

CADWeb – Categorização automática de documentos digitais

Georgia Regina Rodrigues Gomes

Doutora em informática pela Pontifícia Universidade Católica do Rio de Janeiro. Professora adjunta da Universidade Cândido Mendes. Rio de Janeiro, RJ - Brasil

E-mail: geogomes@ig.com.br

Rubens de Oliveira Moraes Filho

Aluno de ciência da computação pela Universidade Cândido Mendes. Rio de Janeiro, RJ - Brasil.

E-mail: rubens_olv@hotmail.com

repositories of electronic works, for example, there is a need for tools that will automatically classify documents, since the classification process (categorizations) is done manually. Such a tool will represent an important resource and support for cataloging. This article presents the development of a tool whose chief objective is to categorize digital documents automatically, using pre-established categories, where each document will belong to one or more categories according to its content, thus making the classification of such documents more efficient and also quicker. Techniques and algorithms of text mining were used to develop and validate the tool; also, some categories were defined in the case study, as well as related terms such as: information technology, law and physics.

Keywords

Information technology. Categorization. Digital libraries. Text mining. Digital documents.

Resumo

Com a evolução da tecnologia da informação e a disseminação de documentos digitais na Web, faz-se necessário criar meios que forneçam um mecanismo de organização de tais documentos, facilitando sua busca e recuperação. Em bibliotecas digitais ou repositórios de obras eletrônicas, por exemplo, existe a necessidade de uma ferramenta que possa classificar automaticamente os documentos, visto que o processo de classificação (categorização) é feito de forma manual. Esta ferramenta será de grande importância no apoio à catalogação. Este artigo apresenta o desenvolvimento de uma ferramenta que tem como objetivo principal classificar automaticamente documentos digitais em categorias preestabelecidas, nas quais cada documento pertencerá a uma ou mais categorias de acordo com seu conteúdo, tornando assim mais eficaz e rápida a classificação. Na elaboração da ferramenta foram utilizadas técnicas e algoritmos de mineração de textos, sendo definidas no estudo de caso algumas categorias e termos relacionados, tais como informática, direito e física, para validar a ferramenta.

Palavras-chave

Tecnologia da informação. Categorização. Biblioteca digital. Mineração de texto. Documentos digitais.

Automatic categorization of digital documents

Abstract

The evolution of information technology and dissemination of digital documents on the Web calls for a mechanism for the organization of such documents in order to facilitate the search and recall processes. In digital libraries or

1. INTRODUÇÃO

Com a evolução da tecnologia para administrar e armazenar grandes quantidades de informações, a digitalização e a produção de conhecimento em meios eletrônicos cresceram de forma vertiginosa. Extrair conhecimento a partir dos dados armazenados deu origem à mineração de dados.

Segundo Amo (2003), a mineração de dados teve início nos anos 80, quando se verificou a grande quantidade de dados armazenados que não eram utilizados de maneira propícia, e que depois de submetidos a um processamento, poderiam fornecer conhecimentos que contribuiriam para a tomada de decisões.

Inicialmente, as pesquisas na área de extração de conhecimento a partir de dados eram focadas nos dados do tipo estruturado (Amo, 2003), utilizando técnicas de mineração, que permitem extrair padrões consistentes e vínculos das variáveis existentes no contexto das informações analisadas (FAYYAD et

al., 1996). Com a necessidade de obtenção de novas informações e análise de tipos de dados diferentes, as pesquisas se voltaram para os dados do tipo semi ou não estruturados, e conseqüentemente foram desenvolvidas técnicas de mineração de textos para que esses tipos de dados pudessem ser explorados. Assim, ferramentas e técnicas que extraíam esse conhecimento de forma automática, sistematizada e interpretável são de grande valia (FELDMAN e HIRSH, 1997). As técnicas de mineração de texto são áreas de pesquisa voltadas para a obtenção dos melhores meios de executar a organização de documentos.

A biblioteca digital, também referenciada como biblioteca eletrônica ou virtual, implica um novo paradigma de utilização, armazenamento da informação e disseminação que independe de sua localização física, do horário de seu funcionamento ou da disponibilidade da informação procurada (CUNHA, 1999).

Com a informatização e popularização da internet, a quantidade de documentos digitais que são disponibilizados na rede cresce exponencialmente (DIAS, 2001). Nas bibliotecas digitais, que funcionam como grandes repositórios, essa realidade se aplica. Essas bibliotecas dispõem de uma quantidade enorme de documentos digitalizados: artigos, teses, dissertações, obras completas, e tais documentos são separados em categorias que já são predeterminadas e competem a um padrão adotado pela maioria das bibliotecas existentes, digitais ou não. Logo, torna-se imprescindível que se utilizem técnicas apropriadas para que tais documentos sejam organizados e estruturados em suas respectivas categorias, de modo que uma possível busca seja feita de maneira mais rápida e eficiente, com resultados estruturados (SAYÃO, 2008). Entretanto, na classificação de tais documentos em categorias, a maioria é executada de forma manual.

Este artigo consiste em apresentar a ferramenta CADWeb (CADWeb, 2010), desenvolvida para

realizar a categorização (classificação) automática de documentos digitais de acordo com categorias predefinidas, nas quais cada documento deverá pertencer a uma ou mais categorias, de acordo com seu conteúdo.

O foco deste trabalho se dá sobre documentos digitalizados e armazenados no formato PDF, pois esse tipo de arquivo pode ser considerado o mais utilizado nas bibliotecas digitais existentes.

O restante do artigo está organizado conforme descrito a seguir. A seção 2 apresenta as etapas do processo de categorização utilizadas na ferramenta. Na seção 3 será detalhada a implementação, na seção 4 será apresentado o estudo de casos para uso da ferramenta, e na seção 5 mostrar-se-á a conclusão do trabalho.

2. ARQUITETURA DA FERRAMENTA

Tratar documentos digitais não é uma tarefa trivial, pois existem vários fatores, como por exemplo, a falta de estruturação que eles apresentam e palavras que são consideradas irrelevantes para o procedimento de categorização, que tornam uma tarefa complexa. Outro fator importante é a quantidade de palavras relevantes que aparecem nos documentos, pois elas irão determinar a dimensão no espaço de representação dos documentos.

Logo, deve-se tratar com bastante cuidado os dados textuais para que o uso indevido deles não comprometa as atividades de categorização.

A ferramenta desenvolvida, nomeada CadWeb, executa quatro etapas para o processo de categorização. A primeira etapa consiste na preparação do documento que será categorizado, extraíndo-se o conteúdo textual bruto sem distinção. Logo após todo o conteúdo do texto haver sido tratado inicia-se a segunda etapa, que consiste no pré-processamento dos dados textuais mediante técnicas de mineração de texto, retirando-se as

stopwords e realizando o agrupamento das palavras, dando origem a um formato padrão para a realização da terceira etapa, que versa na categorização propriamente dita. Nesta categorização o número de repetições de cada palavra do documento e as palavras-chave de cada categoria são analisadas e fornecem a indicação da(s) categoria(s) em que o documento se compatibiliza.

Na quarta e última etapa são realizadas as avaliações dos resultados obtidos, sendo que, se esses resultados não forem satisfatórios, o processo será repetido desde o início, porém com alterações no projeto, para que ao término de todo o processo gere novos resultados com um percentual de acerto melhor do que os resultados obtidos anteriormente. Este refinamento foi repetido até que se obtivessem resultados satisfatórios, ou seja, o ciclo foi interrompido quando os documentos categorizados apresentaram um resultado com bom percentual de acertos, e assim foi criado um algoritmo que atende de forma satisfatória a toda a execução do processo.

As etapas realizadas no desenvolvimento da ferramenta são apresentadas na figura 1, assim como uma visão macro do funcionamento do sistema.

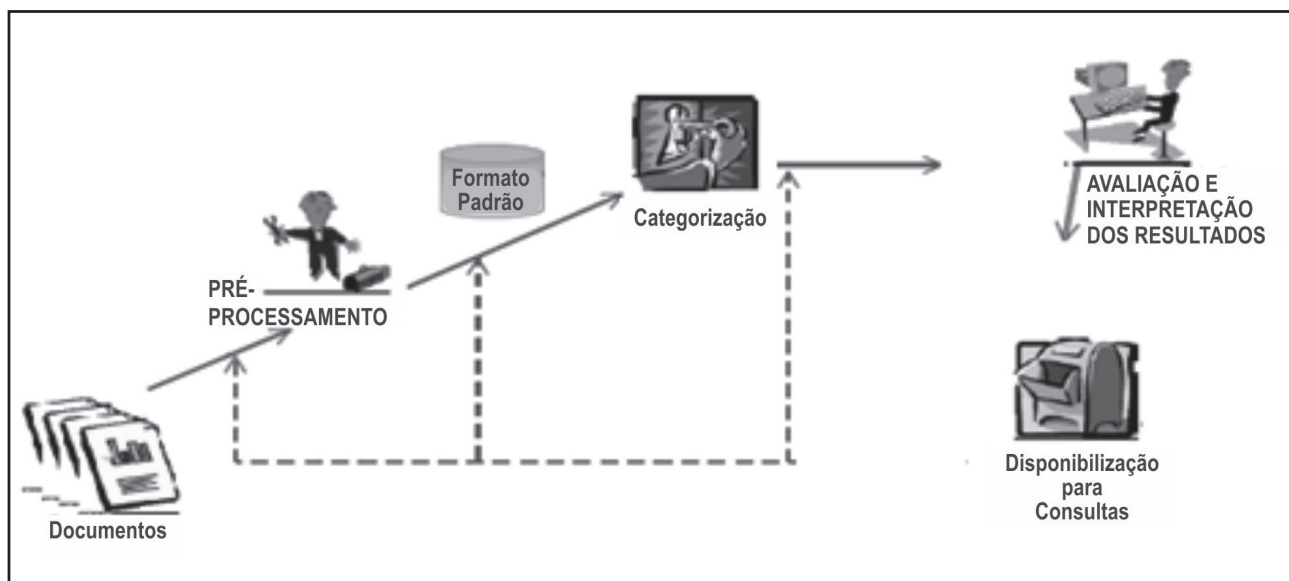
3. DESENVOLVIMENTO DA FERRAMENTA

Serão apresentadas nesta seção as tecnologias utilizadas no desenvolvimento da ferramenta, detalhando-se todas as etapas do desenvolvimento.

3.1 Tecnologias utilizadas

Foram utilizadas tecnologias de programação de código livre e amplamente conhecidas, tais como a linguagem Java (Deitel, 2007), que foi utilizada para a construção da inteligência da aplicação, escolhida para a criação do núcleo da ferramenta devido a características de orientação a objetos, que permite que correções sejam executadas com o mínimo de retrabalho possível, bem como a multiplataforma, que permite que o servidor da aplicação seja hospedado em qualquer sistema operacional, exigindo como pré-requisito a instalação da máquina Java (Java Virtual Machine - JVM).

FIGURA 1
Etapas da ferramenta



Para o desenvolvimento da interface gráfica foi utilizada a tecnologia da fabricante Adobe Systems, conhecida como Flex ou Flash Builder (Schmitz, 2008) em sua versão 3.5, sendo *open source* e de distribuição livre. Esta linguagem permite a construção de páginas RIA (Rich Internet Applications), terminologia utilizada para classificar aplicações desenvolvidas para a internet com grande uso de recursos visuais e interações avançadas com o usuário (VICTORAZZI, 2007). Para a comunicação com o servidor Java (Tomcat), foi necessária a utilização do plugin BlazeDs, também disponibilizado pela fabricante Adobe Systems.

As aplicações Flex necessitam do Adobe Flash Player para que o navegador possa exibir suas funcionalidades; caso o sistema operacional não o tenha instalado, será direcionado ao *site* para *download*.

Como banco de dados foi utilizado o PostgreSQL (PostgreSQL, 2010), devido à sua estabilidade e confiança, encontrando-se há mais de 15 anos em desenvolvimento constante, sendo compatível com os sistemas operacionais mais utilizados para servidores (Unix, Linux e Windows).

Algumas técnicas que compõem a tecnologia de mineração de texto foram implementadas, e o formato de arquivo texto PDF, por ser o formato padrão de documentos digitalizados, é o único a ser processado pela ferramenta, atualmente. Nas seções subsequentes serão detalhadas as etapas do desenvolvimento da ferramenta.

3.2 Preparação do documento

Na implementação desta ferramenta foi utilizado um componente chamado PDFBox e FontBox (PDFBox, 2010). A finalidade em sua utilização foi a extração do conteúdo de um documento no formato PDF para que o processamento pudesse ser realizado.

Através de métodos do PDFBox, toda a informação do texto é armazenada numa variável do tipo *string*.

Como uma variável do tipo *string* não suporta grandes quantidades de caracteres, a solução encontrada foi dividir o conteúdo de um documento grande em várias *strings*, ou seja, a cada 30 páginas uma variável é utilizada para armazenar sua informação. Logo, se um documento é formado por 60 páginas, por exemplo, seu conteúdo será dividido em duas partes.

Para organizar as diversas partes do documento foi utilizada a estrutura conhecida como *HashMap*, tipo de dado composto de chaves e valores, em que cada chave pode ser considerada como um indexador e cada valor um fragmento de texto a ser analisado, que implementa um mecanismo de coleção não fixo (autodimensionável) e não ordenado. A primeira coluna (chaves) é composta de números, desde o número um até o número de parte em que o documento foi dividido, e desse modo, a primeira parte do documento está associada ao número 1, a segunda parte ao número 2, e assim sucessivamente.

3.3 Pré-processamento do texto

Com o documento preparado, pode-se fazer o pré-processamento.

Todo documento de texto é composto de termos (ou palavras). Logo, a palavra é a menor unidade de análise e de acesso nesse tipo de documento. As palavras são os atributos ou características de um texto. Formam elementos primitivos que serão analisados para a descoberta do conhecimento ou extração de um padrão que defina a categoria do documento.

A etapa de pré-processamento é responsável por transformar o documento em uma forma padrão. Para se chegar a esse modelo, o texto passa por quatro processos: o *case folding*, a eliminação de acentos gráficos, a separação das palavras e a eliminação das *stopwords*.

O formato padrão que é gerado ao término desta etapa consiste das palavras relevantes do documento e da quantidade de vezes que cada uma delas ocorreu, o que é chamado de *Bag of Words*.

O formato padrão fica armazenado em outra estrutura *HashMap*, e no caso da ferramenta desenvolvida, a primeira coluna é composta pelas palavras encontradas e a segunda coluna é composta pelo número de ocorrências de cada palavra.

3.4 Processamento do formato padrão

Esta é a etapa de categorização propriamente dita e o processo realizado consiste basicamente na comparação entre as palavras que compõem o formato padrão gerado na etapa anterior com os termos definidos para cada categoria ou subcategoria.

De acordo com Peixoto (2004), categorizar é o mesmo que classificar automaticamente documentos em relação a um conjunto de categorias ou matérias predefinidas.

Vale ressaltar que, para definir os termos para uma categoria ou subcategoria, é necessário o auxílio de um especialista, e que quanto maior for o número de termos definidos, melhor será o desempenho da ferramenta.

Durante o processo são criadas duas variáveis, uma com objetivo de armazenar a quantidade total de termos referentes à categoria que ocorrem no documento, ou seja, a cada termo que é encontrado no documento essa variável é incrementada em um, e a outra variável tem por objetivo armazenar o total de ocorrências desses termos, em que a cada termo observado no documento tem seu total de ocorrências somado à variável.

No processo de comparação a ideia utilizada foi verificar se cada termo, de cada categoria ou subcategoria, está contido em cada palavra pertencente ao formato padrão, ou seja, supondo que um documento passará pelo processo de categorização, foram definidos termos para a categoria Informática, e um destes termos definidos foi Conect.

Pelo processo de comparação proposto, caso ocorressem no documento as palavras Conectado, Conectando, Conectou e Conectar, por exemplo,

todas elas seriam levadas em consideração. Isso diminui consideravelmente a quantidade de termos que deverá ser definida para a mesma categoria, além de otimizar o processo de comparação, pois o radical (*Stemming*) de uma palavra possui abrangência considerável, sem impactar na perda de sentido da palavra, tendo em vista que palavras originadas a partir do radical compartilham, em sua maioria, de um significado comum.

Para que um documento seja associado a determinada categoria, duas regras deverão ser respeitadas: 1) o documento deverá conter, pelo menos, 20% dos termos definidos para a categoria em questão. Se um documento tem 100 termos definidos, por exemplo, ele deverá conter 25 termos; 2) respeitada a primeira regra, a segunda regra é analisada. Esta regra recupera a variável que armazenou a soma das ocorrências de cada termo encontrado no documento e divide pelo total de páginas do documento. Para que a regra seja obedecida, o resultado da divisão efetuada deve ser de no mínimo cinco, ou seja, admite-se que a cada página do documento uma média de cinco ou mais termos da categoria, iguais ou não, ocorra.

4. FUNCIONALIDADES DA FERRAMENTA

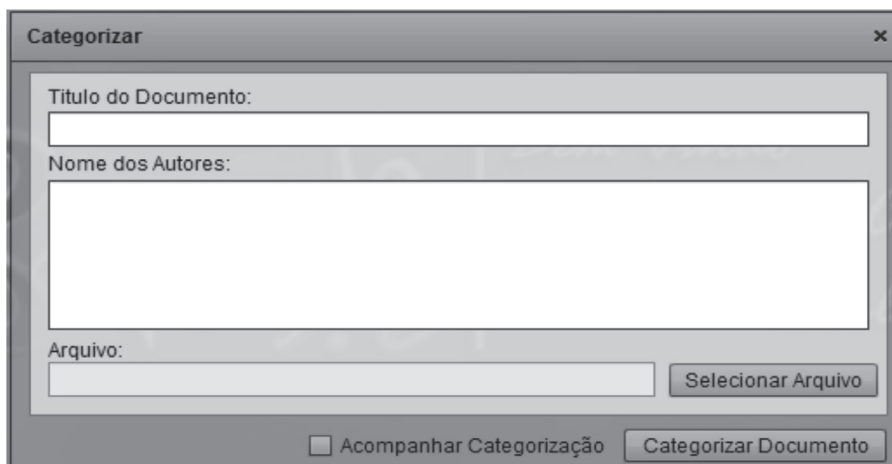
A tela inicial da ferramenta CaDWeb apresentada na figura 2, a seguir, é composta das seguintes funcionalidades: Categorizar, Categorias, Categorizados e Ajuda.

Na opção Categorizar, apresentada na figura 3, a seguir, o usuário informará o título do documento, o nome dos autores e selecionará o arquivo para ser categorizado. Existe a opção de acompanhamento do processo de categorização ou apenas a inicialização do processo. Se o usuário desejar acompanhar o processo, uma tela de processamento é mostrada. Caso contrário, a tela é encerrada e o arquivo continua a ser categorizado pela ferramenta. O CADWeb faz *upload* dos arquivos a serem categorizados e os armazena em diretórios locais, executando a função de repositório de arquivos.

FIGURA 2

Página principal da ferramenta

FIGURA 3

Tela da funcionalidade Categorizar

A manutenção do repositório de categorias é acessada pela opção Categorias, na qual pode-se incluir, alterar e excluir categorias.

Para pesquisa dos documentos categorizados no sistema, o usuário utilizará a opção Categorizados, na qual pode realizar uma busca pelos documentos já categorizados no sistema e obter uma cópia das obras disponibilizadas.

5. ESTUDO DE CASO

5.1 Definindo termos para categorias

Em bibliotecas digitais, existem as categorias que já estão predefinidas. Neste estudo de caso, foram

definidos termos para três categorias: Informática, Direito e Física, e duas subcategorias: Programação e Redes.

Cada uma dessas categorias tem uma lista de termos referentes que a define. Essa lista de termos é constituída pelas palavras utilizadas em cada categoria para a comparação com o documento a ser categorizado.

5.2 Banco de dados

Na construção da ferramenta, foi levado em consideração o princípio de que o acesso aos documentos armazenados no banco de dados da biblioteca digital seja fornecido pelo usuário, visto

que para o correto funcionamento da ferramenta é necessário que o usuário forneça o arquivo no momento da composição da solicitação de categorização. O sistema efetua a cópia deste arquivo local para o servidor, cadastra-o no banco de dados e informa para a aplicação o código e caminho do documento a ser categorizado.

O banco de dados utilizado para testes e vigente no funcionamento da aplicação é o PostgreSQL. A aplicação foi desenvolvida com a utilização de padrões de projeto como o Abstract Factory, que permite a inclusão ou alteração do servidor de forma simples e sem grande retrabalho para os programadores. Lançando mão desse recurso, tanto a gama de servidores que podem ser utilizados é expandida, quanto há facilidade de programar novos acessos a novos servidores. O mesmo padrão também foi aplicado ao tratamento dos documentos, permitindo que novos formatos sejam inseridos de forma simples.

5.3 Testes efetuados

Para avaliar a ferramenta CADWeb, foram utilizados cerca de 50 documentos digitais, sendo que o tempo de análise de cada documento varia quanto ao seu tamanho e o número de palavras contidas. O desempenho da aplicação também é influenciado pela velocidade do servidor e pelo tempo de *upload* do arquivo a ser avaliado.

5.4 Avaliação dos resultados

A métrica utilizada para avaliar os resultados obtidos foi a métrica precisão (Gomes, 2006) que avalia quanto o sistema acerta. O cálculo realizado é feito pela equação (1).

$$\text{Precisão} = \frac{\text{Número de itens corretamente classificados}}{\text{Número total de itens}}$$

Não foi encontrado na literatura um valor mínimo estipulado para determinar se um sistema obteve taxa de precisão alta ou baixa. Assim, essa avaliação

é feita de forma subjetiva, em que o sistema deve atingir uma taxa de precisão considerada satisfatória pelos especialistas do domínio.

Dos 50 documentos que foram submetidos ao processo de categorização realizado pela ferramenta, apenas 32 puderam ser categorizados. Isso ocorreu pelo fato de haver 18 documentos que estavam protegidos.

Após a ferramenta ter realizado o processo de categorização sobre os documentos que puderam ser selecionados, somando-se um total de 32, verificou-se que 23 documentos foram associados à categoria Informática, quatro documentos foram associados à categoria Direito e nenhum documento foi associado à categoria Física.

Dos 23 documentos pertencentes à categoria Informática, cinco também foram associados à subcategoria Redes, e 21 também foram associados à subcategoria Programação.

O quadro 1, a seguir, apresenta as categorias e subcategorias e os resultados da categorização da ferramenta, bem como o percentual de acerto.

6. CONCLUSÃO

Este trabalho descreveu o desenvolvimento de uma ferramenta para classificar documentos digitais automaticamente utilizando categorias preestabelecidas.

Segundo Torres (2005), em janeiro de 2005 já existiam cerca de 600 bilhões de arquivos disponibilizados na Internet. Os portais Terra (Terra, 2010) e Euronews (Euronews, 2010) mostram que a empresa Google inicia processo de digitalização de obras fechando acordo com escritores e grandes bibliotecas, como as de Roma e Florença.

Nesse contexto, o trabalho de processamento automático de documentos é um facilitador tanto na organização quanto no acesso à informação.

QUADRO 1

Resultado da categorização da ferramenta

	Categorias			Subcategorias			Sem Categoria Definida
	Informática	Direito	Física	Redes	Programação		
Documentos Associados	23	4	0	5	21	9	
Percentual de Acerto	86%	75%	100%	80%	90%	77%	
Percentual Total de Acerto				84,6%			

No caso das bibliotecas digitais, será um instrumento que contribuirá para o bibliotecário classificar de forma mais rápida e precisa os documentos, facilitando a busca aos mesmos.

Outro fator importante que merece ser mencionado é que a ferramenta foi desenvolvida para trabalhar com arquivos no formato PDF, pois se tratava de um sistema para ser utilizado por bibliotecas digitais. Por esse motivo foi especificado o formato PDF. Entretanto, já estão sendo desenvolvidos módulos para outros formatos.

Artigo submetido em 15/10/2010 e aceito em 15/09/2011.

REFERÊNCIAS

AMO, Sandra de. Técnicas de Mineração de Dados. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA, 24., 2004, Salvador. Disponível em: <<http://inf.cp.cefetpr.br/ligia/papers/jai-cap5.pdf>>. Acesso em: 16 de ago. 2010.

CADWeb, Disponível em: <<http://www.net.ucam-campos.br/>>. Acesso em: 19 set. 2010.

CUNHA, Murilo Bastos da. Desafios na construção de uma biblioteca digital. *Ci. Inf.*, Brasília, v. 28, n. 3, dez. 1999. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-9651999000300003&lng=en&nrm=iso>. Acesso em: 29 ago. 2010.

DEITEL, H. M.; DEITEL, P. J. *Java: como programar*. Trad. de Edson Furmankiewicz. 6.ed. São Paulo: Pearson Prentice Hall, 2007. 1120p., il.

DIAS, Eduardo Wense Contexto Digital e Tratamento da Informação. *DataGramaZero - Revista de Ciência da Informação*, v. 2, n. 5 out. 2001. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: 11 out. 2011.

EURONEWS. Google vai digitalizar obras de Galileu. Disponível em: <<http://pt.euronews.net/2010/03/12/google-vai-digitalizar-obras-originais-de-galileu/>>. Acesso em: 18 jun. 2010.

FAYYAD, Usama; PIATETSKI-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, v.39, p.27-34, nov. 1996.

FELDMAN, R. e Hirsh, H. Exploiting background information in Knowledge discovery from text. *Journal of Intelligent Information System*, v.9, n.1, jul./aug. 1997, p. 83-97.

GOMES, Geórgia Regina Rodrigues. *Integração de repositórios de sistemas de bibliotecas digitais e de sistemas de aprendizagem*. 2006. 143 f. Tese (Doutorado em Informática). Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. Rio de Janeiro, 2006.

PDFBox. Disponível em: <www.pdfbox.org>. Acesso em: 16 jun. 2010.

PEIXOTO, Maria Dulce Folgado; BATISTA, Maria da Graça Tomás Rodrigues Henriques; CAPELO, Maria João Tavares de Sá Pereira. *Categorização de Textos*. Disponível em: <http://www.di.ubi.pt/~api/text_categorization.pdf>. Acesso em: 20 jun 2010.

PostgreSQL. Disponível em: <www.postgresql.org.br>. Acesso em: 10 abr. 2010

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. *Transinformação*, v. 20, n. 2, 2008. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=215>>. Acesso em: 30 ago. 2010.

SCHMITZ, Daniel Pace, *Adobe Flex Builder 3.0: conceitos e exemplos*. Rio de Janeiro: Brasport, 2008. 180 p.

TERRA. Google fecha acordo com escritores e editores para digitalizar livros. Disponível em: <<http://tecnologia.terra.com.br/interna/0,,OI3290212-EI4802,00-Google+fecha+acordo+com+escritores+e+editores+para+digitalizar+livros.html>>. Acesso em: 10 maio 2010.

TORRES, José Alberto Sousa; OLIVEIRA, Grinaldo; AMORIM, Cláudio Alves. Utilização do Corte Percentual na Categorização de Documentos da Web com o Algoritmo Naive Bayes. In: CONGRESSO DE TECNOLOGIA DA BAHIA, 1., 2005, Ilhéus, Bahia. *Anais...* Disponível em: <[http://torres.eti.br /component/option,com_docman/task,cat_view/gid,13/Itemid,34/](http://torres.eti.br/component/option,com_docman/task,cat_view/gid,13/Itemid,34/)>. Acesso em: 29 ago. 2010.

VICTORAZZI, Nelson Rogério. *Ria - Rich Internet Applications*. 2007. 38 f. Trabalho de Conclusão de Curso (Graduação) Instituto de Informática. Curso de Especialização em Web e Sistemas de Informação. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2007. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/14328/000664860.pdf?sequence=1>>. Acesso em: 20 maio 2010.