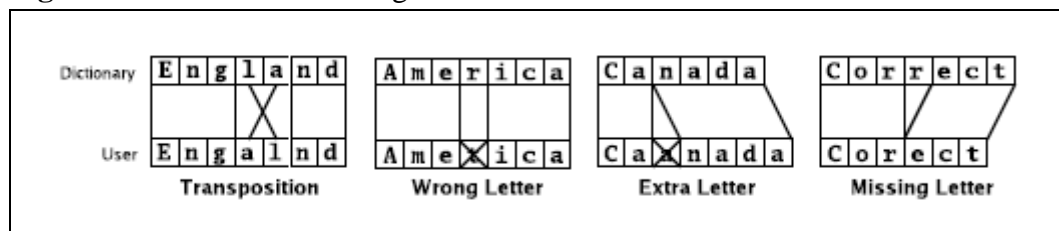
A cada dia temos milhares de consultas sendo executadas em buscadores, bases de dados e um percentual dessas consultas é feito utilizando grafias incorretas. Um exemplo prático é o de pessoas que pesquisam por "Niemaier" ao invés de "Niemeyer" ou "Alejadinho" ao invés de "Aleijadinho". Em diversas bases ao realizar esse tipo de pesquisa o usuário não terá retorno do sistema o que pode levá-lo a achar que não existe informação acerca do objeto procurado. Segundo Choo (2006, p.119) Quando a busca consegue produzir informações úteis, o indivíduo sente satisfação e confiança, mas, se a busca não é bem sucedida, o resultado são sentimentos de decepção e frustração.

Com a utilização desse tipo de ferramenta a tentativa é a que o usuário tenha mais sucesso em suas pesquisas e com isso obtenha maior satisfação.

Para exemplificar o que foi citado foi realizada uma pesquisa no site da UNESCO pesquisando pelo termo "alejadinho" nenhuma informação é retornada. Ao pesquisar por "aleijadinho" são retornados dois resultados onde existem obras do artista tombadas pela UNESCO.

Os erros de grafia mais comuns que podem ocorrer são por Transposição, letra incorreta, a mais ou a menos na grafia da palavra. A seguir vemos uma figura exemplificando esses erros.

Figura 1 - Erros comuns de grafia



Fonte: Martins e Silva (on-line)

Existem três passos básicos a serem considerados, a identificação do termo incorreto, encontrar termos candidatos e ordenar os termos candidatos de forma a identificar a melhor sugestão ao usuário.

PROCEDIMENTOS METODOLÓGICOS

A natureza da pesquisa é aplicada, pois visa à geração de conhecimento para aplicação prática, dirigidos à solução de problemas específicos, que no caso é traduzido no desenvolvimento de um modelo e protótipo para auxiliar na recuperação e gestão de informações acerca de objetos arquitetônicos.

As etapas metodológicas foram geradas tendo em mente a elaboração de um modelo para gerenciamento, disseminação e recuperação de informação referente a objetos arquitetônicos edificados.

COLETA DE DADOS PRÉ-EXISTENTES

Segundo Sapiro (1993, p. 111) uma escolha acertada das fontes de dados é crítica para a coleta e a classificação das informações. Foram escolhidas fontes confiáveis como o IPHAN, IEPHA-MG, Secretaria Municipal de Cultura de Cultura e o Arquivo Público Mineiro para a obtenção de informações acerca dos objetos. Nos institutos e secretaria foram encontrados arquivos que continham referência a livros e artigos antigos, os quais contam a história dos objetos. Isso foi muito interessante para visualizar dentre outras coisas a diferença de posicionamento da população no período em que as obras foram realizadas e o atual. Um exemplo é a Igreja da Pampulha que foi duramente criticada justamente por suas formas e sua decoração e hoje são essas características que a fazem destacar-se.

COLETA DE DADOS NO LOCAL

Além de coleta de documentos nos já citados institutos e secretarias foram tiradas fotos de cada OEA presente na pesquisa. As imagens são importantes para tornar concreto na mente do usuário o que está sendo descrito a partir dos textos. As imagens são documentos importantes do ponto de vista gerencial para demonstrar o atual estado de cada objeto, exibindo seu grau de conservação e características. A partir disso é possível visualizar quais objetos precisam de atenção, também é possível a partir de coletas realizadas em espaços regulares de tempo acompanhar se determinado objeto está sendo degradado ou ainda se alguma característica foi alterada, e em casos de necessidade de reforma ou restauro as imagens podem ser utilizadas como base de pesquisa.

Para capturar as imagens foi utilizada uma câmera Nikon D3100 na sua maior resolução e com o método automático ligado de forma a ter-se sempre uma imagem de qualidade e com certo padrão entre as mesmas. Não foi utilizado nenhum tipo de filtro ou efeito uma vez que apesar da imagem ser utilizada de forma a valorizar o objeto frente aos usuários sua principal função é a de documento arquitetônico, onde a mais valia é a qualidade da imagem no que tange a possibilidade de observação de detalhes.

Além dos importantes fatores já citados, outro ponto que foi levado em consideração para que as fotos fossem de autoria própria são os direitos. Diversas imagens referentes aos OAE foram encontradas na web, inclusive a mesma imagem foi encontrada em vários sites o que dificulta saber quem realmente é o proprietário. Assim sendo todas as imagens utilizadas na pesquisa são de autoria própria ou dos institutos onde a pesquisa foi realizada e com a devida autoria incluída. Para as fotos que foram feitas para o trabalho adotou-se a licença da Creative Commons do tipo Atribuição-NãoComercial-CompartilhaIgual CC BY-NC-SA segundo o próprio site essa licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam ao autor o devido crédito e que licenciem as novas criações sob termos idênticos.

Como exemplo do que foi citado, no caso das imagens, temos abaixo duas fotos do mesmo objeto arquitetônico. No caso é possível visualizar o Arquivo público mineiro. À esquerda temos a foto retirada do registro de tombo do IEPHA-MG, no arquivo as fotos estão datadas como sendo de 1974 e à direita a foto tirada pelo autor em 2013.



DESENVOLVIMENTO DA BASE DE TESTES E PROTÓTIPO

Como o volume de informação coletada na primeira etapa foi muito grande, foi necessário fazer uma seleção do que seria disponibilizado ao usuário. Conforme Choo (2006, p. 419) A variedade de informação deve ser administrada de modo que as informações coletadas reflitam a complexidade do ambiente, sem sobrecarregar os usuários com excesso de informação.

Durante essa pesquisa não foram disponibilizados os arquivos encontrados na íntegra. Foi feito a extração de informações dos diversos arquivos e sua mesclagem de forma a disponibilizar a informação já condensada. Alguns dados deverão estar presente na inclusão de todos os objetos. A definição de grande parte destes metadados para o cadastro inicial foi feita baseando-se nas fichas do SICG do IPHAN, sendo alguns deles: Denominação oficial, outras denominações, Localização (UF, Município, logradouro, número, complemento, código postal), Coordenadas geográficas, Data Construção, Uso Original, Uso atual, Tipologia (Religiosa, civil...), estilo arquitetônico, pessoas envolvidas na construção bem como suas funções à época, responsável pela coleta da informação, responsável pela inserção no sistema.

É permitido a entrada de informações históricas, as quais serão retiradas dos arquivos coletados bem como as imagens tiradas in-loco. Para cada objeto o protótipo deverá gerar uma página web, sem que haja a interferência do usuário utilizando para tal as informações inseridas pelo mesmo. As páginas contém os metadados, as informações históricas inseridas e as imagens. Será gerado um mapa utilizando as coordenadas geográficas do objeto, permitindo que o usuário visualize o ponto onde o objeto se encontra e caso o usuário deseje, poderá procurar por outros objetos próximos fisicamente daquele sobre o qual ele está lendo.

A indexação da base foi realizada de forma automática ao final do processo de preenchimento dos dados. Quando outros objetos forem inseridos na sua indexação será realizada apenas em períodos fixos, como por exemplo, todas as noites. Isso devido

ao volume de processamento que pode ser demandado se várias inserções forem realizadas durante o dia. Ao invés disso todo objeto, seja novo ou alterado, será marcado como "indexação pendente" e a cada noite será executado um processo verificando a necessidade de executar a indexação.

O protótipo foi criado de forma a validar a metodologia criada e a demonstrar como a tecnologia pode ajudar no problema de gerenciamento e divulgação dos Objetos Arquitetônicos Edificados.

Foi utilizado o padrão MVC (Model, View Controller) para o desenvolvimento. Esse é um padrão de arquitetura para aplicativos web, o qual prega o desacoplamento entre as camadas de visão, controle e regra de negócio. Com isso a manutenção e adaptação do mesmo é facilitada uma vez que alterações em uma das camadas não afetarão as demais. Um exemplo disso é que ao alterar a forma como determinados objetos são exibidos nenhuma alteração será necessária nas outras partes.

Outra característica é que a interface de busca utilizada pelos usuários em geral, foi desenhada pensando em um design responsivo. Essa técnica é utilizada para que o site possa se adaptar ao navegador do usuário sem que tenha que ser desenhado um site para cada dispositivo. Um dos problemas em se ter versões diferentes de um site para vários dispositivos é a indexação por parte de buscadores como o google. Faz muito mais sentido ter um único endereço que pode ser acessado por todos do que tablet.site ou tv.site para ser acessado a partir de tabletes ou televisores.

CONCLUSÕES

A pesquisa veio comprovar a interdisciplinaridade uma vez que utilizou conceitos referentes à ciência da informação e à computação. Utilizado em conjunto para chegar ao objetivo de tratar, gerenciar e difundir informação. Isso vem de encontro ao que afirma Gomes (2007, p. 150) sobre a ciência da informação:

sua constituição como área interdisciplinar passa por diferentes caminhos, em diferentes sociedades: o documentalismo europeu; a biblioteconomia, marcadamente na América Latina, onde alarga seu escopo e abrangência pela soma e diálogo com as tradições européias e norte-americanas; a aplicação da computação à palavra, no pós-guerra, e o desenvolvimento da recuperação da informação com o uso de computadores, nos Estados Unidos.

A utilização do índice invertido na pesquisa, bem como as sugestões no caso de digitação errada por parte do usuário não causaram impacto significativo na performance do protótipo, no entanto vale ressaltar que como a pesquisa ainda está em andamento outros dados serão inseridos e com a base mais robusta os testes poderão revelar outros resultados ou confirmar os atuais.

A questão da criação dinâmica de páginas agilizou bastante o processo e retira a necessidade de centralizar a criação de cada página nas mãos do setor de TI que ficaria sobrecarregado.

Fica como sugestão a implementação de outras formas de indexação e classificação no intuito de comparar os resultados na base. Além disso no futuro pode ser estudada a forma de extrair os textos dos arquivos utilizados na pesquisa de forma automática e a utilização dos algoritmos já implementados para disponibilizar o conteúdo dos documentos para consulta.

REFERÊNCIAS

- ALVARENGA, Lidia. A Teoria do Conceito Revisitada em Conexão com Ontologias e Metadados no Contexto das Bibliotecas Tradicionais e Digitais. *DataGramaZero - Revista de Ciência da Informação* - v.2 n.6 dez/01. Acessado em 27/08/2013 em http://www.dgz.org.br/dez01/Art_05.htm
- BAEZA-YATES, Ricardo. RIBEIRO-NETO, Berthier. **Recuperação de Informação: Conceitos e tecnologia das máquinas de busca.** tradução técnica: Leandro Krug Wives, Viviane Pereira Moreira. 2. ed. Porto Alegre : Bookman, 2013.
- CHOO, Chun Wei. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões.** tradução Eliana Rocha. 2ª ed - São Paulo: Editora Senac São Paulo, 2006.
- DAVENPORT, Thomas H.. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação / Laurence Prusak; tradução Bernadette Siqueira Abrão.** — São Paulo : Futura, 1998.
- FERNEDA, Edberto. **Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação.** 2003. 147 fl. **Tese** (Doutorado em Ciências da Comunicação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003.
- GÓMEZ, Maria Nélide Gonçalves. **Ciência da Informação, economia e tecnologias de informação e comunicação: A informação nos entremeios.** In: MACIEL, Maria Lucia. ALBAGLI, Sarita. Orgs. **Informação e Desenvolvimento: Conhecimento, inovação e apropriação social.** Brasília: IBICT, UNESCO, 2007. p.149-184.
- LANCASTER, F. Wilfred. WARNER, Amy J. **Information Retrieval Today.** Information Resources Press. 1993. United States of America.
- LIMA-MARQUES, Mamede. MACEDO, Flávia Lacerda Oliveira de. **Arquitetura da Informação: base para a gestão do conhecimento in TARAPANOFF, Kira. Inteligência, informação e conhecimento.** Org. — Brasília : IBICT, UNESCO, 2006.

SALTON, Gerard. BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, Vol. 24, No. 5, pp. 513-523, 1988.

SAPIRO, Arão. Inteligência empresarial: a revolução informacional da ação competitiva. **Revista de Administração de Empresas**, São Paulo, v. 33, n. 3, p. 106-116, mai/jun., 1993.

SOUZA, Renato Rocha. Sistemas de recuperação da informações e mecanismos de busca na Web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n.2, p. 161-173, maio/ago. 2006.

WILSON, Tom D. Information management. In: **INTERNATIONAL ENCYCLOPEDIA OF INFORMATION AND LIBRARY SCIENCE**. London : Routledge, 1997. p. 187-196.