UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Sérgio Murilo Penedo

UMA TÉCNICA DE RECUPERAÇÃO ADAPTATIVA DE OBRAS EM BIBLIOTECAS DIGITAIS BASEADA NO PERFIL DO USUÁRIO

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação

Prof. Dr. Roberto Willrich
Orientador

Florianópolis, fevereiro de 2005

UMA TÉCNICA DE RECUPERAÇÃO ADAPTATIVA DE OBRAS EM BIBLIOTECAS DIGITAIS BASEADA NO PERFIL DO USUÁRIO

Sérgio Murilo Penedo

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Sistemas de Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

_____

Prof. Dr. Raul Sidnei Walzlawick
Coordenador do Curso

Banca Examinadora

_____

Prof. Dr. Roberto Willrich (Orientador)
INE/UFSC

_____

Prof. Dr. Mário Antônio Ribeiro Dantas
INE/UFSC

_____

Prof. Dr. Vitório Bruno Mazzola
INE/UFSC

_____

Prof. Dr. Pierre de Saqui-Sannes
ENSICA/França

"Há homens que lutam um dia e são bons

Há outros que lutam um ano e são melhores

Há os que lutam muitos anos e são muitos bons

Porém há os que lutam toda a vida

Estes são os imprescindíveis"

(Bertold Brecht)

Agradecer...

A Deus, motivo maior de tudo,

Aos meus Pais, Osvaldo e Inezita Penedo

por estarem sempre ao meu lado me apoiando ,

Ao meu orientador Dr. Roberto Wilrich ,

A todos que contribuíram para

A realização deste trabalho.

Sumário

Lista de ilustrações

Resumo

As bibliotecas digitais oferecem um modo eficiente de busca ao acervo graças à indexação de seu conteúdo. Apesar disto, caso o acervo seja grande, mecanismos tradicionais de busca podem não ser eficientes em recuperar informação mais pertinente às necessidades do usuário. O conceito de Hiperdocumento Adaptativo pode ser adicionado às bibliotecas digitais de modo a considerar o interesse do usuário na busca de informações, representando as suas necessidades de informação através de uma estrutura bem definida, denominada de perfil. Esta Dissertação tem como proposta aplicar os conceitos de hiperdocumento adaptativo às bibliotecas digitais, para que o resultado das consultas sejam apresentados de maneira organizada levando em conta o perfil do usuário da biblioteca, visando facilitar a identificação dos documentos relevantes no conjunto de documentos da Biblioteca. É realizado o agrupamento dos documentos recuperados, atendendo os critérios de busca em níveis de relevância e em cada nível os documentos são ordenados segundo um critério. Tanto o critério de agrupamento quanto o de ordenação são definidos baseados no perfil do usuário. Como base para a implementação da proposta é utilizada a Biblioteca Digital de Literatura Brasileira e Catarinense (LBC), desenvolvida no contexto do projeto CNPq SIDIE (Sistema de Disponibilização de Informações para o Ensino).

Ao final se verificou que ao ser realizada uma busca que retornam muitas obras, o efeito de sobrecarga cognitiva imposta ao usuário diminui, pois o mesmo recebe os itens (autores, obras ou gênero) de sua preferência agrupados antes dos demais, facilitando a identificação dos documentos relevantes ao mesmo dentre os resultados que obedecem ao seu critério de busca.

Palavras-Chave: Bibliotecas Digitais, Perfil do Usuário, Hiperdocumento Adaptativo, Recuperação de Informações.

Abstract

The digital libraries offer a more efficient way of search to the collection thanks to the indexation of your content. In spite of this, in case the collection is big, traditional mechanisms of search cannot be efficient in recovering more pertinent information the user's needs.

Adaptive Hypermedia's concept can be added to digital libraries in way to consider the user's interest in the search of information, representing your needs of information through a very defined structure, denominated of profile.

This Dissertation has as intended to apply the concepts of Adaptive Hypermedia to digital libraries, so that the result of the consultations is presented in an organized way taking into account the user's of the library profile, seeking to facilitate the identification of the important documents in the group of documents of the library. The grouping of the recovered documents is accomplished, assisting the search criteria in levels of relevance and in each level the documents are ordered second a criterion. As much the grouping criterion as the one of ordination they are defined based on the user's profile.

As base for the implementation of the proposal is used the Digital Library of Brazilian Literature and Catarinense (LBC), developed in the context of the project SIDIE (System of Available of Information for the Teaching).

At the end it was verified that when being accomplished a search that a lot of works come back, the effect of cognitive overload imposed the user it decreased, therefore the same receives the items (authors, works or gender) of your preference contained before the others, facilitating the identification of the important documents to the same among the results that obey your search criterion.

Keyword: Digital libraries, User Profile, Adaptive Hypermedia, Recovery of Information.

Lista de Abreviaturas e Siglas

AACR2 – Anglo American Cataloguing Rules

AHAM – Adaptative Hypermedia Application Model

ARL – Association of Research Libraries

ARPA – Advanced Research Projects

ASCII – American Standard Code for Information Interchange

BD – Bibliotecas Digitais

BDB – Biblioteca Digital Brasileira

CEN/ISS – European Committee for Standardization/Information Society Standardization System

CGI – Common Gateway Interface

CIMI - Consortium for the Computer Interchange of Museum Information

CLIR – Council on Library and Information Resources

CNEN/CIN – Comissão Nacional de Energia Nuclear/Centro de Informações Nucleares

DCMI – Dublin Core Metadata Initiative

DIT – Directory Interchange Format

DLF – Digital Library Federation

DLI – Digital Libray Initiative

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

FTP – File Transfer Protocol

GA – Genetic Algoritm

HA – Hiperdocumentos Adaptativos

HTTP – Hypertext Transfer Protocol

IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia

IDF – Inverso Termo freqüência

IGIMA – Intelligent Information Gathering and Filtering System based on Multi-Agent

ISO – International Organization for Standardization

LANL – Los Alamos National Laboratory

LBC – Biblioteca Digital de Literatura Brasileira e Catarinense

LCSH – Library of Congress Subject Headings

MARC – Machine Readable Cataloging Recorde

NASA – National Aeronautics and Space Administration

NCSA – National Center for Supercomputing Applications

NDLP – National Digital Library Project

NDLTD – Networked Digital Library of Theses and Dissertations

NSF – National Science Foundation

NISO – National Information Standards Organization

OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting

OCLC – Online Computer Library Center

OCR – Optical Character Reader

PHP - Pre-Hypertext Processor

RAP – Repertory Access Protocol

RIA-BD – Sistema de Recuperação de Informações Adaptativos Aplicado a Bibliotecas Digitais

SIDIE – Sistema de Disponibilização de Informações para o Ensino

SPARC – Scholary Publishing & Academic Resources Coalition

SRIA – Sistema de Recuperação de Informações

TF – Termo freqüência

USP – Universidade de São Paulo

XML – Extensible Markup Language

ZIG – Z39.50 Implementors Group

Capítulo 1. Introdução

O surgimento das bibliotecas digitais possibilitou uma melhora e inovações em relação às bibliotecas convencionais. Bibliotecas digitais são organizações que provêem os recursos, inclusive o pessoal especializado para selecionar, estruturar, preservar a integridade e assegurar a persistência com o passar do tempo de coleções digitais, de forma que estejam prontas e economicamente disponíveis para uso por uma comunidade ou por um grupo de comunidades *(DLF, 1998).*

Ferreira (1997) e Cunha (1999) indicam que as bibliotecas digitais devem oferecer os seguintes serviços: possuir razoável quantidade de fontes de informação e coleções de qualidade, em versões on-line, integrando-as com os objetos físicos da informação; armazenar e processar informação em múltiplos formatos, texto, imagem, áudio, vídeo; desenvolver interfaces de informações gerais ou especializadas de interesse dos seus usuários; e permitir a utilização simultânea do mesmo documento por duas ou mais pessoas.

Pode-se destacar alguns benefícios em potencial das bibliotecas digitais (IBICT, 2001):

- Ela traz a biblioteca até o usuário, eliminando vários contratempos, como o do usuário ter que se deslocar até a biblioteca, problemas de trânsito, mau tempo, etc.
- A biblioteca é levada diretamente até a mesa de trabalho ou estudo do usuário, aumentando o uso da biblioteca.
- O poder da computação é usado para procurar e folhear os documentos, recuperando documentos mais eficazmente.
- A informação está mais tempo disponível, não existindo restrições de horários de fechamento de prédios, por exemplo.
- As novas versões de textos, trabalhos de referência podem ser imediatamente disponibilizadas.

- Novos formatos de informação são disponibilizados, por exemplo, uma foto de satélite pode ser vista de vários modos diferentes, uma fórmula matemática pode ser manipulada com softwares como Mathematica ou Maple (Arms, 1999).

Diversas bibliotecas digitais já existem hoje. Em especial para este trabalho, o projeto CNPq Sistema de Disponibilização de Informações para o Ensino (SIDIE), (SIDIE, 2001) vem utilizando o conceito de bibliotecas digitais para criar um sistema oferecendo informações para o ensino. No SIDIE, foi definida e implementada uma biblioteca digital usando softwares de domínio público (PHP, MySQL, Apache e Sistema Operacional Linux).

A principal técnica de recuperação da informação usada nas bibliotecas digitais é a pesquisa usando metadados, que são informações utilizadas para descrever um objeto, ou seja, os seus atributos. A crescente complexidade e volume de informações disponibilizadas nas bibliotecas digitais exigem processos de recuperação de informações cada vez mais eficientes. Diante desse quadro, o processo de recuperação de informações apresenta a cada dia novos desafios e se configura como uma área de extrema importância para as bibliotecas digitais.

Diversos trabalhos , (Chen, 2002), (Crestani,1999), (Deniman, 2003), (Chen,1998), (Lui, 2004), (Dizagi, 2004), dentre outros, já levam em consideração o perfil do usuário (suas preferências e conhecimentos) para melhorar a eficiência do SRI, criando os SRI Adaptativos. Os SRI adaptativos podem também ser aplicados a bibliotecas digitais, onde o sistema de busca leva em consideração, além dos metadados do acervo e dos critérios de busca, o perfil do usuário. Neste caso, os conhecimentos e interesses do usuário relativos ao tema da biblioteca serão muito úteis para a busca de informações.

1.1 Objetivos

Esta dissertação tem por objetivo principal a proposição de um sistema de recuperação de informações adaptativo (SRIA) aplicado a bibliotecas digitais, chamado de RIA-BD. Nesta proposta, o RIA-BD apresenta o resultado da consulta de uma maneira organizada levando em conta o perfil do usuário, visando facilitar a identificação dos documentos relevantes no conjunto de documentos da Biblioteca. Este trabalho considera a possibilidade que o interesse do usuário pode variar com o tempo, sendo que o perfil do usuário é constantemente atualizado. Esta

atualização é determinada levando em consideração que documentos da biblioteca o usuário acessa e com qual freqüência. O RIA-BD realiza o agrupamento dos documentos atendendo os critérios de busca em níveis de relevância e em cada nível os documentos são ordenados segundo um critério. Tanto o critério de agrupamento quanto o de ordenação são definidos baseados no perfil do usuário.

A Biblioteca Digital de Literatura Brasileira e Catarinense (LBC), desenvolvida no contexto do projeto SIDIE (SIDIE, 2001) é utilizada como cenário de uso do sistema de recuperação proposto.

Os objetivos específicos deste trabalho são os seguintes:

- Estudar a arquitetura da biblioteca digital desenvolvida no contexto do projeto SIDIE.
- Levantar as possibilidades de implantação de um sistema de recuperação adaptativo aplicado a biblioteca digitais.
- Estudar as técnicas de recuperação de documentos relativas a Hiperdocumentos Adaptativos que podem ser utilizadas em uma biblioteca digital, principalmente modelagens de usuários;
- Definir o sistema de recuperação adaptativo aplicado a bibliotecas digitais;
- Estender a arquitetura da biblioteca digital do projeto SIDIE para a inclusão do sistema de recuperação adaptativo
- Implantar um protótipo da biblioteca digital já munida do sistema proposto e realizar testes para validação da proposta.

1.2 Justificativa

O uso de Hiperdocumentos Adaptativos (HA) ainda não está bem difundido, sendo que existem poucos trabalhos nesta área, sendo mais raros ainda aplicados à Bibliotecas Digitais. Segundo Silva (2003), a utilização do hiperdocumento adaptativo em bibliotecas digitais pode minimizar o tempo de pesquisa do usuário, fazendo com que somente as informações relevantes,

de acordo com o seu perfil, cheguem até ele. Além disso, a HA pode ser utilizada para alimentar o perfil do usuário, que deverá ser criado no momento da inscrição na biblioteca, onde o usuário deixará registradas suas preferências e terá, na interface personalizada, *links* para documentos do acervo que poderão ser de interesse do usuário".

1.3 Estrutura do trabalho

O restante desta dissertação está organizado na forma que segue. O capítulo 2 apresenta os principais conceitos relacionados às Bibliotecas Digitais. Em seguida, o capítulo 3 aborda o tema de Hiperdocumentos Adaptativos, descrevendo os tipos, os métodos e técnicas de adaptação, a modelagem do usuário, a obtenção do perfil do usuário e a arquitetura de HA. Na seqüência, o capítulo 4 apresenta os conceitos relacionados com Sistemas de Recuperação de Informação, apresentando alguns trabalhos relacionados. Em seguida, no capítulo 5 é apresentada a técnica de recuperação adaptativa aplicada a bibliotecas digitais proposta neste trabalho. O capítulo 6 apresenta o sistema RIA-BD integrado à LBC. Finalmente, no capítulo são 7 apresentadas as conclusões e os trabalhos futuros.

Capítulo 2. Bibliotecas Digitais

Nos últimos tempos a Web vem tornando-se um meio eficaz de disseminar tecnologias e conhecimento, disponibilizando artigos, tutoriais, e várias formas do conhecimento humano, diminuindo o tempo entre o momento em que o arquivo é escrito e o tempo em que ele pode levar para ser editado, publicado e disponível ao público alvo. As tecnologias utilizadas na Web já estão incorporadas a vários setores da sociedade, como lojas, bancos, hotéis, etc. É o que está acontecendo no caso das Bibliotecas, com o surgimento das Bibliotecas Digitais (BD), onde se pode observar o resultado de um processo que vem se desenvolvendo ao longo dos anos, acompanhando o desenvolvimento da informática.

A construção da Biblioteca Digital ocorreu de duas maneiras: de forma *off* e *online* (Levacov, 1997). A forma *off* teve como inicio o controle do inventário e da circulação, e depois com criação de catálogos eletrônicos e a automação das atividades de indexação. Em seguida vieram as versões eletrônicas de obras de referência, a maioria em Cd-*roms*, até chegar ao armazenamento e a recuperação das obras. A evolução tecnológica das comunicações foi disponibilizando recursos *online* como FTP (*File Transfer Protocol*) a própria Web, e que foram sendo incorporados às Bibliotecas, os quais foram se integrando com os recursos *offline*.

Este capítulo apresenta diversos conceitos relacionados às Bibliotecas Digitais (BDs). Além disso, são apresentadas algumas arquiteturas, em especial a da biblioteca do projeto CNPq - SIDIE que servirá de base para o desenvolvimento da proposta apresentada nesta dissertação.

2.1 Definição de Biblioteca Digital

Não existe um conceito único para definir Bibliotecas digitais, as definições ocorrem mais em nível dos serviços e tarefas oferecidos. A seguir serão apresentadas alguns definições e requisitos apresentados por diversos autores.

A *Digital Library Federation (DLF, 1998)* define bibliotecas digitais como organizações que provêm os recursos, inclusive o pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade, e assegurar a constância com o passar do tempo, de coleções de trabalhos digitais de forma que eles estejam prontamente e economicamente disponível para uso por uma comunidade ou por um grupo de comunidades".

A *Digital Libraries Initiative* (DLI, 1998) afirma que uma biblioteca digital não é somente o equivalente a conjuntos digitalizados com métodos de gestão da informação, mas também um conjunto de coleções, serviços e pessoas que favorecem o ciclo completo da criação, difusão, uso e preservação dos dados, para a informação e para o conhecimento.

Segundo Lucier(1995) e Fox (1995), uma Biblioteca Digital deve ter incorporada as seguintes funcionalidades: serviços humanos como publicações eletrônicas, pessoal especializado e educação a distancia; conteúdo, fontes primárias, comunicação informal e textos eletrônicos; ferramentas para uso no browser escolhido; novos tipos de recursos de informação; novas propostas de aquisição; novos métodos de armazenagem e preservação; novas formas de classificação e catalogação; e novas formas de interação com os usuários.

O que define uma biblioteca como sendo digital é o fato dela consistir em várias bibliotecas e possuir tarefas básicas as quais são responsáveis por seu caráter transformador. As principais características de uma Biblioteca Digital apontadas por Ferreira (1997) são:

- Possuir um ambiente compartilhado onde os usuários tenham acesso a coleções de informação pessoal, coleções encontradas em bibliotecas convencionais e coleções de dados utilizados por cientistas;
- Prover acesso a um grande número de fontes de informação e coleções de qualidade, em versões on-line;
- Desenvolver interfaces de informação gerais ou especializadas, úteis para os usuários;
- Disponibilizar um ambiente que permita a experimentação e incorporação de novos serviços e produtos;
- Armazenar e processar informação em múltiplos formatos, incluindo texto, imagem, áudio, vídeo, dentre outras, e;

- Intensificar a comunicação e colaboração entre os sistemas de informação para benefício da sociedade em geral.

2.2 Metadados em bibliotecas digitais

Metadados, ou seja, dados sobre dados, são elementos chave para qualquer biblioteca, seja tradicional ou digital. Eles são usados para descrever e organizar os recursos da biblioteca e para procurar e folhear esses recursos depois de salvos (Geisler, 2002). O conceito de metadado aplicado ao contexto de bibliotecas, pode ser entendido como registros em um catálogo, semelhante aos usados em bibliotecas tradicionais. Mais formalmente, metadado trata-se de um conjunto de descritores de documentos, que são usados para a sua indexação e localização (Smith, 1996).

Para Baeza-Yates (1999), metadados são os atributos de dados ou de documentos, que geralmente descrevem informações sobre autor e conteúdo, que são subdivididas em categorias, como as encontradas em um catálogo de biblioteca. Os metadados em bibliotecas digitais geralmente são anotadas de acordo com algum padrão como Marc (*Machine Readable Cataloging Record*), (MARC, 2005) ou Dublin Core (*Dublin Metadata Core Element Set*), (DCMI, 2005) e que serão vistos mais adiante neste capítulo.

Segundo Smith (1996), no contexto de bibliotecas digitais, o conceito de metadados geralmente está associado ao tipo de informação que: produz uma breve caracterização do objeto de informação nas coleções de uma biblioteca; é armazenado como os conteúdos dos catálogos digitais nas bibliotecas tradicionais, e; é usado para ajudar os usuários a acessar informação pertinente ao seu interesse.

Um dos principais objetivos do uso de metadados é o de prover a interoperabilidade entre os sistemas, e é largamente utilizado na Web em sistemas de comercio eletrônico, máquinas de busca, e entre Bibliotecas digitais.

A seguir será visto a questão da interoperabilidade, com a descrição dos principais padrões de metadados associados à Bibliotecas Digitais.

2.3 Interoperabilitade em Bibliotecas Digitais

Segundo Pistori (2000), as bibliotecas digitais podem ter uma arquitetura centralizada ou distribuída. Em uma arquitetura centralizada, as coleções podem estar armazenadas em vários servidores, mas o gerenciamento e a busca são realizados de forma centralizada. Já em uma arquitetura distribuída, existem diversas bibliotecas que podem ser acessadas através da mesma interface única do lado cliente. Neste tipo de arquitetura, o gerenciamento de seu acervo é feito de maneira centralizada, porém, a busca e o armazenamento são distribuídos.

No caso de bibliotecas distribuídas, e para possibilitar uma disseminação da informação mais eficiente, se faz necessário a utilização de padrões de interoperabilidade entre as bibliotecas participantes de um sistema de bibliotecas, também denominados de protocolos de interoperabilidade.

Em todas as soluções de interoperabilidade, um requisito básico é a padronização em nível de metadados. Esta seção apresenta alguns padrões de metadados aplicados em bibliotecas digitais e protocolos garantindo a interoperabilidade.

2.3.1 Padrão Marc – (*Machine-Readable Cataloging)*

O formato MARC (ISO 2709, 1996) (Z39.2, 1996) foi criado na década de 60 pela biblioteca do Congresso Americano (*Library of Congress*), e a qual é a responsável pela sua documentação. Trata-se de um sistema que utiliza números, letras e símbolos para fornecer diferentes tipos de informação, e tem como objetivo principal, promover a comunicação de registros bibliográficos. (Vosgrau, 2005).

O MARC se utiliza de dois conceitos:

- *Machine-readable,* que significa um tipo particular de computador que pode ler e interpretar a informação contida em um registro bibliográfico. Este computador deve estar munido de um software que possa interpretar os registros MARC.
- *Cataloging-record* , que se trata de um registro bibliográfico, ou informação do tipo que geralmente é anotada em um cartão de um catálogo de biblioteca.

Um registro do tipo MARC inclui, segundo MARC (2005):

- Uma descrição do item - para compor a descrição bibliográfica de um item da biblioteca o sistema MARC utiliza-se das regras do *Anglo-American Cataloguing Rules* (AACR2). A descrição é representada nas seções de parágrafo de um cartão MARC, e inclui o título, uma declaração de responsabilidade, edição, detalhes materiais específicos (como informação de publicação, descrição física, série, notas e números padrões);

- Entrada principal e entradas adicionais - a AACR2 também contém regras para determinar "pontos de acesso" ao registro (são denominados de "entrada principal" e mais outras entradas adicionais) e determinar a forma que estes pontos de acesso devem ter. Esses pontos são pontos de recuperação no catalogo da biblioteca por onde se pode acessar o item.

- Cabeçalhos de assuntos - são utilizados para selecionar os assuntos sob os quais o artigo será listado. O MARC utiliza a lista de marcadores de cabeçalhos de assuntos fornecida pela *Library of Congress Subject Headings* (LCSH).

- Número de chamada - tem como propósito de colocar artigos de assuntos iguais juntos na mesma estante da biblioteca. A maioria dos itens são sub-organizados por ordem alfabética de autor. A segunda parte do número de chamada normalmente referencia o nome do autor, para facilitar a sub-organização. Para selecionar o número de chamada para um item, é usado no sistema MARC o *Decimal Dewey* ou *Library of Congress Classification Schedule* (MARC, 2005).

A tabela 1 apresenta um registro MARC, rotulado com indicadores que são denominados de: campos, etiquetas, indicadores, sub-campos, código do sub-campos e volumes.

Tabela 1-Exemplo de um Registro MARC e suas Tags, fonte (MARC, 2005)

INDICADORES	DADOS
100 1# $a	Arnosky, Jim.
245 10 $a	Raccoons and ripe corn/
$c	Jim Arnosky.
250 ## $a	1st ed.
260 ## $a	New York:
$b	Lothrop, Lee & Shepard Books,
$c	C1987.
300 ## $a	25 p.:
$b	Col. ill.;
$c	26 cm.
520 ## $a	Hungry raccoons feast at night in a field of ripe corn.
650 #1 $a	Raccoons.
900 ## $a	599.74 ARN
901 ## $a	8009
903 ## $a	$15.00

A seguir será detalhado o conteúdo do registro MARC apresentado na Tabela1:

- Campos: são marcados através de etiquetas. Cada registro bibliográfico é dividido logicamente em campos. Existem campos para: autor, para a informação dos títulos, etc. Esses campos são subdivididos em um ou mais sub-campos. Os nomes textuais dos campos são muito longos para serem reproduzidos dentro de cada registro MARC, por esse motivo eles são representados por etiquetas de três dígitos. A etiqueta é sempre os primeiros três dígitos do registro MARC. O quadro a seguir apresenta as etiquetas mais usadas.

Tabela 2-Etiquetas mais usadas em um registro MARC, fonte (MARC, 2005)

010 tag		Número de Controle da Biblioteca do Congresso (LCCN)
020 tag		Padrão Internacional de numeração de livros (ISBN)
100 tag		Entrada Principal de Nome Pessoal (Autor)
245 tag		Informação de Título (que inclui o título, outras informações do título, e uma declaração de responsabilidade
250 tag		Edição
260 tag		Informação sobre a Publicação
300 tag		Descrição Física
440 tag		Declaração de série/Entradas Adicionais
520 tag		Anotação ou Nota de Resumo
650 tag		Título de Assunto ou Tópico
700 tag		Entrada Adicional de Nome Pessoal (outros autores, editor ou Ilustrador)

Alguns campos são definidos por indicadores, que são caracteres que vem em seguida a uma etiqueta. Quando não existe um valor para o indicador, ele é anotado pelo símbolo #. Cada valor do indicador é um número de zero a nove, ou uma letra, mas o mais usual é um número. No exemplo a seguir, apresentado por MARC (2004), os primeiros três dígitos são a etiqueta, 245 é definido como um campo de titulo, e os próximos dois dígitos (1 e 4) são os valores de indicador. O número um é o primeiro indicador e o número 4 o segundo, com o valor um no campo de titulo indicando que deve haver uma entrada de titulo separada no catalogo.

```
245 14  $a The empero's new clothes /
        $c adapted from Hans Christian Andersen
          and illustrated by Janet Stevens.
```

- Sub-campos: são marcados por códigos, códigos de sub-campos e delimitadores. A maioria dos campos contem vários pedaços relacionados de dados, cada tipo de dado dentro de um campo é denominado sub-campo, sendo cada sub-campo precedido por um código de sub-campo.

- Código de sub-campo: são representados por uma letra maiúscula precedido por um delimitador. Um delimitador é um caractere que separa os sub-campos. Cada código de um sub-campo indica que tipos de dados o sub-campo contém. O delimitador a ser utilizado depende do software utilizado para gerenciar os registros MARC, no exemplo é utilizado o $.

- Designadores de volume: é um termo abrangente, utilizado para se referir as etiquetas, indicadores e códigos dos sub-campos. Que é a chave para o sistema de anotação do MARC. Segundo MARC (2005), os três tipos de designadores de volume, são os símbolos que etiquetam e explicam o registro bibliográfico.

Este padrão é utilizado, por exemplo, nas seguintes bibliotecas digitais: biblioteca nacional do Canadá, biblioteca nacional da África do Sul, biblioteca Nacional da Escócia e na biblioteca da USP (mais especificamente no banco de dados bibliográfico Dedalus), entre outras (Vosgrau, 2005).

2.3.2 Padrão DCMI

O *Dublin Core Metadata Initiative* (DCMI) (DCMI, 2005) foi criado em 1995, como resultado de um *workshop* de metadados em Dublin, Ohio, patrocinado pelo *Online Computer Library Center* (OCLC), (OCLC, 2004) e pelo *National Center for Supercomputing Applications* (NCSA), (NCSA, 2004), com o objetivo de identificar e definir um conjunto simples de elementos metadados para descrever recursos na Internet.

O DCMI é composto por vários grupos de trabalho, cada qual responsável por discutir e apresentar soluções para uma área especifica de atuação. O grupo de trabalho em Bibliotecas, DCMI- *Libraries Working Group,* pesquisa a utilização do Dublin Core em bibliotecas digitais, e identificou os possíveis usos do Dublin Core em Bibliotecas Digitais (DCMI, 2005):

- Servir como um formato de intercâmbio entre os vários sistemas que usam formatos diferentes de metadados;
- Usar o Dublin Core para colher metadados de fontes de dados dentro e fora dos domínios da biblioteca;
- Apoiar a criação simples de registros de catálogos de biblioteca para recursos, dentre uma variedade de sistemas;
- Expor dados MARC a outras comunidades (convertendo o dado para o padrão DC);
- Permitir a aquisição de metadados de outras comunidades, que não sejam bibliotecas, mas que usam o DC.

Atualmente o padrão *Dublin Core* define um conjunto de 15 elementos metadados, que podem coexistir com outros padrões de metadados (NISO, 2001). Esses elementos são apresentados na tabela 3, e são documentados também pelos seguintes órgãos internacionais: ISO (ISO, 2003), NISO (NISO,2001) e CEN/ISS (*European Committee for Standardization/Information Society Standardization System*), (CEN, 2004).

Tabela 3-Descritores do padrão Dublin Core

Rótulo	**Definição**
Title	Um título dado ao recurso.
Subject	O assunto referente ao conteúdo do recurso, definido com palavras-chave ou tópicos.
Description	Uma breve descrição sobre o conteúdo do recurso.
Creator	Uma entidade principal responsável pelo conteúdo do recurso
Publisher	Um agente ou agência responsável pela disponibilização do recurso em sua forma atual, geralmente uma editora.
Contributor	Pessoas, além dos autores que contribuíram substancialmente para o conteúdo do recurso.
Date	Data da disponibilização do recurso em sua forma descrita.
Type	Tipo do objeto, por exemplo, um livro ou páginas Web.
Format	Formato que o recurso assume, geralmente um formato de arquivo, como PDF, HTML, MPEG.
Identifier	Uma cadeia de caracteres que identifica exclusivamente o recurso, por exemplo, ISBN e URLs.
Relation	Relacionamento, se existir, do recurso com outros recursos, normalmente descrito como parte de um conjunto maior.
Source	Outras fontes, se existirem, das quais o recurso se origina.
Language	O idioma no qual o recurso foi desenvolvido.
Coverage	A área geográfica que o recurso se engloba, se aplicável.
Rights	Direitos ou outras propriedades intelectuais especificando as condições através das quais o recurso pode ou não ser usado.

O Dublin Core é utilizado atualmente pelas seguintes instituições, entre outras: *Networked digital Library of Theses and Dissertations* (NDLTD, 2004), *Consortium for the Computer Interchange of Museum Information* (CIMI, 2004), e em nível nacional a biblioteca digital de Teses e Dissertação da USP (Universidade de São Paulo), (Rossetto, 2005).

2.3.3 Padrão Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

O padrão OAI-PMH (OAI, 2004), emergiu de uma reunião promovida em outubro de 1999, em Santa Fé, Novo México, por Paul Ginsparg, Rick Luce e Herbert Van de Somple, e patrocinada pelo *Council on Library and Information Resources* (CLIR), *Digital Library Federation* (DLF), *Scholarly Publishing & Academic Resources Coalition* (SPARC), *Association of Research Libraries* (ARL) e do *Los Alamos National Laboratory* (LANL).

O objetivo inicial da Convenção de Santa Fé, como ficou conhecida, foi o de promover e desenvolver normas de interoperabilidade visando facilitar a difusão eficiente de conteúdos na Internet. Surgiu como um esforço para melhorar o acesso a repositórios de publicações eletrônicas e e-prints[1], mas em seguida foi observado que havia necessidade da criação de um protocolo que proporcionasse o intercambio de formatos bibliográficos entre máquinas distintas (Weitzel, 2003).

Segundo (OAI, 2004), o protocolo OAI-PMH, é composto por duas classes de participantes:

- *Data Providers*, ou instituições provedoras de dados, que são bancos de documentos eletrônicos que oferecem acesso para publicação e armazenamento desses documentos e a sua disponibilização em um servidor conectado à Internet, e;
- *Service Provider*, ou instituições provedoras de serviço, que coletam metadados de um ou mais provedores de serviço.

A troca de mensagens entre o servidor do provedor de dados e o servidor do provedor de serviços para a transferência de metadados ocorre somente em uma direção de cada vez, ou seja, o provedor de serviços faz uma requisição ao provedor de dados, que responde enviando metadados. As requisições são enviadas através do protocolo http, usando comandos CGI, utilizando o método *Get* ou *Post*. Os metadados da resposta são codificados em XML (Marcondes, 2002).

[1] e-print se refere às versões eletrônicas de trabalhos de pesquisa que foram submetidos à revisão entre os pares, trabalhos apresentados em conferências, ou manuscritos que ainda não foram publicados (Weitzel, 2003).

O OAI-PMH, define o conjunto de metadados Dublin Core (*Dublin Core Metadata Element Set*) como um conjunto mínimo de metadados que devem ser suportado pelos provedores de dados em resposta a uma requisição de um provedor de serviços, embora o provedor de serviços possa oferecer outros formatos de metadados mais complexos como o formato MARC (Weitzel, 2003).

Segundo Marcondes (2002), cada documento armazenado no provedor de dados possui um registro formado por metadados. Cada registro está associado à um identificador único, que é formado por um identificador do provedor de dados mais um identificador do registro. O registro possui uma parte denominada *datestamp*, que indica a data da criação ou da última alteração sofrida pelo registro, permitindo a coleta automática dos metadados do provedor de dados a partir de uma determinada data,.

O protocolo define seis verbos que um provedor de serviços pode enviar a um provedor de dados para coletar metadados de documentos armazenados, eles são descritos na tabela 4.

Tabela 4-Verbos do padrão OAI-PMH

RÓTULO	**DESCRIÇÃO**
Identify	Obtém dados administrativos sobre o provedor de dados, política de publicação de documentos, etc.
ListSets	Lista as classificações sob as quais os documentos são organizados no provedor de dados.
ListMetadataFormats	Lista em que formato os metadados do provedor de dados podem ser apresentados.
ListIdentifiers	Lista os identificadores de registros do provedor de dados.
List Records	Lista os metadados dos registros por set ou por data.
GetRecords	Dado um identificador de registros, obtém os metadados armazenados neste registro.

2.3.4 Padrão ANSI/NISO Z39.50

O padrão Z39.50 (ISO, 29950 de 1997), foi proposto inicialmente, em 1984 pela NISO, com o propósito de prover um padrão de pesquisas em bases de dados bibliográficos (Rosetto, 1997). Segundo NISO/Z39.50 (2003), a norma é mantida pelo Z39.50 *Implementors Group* (ZIG). Trata-se de um protocolo de comunicação entre computadores, baseado na arquitetura cliente/servidor, e que permite pesquisa e recuperação de informação em redes de computadores.

O protocolo utiliza o conceito de sessão. O qual consiste inicialmente em uma negociação entre o cliente e o servidor, sendo em seguida enviada uma requisição ao servidor, que quando a recebe, realiza uma pesquisa nas suas bases de dados, originando um resultado (result Sets). Este resultado da pesquisa é mantido no servidor. Após a pesquisa o servidor envia em forma de resultado (Report) o número de resultados obtidos, os quais ficaram a disposição do cliente para posterior recuperação, e apresentação, após processados, ao usuário do sistema.

O Z39.50 é dividido em onze blocos estruturais básicos, denominados Facilidades, que são: *Initialiation, Search, Retrieval, Result-Set-Delete, Browse, Sort, Access Control, Accounting/Resource Control, Explain, Extended Services,* e *Termination*. Essas Facilidades podem se dividir em um ou mais serviços. Um serviço facilita um tipo de operação entre origem e destino, sendo selecionados pelo protocolo, de acordo com a necessidade para cumprir a sua função. Os três serviços mais básicos são apresentados na tabela 5 (Pistori, 2000).

Tabela 5-Os serviços mais básicos do Z39.50

Nome	**Descrição**
Initialiation	Primeira etapa em todo processo, é utilizado para negociar os outros serviços durante a sessão
Search	Permite a origem submeter pedidos ao destino, podem variar de requisições simples a pedidos booleanos.
Retrieval	Utilizado para controlar a maneira com que os resultados são retornados ao usuário.

A criação e o desenvolvimento de protocolos de interoperabilidade tém contribuído muito para o desenvolvimento das bibliotecas digitais. Conforme descrito por Rosetto (1997), ao se referir ao desenvolvimento dos protocolos: “Todos esses facilitadores têm como finalidade única e vital para sociedade a cooperação e o compartilhamento de recursos informacionais”.

2.4 Projetos de Bibliotecas Digitais

Esta seção apresenta alguns projetos de bibliotecas digitais.

2.4.1 Arquitetura para Informações em Bibliotecas Digitais de (ARMS, 1997)

Esta arquitetura foi proposta por Arms (1997), tomando por base o Programa de desenvolvimento da Biblioteca Digital Nacional (NDLP) da Biblioteca do Congresso americano, que se trata de um projeto para converter coleções históricas para a forma digital e as tornar disponíveis na Internet.

Um protótipo foi apresentado aos membros da NDLP em julho de 1996. Este protótipo incluía uma variedade de fotografias e textos, compilados do próprio NDLP e convertidos para a forma digital.

A figura 1 apresenta os componentes básicos da arquitetura proposta (ARMS, 1997).

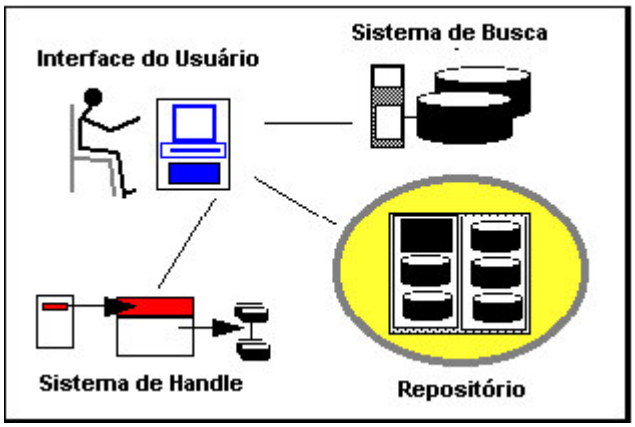

Figura 1-Principais Componentes da Arquitetura da Biblioteca Digital (ARMS, 1997)

Nesta arquitetura pode-se verificar os seguintes componentes:

- Interfaces com o usuário: são duas interfaces, uma para o usuário da biblioteca e outro para os administradores da biblioteca que gerenciam a coleção. Elas são páginas *Web* interpretadas por qualquer navegador *Web*. Os navegadores conectam aos "serviços clientes", que fornecem funções intermediárias entre o navegador e as outras partes do sistema. Os serviços-cliente permitem ao usuário decidir onde buscar e o que acessar; ele interpreta informações estruturadas como objetos digitais (componentes da coleção); ele negocia termos e condições, gerencia relacionamentos entre objetos digitais, lembra o estado da interação, e converte entre os protocolos usados pelas várias partes do sistema.

- Repositório: armazena e gerencia objetos digitais e outras informações. Uma grande biblioteca digital pode ter vários repositórios de vários tipos, incluindo repositórios modernos, BD, servidores *Web*. A interface para este repositório é chamada protocolo de acesso ao repositório (RAP - *Repertory Access Protocol*). Características do RAP são reconhecimento explícito de direitos e permissões que necessitam ser satisfeitos antes do cliente acessar o objeto digital, suportar uma grande faixa de disseminações de objetos digitais, e uma arquitetura aberta com interfaces bem definidas.

- Sistema *Handle*: *Handles* são identificadores únicos de propósito geral que podem ser usados para identificar objetos digitais e gerenciar objetos armazenados em qualquer repositório ou BD. Um *handle* faz parte do metadado que descreve o objeto digital. O sistema *handle* é um sistema computacional que fornece um serviço de diretório distribuído para identificadores (*handles*) para recursos Internet. Quando usado com repositórios, o sistema *handle* recebe como entrada um identificador para um objeto digital e retorna o identificador do repositório onde o objeto está armazenado.

- Sistema de Busca: o projeto do sistema da biblioteca digital assume que haverá vários índices e catálogos que podem ser procurados para descobrir a informação antes de obtê-la de um repositório.

Para entender as funções destes componentes, será apresentado um exemplo de busca de uma informação:

- O primeiro passo é procurar a informação, neste caso uma determinada fotografia digitalizada. Os serviços-cliente fornecem ao usuário um formulário para busca via navegador. O usuário preenche o formulário com uma consulta de busca (*search query*), perguntando pela fotografia. O formulário completado é enviado aos serviços-cliente. Estes transladam a questão nos formatos e protocolos requeridos pelo sistema de busca. O sistema de busca pode usar o Z39.50, por exemplo. Os serviços-cliente conduzem uma seção Z39.50 com o sistema de busca e obtêm a lista dos objetos digitais que satisfazem a pergunta. Cada objeto digital é identificado por seu *handle*.

- O segundo estágio é a seleção, pelo usuário, de uma fotografia digitalizada para ver. Os serviços-cliente apresentam ao usuário, via navegador, a lista de objetos digitais encontrados através do sistema de busca (atualmente como uma página HTML com *links* selecionáveis por mouse). O usuário seleciona a fotografia desejada.

- O terceiro estágio é a recuperação da fotografia digitalizada. Os serviços-cliente enviam o *handle* da fotografia escolhida para o sistema *handle*, que retorna ao endereço do repositório. Os serviços-cliente passam o *handle* para o repositório usando o protocolo RAP. Várias versões da fotografia podem estar armazenadas no repositório como um conjunto de objetos digitais, identificados pelo *handle*. Os serviços-cliente selecionam um, talvez um pequeno *preview* e pedem este ao repositório. Todas as transações RAP passam através de termos explícitos e passos de condições. Verificação em termos e condições associadas com esse objeto digital podem necessitar negociação entre o serviços-cliente e o repositório, ou interação direta com o usuário.

- Finalmente, a fotografia digitalizada que foi escolhida é transmitida pelo repositório, via serviços-cliente, para o navegador do usuário e apresentado na sua tela.

2.4.2 Biblioteca Digital de Berkeley (OGLE, 1996)

A Biblioteca Digital de Berkeley é um dos principais projetos da Biblioteca Digital da Universidade da Califórnia, integrante da NSF/NASA/ARPA – *Digital Library Initiative* (DLI, 1998).

Em Ogle (1996), é apresentada a arquitetura da Biblioteca, onde todo o acesso é provido via protocolo HTTP. Esta biblioteca digital não permite a transferência tempo-real de áudio e vídeo. Como é mostrado na figura 2, o mecanismo CGI (*Common Gateway Interface*) é usado para permitir a interação entre os clientes *Web* e os sistemas. Entre estes sistemas está um servidor de banco de dados relacional, que permite o acesso baseado em formas a quase todos os dados da biblioteca digital. Outros métodos além de formas são disponíveis para acessar o dado, como *links* e listas organizadas. Além deste método muitos outros são disponíveis via matriz de acesso, que fornece um ponto de acesso de alto nível para todos os dados da biblioteca.

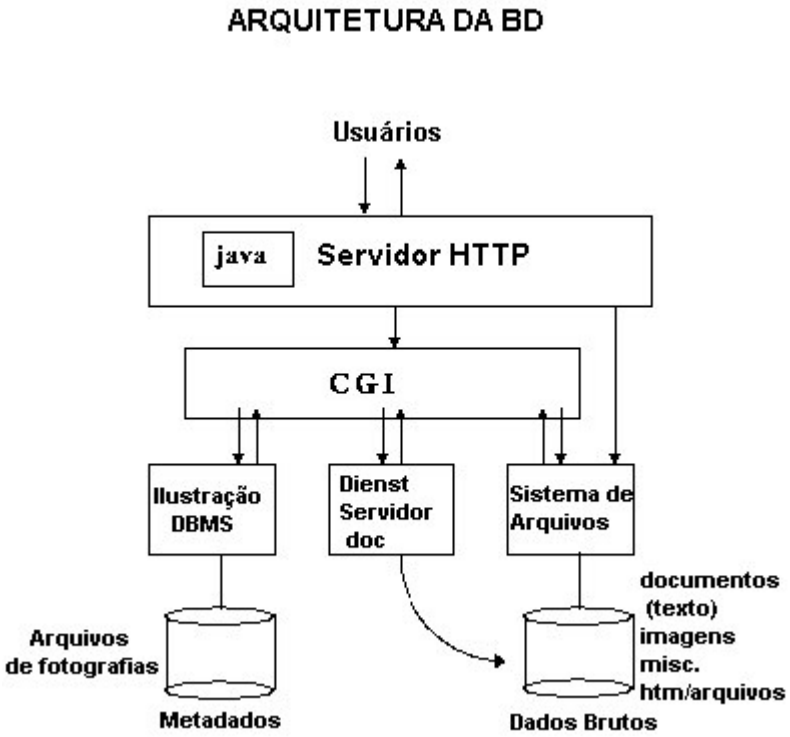

Figura 2-Arquitetura da Biblioteca Digital de Berkeley (OGLE, 1996)

Na biblioteca digital de Berkeley, os documentos são recebidos em papel. Destes são extraídos metadados bibliográficos tais como título, autor e data de publicação. O documento é

escaneado para obter a imagem do papel. Um software de OCR (*Optical Character Reader*) é usado nas imagens para obter um texto ASCII (*American Standard Code for Information Interchange*) junto com informação de localização de palavra (arquivos XDOC). Os metadados são armazenados em um banco de dados relacional e então as imagens da página, texto e XDOC são arquivados no sistema de arquivos.

2.4.3 Biblioteca Digital Brasileira (BDB)

Segundo Marcondes (2002), a Biblioteca digital Brasileira tem dois objetivos principais. O primeiro é estimular a publicação da Ciência e Tecnologia (C&T) na Internet, e dentro deste objetivo permitir que a comunidade de C&T brasileira tenha os meios para a publicação diretamente na Internet e dar mais visibilidade tanto nacionalmente como internacionalmente a esta comunidade. O outro objetivo é o de viabilizar o acesso rápido e integrado aos recursos disponibilizados pela comunidade cientifica, seja facilitando a descoberta de recursos de informação brasileiros em C&T na Internet de forma integrada, seja encurtando o ciclo de comunicação entre a comunidade de C&T.

A figura 3 apresenta a arquitetura da Biblioteca Digital Brasileira, e em seguida é feita uma descrição dos seus principais componentes.

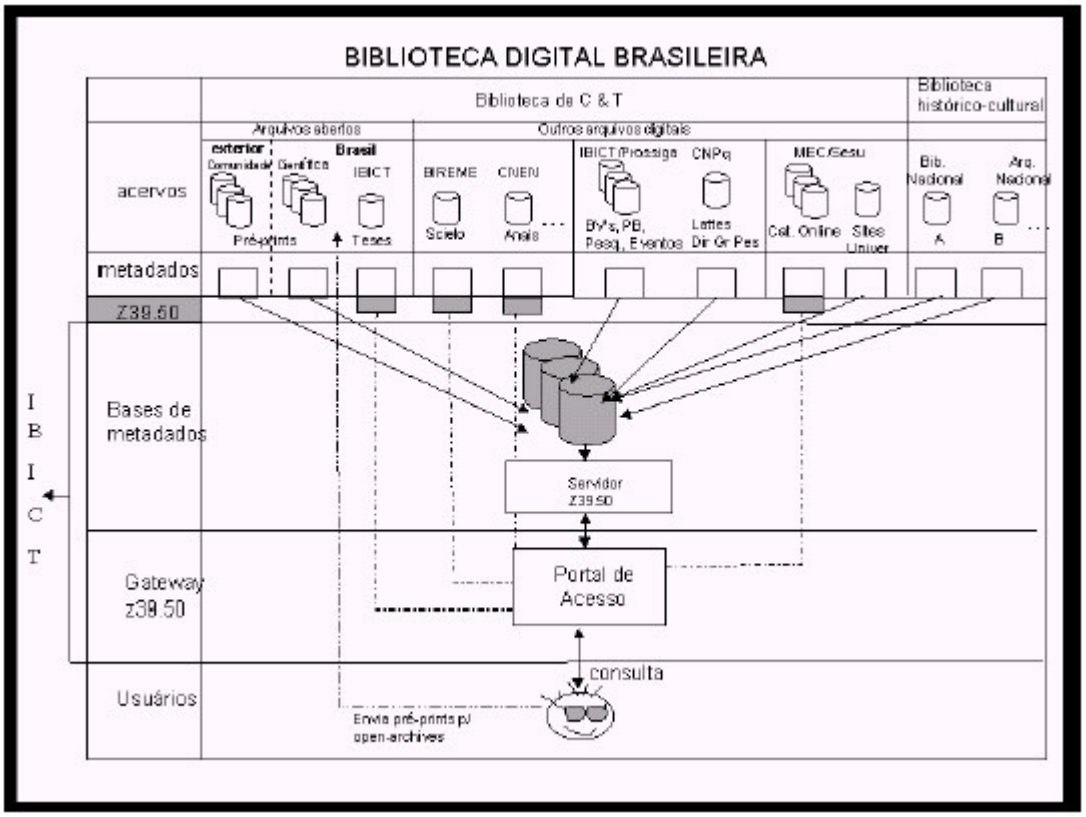

Figura 3-Arquitetura da BDB (IBICT, 2001)

O acervo da BDB deve ser composto por documentos digitais formados por textos completos, imagens, sons etc. Os acervos são agrupados em duas bibliotecas, de acordo com a sua natureza (IBICT, 2001):

- Biblioteca de C&T: formada por organizações que armazenam ou disponibilizam documentos e serviços da área da ciência e tecnologia. Como exemplos temos a FAPESP, com a base de dados de periódicos brasileiros, a CNEN/CIN, base de dados de textos completos de trabalhos de congressos, o MEC, unificação de Catálogos de Acesso Público On-line das bibliotecas universitárias brasileiras com catálogos na Internet, segundo algum protocolo padronizado.
- Biblioteca Histórico-Cultural: formada pelas instituições que atuam na guarda e preservação de documentos históricos e das que armazenam documentos artísticos e culturais da sociedade brasileira. Esta biblioteca é de responsabilidade da Biblioteca Nacional e do Arquivo Nacional.

Quanto à interoperabilidade, a BDB define duas formas pelas quais as instituições provedoras de dados ou de serviços de informação se integrarão aos seus acervos: via base comum de metadados, as instituições que não utilizam o Protocolo Z39.50 devem gerar metadados, que devem ser ao menos compatível com o padrão Dublin Core; e via Protocolo Z39.50.

Segundo Marcondes (2002), a BDB adota os seguintes padrões:

- Acesso dos usuários aos recursos da BDB via Protocolo Z39.50;
- Formato da base comum de metadados em Dublin Core, Open Arquivos;
- Intercâmbio de documentos eletrônicos via XML, e;
- Para digitalização de documentos o padrão TIFF de 24 bits.

A arquitetura também prevê um Gateway de serviços Z39.50, para acesso aos serviços. Trata-se de um portal de acesso a serviços e dados via Web, através do qual são combinados recursos de busca de diferentes provedores.

2.4.4 Biblioteca Digital de Literatura do Projeto SIDIE

O projeto SIDIE (SIDIE, 2001), tem como objetivo a construção de uma infra-estrutura de digitalização, organização e disponibilização de conteúdos digitais voltados ao ensino e pesquisa. O sistema constitui-se de um sistema de biblioteca digital distribuída, composto de diversas bibliotecas digitais, operando de forma integrada, formando um grande centro de documentação, direcionado ao ensino. Trata-se de uma evolução da biblioteca digital Multimídia (BDMm) (Pistori, 2000) e da mesma forma, foi implementada usando a plataforma web código aberto LAMP (Linux, Apache, MySQL e PHP). Esta biblioteca adotou os metadados Dublin-Core (DCMI, 2005) e o protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) (OAI, 2004) como padrões de bibliotecas digitais.

Como parte do projeto SIDIE, foi desenvolvida a Biblioteca Digital de Literatura Brasileira e Catarinense (LBC). Esta biblioteca oferece um vasto acervo de obras literárias brasileiras. A arquitetura da LBC é apresentada na figura 4. Os principais componentes desta arquitetura são:

- Módulo de Interface: composto pelas Interfaces de Usuário, Interface de Catalogação de obras/autores, e Interface de Administração. A interface do usuário permite ao usuário da LBC realizar buscas e navegações no acervo.
- Gerenciador de Biblioteca Digital (GBDM): responsável pela geração dinâmica das interfaces da biblioteca e implementação de mecanismos de busca e acesso às obras, além de fornecer ferramentas para a administração da biblioteca.
- Base de Metadados: metadados Dublin Core e outros relativos às obras são armazenados neste módulo, sendo implementado por meio de um SGBD. Observa-se que além dos metadados Dublin-Core, a LBC conta com metadados adicionais, que visam atender às necessidades específicas da biblioteca LBC.
- Servidores de Mídia: As mídias das obras catalogadas na biblioteca são armazenadas nos Servidores de Mídia.
- Data Provider: Mecanismo responsável por tratar as requisições OAI-PMH feitas pelo harvester, e conseqüentemente a realização da busca no Repositório de Metadados e geração das respostas usando metadados Dublin-Core.

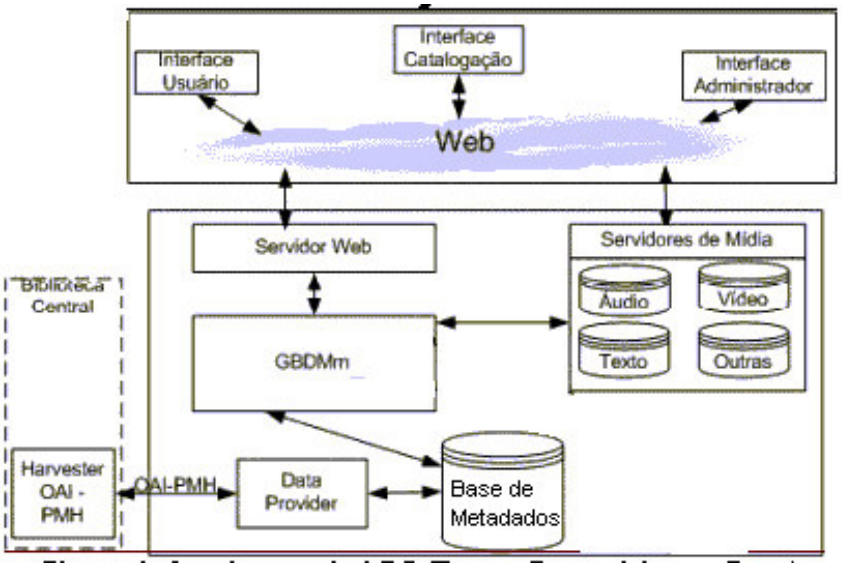

Figura 4-Arquitetura Biblioteda SIDIE (SIDIE, 2001)

Com relação ao sistema de recuperação, implementado no GBDM, a busca de obras é feita através de três modos oferecidos pela Interface Usuário: busca simples, busca por obra e busca por autor. Em todos os modos o usuário envia um conjunto de palavras que representam sua

consulta, e o sistema pesquisa na base de metadados, tentando localizar documentos que contenham as palavras informadas para consulta, sendo gerada uma página de resposta com os resultados obtidos.

A figura 5 apresenta a interface de consulta simples da BD, onde pode ser escolhido para pesquisar por qualquer palavra, frase exata, todas as palavras (disponível por intermédio de menus *pop-down*). Podem ser realizadas buscas por autor, selecionado através de um menu, por gênero ou uma combinação das três opções, palavra-chave, autor e gênero.

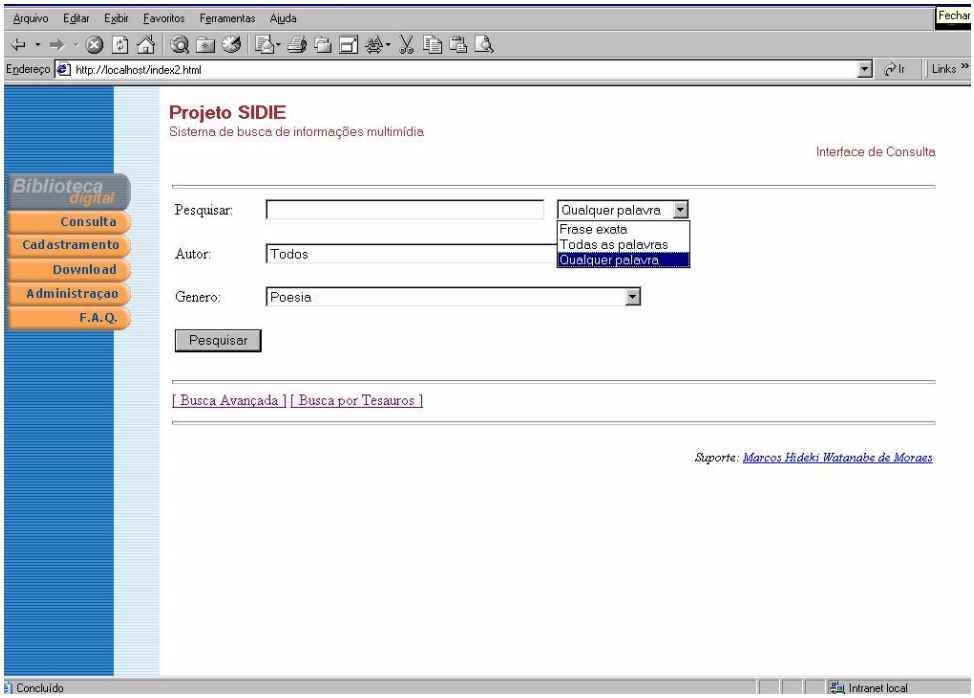

Figura 5-Tela de Busca simples da BD-SIDIE

A figura 6 apresenta a tela de Busca Avançada, onde pode ser escolhido realizar busca por palavra-chave em título, personagem, descrição , fatos históricos ou somente por uma palavra-chave, combinadas por operadores booleanos "ou" e "e", com mais três opções de palavras - chaves com as mesmas opções. Pode ainda, ser selecionado um autor, um gênero um idioma, a ordenação (por título, autor ou data) e a quantidade de resultado por página (5, 10, 20 ou 30). Todas as opções podem ser também, combinadas em uma única consulta.

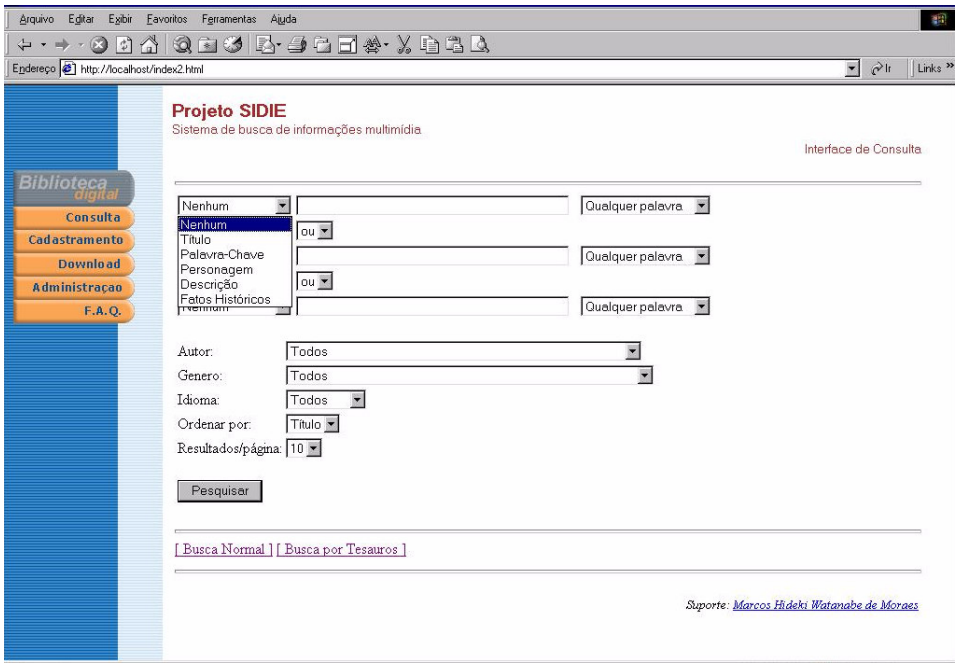

Figura 6-Tela de Busca Avançada – BD-SIDIE

A figura 7 é a tela de Busca por assuntos, segundo um thesaurus temático. A busca é realizada no dicionário temático em cinco níveis. Seleciona-se o tema do primeiro nível (entre Vida Interior, Opiniões, Valores Morais, Vida Social e Mundo Exterior) e os demais níveis são derivados da escolha inicial por um deles.

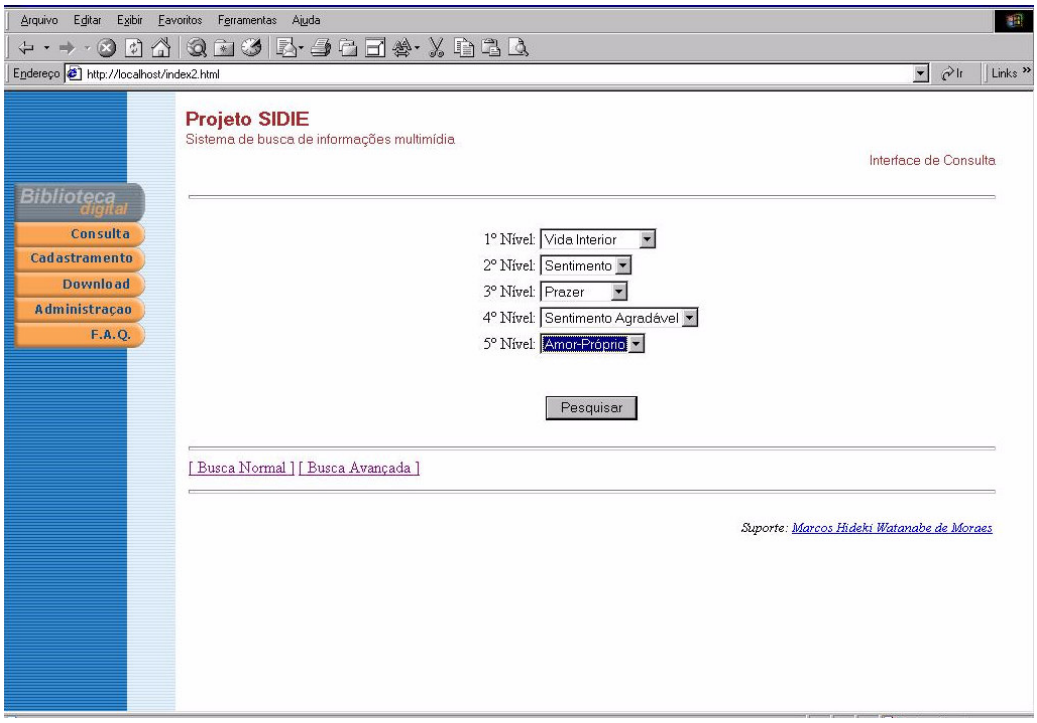

Figura 7-Tela de Consulta por Thesaurus

A Figura 8 apresenta a tela de resultado de busca. A partir desta interface, pode-se consultar as obras e os respectivos autores que atenderam ao critério de busca.

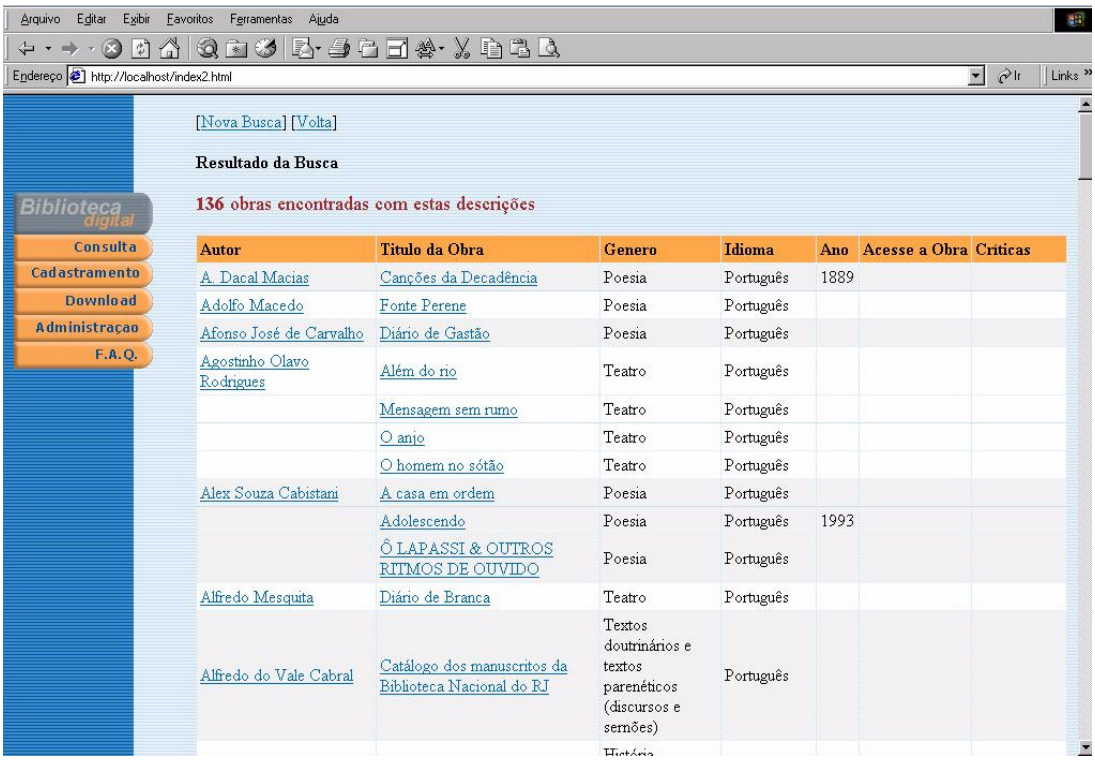

Figura 8-Tela de Resultado da Busca simples

2.4.5 Trabalhos Relacionados Realizados no Contexto do Projeto SIDIE

O Projeto SIDIE tem sido a base para a realização de várias pesquisas na área de Bibliotecas Digitais dentre os principais e que serviram de base para este trabalho podem-se relacionar:

- Silva (2003), faz uma análise da evolução das bibliotecas tradicionais até as bibliotecas digitais, dos problemas envolvidos na digitalização e recuperação de documentos, e faz uma análise de requisitos funcionais básicos para uma biblioteca Digital adaptativa.
- Antunes (2003), pesquisa a parte de recuperação de documentos em Bibliotecas Digitais armazenados em bancos de dados relacionais, por intermédio de um tradutor de consultas XML/ Dublin Core para sentenças SQL. A ênfase deste

trabalho está na interoperabilidade entre bibliotecas digitais, através da adoção de padrões para metadados.

- Letícia (2003) traz a proposta de uma arquitetura de recuperação de informações usando os recursos de hiperdocumentos adaptativos aplicado à bibliotecas digitais.

2.5 Resumo

O capítulo procurou apresentar uma visão geral das Bibliotecas Digitais. Foram apresentados alguns dos principais protocolos de interoperabilidade, um fator de suma importância para a atualização da biblioteca, e para os serviços a que ela se destina. Também foram apresentadas algumas arquiteturas clássicas de BD citadas na maioria dos trabalhos sobre BD e que devem ser tomadas como ponto de partida para o estudo mais aprofundado da mesma, bem como a arquitetura da BD brasileira, e um pouco mais detalhadamente a BD do Projeto SIDIE, que foi a base para esta pesquisa.

O próximo capitulo trata de Hiperdocumentos adaptativos, uma técnica utilizada para prover personalização de conteúdos aos usuários em sistemas de informação.

Capítulo 3. Hiperdocumentos Adaptativos

A Web vem tornando-se uma fonte inesgotável de conteúdos nas mais diversas áreas do conhecimento humano. Estes conteúdos encontram-se muitas vezes de forma desorganizada, dificultando a busca por informações. Deve-se ainda levar em conta que muitos usuários possuem diferentes níveis de conhecimento, sobre a própria tecnologia web e sobre o conteúdo procurado. Maior parte dos sistemas de busca e de navegação não leva em consideração os conhecimentos e anseios do usuário utilizando o sistema de busca. Com isto é apresentado um resultado da busca baseado exclusivamente no critério de busca.

Neste contexto surge a idéia do desenvolvimento de sistemas de Hiperdocumento Adaptativo (HA), que busca adaptar os recursos originados de uma fonte de informação ao perfil de seus usuários.

Neste capítulo apresenta uma revisão bibliográfica sobre HA, por ser tratar da tecnologia que é utilizada no desenvolvimento do sistema de recuperação de informações aplicado a bibliotecas digitais proposta nesta dissertação.

3.1 Definição de HA

Segundo Palazzo (2002), Hiperdocumento Adaptativo é a área da ciência da Computação que se ocupa do estudo e desenvolvimento de sistemas, arquiteturas, métodos e técnicas capazes de promover a adaptação de hiperdocumentos e hipermídia em geral às expectativas, necessidades, preferências e desejos de seus usuários.

De uma forma geral, os sistemas e modelos de HA têm por objetivo disponibilizar informações atualizadas a seus usuários, com um conteúdo interessante, em um tamanho e profundidade adequados ao contexto e em correspondência direta com um modelo do usuário. Este modelo funciona como referência para o sistema, que busca adaptar o hiperespaço, muitas vezes caótico, aos anseios e expectativas particulares de seus usuários (Brusilovsky, 1996). Assim, usuários com metas e conhecimentos diferentes podem se interessar por partes diferentes da informação apresentada em um documento os quais podem oferecer links diferentes de

navegação de acordo com o nível do usuário.O objetivo de um sistema de HA é portanto oferecer a cada usuário uma interface modelada de acordo com suas características específicas. Usuários devem acessar interfaces cujo estilo, conteúdo, recursos e links serão dinamicamente selecionados, entre diversas possibilidades, reunidos e apresentados a eles conforme seus objetivos, necessidades, preferências e desejos.

Palazzo(2002) acrescenta o fato de que a pesquisa em HA situa-se na fronteira dos estudos em hiperdocumentos e modelagem do usuário, sendo estes os dois pilares básicos que sustentam o desenvolvimento de aplicações nesta área. O conceito de HA está intimamente ligado as tecnologias de modelagem de usuários e grupos, bancos de dados, programação distribuída na Web, métodos colaborativos e interface dinâmicas adaptativas.

Um sistema de HA deve satisfazer a três critérios básicos: ser um sistema hiperdocumento (hipertexto ou hipermídia); possuir um modelo do usuário; e ser capaz de adaptar o hiperdocumento usando tal modelo.

A figura 9 ilustra a visão clássica do laço de adaptação do sistema ao modelo do usuário.

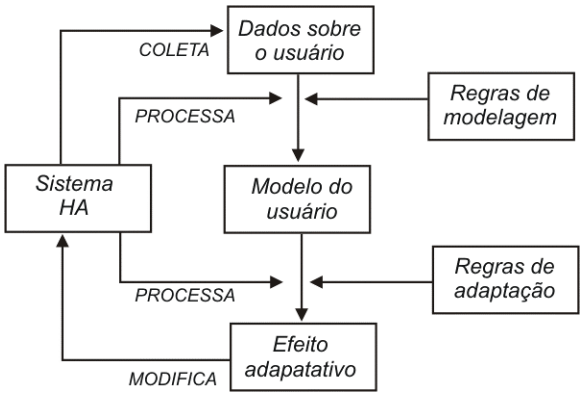

Figura 9-Laço Clássico - Modelo do Usuário - Adaptação (Palazzo, 2000)

Segundo Brusilovsky (1996), os sistemas de HA podem ser encontrados em seis áreas de aplicação: sistemas de recuperação de informações, sistemas de informações, sistemas de ajuda (*help*) online, sistemas educacionais, hipermídia institucional e personalização de visões em espaços de informações.

3.2 Tipos de adaptação

Os sistemas hipermídia são constituídos, basicamente, por um conjunto de nodos ou hiperdocumentos conectados por links. Cada nodo contém alguma informação local e links para outros nodos relacionados. Neste contexto, a adaptação pode ocorrer em nível do conteúdo dos nodos ou em nível dos links. Estes níveis compõem duas classes diferentes de HA(Brusilovsky, 1996): Apresentação Adaptativa e Navegação Adaptativa, conforme apresentado na Figura 10.

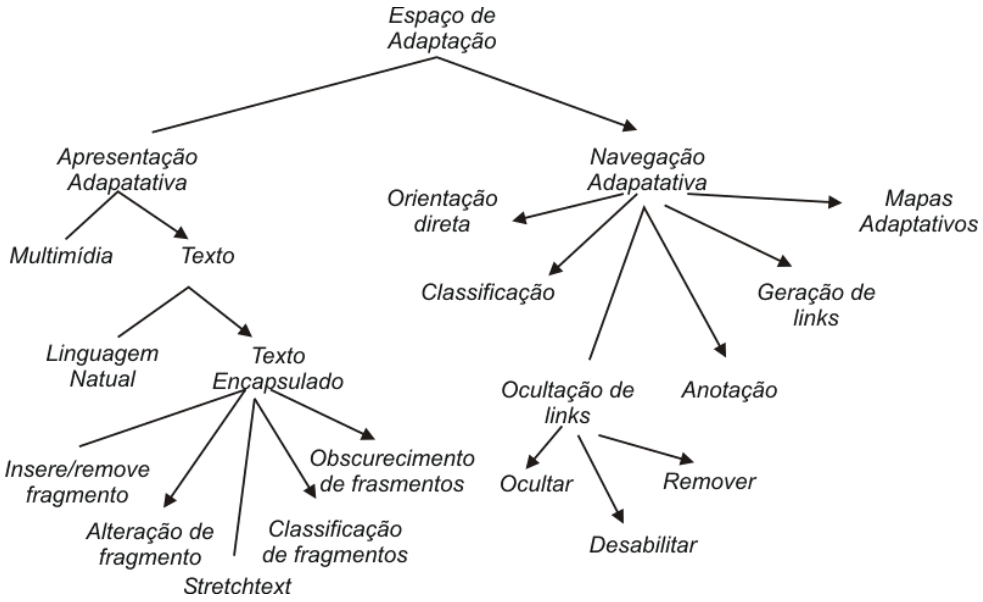

Figura 10-Espaços de Adaptação em HA (Brusilovski, 2001)

3.2.1 Apresentação Adaptativa

O objetivo desta classe de HA é a adaptação do conteúdo do nodo ao conhecimento, objetivos e demais características, de acordo com o modelo do usuário. Existem técnicas de apresentação adaptativa de textos e de apresentação adaptativa de objetos multimídia. Geralmente em textos a adaptação ocorre em termos das possíveis modificações a que estes podem ser submetidos antes de serem apresentados ao usuário

No caso de objetos multimídia geralmente o usual é selecionar os objetos a serem apresentados dentre certa quantidade de opções. A adaptação aqui refere-se a escolha de um meio específico vídeo, áudio, animação, etc, entre os diversos possíveis para apresentar um mesmo conteúdo.

As técnicas mais conhecidas de Apresentações adaptativas são (Palazzo, 2002):

- Explicação Adicional: é um método muito difundido para apresentação de conteúdos adaptados e, consiste em ocultar do usuário alguma parte da informação sobre um determinado conceito, que não seja importante para o nível de conhecimento ou interesse do usuário.
- Explicação requerida: funciona na forma de pré-requisitos onde a informação dos assuntos é previamente ordenada, e a apresentação da próxima depende da apresentação da anterior. Assim, ao apresentar explicação de um conceito, o sistema insere a explicação de todos os conceitos requeridos para o seu entendimento.
- Explicação comparativa: verifica se os conceitos apresentados possuem semelhanças entre si, e quando da apresentação de um conceito semelhante, o usuário recebe uma explicação comparativa entre os dois, realçando os pontos em comum e as diferenças.
- Explicação Variante: assume que ocultar ou mostrar partes da informação nem sempre é adequado para uma adaptação, visto que os usuários podem necessitar de informações diferentes. Esse método armazena diversas variáveis para alguns dos conteúdos de uma pagina e o usuário obtém a apresentação de acordo com o seu modelo previamente determinado.
- Classificação de Fragmentos: leva em conta o nível de conhecimento e a experiência do usuário, e consiste em ordenar fragmentos de informação sobre o conceito, de maneira que a informação mais importante para o usuário, de acordo com o seu modelo, seja apresentada primeiramente.
- Representação por *frames*: utiliza um *frame* para apresentar a informação sobre um determinado conceito. Os *frames* podem ser formados por explicação variante (EV) sobre o conceito, *links* para outros *frames*, exemplos, etc, e são selecionados através de regras especiais de apresentação a partir do modelo do usuário.

3.2.2 Navegação Adaptativa

A Navegação Adaptativa tem por objetivo ajudar os usuários a encontrar seus caminhos no hiperespaço, através da adaptação da forma de apresentar os links na rede hipermídia.

Os principais métodos de suporte à navegação são os seguintes:

- Condução Global: consiste em ajudar o usuário a encontrar o caminho mis curto até a informação que ele procura, minimizando os possíveis desvios. A Condução Global ocorre quando o objetivo de informação do usuário é global, isto é, se encontra em um ou mais nodos que estão em algum lugar do hiperespaço e o sistema conduz o usuário até este local.
- A maneira mais direta é oferecer ao usuário em cada passa da navegação, *links* mais apropriados para que ele chegue a informação que procura a partir do nodo em que se encontra.
- Condução Local: O objetivo da Condução Local é semelhante ao da Condução Global, porém de alcance muito menor. Enquanto que a Condução Global preocupa-se com seqüências de links que conduzem ao objetivo desejado, a Condução Local ocupa-se de um único passo e tenta sugerir ao usuário os links mais relevantes considerando suas preferências, conhecimento e experiência. Por exemplo, um método para condução local é classificar os links de acordo com as preferências do usuário.
- Apoio a Orientação Local: consiste em ajudar o usuário a entender sua localização na rede do hipertexto local. Geralmente são utilizados dois métodos: informação adicional sobre os nodos, que podem ser acessados a partir do nodo corrente, ou limitando a navegação do usuário.
- Esses métodos geralmente são baseados na técnica de ocultação, ou seja, removem da visão do usuário toda a informação que não seja relevante aos seus objetivos imediatos.
- Apoio a Orientação Global: consiste em ajudar o usuário a entender a estrutura do hiperespaço que forma o domínio da navegação do sistema.

- Nos sistemas que não são adaptativos, por exemplo, são oferecidos mapas globais, com a localização do usuário em relação ao contexto global.
- Já nos sistemas de HA são usadas as técnicas de ocultação e de anotação, para ajudar o usuário a se orientar. O método mais utilizado é aumentar gradualmente o número de *links* visíveis a metida em que a experiência do usuário no hiperespaço vai aumentando.
- Gerenciamento de Visões Personalizadas: consiste em uma forma de construir interfaces de trabalho personalizadas por meio de adaptação. Estas visões são necessárias em ambientes dinâmicos na Web, onde links podem aparecer, desaparecer e evoluir.

A seguir são apresentadas as técnicas de Navegação Adaptativa, que são utilizadas para implementar os métodos de Suporte à Navegação visto anteriormente:

- Orientação Direta: Consiste em decidir, em cada ponto da navegação, qual o melhor nodo a ser visitado a seguir, levando em conta os objetivos, preferências, conhecimento e outros parâmetros representados no modelo do usuário. Para oferecer orientação direta o sistema pode destacar visualmente o link para o melhor nodo ou apresentar um link dinâmico adicional conectado ao melhor nodo selecionado, como os botões *Done*, ou Próxima.
- Técnica de Classificação Adaptativa: Trata-se da classificação de todos os links partindo de um nodo de acordo com a sua relevância, calculada sobre o modelo do usuário. Os links são apresentados então em ordem decrescente desta relevância. Segundo Brusilovsky(1996), esta técnica possui uma aplicação limitada, pode ser aplicada satisfatoriamente com links não contextuais, mas seu uso com índices e tabelas de conteúdos é muito difícil, e não podendo ser usada para links contextuais e para mapas. Essa técnica se presta para recuperação de informações e em sistemas de documentação on-line. (Palazzo, 2002).
- Técnica de Ocultação: Consiste em limitar o espaço de navegação ocultando os links menos relevantes ao usuário de acordo com o pré-estabelecido no seu modelo.

- Técnica de anotação: Mantém uma ordenação estável dos links deixando os mais relevantes destacados.
- Mapas adaptativos: Consistem em adaptar de várias maneiras os mapas de hipermídia global e local que podem ser apresentados ao usuário.

3.3 Modelagem do usuário

Nos primeiros sistemas de HA a modelagem do usuário era executada de forma integrada aos demais componentes, não havendo uma distinção entre os componentes do sistema que realizavam a modelagem do usuário e aqueles que executavam outras tarefas. A separação entre os processos de adaptação e modelagem do usuário somente foi estabelecida a partir de 1985 com as publicações de Kobsa (1985), Sleeman (1985), Kass (1988) e Allgayer et al. (1989).

Em 1990, com o artigo de Kobsa (1990), surgiu a expressão sistema de modelagem de usuários, derivada dos sistemas especialistas da IA (inteligência artificial). Esses sistemas utilizavam uma interface de programação (*Shell*) que permitia a configuração de modelos de usuários genéricos para diversas finalidades.

Os principais serviços disponibilizados por um Shell para a modelagem do usuário e que devem ser levados em contas em toda modelagem de usuário são (Kobsa, 1995):

- Representação de fatos sobre um ou mais tipos de características de usuários em modelos de usuários individuais, como fatos sobre seu conhecimento, objetivos, planos, preferências, tarefas e habilidades;
- Representação de características comuns a usuários participantes de subgrupos (estereótipos) no escopo do sistema;
- Classificação dos usuários em subgrupos, e inserção das características do grupo no modelo de cada usuário;
- Registro do comportamento do usuário, sua interação com o sistema;
- Formulação de hipóteses sobre o usuário com base no seu histórico de interações com o sistema;

- Formulação de estereótipos de usuários;
- Inferir novas hipóteses sobre os usuários baseado em fatos iniciais;
- Manutenção da consistência do modelo;
- Apresentar justificativas para as hipóteses assumidas;
- Verificar novas informações presentes no modelo do usuário comparando com informações anteriores.

Para poder oferecer esses serviços um Shell usado para modelagem de usuários deve satisfazer os requisitos de generalidade, isto é deve poder ser usado pelo maior número de aplicações possíveis e em cada aplicação oferecer várias possibilidades de configuração para a modelagem do usuário. Também deve satisfazer requisitos de expressividade, devendo ser capaz de expressar tantos tipos de fatos e hipóteses sobre o usuário quanto possível e ao mesmo tempo. Deve ter capacidade inferêncial, ou seja, conseguir realizar todos os tipos de raciocínio clássicos da inteligência artificial e da lógica formal, como cálculo de predicados de primeira ordem, raciocínio modal complexo, raciocínio com incerteza, raciocínio plausível (quando da ausência de informações) e resolução de conflitos, quando forem detectadas hipóteses contraditórias.

3.4 Obtenção do Perfil do Usuário

Conforme Brusilovsky (1998), existem ao menos cinco características associadas a um usuário que podem ser levadas em conta por um sistema adaptativo: seu conhecimento; objetivos; história; experiência; e preferências. Essas características são de alguma maneira dinâmicas, forçando o modelo do usuário a ter que se ajustar continuamente para garantir sua permanente atualização. O restante desta seção descreve cada uma destas características.

3.4.1 Conhecimento

O conhecimento do assunto do usuário sobre o assunto representado no hiperespaço é uma das características mais importante do usuário para os sistemas HA. Quase todas as técnicas de apresentação adaptativas se baseiam no conhecimento do usuário como fonte de adaptação. O

conhecimento do usuário é uma variável particular de cada usuário. Assim, um sistema de HA baseado no conhecimento de usuário deve estar apto a reconhecer mudanças no estado de conhecimento do usuário e atualizar o seu modelo de usuário de forma adequada com o que está acontecendo.

As técnicas mais comumente usadas para modelar o conhecimento do usuário são a sobreposição conceitual e os estereótipos. No modelo de sobreposição conceitual para cada conceito do modelo de domínio, uma camada no modelo individual é acrescido com um valor estimado do nível de conhecimento do usuário sobre este conceito. Podendo ser um valor binário (não-conhecido conhecido), uma medida qualitativa (bom-médio-pobre), ou uma medida quantitativa, como uma probabilidade do que o usuário sabe sobre o assunto (Brusilovisky, 1996)

No modelo de estereótipo existe uma tentativa de generalizar o conhecimento do usuário. O sistema pode ter um conjunto de esteriótipos para um determinado conjunto de conhecimentos de usuários, como o exemplo citado em Brusilivisky (1996) do MetaDoc que usa estereótipos como: Novato, intermediário e perito. O modelo de estereótipos é mais simples e menos poderoso, mas em contrapartida é mais fácil de inicializar e manter.

Podem ser alcançados bons resultados combinando estereótipo e modelo de sobreposição conceitual, utilizando o modelo de estereótipo para classificar um usuário novo e determinar valores iniciais e em seguida aplicar um modelo de sobreposição conceitual.

3.4.2 Objetivos

O objetivo do usuário é uma característica relacionada com o contexto do trabalho de hipermídia que o mesmo está realizando. Dependendo do tipo de sistema, pode ser o objetivo do trabalho (em sistemas de aplicação) uma meta de busca (em sistemas e recuperação de informação) ou a resolução de um problema ou aprendizado (sistemas educacionais). Em todos estes casos, a meta é uma resposta para a pergunta: Por que o usuário está usando o sistema de hipermídia e o que o usuário deseja alcançar? A meta do usuário é a característica mais mutável do usuário e quase sempre muda de sessão para sessão, geralmente mudando várias vezes dentro de uma mesma sessão de trabalho (Busolovisky, 1996).

3.4.3 História e Experiência

História e experiência são duas características semelhantes ao conhecimento do usuário, mas que diferem funcionalmente (Pallazzo, 2002): história significa a experiência anterior do usuário não relacionada ao assunto tratado pelo sistema HA, mas que deve ser levada em conta; a experiência do usuário está relacionada com a familiaridade do usuário em navegar pelo hiperespaço em questão.

Estas características individuais de um usuário como história ou experiência normalmente são modeladas também por um modelo de estereótipo de usuário.Um modelo de estereótipo de usuário é um modelo que trata o usuário através da classe que ele pertence, por exemplo um usuário poderia ser classificado como iniciante, intermediário ou avançado, sem ser considerado individualmente.

O estereótipo pode ser um modelo de experiência do usuário ou um estereótipo de história do usuário tais como profissão, ou idioma nativo (Rosis, 1993), (Beaumont, 1994) e (Kay, 1994).

3.4.4 Preferências

O usuário pode preferir certas associações de links ou conteúdos ao invés de outros que poderiam ser indicados por outras características de seu modelo. Em geral, as preferências do usuário são difíceis de serem deduzidas pelo sistema, necessitando serem declaradas ou informadas por meio de algum sistema de feedback. Os sistemas de Hiperdocumentos Adaptativos podem generalizar as preferências dos usuários e aplicar esta generalização para promover adaptação em novos contextos (Kobsa, 2001).

3.5 Arquitetura de Sistemas de Hiperdocumento Adaptativo

Na figura 3 é apresentada a arquitetura básica de um sistema de Hiperdocumento Adaptativo proposta por Brusilovsk (2001). Esta arquitetura é composta por três elementos fundamentais: uma interface Adaptativa (IA), Base de Modelos do Usuário (BMU) e a Fonte de hipermídia.

Para efetuar uma adaptação os Hiperdocumentos Adaptativos, necessitam capturar as características dos seus usuários. Essa captura pode se dar de várias maneiras, através de um cadastro ou até observando a navegação do usuário no sistema.

Essas características capturadas compõem o Modelo do Usuário (UM), que fica armazenado em uma Base de Modelos de Usuários (BMU). O Modelo do Usuário atua como um filtro, definindo o conteúdo e a estrutura de navegação para o Hiperdocumento Adaptativo e deve evoluir a medida que o usuário for usando o sistema.

A interação do usuário com o Hiperdocumento Adaptativo se dá por intermédio da Interface Adaptativa (IA), que executa dois processos: a apresentação de conteúdos e *links* adaptados ao modelo do usuário e coleta as informações para atualizá-lo. A interface é construída a partir das informações obtidas sobre o usuário e armazenadas na BMU.

Uma vez identificado o usuário, o MU é carregado, permitindo ao sistema construir a estrutura básica da interface, que será composta por componentes selecionados da Fonte de Hipermídia (FH). O MU é então, atualizado a cada sessão.

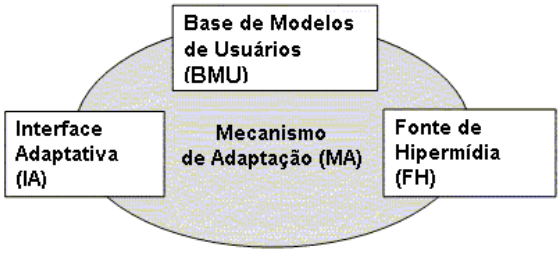

Figura 11-Arquitetura Básica de um sistema de HA (Brusilovsk, 2001)

3.6 Resumo

O capitulo oferece um embasamento na área de Hiperdocumentos Adaptativos, seus principais conceitos, modelos e técnicas, visando subsídios para alcançar os objetivos desta dissertação, que é a proposição de um sistema de recuperação adaptativo aplicado a bibliotecas digitais. No próximo capítulo serão apresentados conceitos importantes na área de sistemas de recuperação de informações.

Capítulo 4. Recuperação de Informações

O processo de recuperação de informação consiste em identificar no conjunto de documentos de um sistema, quais atendem à necessidade de informação do usuário. Para se obter uma recuperação eficiente deve-se ter uma boa representação dos documentos. O processo de representação descreve o conteúdo de um conjunto de documentos através de uma descrição ou identificação. Geralmente é usado um processo de indexação para se alcançar tal representação. O ponto principal da indexação é extrair conceitos do documento que facilitem sua identificação. São utilizados cabeçalho de assunto, um tipo de índice com resumos dos assuntos tratados, Thesaurus, etc., que identificam o documento e definem seus pontos de acesso para a busca (Ferneda, 2004).

As técnicas clássicas de recuperação estão baseadas em modelos de recuperação e podem ser divididas em dois grupos: os modelos quantitativos, que são mais antigos; e os modelos dinâmicos, que são baseados em técnicas modernas de Inteligência Artificial (IA). Os dois modelos serão analisados mais detalhadamente a seguir.

Uma das principais funções de um sistema de recuperação de informação é a de busca, que compara a identificação do documento com a expressão de busca dos usuários e fornece a informação solicitada. Na maioria das vezes essa recuperação pode incluir documentos que não são relevantes ao usuário, embora contenham expressões semelhantes às buscadas. Neste contexto, desenvolve-se o conceito de modelos de recuperação de informação adaptativos, que acrescentam ao conjunto de técnicas de recuperação clássicas, o conceito de modelos de usuário.

Este capítulo tem por objetivo pesquisar as principais técnicas de recuperação de informação utilizadas, visando formar uma base para entender os processos utilizados em BD.

4.1 Modelos Quantitativos

A maioria dos sistemas de recuperação de informação se baseia em conceitos da lógica, da estatística e da teoria dos conjuntos, que denotam uma natureza quantitativa. Neste modelo de representação, os documentos são representados por um conjunto de termos de indexação, que geralmente é uma palavra que representa um conceito ou significado presente no documento (Ferneda, 2004).

A seguir serão apresentados os principais modelos quantitativos.

4.1.1 Modelo Booleano

Neste modelo os documentos são representados por um conjunto de termos de indexação, obtidos geralmente através de especialistas na área de documentação, as buscas são formuladas através de uma expressão booleana composta por termos ligados através dos operadores lógicos *AND*, *OR* e *NOT*, e os documentos recuperados são aqueles que satisfaçam as restrições lógicas da busca.

Outros operadores que são utilizados no modelo Booleano são os operadores de proximidade, que permitem especificar condições relacionadas à distância e a posição dos termos no texto. A representação de um operador de proximidade é a seguinte: t_1 n unidades de t_2, onde a distância n é um número inteiro e unidades podem ser palavras, sentenças ou parágrafos. Pode-se destacar alguns problemas principais do modelo booleano:

- O resultado da busca, neste modelo, se caracteriza por uma simples divisão do conjunto de documentos do sistema em duas partes: os documentos pertinentes, que atendem as expressões de busca e os não pertinentes.
- Para cada termo em uma expressão booleana não se consegue atribuir um peso maior que o outro.

Observa-se que em um modelo de recuperação de informação adaptativo, os documentos recuperados devem ser ainda, colocados contra um perfil de usuário, conseguindo-se assim, uma

ordenação de relevância, e uma distribuição de pesos, que são uma evolução ao modelo clássico, e uma atenuante as suas principais deficiências.

4.1.2 Modelo Vetorial

No modelo baseado em vetor-espaço, um documento é conceitualmente representado por um vetor de palavras-chave extraídas do documento, aos quais são associados pesos, que representam a importância dessas palavras-chave no documento e na coleção de documentos (Lee, 1997). Os pesos servem para calcular o grau de similaridade entre a expressão de busca do usuário e os documentos da coleção, obtendo-se como resultado um conjunto de documentos ordenados pelo grau de similaridade em relação a busca.

São atribuídos pesos tanto aos termos de indexação quanto aos termos da expressão de busca, o que permite obter-se documentos que correspondem parcialmente a expressão de busca. O usuário pode atribuir a cada termo da busca um número (peso) que representa a importância relativa do termo para sua necessidade de informação.

Segundo Lee (1997), o peso de um termo em um vetor de documento pode ser determinado de várias formas. A forma mais usual é denominada método tf x idf, no qual o peso de um termo é determinado por intermédio de dois fatores: com que freqüência o termo j ocorre no documento i (termo freqüência $tf_{i,j}$) e com que freqüência ele ocorre na coleção de documentos (documento frequência df_j). O peso de um termo j no documento i é obtido através da seguinte fórmula:

$$W_{i,j} = tf_{i,j} \times idf_j \times \log N / df_j$$

Onde N é o número de documentos na coleção de documentos e idf (*Inverse Document Frequency*) caracteriza o termo em relação ao conjunto de documentos.

Assim como nos documentos, a expressão de busca também é representada por um vetor. Dessa forma o usuário pode atribuir a cada termo da busca um número que representa a importância relativa do termo para a sua necessidade de informação.

A padronização na representação dos documentos e das expressões da busca permite que seja realizado o cálculo do grau de similaridade entre a busca e os documentos.

Segundo Ferneda (2004), a similaridade (sim) entre dois vetores x e y em um espaço vetorial é obtido através do coseno do ângulo formado por estes vetores, através da seguinte fórmula:

$$sim(x, y) = \frac{\sum_{i=1}^{t} (w_{i,x} \times w_{i,y})}{\sqrt{\sum_{i=1}^{t} (w_{i,x})^2} \times \sqrt{\sum_{i=1}^{t} (w_{i,y})^2}}$$

Onde $W_{i,x}$ é o peso do i-ésimo elemento do vetor x . $W_{i,y}$ é o peso do i-ésimo elemento do vetor y.

Outra fórmula apresentada em (Aldenderfer,1984), é a da distância Euclidiana, obtida por:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{i,k} - x_{j,k})^2}$$

Onde d_{ij} é a distância entre os objetos i e j, $x_{i,k}$ é o valor do k-ésimo termo do documento i.

Quanto mais próximo os objetos estiverem entre si, mais próximo de zero será o valor de d_{ij}, e para que a semelhança não aponte um valor nulo é utilizada a fórmula S= 1-d, onde S é o grau de similaridade resultante.

Uma desvantagem do modelo vetorial é não permitir a implementação de buscas booleanas, o que torna a sua implementação pouco flexível em sistemas de recuperação de informações.

O modelo vetorial foi a base para a implementação de um sistema pioneiro na recuperação de informação o SMART (*System for the Manipulation and Retrieval of Text*) desenvolvido na Universidade de Cornell em 1965 por Gerard Salton (Sumner, 1997).

4.1.3 Modelo Probabilístico

No modelo probabilístico, a exemplo da teoria das probabilidades matemáticas, procura-se estimar a probabilidade de um documento específico ser pertinente a uma consulta, caso os termos da consulta estejam presentes no documento. A probabilidade de um documento satisfazer a expressão de busca geralmente, é obtida através da função de Bayes (Nallapati, 2004). Em uma busca, o resultado ideal seria apresentar somente documentos relevantes à pesquisa. Mas, como o sistema não conhece as características que distinguem o documento pertinente dos outros

documentos irrelevantes, ele tenta prever quais documentos são relevantes, através de uma expressão de busca, gerando uma primeira descrição probabilística da busca. Através de sucessivas iterações com o usuário, o sistema tende a melhorar os seus resultados.

Através do cálculo de probabilidade obtém-se os documentos relevantes a consulta, dentre o conjunto dos documentos. Estes documentos são ordenados de acordo com sua provável relevância. Depois de ordenado, o conjunto resultante da primeira busca, o usuário pode selecionar os documentos que considera mais relevante, e o sistema utiliza esta informação para tentar otimizar os resultados posteriores.

4.1.4 Modelo Fuzzy

Segundo Barreto (2001), a teoria de conjuntos fuzzy (ou nebulosos) trata-se de uma teoria dos conjuntos onde não existem valores absolutos, tal como se um elemento pertença ou não a um conjunto determinado. Um valor entre zero e um indica o quanto o elemento pertence a determinado conjunto. Um conjunto *Fuzzy* é composto por elementos os quais a transição dos mesmos de não pertencentes para pertencentes ao conjunto é gradativa. Tal grau de imprecisão de um elemento pode ser definido como sendo uma métrica de possibilidade de um determinado elemento pertencer a um conjunto.

Neste paradigma, um documento pode ser visto como um conjunto *Fuzzy* de termos, cujos pesos dependem do documento e do termo em questão. A representação *Fuzzy* de um documento é baseada na definição de uma função que produz um valor numérico que representa o peso desse termo para o documento em questão. O peso associado a um termo expressa o quanto esse termo é significativo na descrição do conteúdo do documento.

A qualidade da recuperação depende em grande parte da função adotada para calcular os pesos dos termos de indexação (Ferneda, 2004).

4.2 Modelos Dinâmicos

Os modelos dinâmicos têm como foco principal o usuário para ajudar a definir a representação dos documentos. São baseados em técnicas recentes de Inteligência Artificial (IA).

Para se obter a medida de similaridade entre os documentos nestes modelos são usados geralmente coeficientes de associação. Segundo Aldenfer (1984), coeficientes de associação são utilizados para estabelecer similaridade entre objetos descritos por variáveis binárias (variáveis que possuem dois estados ou valores).

Os coeficientes de associação mais utilizados são: Coeficiente de Jaccard, coeficiente de associação simples e coeficiente de Gower (Wives, 2004). Por exemplo, no Coeficiente de Jaccard o grau de associação entre dois objetos (X e Y) é obtido através da fórmula: S = a/(a+b+c), onde *a* é a quantidade de características presentes em ambos os objetos sendo testados, *b* é a quantidade de características que o objeto X possui e o Y não possui, c é a quantidade de características que Y possui e X não.

A seguir serão apresentados os principais modelos Dinâmicos.

4.2.1 Sistemas Especialistas

Segundo Fernandes (2003), sistemas especialistas são sistemas que buscam modelar o conhecimento de um especialista humano em uma determinada área de forma a solucionar problemas intrínsecos a esta área. Eles são também conhecidos como Sistemas Baseados em Conhecimento.

Um Sistema Especialista é composto por uma base de conhecimentos e um motor de inferência. Na base de conhecimentos armazena os fatos e as regras de inferência do especialista humano. O motor de inferência mantém um conjunto de métodos que tem a função de manipular o conhecimento contido na base de conhecimento. A base de conhecimento é separada da máquina de inferência, favorecendo que o conhecimento contido nela seja facilmente modificado quando necessário, para ser expandido, por exemplo. Uma mudança na base de conhecimento é realizada simplesmente através da adição de novas regras ou exclusão ou alteração das já existentes (Ferneda, 2004).

O conhecimento é representado na base de conhecimento por regras de produção ou regras simples, através de pares condição–ação, sob a forma de SE–ENTÃO, seguindo a estrutura: SE <CONDIÇÃO> ENTÃO <AÇÃO>.

A máquina de inferência é a responsável por procurar as respostas na base de conhecimento. Ela busca as regras necessárias a serem avaliadas, ordena-as de maneira lógica, e direciona o processo de inferência. A máquina de inferência toma decisões e julgamentos baseada em dados simbólicos da base de conhecimento. (Fernandes, 2003)

Em um sistema de informação que utiliza o modelo de Sistemas Especialistas, uma base de conhecimento pode ser construída através da identificação dos principais conceitos contidos na coleção de documentos deste sistema. Essa identificação pode ser realizada por intermédio de pessoas especialistas na área dos documentos, ou ser usado algum sistema automático de computação.

A base de conhecimento resultante contém um conjunto de conceitos, que são representados por palavras extraídas do conjunto de documentos, e que caracterizam a idéia contida neste conjunto.

Com essas palavras podem-se aplicar as regras de produção (se–então) para localizar os documentos a partir das expressões de busca dos usuários.

4.2.2 Redes Neurais

Conforme indicado anteriormente, um sistema de Recuperação de Informações é composto basicamente, por um conjunto de expressões de busca, os termos de indexação e os documentos. Esta estrutura pode ser representada como uma rede neural de três camadas como o exemplo da figura 12, onde se pode observar a camada de busca representando a camada de entrada da rede neural, a camada de documentos representando a saída da rede e os termos de indexação a camada central.

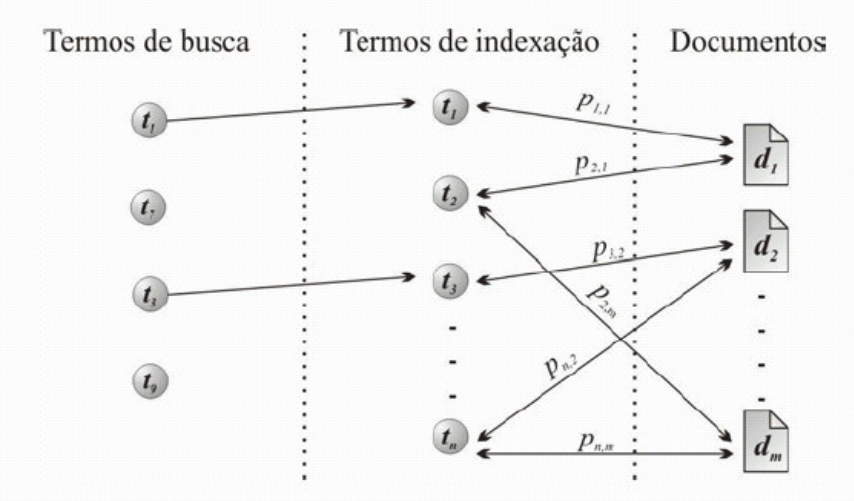

Figura 12-Rede Neural utilizada em um processo de Recuperação de Informação segundo (Ferneda, 2004)

Os termos de busca utilizados ativam os seus correspondentes na camada central, que são os termos de indexação. Os termos de indexação associados aos seus pesos enviam sinais aos documentos relacionados. Esses documentos ativados enviam sinais de volta aos termos de indexação, que ao receberem estes estímulos enviam de volta novos sinais aos documentos, repetindo o processo. A cada iteração, os sinais tornam-se mais fracos e o processo de propagação para após um certo período. O resultado final será um conjunto de documentos que foram ativados, com seus respectivos níveis de ativação, correspondente ao grau de relevância do documento em relação à busca. (Ferneda, 2004).

4.2.3 Algoritmos Genéticos

Algoritmos genéticos são métodos adaptativos que podem ser usados para resolver problemas de busca e otimização. Eles são inspirados no processo genético e evolutivo dos organismos vivos (Fernandes, 2003). Trata-se de um processo repetitivo que mantém uma população de indivíduos que representam as possíveis soluções para um determinado problema.

A cada geração os indivíduos passam por uma avaliação para verificar a sua capacidade em oferecer uma solução satisfatória. Essa avaliação é realizada através de uma função de adaptação denominada Função de *Fitness* seleciona os indivíduos de acordo com uma regra probabilística, e os submete a um processo de reprodução, que gera uma nova população de possíveis soluções.

Um documento é representado como uma espécie de código genético, no qual um cromossomo é representado por um vetor binário, onde cada elemento armazena um valor de zero ou um, correspondendo à ausência ou presença do termo na representação do documento, respectivamente.

4.3 Sistemas de Recuperação de Informação Adaptativos

Esta seção apresenta alguns trabalhos na área de Sistemas de Recuperação de Informação adaptativos, com o objetivo de se averiguar algumas técnicas de modelagem de usuário utilizadas na prática atualmente.

4.3.1 Catálogo Eletrônico de (Liu, 2001)

Em (Liu, 2001) é apresentado um sistema que integra a tecnologia de agentes de software com modelos de metadados e modelos de usuários para desenvolver um sistema de catalogo eletrônico personalizado. A função primária de um sistema de catalogo eletrônico, é o de dar suporte a procura por produtos baseado na Internet.

Foram criados metadados descritos através de uma DTD (Descrição de Tipo de Documento) de XML, para fornecer informação sobre as lojas virtuais. As lojas são agrupadas em uma estrutura hierárquica em forma de árvore. Por exemplo, um shopping sendo a raiz, um segundo nível sendo as lojas, e outros níveis para categoria de mercadorias e outros para descrições de mercadorias. O sistema utilizado para formar a árvore foi o Diretório de Árvore de Informação DIT (NASA, 2004). A vantagem básica em usar um DIT é simplificar a busca por um determinado produto, devido a relação hierárquica que ela implementa, ao invés de seguir todos os nodos, a busca pode se dar em apenas uma árvore especifica.

O modelo de usuário, formado pelo perfil do usuário, descreve os interesses de compra dos usuários o mais detalhado possível. Para isso o perfil do usuário guarda as mercadorias do seu interesse com a descrição e um valor que determina o nível de preferência do usuário pela mesma. Por exemplo, um usuário indica para um atributo cor de um produto o peso (valor) de 0.3, indicando que a importância da cor de um produto para o usuário é de 0.3. o atributo cor

pode ter o valor de vermelho, azul, ou verde, os quais têm os pesos 0.1, 0.1 e 0.8 respectivamente. Indicando que o usuário prefere o produto de cor verde aos outros. Como resultado, o produto de cor verde, tem uma importância de 0.3 * 0.8 = 0.24, significando em 24% de influencia da cor na compra do produto.

Uma vez determinada as informações dos produtos, que já estão disponíveis em forma de árvore, e as preferências dos usuários, guardadas no seu perfil, são empregados agentes para ordenar os produtos por ordem de preferência do usuário e efetuar as buscas nas lojas virtuais de acordo com o seu perfil, e entregar a informação solicitada ao usuário por intermédio das interfaces apropriadas, entre outros agentes que tem outras funções que não são de interesse do trabalho.

4.3.2 Mecanismo de aprendizagem de (Park, 1998)

Em (Park, 1998) é proposto um mecanismo de aprendizagem que é usado para construir um perfil de usuário que aceita os termos relevantes dos documentos. Estes documentos são propostos por uma avaliação (*fedback)* de relevância do usuário e são selecionadas condições - chaves nos documentos para formar o perfil. Este mecanismo foi aplicado em um sistema autônomo de filtro de informação denominado IGIMA (*Intelligent Information Gathering and filtering System based on Multi-Agent*).

Inicialmente o usuário realizada uma avaliação dos documentos como positivos e negativos, para selecionar aqueles que têm um *feedback* relevante. Estes documentos são, então, otimizados utilizando uma função de Jaccard.

Nesta proposta, para adquirir os interesses dos usuários é usado um método de avaliação de relevância baseado em um modelo de vetor-espaço. Inicialmente os documentos são recuperados usando uma máquina de procura comum. Então o usuário dá uma avaliação de relevância para esses documentos. Também são extraídos termos-chave destes documentos usando avaliação positiva, através da aplicação da fórmula de TF (Termo Freqüência) que é uma heurística de pesagem de palavra. Também é usado o IDF (Inverso-documento-frequência), que quer dizer que quanto mais um termo aparece em vários documentos, o menos provável é que o mesmo seja útil para diferenciar entre estes documentos. A fórmula de TFDF é utilizada para eliminar essas

condições sem importância nos documentos de avaliação negativa. Na geração de um novo perfil, são utilizados os termos-chave selecionados em cada documento.

É necessário normalizar os termos dos documentos para usar os mais relevantes, e conhecer os interesses dos usuários. Obtendo-se a relevância de cada documento os mesmos são normalizados. Para esta normalização é utilizado a função de máquina de Jaccard para calcular a similaridade entre os documentos. Um escore mais alto da função de Jaccard indica uma similaridade mais forte entre os dois documentos. Conhecendo a similaridade entre cada documento, pode-se saber quais documentos são mais relevantes aos demais.

Finalmente multiplicando-se o peso de cada termo-chave pela relevância de todos os documentos, determinam-se a normalização dos mesmos.

Do resultado da aplicação da função de Jaccard tem-se um conjunto de condições. Este sendo construído para cada termo de todos os documentos. Este jogo, possui a seleção dos termos de maior peso, e são usados para construir um conjunto de termos relevantes, formando um semi-perfil para que possa ser construído um novo perfil.

O novo perfil gerado reflete mais os interesses do usuário atual, sendo incluído na fórmula, um valor de peso associado ao mesmo, que pode ser de zero ou um.

Finalmente, para filtrar os documentos recobrados, é usado o cálculo de semelhança entre o perfil e o documento. A similaridade mais alta reflete que este documento está mais próximo ao documento que o usuário tem preferência.

O perfil e o documento são considerados como um vetor espacial, e é utilizado uma fórmula que calcula a diferença dos coeficientes de dois vetores, para verificar a diferença entre os mesmos.

Outra observação é que quando um usuário faz uma primeira requisição ao sistema, e ele ainda não possui um perfil, é usado um perfil social. O perfil social é construído através dos perfis de outros usuários que tenham os mesmos interesses.

4.3.3 ALIPES (Widyantoro, 1999)

O ALIPES (Widyantoro, 1999) é um agente de notícias personalizado que junta periodicamente artigos de várias fontes de notícias on-line da WWW e os filtra para seus usuários de acordo com o seu perfil. Este agente filtra e recupera informação de acordo com os interesses de usuários da Internet.

O ALIPES consiste em três descritores, um perfil de usuário e seu algoritmo de aprendizagem. Um descritor mantém os interesses de longo prazo, e outros dois descritores, positivo e negativo mantém interesses de curto prazo. A estrutura básica da representação de uma categoria de interesses é um vetor de características. Este vetor contém uma lista de palavras-chave que são pesadas de acordo com seu grau de importância.

O esquema de pesagem de palavras utilizado é o TF-IDF. Baseado em TF-IDF, a importância da palavra-chave é proporcional a freqüência de ocorrência de cada termo em cada documento e inversamente proporcional ao número de documentos em uma coleção de documentos dentro do qual o termo acontece, isto quer dizer, que uma palavra-chave aparecendo em menos documentos discriminam melhor do que as que aparecem em mais documentos. Em seguida é utilizada uma fórmula para calcular o peso da palavra-chave em um documento, que leva em conta o número de documentos e a freqüência do termo no documento.

O descritor positivo e negativo mantém um vetor de características aprendido de documentos, com avaliação positiva e avaliação negativa, respectivamente, enquanto o de longo prazo mantém um vetor de características de um documento com ambos os tipos de avaliação. Cada descritor tem também um peso de interesse para representar o interesse em nível do descritor da categoria de interesse correspondente.

Para a medida de semelhança entre os dois vetores (vetor de características de documentos e o vetor de característica positivo) é utilizada a fórmula do coseno dos vetores.

O algoritmo de aprendizagem confia na avaliação do usuário, modificando o seu perfil de acordo com a mesma. A avaliação pode ser positiva ou negativa (indicando se o usuário aprova ou não do conteúdo do documento recuperado como pertinente a busca). A taxa de aprendizagem representa a força do usuário (por exemplo, muito interessante, interessante, não ruim, desinteressante, etc.) sendo sua gama (0,1).

4.3.4 Modelo de Referência de (Wu, 2001)

Em (Wu, 2001) é proposto um modelo de referência para a arquitetura de aplicações de Hipermídia adaptáveis, aplicado em um sistema denominado AHAM (*Adaptive Hypermedia Application Model*). Trata-se de uma máquina adaptativa de propósito geral. A máquina executa a adaptação e atualiza um modelo de usuário de acordo com um jogo de regras de adaptação especificado em um modelo. Em AHAM existem três fatores para que a adaptação seja eficaz:

- A adaptação deve estar baseada em um modelo de domínio, descrevendo como o conteúdo da informação (HA) é estruturado (usando conceitos e relações de conceitos). É usado o termo conceito e relacionamento de conceito para generalizar nodos e *links*;
- O sistema tem que construir e tem que manter um bom modelo de usuário que represente as preferências de um usuário, conhecimento, metas, história de navegação e outros aspectos pertinentes, e;
- O sistema deve poder executar a adaptação do conteúdo e da estrutura dos *links*, baseado no modelo de domínio e modelo do usuário.

Para alcançar esse propósito são criadas regras de adaptação. As regras definem o processo de gerar a apresentação adaptável e de atualização do modelo do usuário. Foi criado um idioma de definição de regras e uma máquina e adaptação (AE) para executar as regras de adaptação.

O modelo do usuário (UM) representa a relação entre o usuário e o modelo de domínio. O usuário pode especificar atualização para o seu modelo e o sistema também o atualiza monitorando o comportamento do usuário.

Para executar a adaptação baseada em DM e UM é preciso especificar como a interação do usuário com o sistema influencia a apresentação da informação no DM. Em AHM isto é feito por meio de um modelo de adaptação, que consiste em regras de adaptação. Uma máquina de adaptação (AE) usa estas regras para manipular *links* âncoras e gerar as especificações de apresentação.

O modelo de domínio consiste em conceitos e relações de conceitos. Conceitos são objetos com um único identificador de objetos e uma estrutura que inclui pares de atributo - valor e *links*

âncoras. Um conceito representa um item de informação abstrata no domínio da aplicação. Uma relação de conceitos é um objeto (com um único identificador e pares de atributo - valor), que relaciona uma sucessão de dois ou mais conceitos. Cada relação de conceito tem um tipo, sendo o mais comum uma ligação de hipertexto.

O modelo do usuário consiste em entidades, as quais armazenam um número de pares atributos - valor. Para cada entidade pode haver atributos diferentes, mas a prática a maioria tem os mesmo atributos. Para representar o modelo do usuário é usada uma estrutura de tabela, na qual, para cada entidade os valores do atributo para aquele conceito, são armazenados. A maioria das entidades em um modelo de usuário, representa conceitos do modelo de domínio. AHAM considera modelo de atributos do usuário os conceitos típicos relatados em DM, sejam: conhecimento (sobre o conceito), leitura (que denota a página, ou o fragmento de página que foi lido pelo usuário) e o que já foi lido (aquilo que o usuário terminou de ver, as informações sobre este conceito ou se já leu a página em questão).

O modelo de adaptação descreve como o AHAM deve executar sua adaptação, que inclui a adaptação do conteúdo, adaptação de *links*, e atualização do modelo do usuário. Trata-se de um conjunto de regras de adaptação que formam a conexão entre DM, UM e a apresentação a ser gerada.

A máquina de adaptação é um ambiente de software que executa as seguintes funções:

- Oferece seletores de páginas genéricos e construtores. Para cada conceito composto o seletor correspondente é usado para determinar quais páginas exibir quando o usuário seguir o *link* do conceito. Para cada página um construtor é usado para construir a apresentação adaptativa daquela página. Construtores de página fornecem conteúdo dinâmico semelhante a uma lista de *links*.
- Opcionalmente oferece uma linguagem para descrever novos seletores de página e construtores. Em AHAM um construtor de página consiste em comandos simples para a inclusão condicional de fragmentos.
- Executa adaptação executando os seletores de página e os construtores. Significa selecionar uma página enquanto seleciona fragmentos, organiza e os apresenta dentro de um modo específico. A adaptação também envolve adaptação de *links*,

alterando links âncoras, dependendo do estado do mesmo(habilitado, desabilitado e oculto).

- . Atualiza o modelo do usuário a cada visita do mesmo a uma página. A máquina mudará alguns valores de atributo para cada conceito atômico de fragmentos exibidos de uma página, da página como um todo e possivelmente de algum conceito dependendo das regras de adaptação.

A máquina de adaptação fornece desse modo, uma implementação dependente de aspectos enquanto DM, UM e AM descrevem a informação e a adaptação no conceitual.

4.3.5 Sistema de Recuperação de (Dizaji, 2003)

Em (Dizaji, 2003) é apresentado um sistema de recuperação de informação da Internet, que aprende as necessidades do usuário através de sua avaliação de relevância para os documentos recuperados. O método apresentado combina uma medida de avaliação qualitativa, inferência fuzzy e avaliação quantitativa usando algoritmos genéticos (GA).

O sistema consiste da combinação de processos de indexação de documentos, aprendizado estratégico por avaliação de relevância, modelagem de usuário por algoritmos genéticos, filtros e avaliação dos documentos recuperados baseado no modelo de usuário. O sistema proposto consiste em vários agentes que cooperam entre si, trabalhando em paralelo para executar a sua meta. A interação do sistema com o usuário se dá através de um agente de interface com dois objetivos principais: prover palavras - chave de seu interesse, e prover avaliação de relevância dos documentos recobrados. O agente de procura é uma máquina de meta-busca de qualquer fonte da Internet que usa palavras - chave para recobrar documentos.

Um GA é usado para evoluir e adaptar um vetor de consultas, que são os modelos das necessidades de informação do usuário. Com GA o modelo do usuário representa um conhecimento hipotético sobre a necessidade do usuário, codificado em um cromossomo. Esses cromossomos são expressos em termos e seus pesos. Cada cromossomo assim, é uma hipótese em como avaliar a relevância de um documento, e compete contra outros cromossomos para prever a satisfação do usuário aos documentos recobrados.

Um papel chave de GA na modelagem das necessidades do usuário é a modificação contínua da representação das necessidades do usuário, em uma métrica de relevância quantitativa e qualitativa. A métrica quantitativa é proporcional a média de similaridade entre o vetor de consulta e todos os documentos recuperados. A métrica qualitativa confia em uma avaliação de relevância *fuzzy* do usuário. Assim o usuário provê um valor linguistico de avaliação da recuperação, o qual é usado em um sistema *fuzzy* de inferência, que obtém uma medida de relevância concisa. A avaliação concisa é combinada com a medida de semelhança entre o vetor da questão e o documento recuperado para determinar um quantitativo métrico usado para ajustar a aptidão dos cromossomos na população. A saída do sistema de inferência *fuzzy* é usada para ajustar a aptidão dos cromossomos que competem para desenvolver o modelo ótimo das necessidades dos usuários.

4.4 Resumo

Este capítulo apresentou os conceitos envolvidos com a recuperação de informação clássica e a recuperação de informação adaptativa. A partir dos conceitos básicos e de como foram utilizados em alguns sistemas que embora, possam ter objetivos diferentes, utilizam-se dos conceitos de recuperação e de modelagem do usuário de forma semelhante a proposta neste trabalho, bem como procurou-se identificar as diversas formas de recuperar e adaptar o perfil do usuário existentes.

No próximo capítulo é apresentado o sistema de recuperação adaptativo aplicado à biblioteca digital proposto nesta dissertação.

Capítulo 5. Técnica de Recuperação de Informação Adaptativa Aplicada a Biblioteca Digital

Conforme levantamento realizado no capítulo anterior, existem várias técnicas de recuperação de informação adaptativas propostas na literatura. A maioria é aplicada a recuperação de informação na Internet (conteúdo Web) ou sistemas de vendas virtuais. Além disto, a adaptação é realizada tentando prever os próximos passos do usuário (pró-ativa) e não procurando adaptar a sua forma de atuar de acordo com a necessidade e anseios do usuário (reativa). Os trabalhos na área de bibliotecas digitais adaptativas são muito raros e quando existentes partem de uma biblioteca criada desde o inicio para o propósito adaptativo.

Este capítulo propõe uma técnica de Recuperação Adaptativa aplicada às Bibliotecas Digitais, chamado aqui da RIA-BD. Esta técnica de adaptação se caracteriza pela simplicidade, sem aplicar técnicas complexas de inteligência artificial e que pode ser aplicada diretamente em uma biblioteca digital baseada em metadados que já esteja em operação.

5.1 Requisitos Funcionais da RIA-BD

Um Sistema de Recuperação de Informações Adaptativo para cumprir o seu objetivo deve ser formado pelo menos por:

- **Interface do Usuário (IU)**: uma série de páginas web na qual o usuário tem acesso a biblioteca. A partir desta, o usuário pode realizar a autenticação, indicando a identificação do usuário (*ID Usuário*) e sua senha, configuração inicial de seu *Perfil* e sua edição; realização de buscas definindo um Critério de Busca (*Consulta*), visualização do *Resultado* da busca, e navegação para acesso aos documentos.
- **Base de Perfis dos Usuários (BPU)**: que mantém informações acerca das preferências, conhecimentos e ranking de consultas de cada usuário da biblioteca.

- **Base de Metadados dos Documentos (BMD)**: mantém informações acerca dos documentos.
- **Recuperação de Informações Adaptativas (RIA)**: baseado no Critério de Busca (*Consulta*) e na identificação do usuário (*ID do Usuário*) o RIA recupera o *Perfil* deste usuário, efetua a consulta e a apresentação do *Resultado*. Este resultado contém a lista dos documentos que atendem ao critério de busca, sendo que eles são apresentados em grupos cuja regra de agrupamento e de ordenamento são baseadas nas preferências do usuário.
- **Atualização do Perfil**: atualiza o *Perfil* do usuário baseado nos metadados associados ao documento acessado pelo usuário. Os metadados do documento são recuperados da BMD via o identificador do documento (*ID Documento*).

A figura 13 apresenta a arquitetura do Sistema de Recuperação Adaptativo proposto, identificando os principais componentes descritos acima:

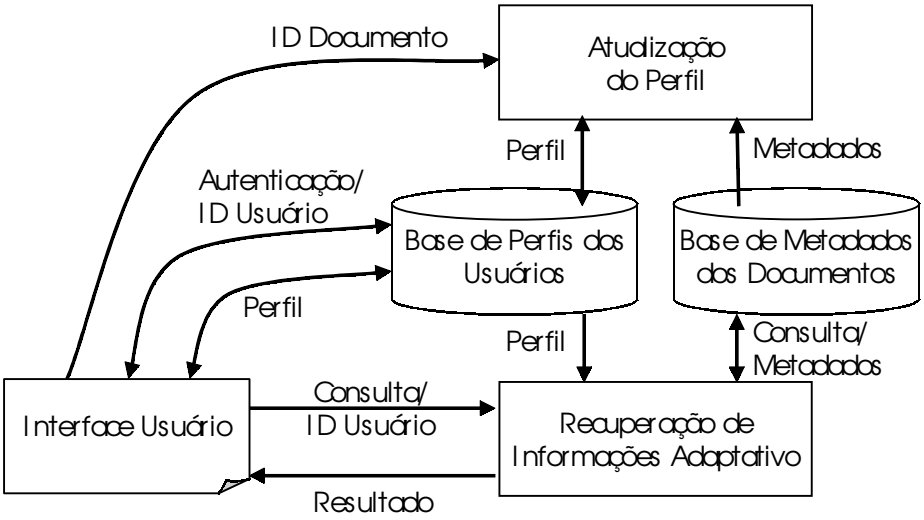

Figura 13-Arquitetura proposta da RIA-BD

As próximas seções apresentam os elementos da RIA-BD e as suas interações.

5.2 Visão Geral da RIA-BD

Para apresentação do sistema de recuperação adaptativo, a biblioteca digital é definida como: BD = (D,M,P), onde:

- $D=\{d_1, d_2, d_3, ..., d_n\}$ é o conjunto de documentos disponibilizados pela biblioteca
- $M=\{md_1, md_2, md_3, md_4,.., md_m\}$ é o conjunto de metadados usados para indexação dos documentos. Usaremos a notação $d_i.md$ para indicar o valor do metadado md para o documento d_i.
- $P=\{p_1, p_2, p_3, ..., p_k\}$ é o conjunto de perfis que identificam as preferências dos usuários da biblioteca. Todo $p \in P$ é definido por p=(IU, PG, PE, PV), onde IU mantém a Identificação do Usuário, PG as Preferências Gerais, PE as Preferências Específicas e PV as Preferências de Visualização do resultado da busca.

Para facilitar o entendimento da proposta, este capítulo utiliza como cenário de uso do sistema de recuperação proposto uma biblioteca digital de literatura fictícia.

5.2.1 Base de Metadados dos Documentos (BMD)

Na Base de Metadados dos Documentos (BMD) ficam armazenados os metadados dos objetos digitais. A partir destes metadados é possível realizar a catalogação das obras disponibilizadas. A fim de permitir a interoperabilidade, é importante adotar metadados padronizados, como aqueles definidos pela iniciativa Dublin Core (DCMI, 2005). Outros metadados podem ser definidos a fim de catalogar informações específicas da biblioteca. Estes metadados servem como suporte a recuperação de informações e também permitem atualizar o perfil do usuário baseado nas obras consultadas por um usuário.

5.2.2 Base de Perfis dos Usuários

A Base de Perfis dos Usuários (BMU) contém os perfis de todos os usuários cadastrados da biblioteca. As necessidades de informação dos usuários podem ser representadas através de uma estrutura bem definida, chamada de Perfil. Nesta proposta, o perfil visa representar explicitamente cada usuário de forma individual. Ele permite ao sistema de recuperação de informações realizar um ordenamento do resultado da consulta baseado nos conhecimentos e preferências de cada usuário.

Todo perfil de usuário p_i de P é definido por $p_i=(IU_i, PG_i, PE_i, PV_i)$, onde:

- Identificação do usuário (IU): essas informações servem geralmente para identificar o usuário tais como seus dados pessoais (nome, login, senha, homepage) e informações de contato (correio eletrônico). Essas informações têm de ser informadas pelo usuário. A identificação do usuário tem por função somente a identificação do usuário e não necessariamente como meio de restringir o acesso a biblioteca digital.

- Preferências Gerais do Usuário (PG): contendo informações sobre suas preferências e conhecimentos gerais do usuário. Por exemplo, preferência em termos de interface e navegação (formato e disposição das informações, cores, etc.) e os idiomas conhecidos. Estas preferências são gerais no sentido que independem do tipo de conteúdo disponibilizado pela biblioteca digital. Estas informações também devem ser fornecidas pelo usuário.

- Preferências Específicas do Usuário (PE): são preferências do usuário acerca do(s) tema(s) que tratam os conteúdos disponibilizados pela biblioteca digital.

- Preferências de Visualização do Resultado da Busca (PVRB): é utilizada pelo usuário para determinar a forma de agrupamento e o método de ordenação dos resultados da busca. Deve ser informado pelo usuário.

Para o contexto deste trabalho, apenas os dois últimos tipos de informações são relevantes, sendo elas descritas mais em detalhe nas seções que seguem. Em especial, as PE devem ser previstas pelo sistema via as características dos documentos acessados.

Preferências Específicas

Para caracterizar as preferências específicas de um usuário é considerado apenas um subgrupo dos metadados de M. Por isso é necessário determinar quais metadados serão usados para determinar as preferências do usuário. Esses metadados são chamados Metadados Observados (MO), onde MO={mo_1, mo_2, .. mo_n}.

Em uma biblioteca de literatura, por exemplo, os metadados observados poderiam ser:

- Autor: o qual os valores possíveis são os nomes dos autores dos documentos disponíveis na Biblioteca Digital de Literatura;
- Gênero Literário: o qual os possíveis valores são poesia, teatro, romance e novela, conto críticas (ensaio, artigos), crônica, história (literatura de viagens e informação), textos doutrinários e textos parenéticos (discursos e sermões), epistolografia, biografia e memórias.
- Assuntos: com os possíveis valores definidos em um vocabulário restrito usando um Thesaurus Temático de Literatura.

As Preferências Específicas (PE) do usuário são definidos como PE = (MO_p, MO_v), onde:

- MO_p={mo_{p1}, mo_{p2}, ..., mo_{pm}} e MO_v={mo_{v1}, mo_{v2}, ..., mo_{vm}} são os conjuntos de valores com peso dos Metadados Observados de preferência do usuário e de documentos visitados, respectivamente. MOp mantém os valores considerados de preferência do usuário. MOv mantém os valores dos documentos acessados pelo usuário mas não considerados de sua preferência.
- mo={(vo_1, w_1,), (vo_2, w_2), ..., (vo_m, w_n)}, onde vo_i é um valor do metadado observado, wi indica um grau (peso) de relevância do valor do metadado para um usuário.
- VP={vo_{p1}, vo_{p2},..., vo_{pk}} é o conjunto de Valores de Preferência dos metadados observados.
- VV={vo_{v1}, vo_{v2},..., vo_{vl}} é o conjunto de Valores Visitados dos metadados observados.

Por exemplo, no caso da biblioteca de literatura, um usuário em particular poderia ter suas preferências específicas definidas por: (((Aa, 5), (Ab, 4)), ((Ac, 3), (Ad, 1)))(((Ga, 7), (Gb, 3)), ((Gc, 1), (Gd, 1))). Sendo assim, os autores preferidos pelo autor são Aa e Ab, e os autores cujas obras já foram visitadas são . O usuário também prefere os gêneros Ga, Gb, mas também já leu obras de Gc e Gd. Sendo {Aa, Ab} e {Ga, Gb} os conjuntos de valores de preferência para os metadados observados autor e gênero.

Definição do Peso de Relevância

Conforme observado em (Chen, 2002), pode-se utilizar pelo menos quatro modos de análise do comportamento do usuário baseado nos metadados de preferência dos documentos acessados:

- Análise da existência: todas as características dos documentos acessados, ou seja valores de seus metadados observados, têm igual peso de preferência sem considerar quando e com freqüência o usuário acessou eles.
- Análise da freqüência: mantém um peso em cada valor de metadado observado. Este peso é definido pela freqüência de acesso a documento com metadados observado. Nesta análise, o peso de relevância é a freqüência de acesso, sendo que esta freqüência pode ser medido durante todo o histórico do usuário (WH – *Whole History*) ou medido sobre um determinado número de acessos (LAN - *Last-N Access*).
- Análise por idade: associa um peso de acordo com a idade do acesso. Esta análise leva em consideração que um acesso recente é mais importante para definir o perfil que um acesso mais antigo. (Chen, 2002) define uma equação para determinar o peso baseado na idade de acesso.
- Análise seqüencial: assume que um usuário repete caminhos similares de tempos em tempos. Neste caso, um acesso do usuário após uma seqüência de acessos pode ser inferido a partir de caminhos antigos conhecidos que incluem a seqüência. Esta análise não se aplica ao contexto deste trabalho.

A definição do método mais apropriado para a determinação de peso de relevância para uma Biblioteca Digital depende do tipo da coleção. Por isso, deveriam ser respondidas algumas perguntas: se os interesses do usuário trocam com o passar do tempo; se uma preferência pode

deixar de existir depois de um certo tempo; se um acesso mais recente deveria ser considerado mais importante para estimar o perfil do que um mais antigo. As respostas destas perguntas dependem de tipo de coleção mantido pela Biblioteca Digital.

Por se tratar de uma coleção de literatura, foi assumido que as preferências do usuário não mudam com o tempo. Acessos mais antigos são tão pertinentes quantos acessos mais recentes. Além disso, foi assumido também que as preferências do usuário são perenes e em conseqüência elas não deixam de ser preferidas, a menos que o usuário defina explicitamente. Estas considerações foram feitas de um modo empírico, a continuidade dos testes aplicados ao protótipo deverá validar esta suposição.

Devido a estas considerações este trabalho utiliza a análise de freqüência e a história total do usuário (WH) para determinar o peso de relevância dos valores dos metadados, onde o peso de relevância corresponde ao número total de acessos durante a história total de acessos do usuário, sendo desta forma, aprimorado com o passar da interação do usuário com a biblioteca.

Uma atenção especial deve ser observada quando a Biblioteca Digital adota informação categorizada para os metadados observados. É o caso do metadado de assuntos da BD, onde foi adotado um Thesaurus temático. O peso de relevância de um certo nível de assunto deve ser a soma dos pesos de relevância do assunto e dos pesos de relevância de todos os assuntos abaixo dele.

Definição da visualização do resultado da busca (PVRB)

PVRB são preferências, indicadas pelo usuário, relativas a forma de apresentação do resultado da busca dos documentos. Estas preferências são definidas por: PVRB= (moa, og)**,** onde moa é a indicação de qual metadado será utilizado como metadado de agrupamento, e og é a indicação do modo de ordenamento do resultado nos grupos.

No caso da LBC, **moa** pode assumir o valor Autor ou Gênero. Considerando que o usuário selecione o agrupamento por autor, os **GNRs** para este usuário seriam:

- Preferidos: composto de obras que atendem o critério de busca e são de autoria de um autor de preferência do usuário.
- Visitados: composto de obras que atendem o critério de busca e são de autoria de um autor cuja uma obra já foi consultada anteriormente.

- Outros: composto de outras obras que não pertencem aos dois grupos anteriores, mas satisfazem o critério de busca.

Como na forma de agrupamento, o critério de ordenamento dos documentos no GNR é escolhido via og. Existem diversos critérios possíveis de ordenamento do documento nos GNR.

Em uma biblioteca de literatura, o usuário pode escolher entre:

- Autor: documentos serão ordenados em ordem alfabética de autor
- Gênero: documentos são ordenados em ordem alfabética de gênero literário.
- Assunto: documentos são ordenados em ordem alfabética de assunto.
- Peso: documentos são ordenados segundo um peso de relevância.

5.2.3 Interface do Usuário

A Interface Usuário é um conjunto de páginas web pela qual o usuário interage com o sistema. Ela oferece quatro funções básicas: Autenticação, Cadastramento do Usuário, Edição do Perfil, Pesquisa de Informações, Visualização do Resultado, e Navegação no Acervo.

Autenticação do Usuário

Para possibilitar a identificação do usuário que está usando a biblioteca digital e para possibilitar a adaptação da recuperação de informações ao seu perfil, é necessário que a biblioteca digital implemente um sistema de autenticação do usuário.

O processo de autenticação permite ao usuário da biblioteca se autenticar (login e senha) no sistema. Com isto, o sistema poderá associar um identificador para cada Usuário (ID Usuário).

ID Usuário trata-se na realidade de um identificador de cookie que identificará o usuário na biblioteca. O cookie foi adotado nesta proposta, pois evita que usuário realize a autenticação a cada novo acesso a biblioteca. Esta operação de autenticação é necessária apenas quando não exista um cookie ativo para o site da biblioteca.

Uma dos principais fatores apontados como oposição ao uso de Cookies é o fato de que ele é o principal mecanismo para suportar o rastreamento de usuários desconhecidos através de sites Web. Isto geralmente ocorre quando se navega em páginas desconhecidas, ou que contém faixas

de anúncio que enviam Cookies com o intuito de monitorar os endereços que o usuário utiliza pela Web.

Um Cookie é enviado de um servidor para um navegador adicionando uma linha aos cabeçalhos do protocolo http, com a informação set-cookie. Este cabeçalho é uma string que contém caracteres com parâmetros que devem ser passados do servidor ao cliente e armazenados na máquina do cliente.Após isso, a cada solicitação http para o servidor que armazenou o cookie os valores armazenados no mesmo são enviados de volta ao servidor. Os parâmetros Path e Domain determinam para quais sevidores, ou domínios os cookies devem ser enviados de volta. Se o domínio não estiver definido explicitamente, ele usa como padrão o domínio completo do documento que está criando o cookie (Conallen, 2003).

No caso da RIA-BD, a quebra de segurança via Cookie não é um atrativo para *Hackers*, pois o mesmo não guarda informações com importância comercial, além de que o usuário pode saber para quem o cookie está enviando os dados, no caso para a BD.

O objetivo da autenticação é apenas identificar o usuário para os efeitos adaptativos propostos, sendo opcional.

Cadastramento de Usuário

O cadastramento do usuário na biblioteca é realizado através da Interface de Cadastramento de Usuário. A partir desta interface o usuário poderá fornecer as informações que irão compor o seu perfil inicial. Como visto mais adiante, este perfil será atualizado baseado nos documentos acessados pelo mesmo.

Neste momento, o usuário deve informar os seus dados pessoais e suas preferências, sendo esta inicialização opcional. Se o usuário não iniciar o seu perfil, o sistema assume um perfil tomando como base os documentos acessados por ele.

É através dessa interface que o usuário também pode definir os seus dados pessoais e suas preferências particulares, como critério de agrupamento, método de ordenação, conjunto inicial de autores, gêneros literários e assuntos de sua preferência.

Edição do Perfil

A edição do perfil do usuário é realizada através da interface de edição de perfil da RIA-BD. Esta interface permite a edição do perfil pelo usuário, oferecendo ao mesmo a possibilidade de atualizar seu perfil baseado no histórico de documentos consultados.

Pesquisa de Informações

A pesquisa de informações é realizada através da interface de consulta. Esta interface permite o usuário preencher e submeter o formulário de consulta. Esta interface pode apresentar as mesmas funcionalidades de uma interface de busca convencional, onde pode ser escolhido para pesquisar por qualquer palavra, frase exata, todas as palavras (disponível por intermédio de menus *pop-down*).

Visualização do Resultado

A visualização dos Resultados da pesquisa permite ao usuário visualizar o resultado da consulta da forma especificada em PV.

Navegação no acervo

A navegação no acervo se dá por intermédio de uma interface especifica, acessada através de *links* na página de resultado da pesquisa, por onde o usuário pode navegar no acervo para consultar informações sobre os autores, obras e o acesso a obra.

5.2.4 Recuperação de Informações Adaptativa

Neste módulo é implantada a Recuperação de Informações Adaptativa RIA-BD proposta neste trabalho. Ele tem como recurso de adaptação a personalização da apresentação do resultado da consulta baseado no perfil do usuário. O objetivo aqui é aumentar a eficiência da recuperação da informação via definição da relevância do documento baseado não somente no atendimento dos critérios de busca e metadados do documento, mas também no perfil do usuário.

A ordenação simples do resultado da consulta baseado em um grau (nível) de relevância único pode não ser suficiente para agilizar o processo de busca. Nesta proposta, estes documentos são organizados em Grupos de Relevância (GR), sendo que os documentos pertencentes a um GR

apresentam um mesmo nível de relevância para o usuário. Na apresentação do resultado da consulta, os GR são dispostos de maneira ordenada dependendo do grau de interesse previsto, sendo que o GR mais relevante é apresentado primeiro.

Grupos de Relevância (GNR)

Os documentos resultantes da busca são organizados em Grupos de Relevância (GR). A definição do critério de agrupamento deve ser definida pelo usuário. O usuário deve escolher quais Metadados observados serão usados como Metadado de agrupamento (MG). Foram definidos três Grupos de Relevância (RG):

- Preferidos: para pertencer a este grupo, o documento deve ter um valor do metadado observado que pertence ao conjunto de valores de preferência (VP).
- Visitados: para pertencer a este grupo, o documento tem um valor de metadado observado que pertence ao conjunto de valores visitados (VV).
- Outros: para pertencer a este grupo, o documento tem um valor de metadado observado que não pertença ao conjunto VP nem a VP.

Além do critério de agrupamento, o usuário deve decidir o método de ordenação, que é definida em PVRB.

Visualização do resultado da busca (PV)

Estas preferências são definidas por PV=(moa, og), onde moa é a indicação de qual metadado será utilizado como metadado de agrupamento, e og é a indicação do modo de ordenamento do resultado nos grupos.

No caso da biblioteca de literatura, moa poderia assumir o valor Autor, Gênero ou Assunto, e og pode ser peso de relevância, autor, gênero literário ou assunto, onde:

- Autor: os documentos são ordenados em ordem alfabética por autor;
- Gênero Literário: os documentos são ordenados em ordem alfabética de gênero literário;
- Assunto: os documentos são ordenados em ordem alfabética de assunto, e;
- Peso de relevância: os documentos são ordenados com base no peso de relevância.

5.2.5 Atualização do Perfil do Usuário

Nesta proposta, as preferências específicas do usuário podem ser iniciadas pelo usuário ou determinadas e atualizadas pelo sistema. Esta atualização é realizada com base nos metadados observados dos documentos acessados pelo usuário. Não sendo necessário que o usuário faça uma inicialização prévia de suas preferências específicas.

Este módulo é atualizado com base no perfil do usuário, levando em conta os Metadados observados dos documentos acessados. Além disso, este modulo permite ao usuário iniciar e revisar o seu perfil.

No momento em que um documento é acessado por um usuário, o modulo de atualização do perfil de usuário é ativado. Baseado na identificação do documento e na identificação do usuário, o modulo captura os valores do Metadado observado do documento e atualiza o perfil do usuário. Se este valor não estiver no perfil do usuário, significando que este Metadado nunca foi acessado pelo mesmo, ele é incluído em visitados (VV) com o peso de relevância igual a 1. De outra forma, se um valor de Metadado observado estiver presente em VV ou VP, seu peso de relevância incrementado com mais 1.

Foi considerado que qualquer acesso ao documento contribui para a construção do perfil do usuário. Também é considerado que o acesso a um documento é uma indicação de que o usuário está interessado em alguma característica desse documento. Um possível problema que pode ocorrer é o acesso a um documento de um modo precipitado, onde o usuário depois de verificar o mesmo, pode considerar que ele não é interessante. Para reduzir este problema, a RIA-BD não oferece acesso direto a obra quando da apresentação do resultado da busca. É oferecida inicialmente ao usuário, uma apresentação das características do documento, visando com isto minimizar o acesso precipitado. Outras soluções estão sendo investigadas, como o tempo de permanência do usuário na página que apresenta o documento, por exemplo.

A conversão de um Metadado observado para preferido, também é uma operação crítica para a atualização do perfil do usuário. Existem duas opções:

- Atualização manual: onde o usuário pode editar o seu perfil manualmente e passar o Metadado de observado para preferido. O usuário também pode definir novos Metadados como preferidos se estes não estiverem em seu grupo de visitados, e;

- Atualização automática: onde se define um valor limiar que deve ser alcançado pelo peso de relevância do Metadado observado, para que este seja incluído no grupo de Metadados preferidos. Esta opção deve ser melhor investigada, através dos resultados dos testes com o protótipo.

5.3 Resumo

O sistema de recuperação adaptado proposto neste artigo visa permitir que o resultado de uma consulta seja apresentado ao usuário de maneira a reduzir o tempo de recuperação do documento procurado. Para tal, é levado em conta o histórico do perfil do usuário quando da apresentação dos resultados da consulta.

É utilizado o agrupamento dos documentos resultantes de uma busca em grupos de níveis de relevância (GNRs) e a atualização do perfil do usuário baseado nas características dos documentos acessados, tornando a atualização do perfil do usuário mais simples e não envolvendo técnicas de aprendizado de inteligencia artificial (IA) , como em (Crestani, 1999) e (Chaves, 2002).

Um problema que deve ser observado na utilização de recuperações adaptativas, é que, a base de perfis do usuário pode levar o sistema a conclusões equivocadas. Isto pode ocorrer quando o usuário procura informações num contexto diferente daqueles que estão armazenados. De uma outra forma, o assunto pesquisado é diferente das consultas anteriores. Uma das alternativas possíveis de se resolver esse problema, que é utilizada nesse trabalho, é a possibilidade do usuário acessar os dados armazenados em seu perfil. Assim, o usuário pode editar o perfil de acordo com suas preferências atuais eliminando quaisquer erros que tenham ocorrido na construção automática deste. Nesta arquitetura a proposta é de uma modelagem colaborativa do perfil do usuário, pois o mesmo fica envolvido no processo de construção e atualização desse perfil.

Capítulo 6. Sistema RIA-BD Aplicado à LBC

Para avaliar a proposta, a biblioteca digital de Literatura Brasileira e Catarinense (LBC) do projeto SIDIE foi alterada de modo a incluir o sistema RIA-BD. A figura 14 apresenta a arquitetura da LBC acrescida dos módulos para a implementação do RIA-BD. Todos os componentes deste sistema já foram descritos no capítulo anterior e os demais no capítulo que trata da Biblioteca Digital do Projeto SIDIE.

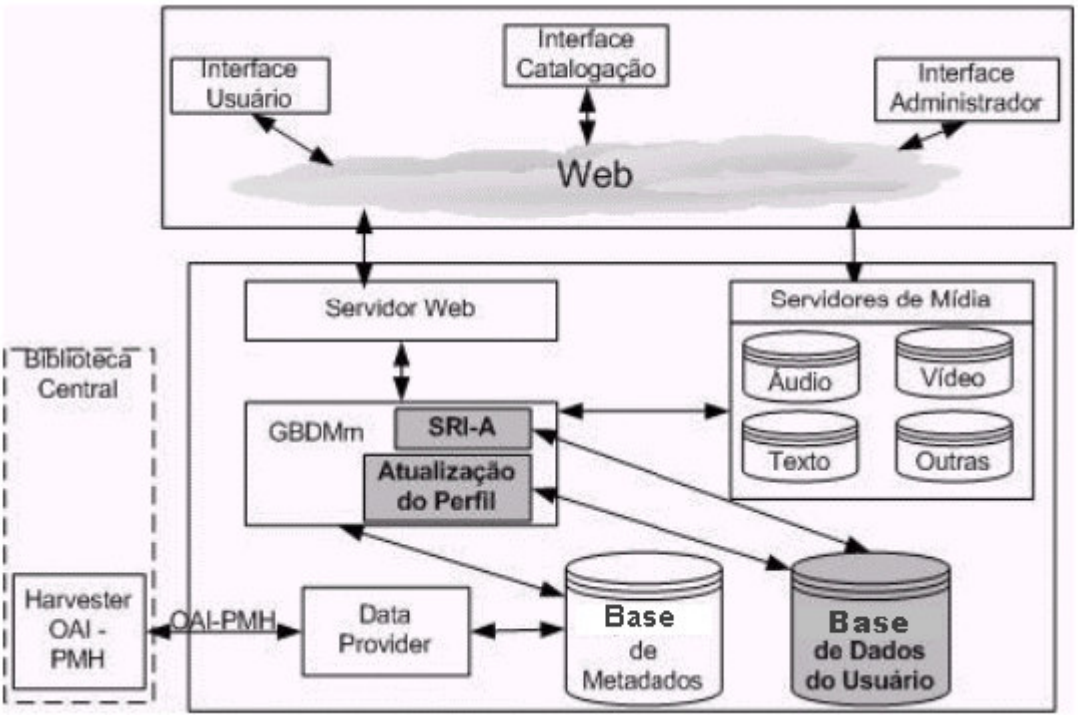

Figura 14-Arquitetura da LBC estendida

A seguir serão descritas as alterações ocorridas na LBC, a descrição será realizada em forma de um roteiro de navegação .

6.1 Cadastramento de Usuário

No momento que o usuário realiza o primeiro acesso à LBC, ele é direcionado para uma Interface de Autenticação, apresentada na figura 15. Nesta, o usuário tem diversas opções de navegação. As mais importantes neste contexto são:

- Clique do botão "Consulta": para realizar consultas sem a opção de adaptação
- Preenchimento de login e senha: quando o usuário já se cadastrou em algum outro momento. A partir daí, o usuário é direcionado para interface de busca simples (visto a seguir).
- Cadastramento do usuário: quando o usuário não está atualmente cadastrado e deseja utilizar os recursos de adaptação.

A figura 15 apresenta a Interface de autenticação.

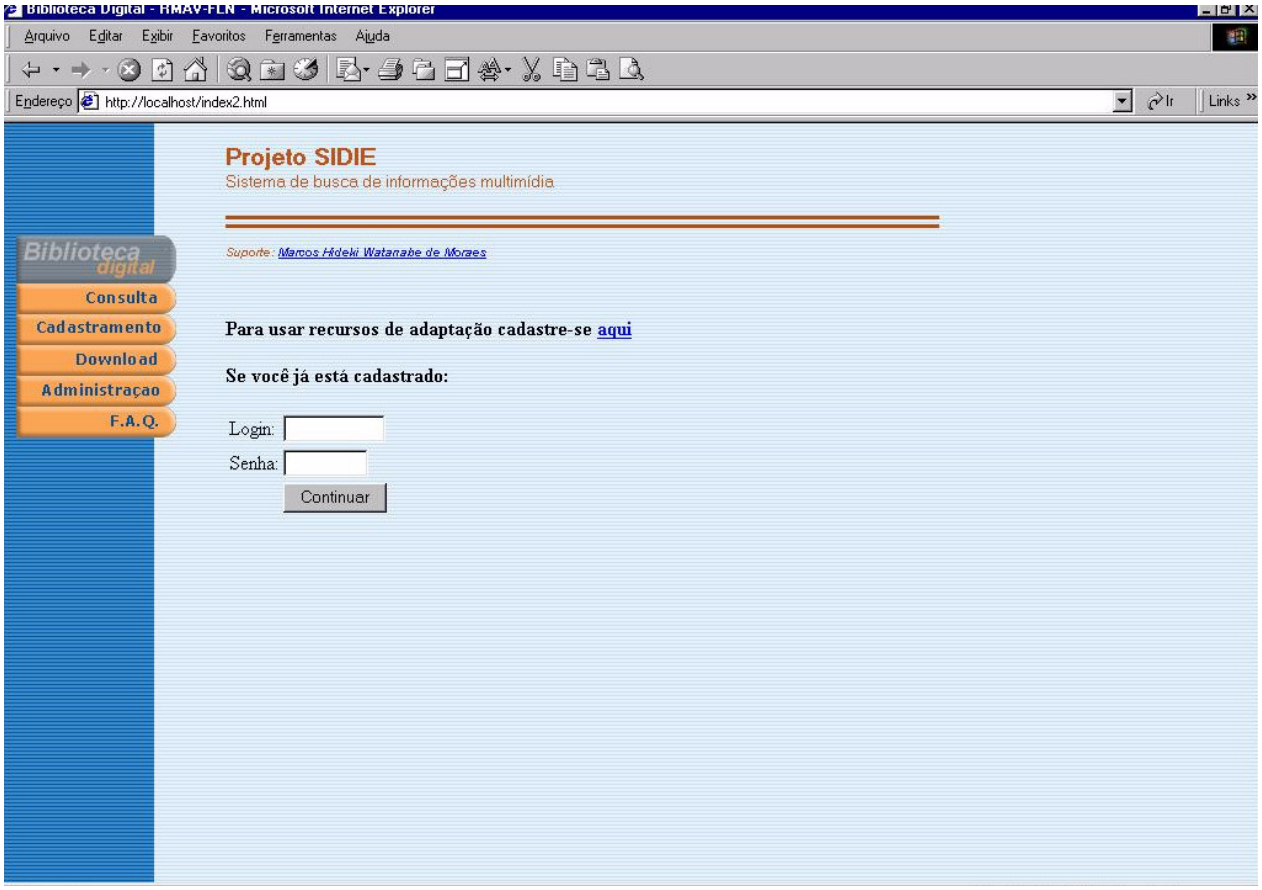

Figura 15-Tela de Autenticação de Usuário da RIA-BD

Caso o usuário não seja cadastrado e queira realizar o cadastro o "*link*" apropriado na Interface de autenticação o leva para a Interface de Cadastro, apresentada na figura 16. Esta Interface procura colher informações sobre o usuário visando implementar as adaptações sugeridas para facilitar a interação do mesmo com o sistema. Nesta Interface o usuário preenche seus dados pessoais, e pode escolher um Login e uma senha, o tipo de ordenação, os idiomas que ele domina, um endereço de e-mail para um contato mais rápido da biblioteca caso necessário e os autores preferidos, estes dados podendo ser editados a qualquer momento.

Uma vez cadastrado, nos próximos acessos o usuário pode ser identificado automaticamente através de um *Cookie* gravado no seu computador, como visto anteriormente.

Figura 16-Interface de Cadastramento do Usuário na RIA-BD

6.2 Edição do Perfil

Uma vez cadastrado e autenticado na RIA-BD, o usuário pode passar para a Interface de edição do perfil (figura 17) e cadastrar as suas preferências. Na realidade, o usuário pode alterar as configurações do seu perfil por intermédio do *link* apropriado <Edição do Perfil> nas Interfaces disponíveis a qualquer momento da navegação.

A figura 17 apresenta a interface de Edição do Perfil do usuário. Nesta tela o usuário pode alterar o seu endereço eletrônico, o endereço de uma página pessoal, que porventura ele possua, editar as preferências gerais e especificas, o idioma, acrescentar ou retirar autores preferidos, passar autores ou gêneros visitados para os preferidos e vice-versa.

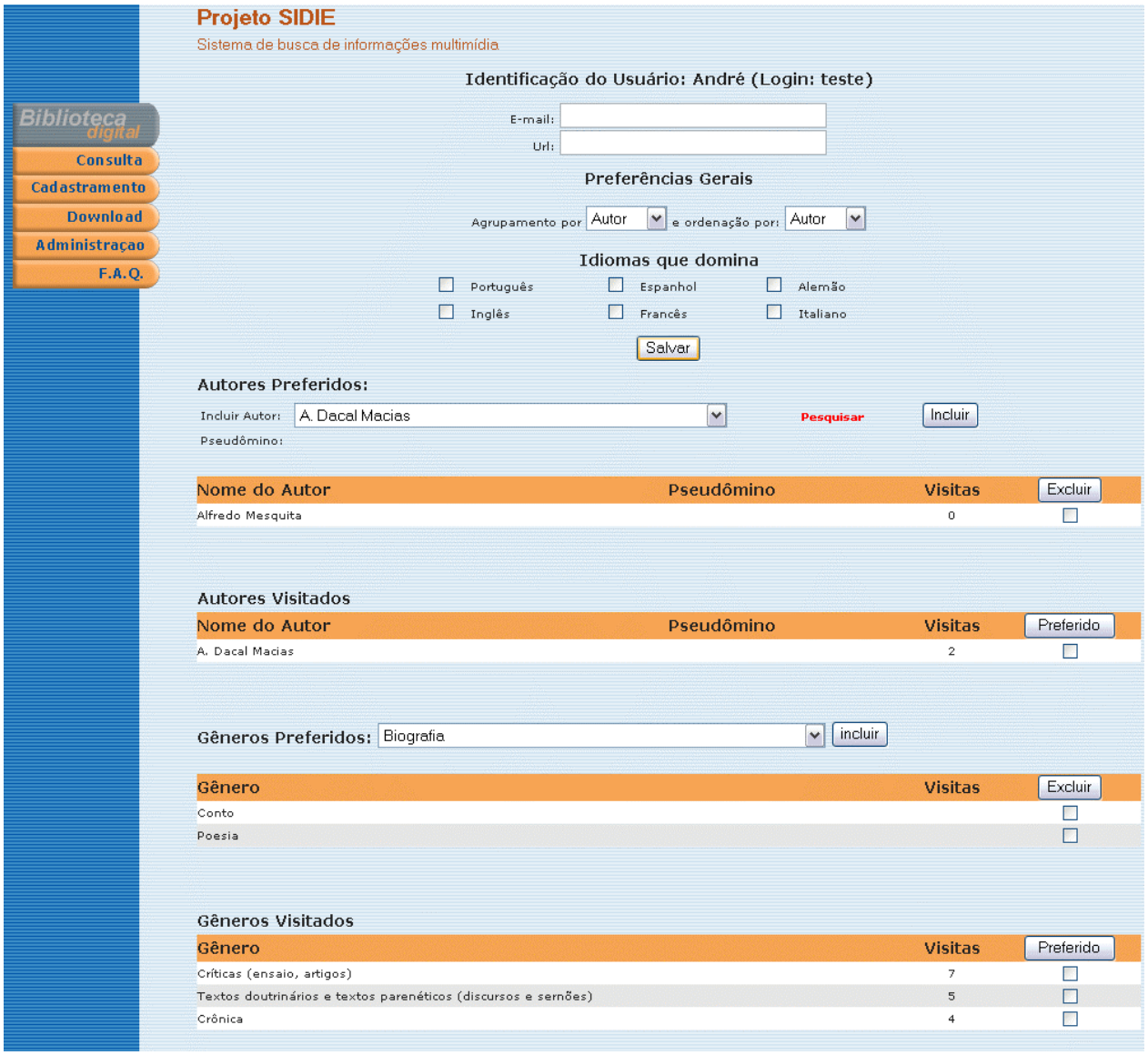

Figura 17-Tela de Edição do Perfil do Usuário

6.3 Interfaces de Busca simples e Apresentação dos Resultados

A figura 18 apresenta a interface de busca simples da LBC. Note que o usuário tem também a opção de utilizar a opção de busca avançada e busca usando um thesaurus temático de literatura.

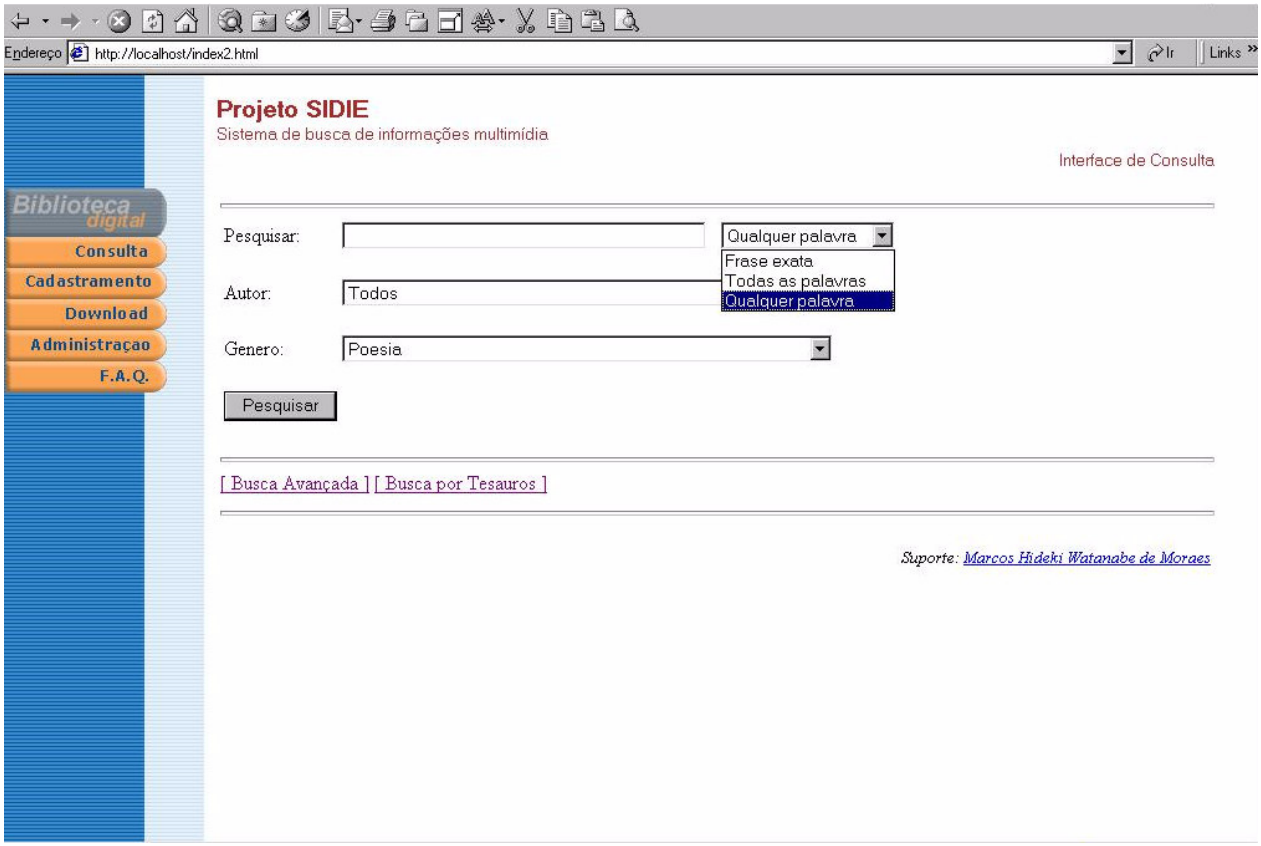

Figura 18-Tela de consulta simples da RIA-BD

O resultado da consulta é apresentado de acordo com as opções anotadas no perfil do usuário e de acordo com as observações efetuadas pelo RIA-BD. A figura 19 apresenta um resultado de busca personalizado onde o usuário optou por agrupar por autor e ordenar por autor.

Figura 19-Tela do resultado da busca personalizado

A partir da tela de resultados, o usuário pode escolher navegar pelo acervo, o que segue da mesma forma da LBC, com a diferença de que as características da obra acessada serão utilizadas para atualizar o perfil do usuário.

Neste tela as opções de navegação mais importante são:

- Acesso ao nome do autor: que tem como informações mais relevantes o nome do autor, ano de nascimento, ano de morte, o gênero(s) que do autor , as obras cadastradas, com um link para acessar detalhes da obra e um link para acessar a obra, se esta já estiver digitalizada e disponível na LBC.
- Acesso ao título da obra: nesta tela as principais informações são : titulo da obra, sub-titulo da obra, gênero literário, idioma da obra, nome da editora da 1ª edição, personagens da obra, fatos históricos relevantes sobre a obra e o acesso a obra, se ela já estiver digitalizada.

A figura 20 mostra o exemplo de uma possível navegação com informações sobre uma obra, nota-se nesta tela o botão "clique aqui" em "Acesse a Obra" , é no momento em que o usuário clica neste link que o seu perfil é atualizado, passa a contar mais um acesso a obra, pois é neste momento que o usuário efetivamente acessou a obra.

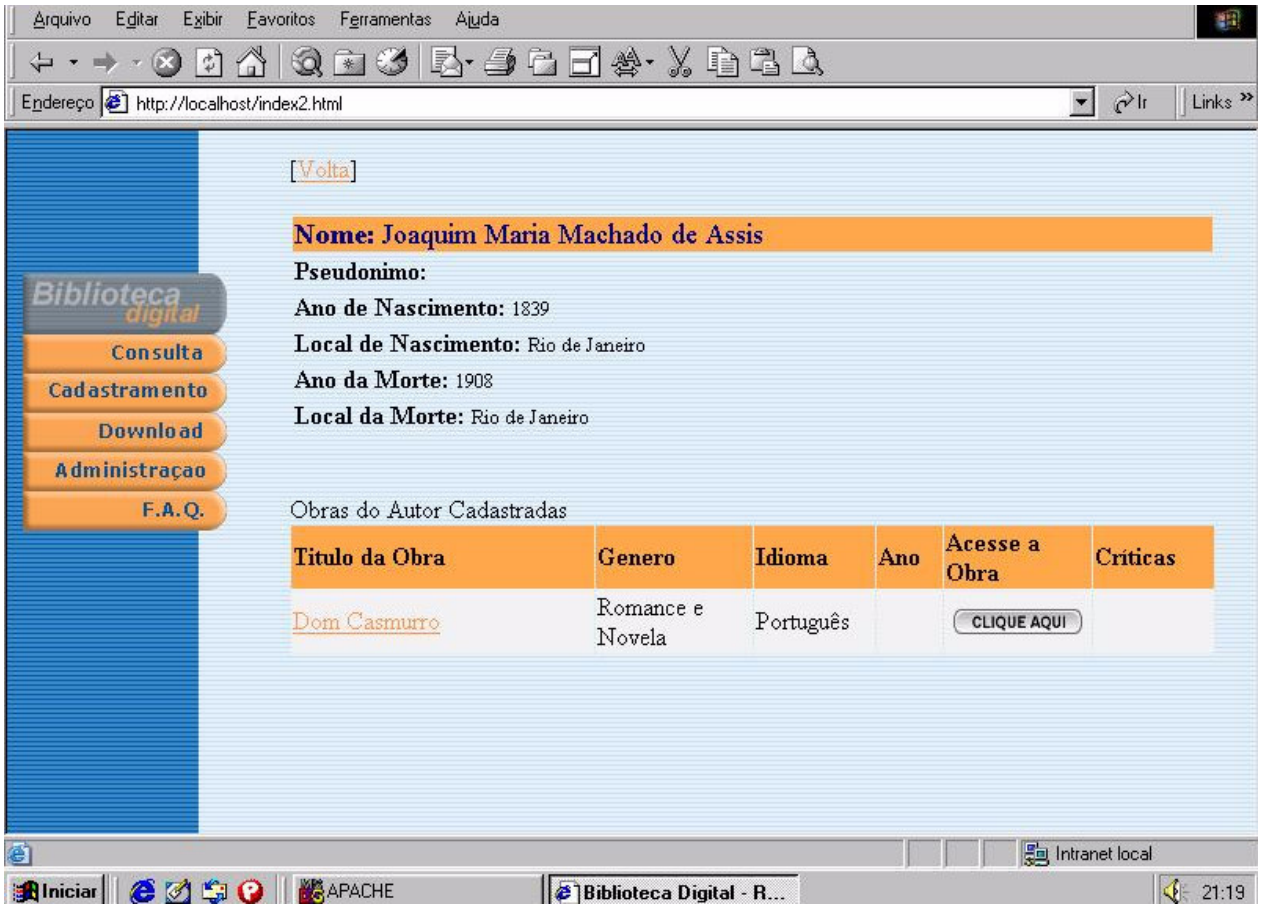

Figura 20-Tela de exemplo de navegação no acervo da RIA-BD

6.4 Protótipo Implementado e Testes Realizados

Os testes foram realizados em um protótipo da LBC, com as seguintes características:

- Autores: mais de 13000;

- Obras cadastradas: mais de 18000, e;
- Obras disponíveis: mais de 200.

A seguir é apresentado uma busca utilizando adaptabilitade e uma busca sem o uso da adptabilidade, validando a proposta no protótipo.

O usuário cadastrado efetua o *logon* na biblioteca através da interface de autenticação, conforme apresentado na figura 21.

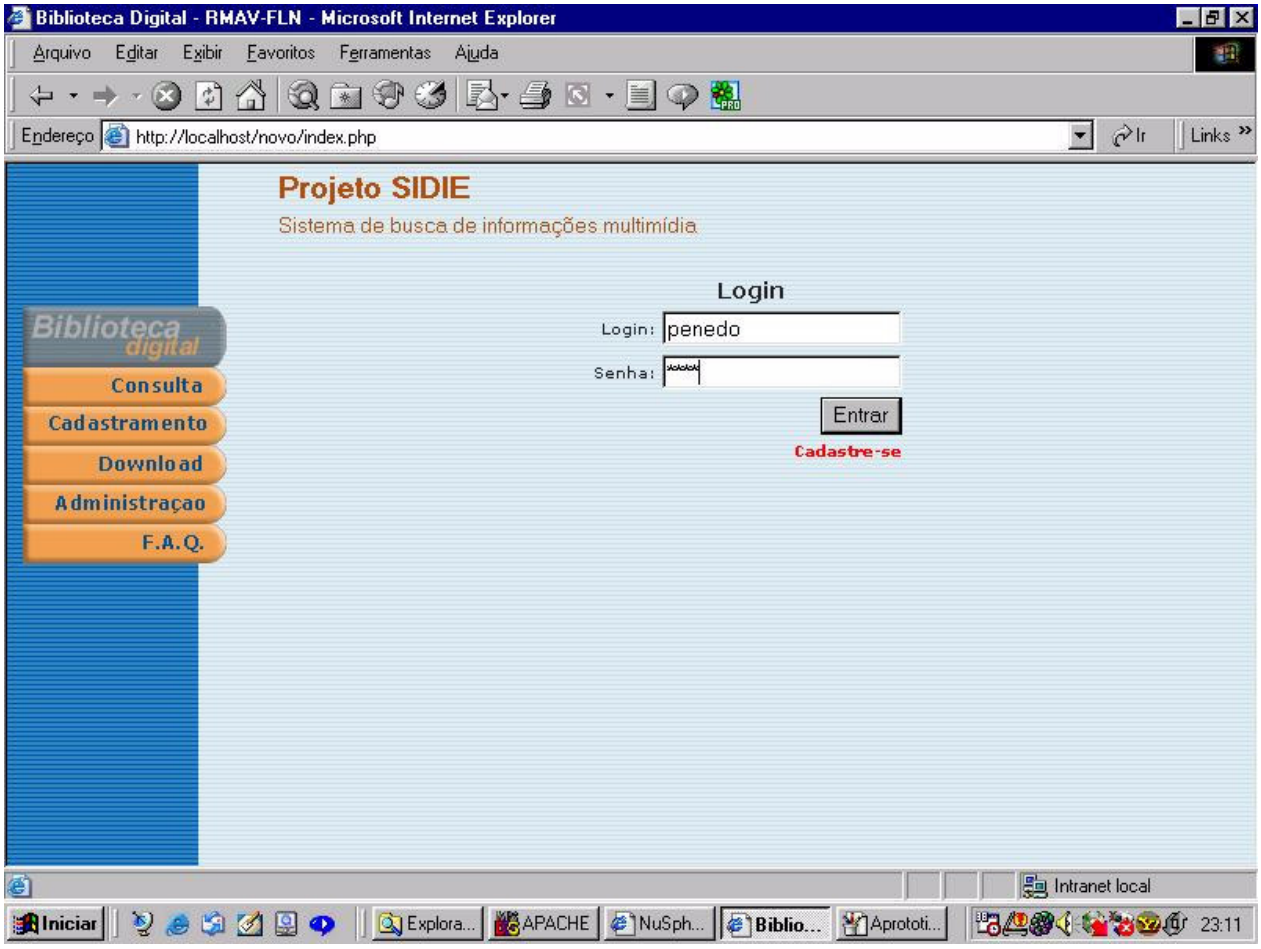

Figura 21- Protótipo, Autenticação do Usuário na Biblioteca

Uma vez logado o usuário tem a disposição a interface onde pode realizar a edição do seu perfil. A tela de edição do perfil foi dividida nas figuras 22, 23 e 24 para possibilitar uma melhor visualização.

Pode-se observar na figura 22, que o usuário tem como preferências o agrupamento por autor e a ordenação neste agrupamento, também por autor.

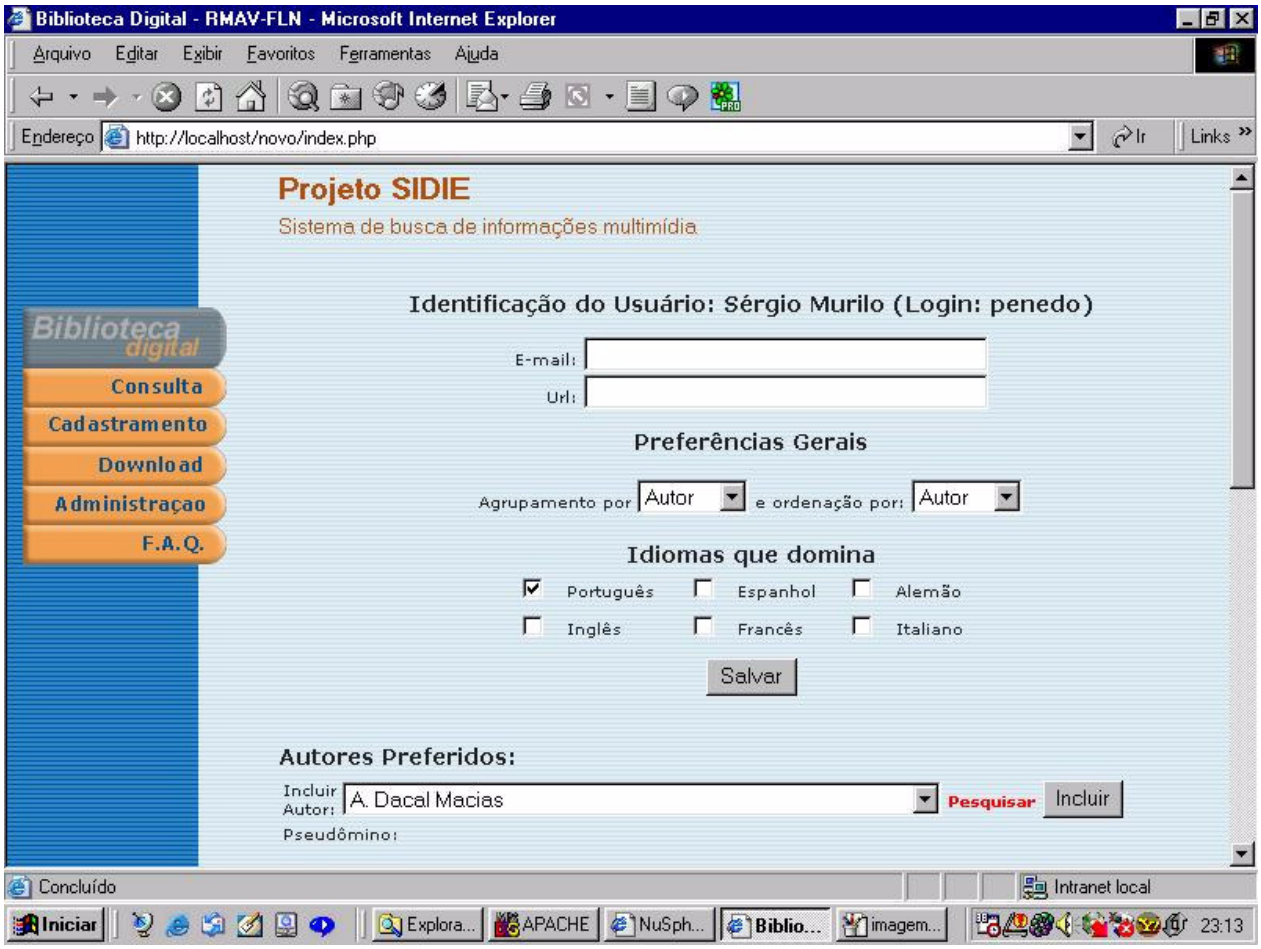

Figura 22 – Protótipo, Tela de Edição do Perfil, primeiro segmento

Na figura 23 observa-se que o usuário Sérgio Murilo tem como Autor Preferido José Carlos Abbate, e também já visitou obras de Joaquim Maria Machado de Assis e de Alex Souza Cabistani.

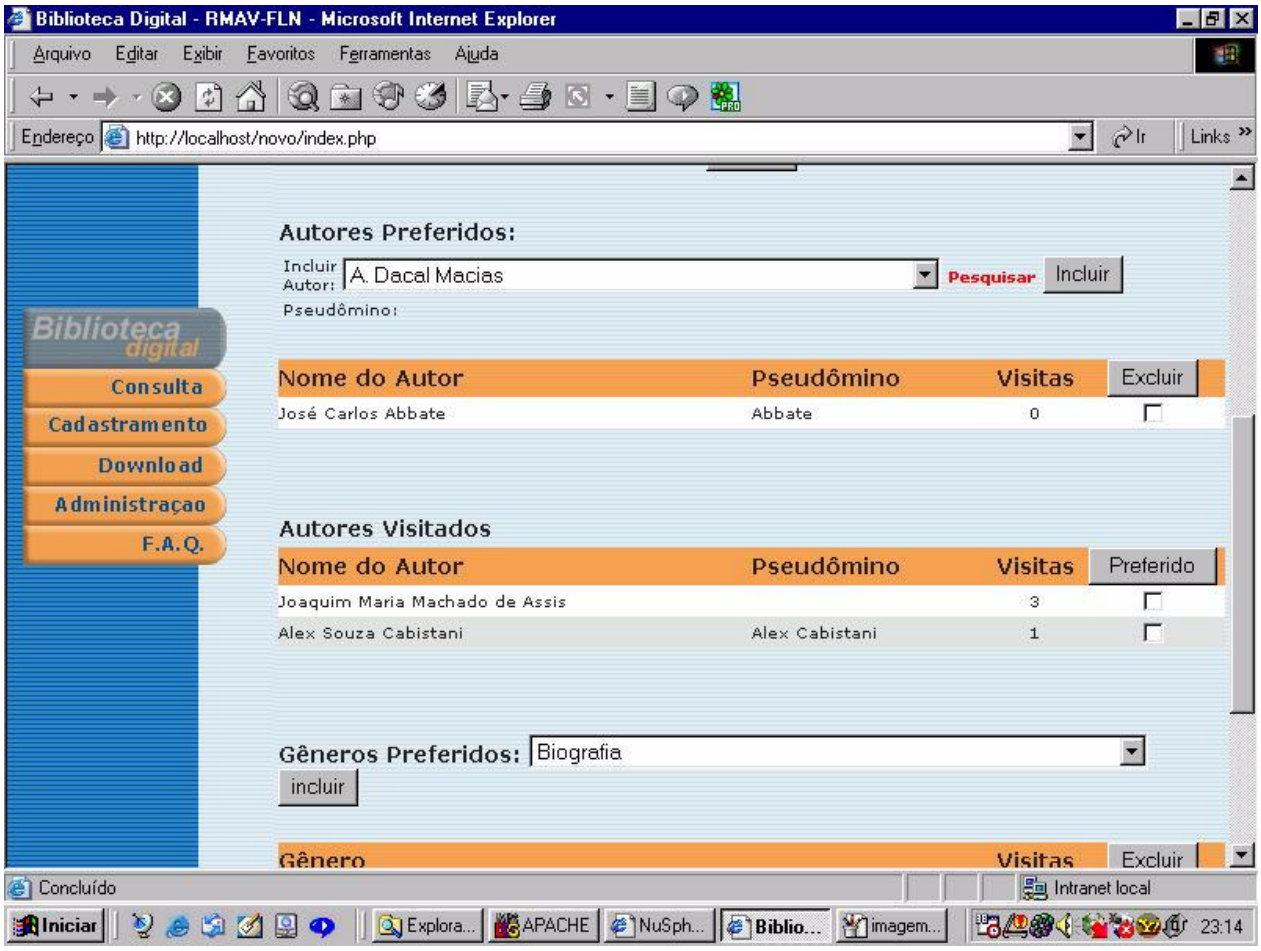

Figura 23 – Protótipo, Tela de edição do Perfil, segundo segmento

Na figura 24 mostra que o usuário já visitou obras dos gêneros Poesia e Conto.

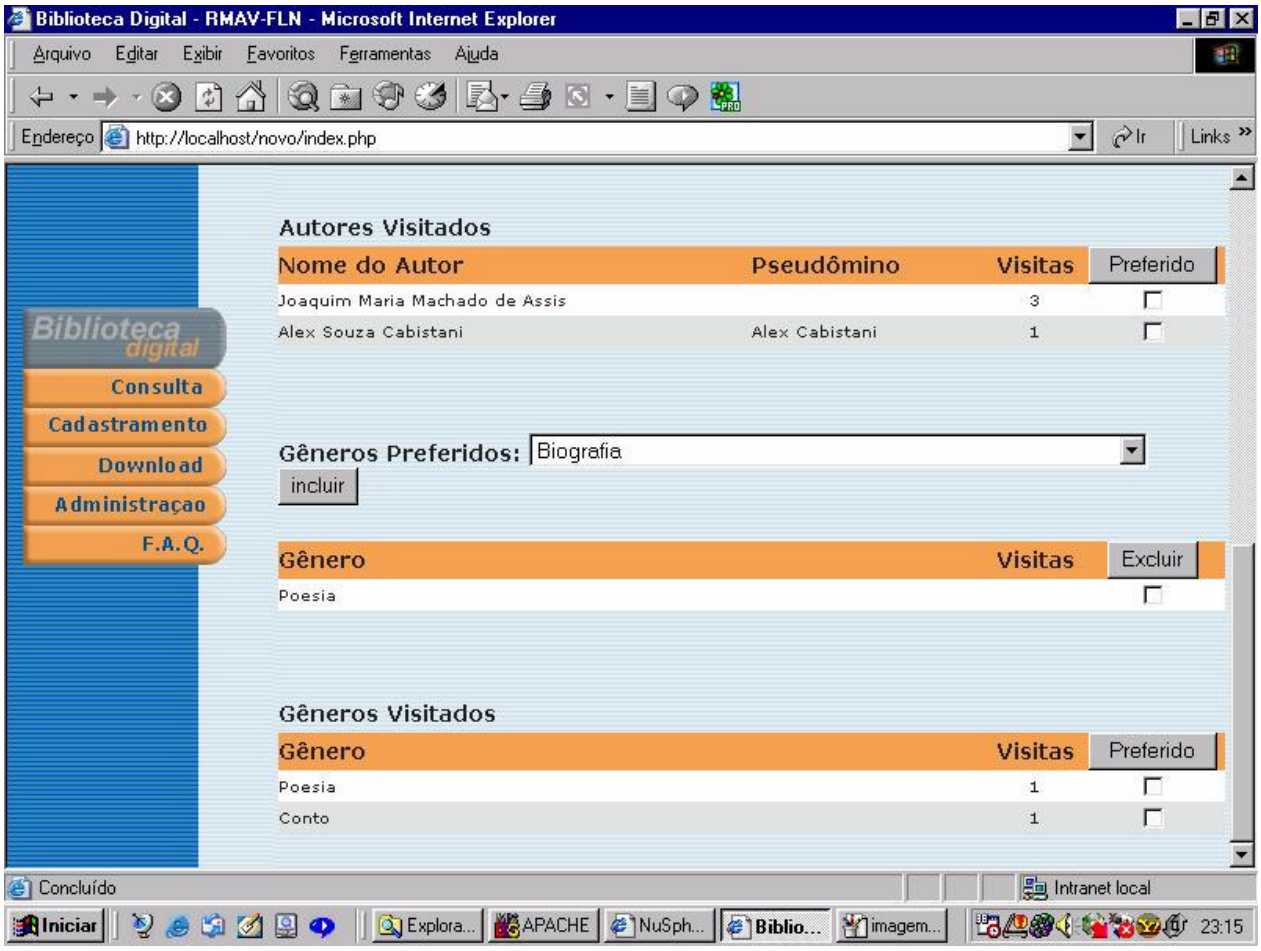

Figura 24 – Protótipo, Tela de Edição do Perfil, terceiro segmento

Da edição do perfil o usuário, através dos botões à esquerda, pode realizar uma consulta, sendo apresentado a tela de consulta, conforme a figura 25.

Nesta tela optou-se em realizar uma busca por uma palavra bem genérica e que aparecem na maioria das obras, isto com o intuito de demonstrar o sistema quando operando em condições extremas, quer dizer quando deve apresentar um resultado com muitas obras. No caso foi escolhida a letra “a”.

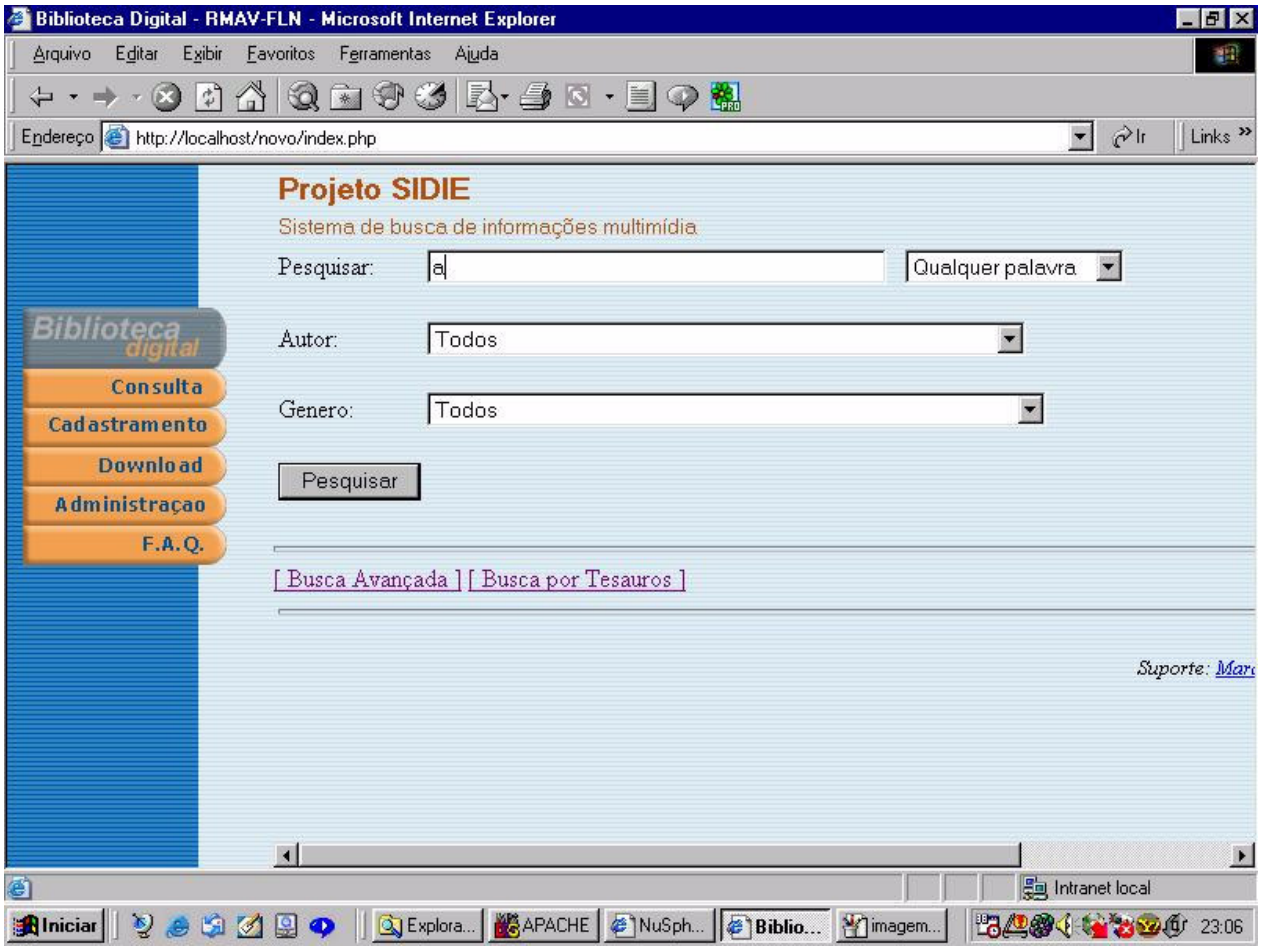

Figura 25 – Protótipo, Consulta Simples

A figura 26 apresenta o resultado da busca com adaptabilidade. Como o usuário Sérgio Murilo optou, o resultado apresenta no primeiro grupo os autores preferidos, ordenados em ordem alfabética, a seguir os autores visitados, e finalmente os demais autores, também ordenados em ordem alfabética.

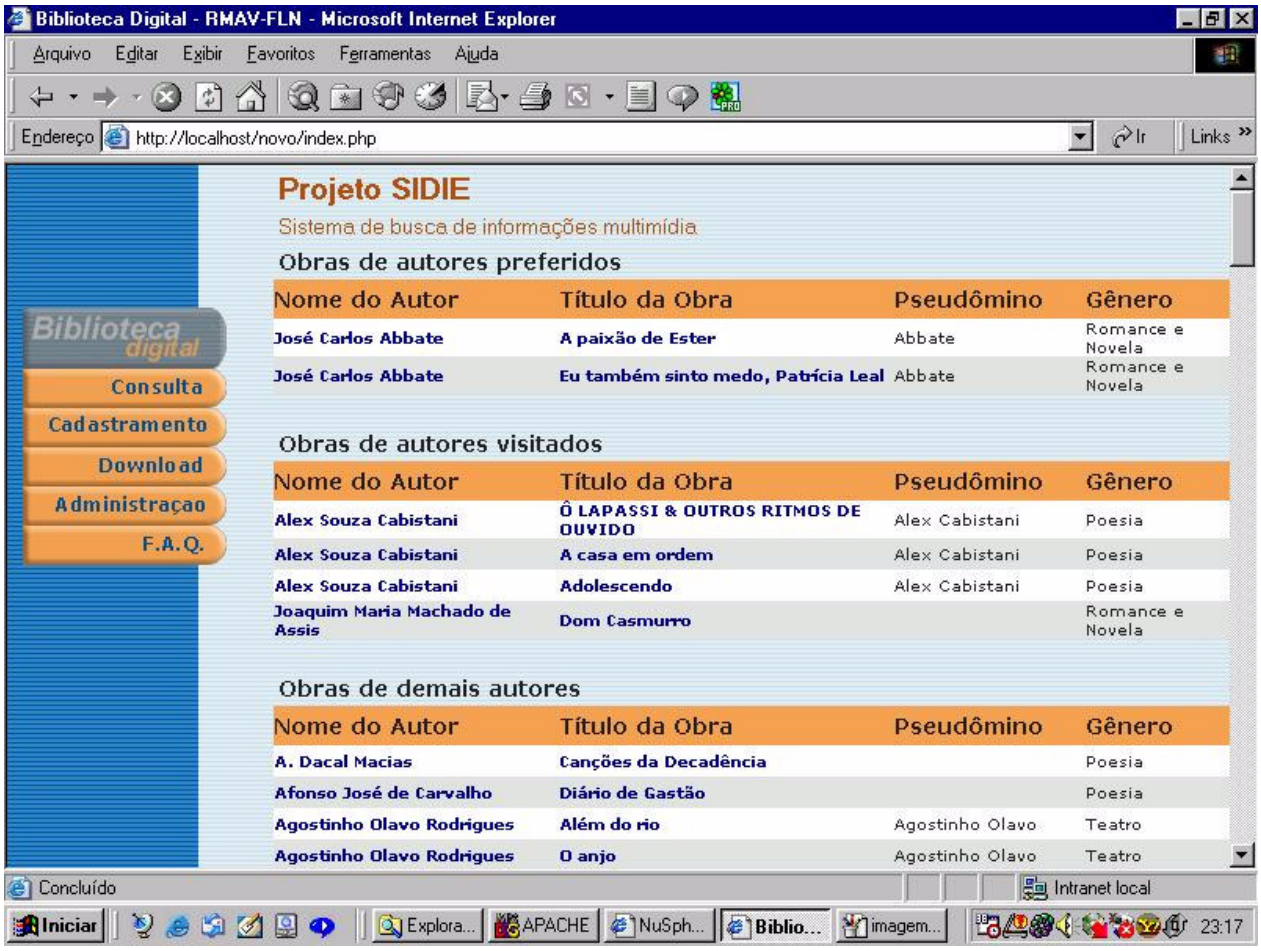

Figura 26 – Protótipo, Resultado da Busca com Adaptabilidade, primeiro segmento

A figura 27 é a continuação da figura 26, onde é apresentado a continuação dos demais autores.

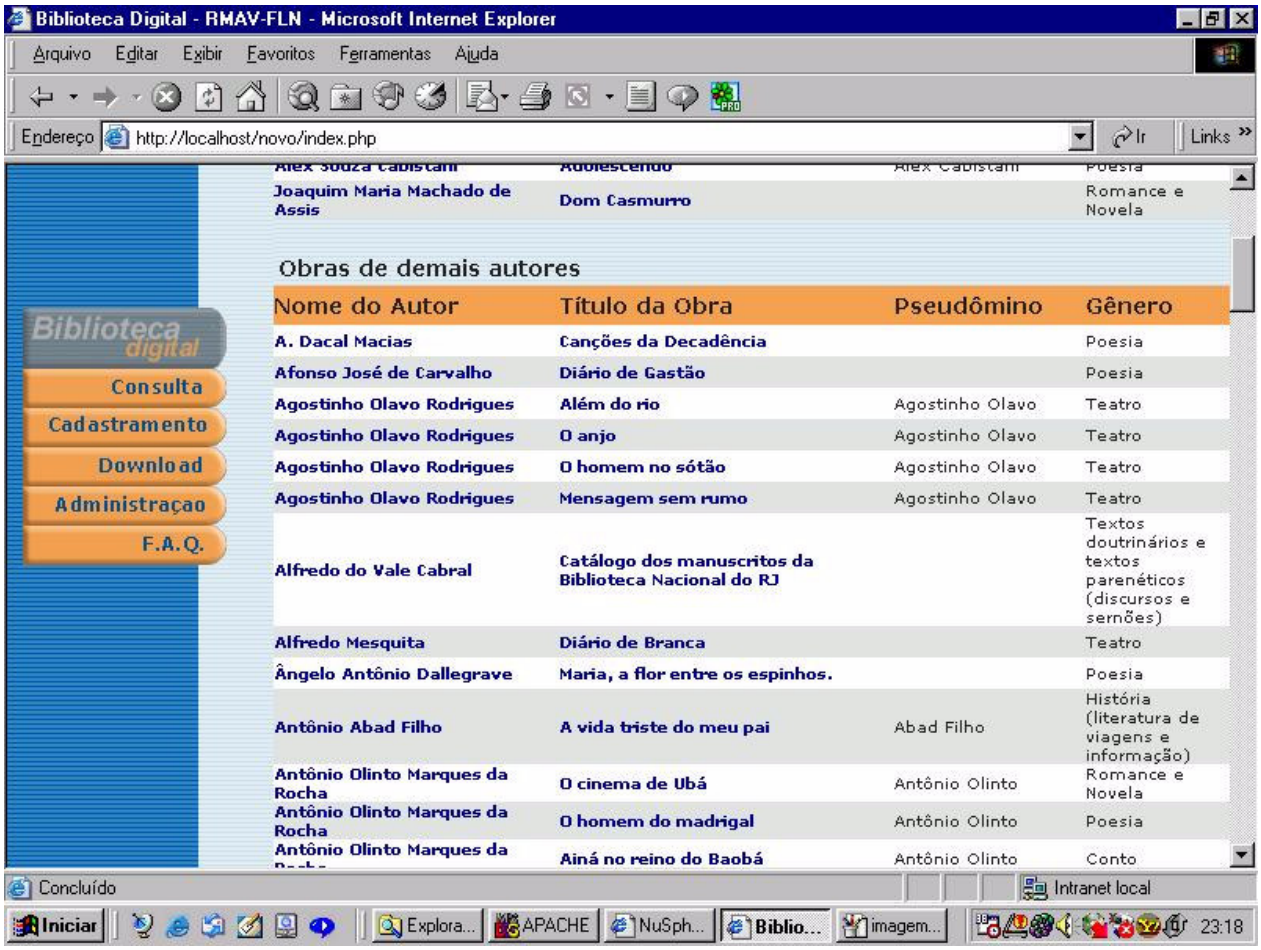

Figura 27 - Protótipo, Resultado da Busca com Adaptabilidade, segundo segmento

A figura 28 apresenta o resultado com o mesmo critério de busca, só que sem o uso da adptabilidade, isto é, o usuário não realizou o longon no sistema. Nota-se que foram localizadas 139 obras que atendem ao critério de pesquisa.

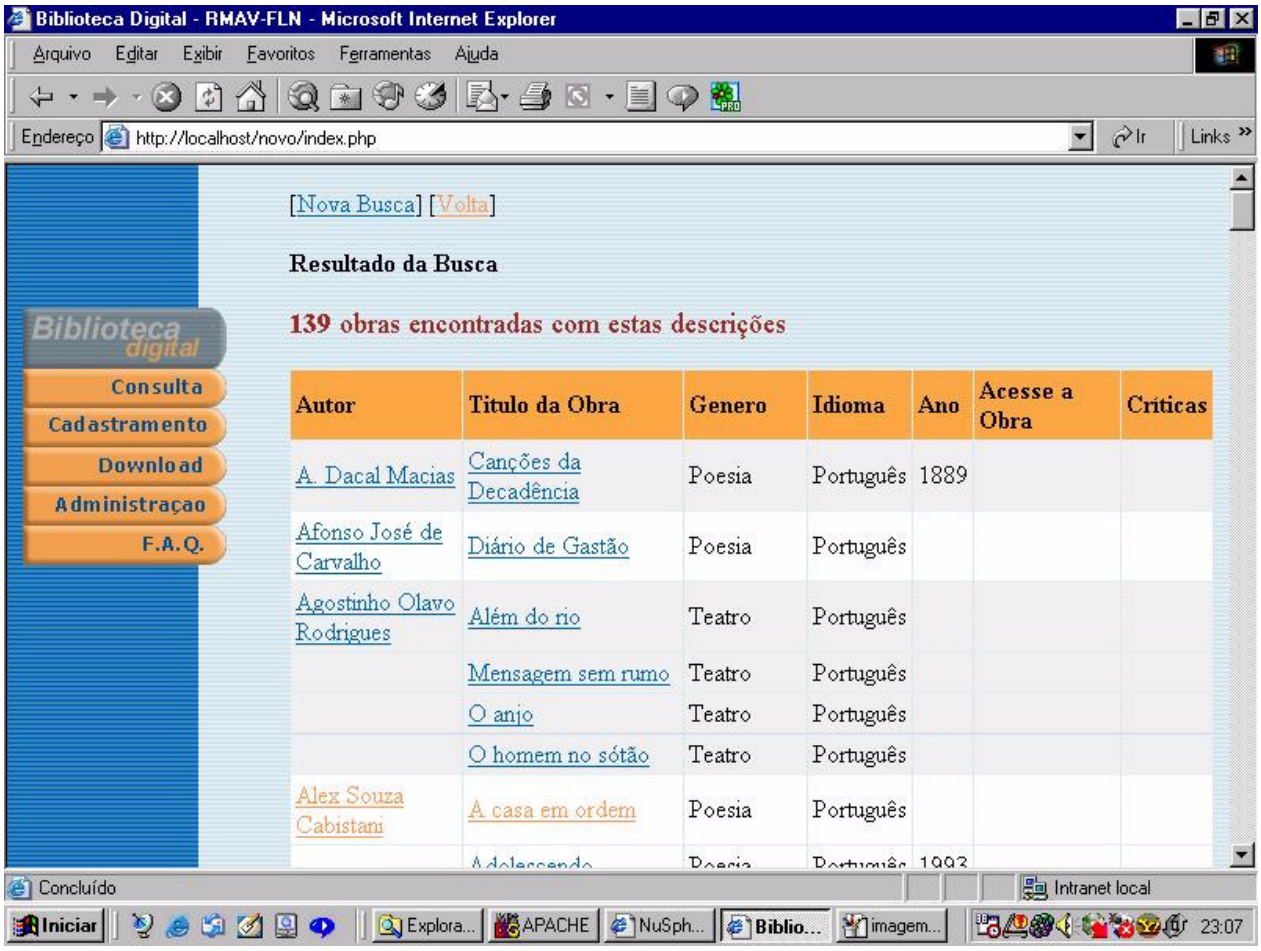

Figura 28 - Protótipo, Resultado da Busca sem Adaptabilidade, primeiro segmento

A figura 29 é a continuação da tela da figura 28. Observa-se na Barra de Rolagem vertical à direita, que os nomes dos autores preferido, José Carlos Abbate e do Visitado Joaquim Maria Machado de Assis, aparecem abaixo da metade dos resultados, obrigando o usuário a procurar e ainda poder passar sem perceber por o nome de sua preferência.

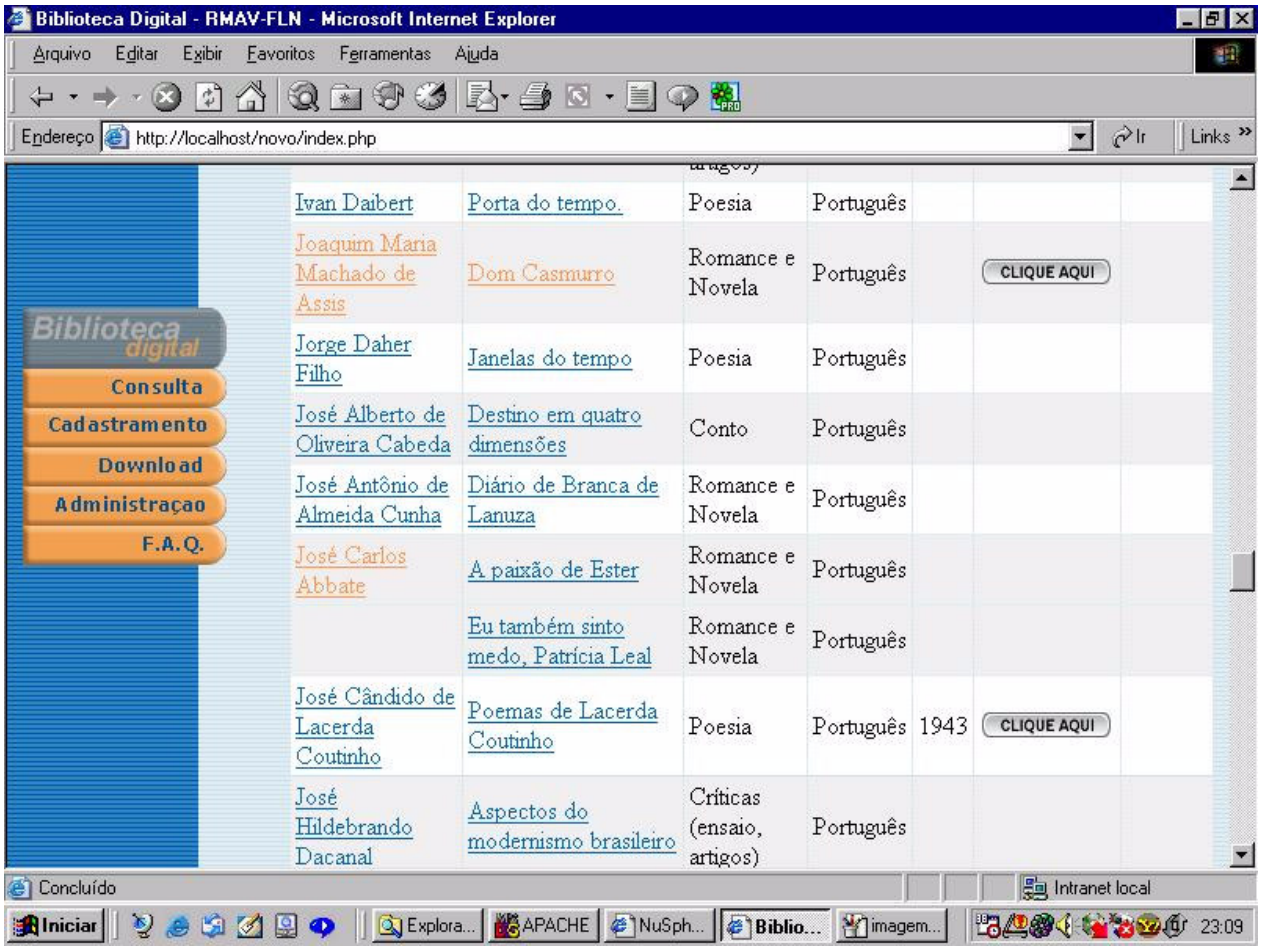

Figura 29 - Protótipo, Resultado da Busca sem Adaptabilidade, segundo segmento

Ao final pode-se verificar que os testes validaram a proposta apresentada. Ao ser realizada uma busca que retorna muitas obras, o efeito de sobrecarga cognitiva imposta ao usuário é praticamente extinto com o sistema RIA-BD, pois ele se dirige diretamente aos preferidos ou visitados com mais rapidez e neste caso a própria apresentação do resultado já facilita a interação do usuário com a biblioteca.

Um problema observado foi quanto aos autores visitados. Nos testes observou-se que os autores visitados começaram a aumentar muito e a sobrecarregar a apresentação dos resultados. Procurou-se levar o sistema a um limite de apresentação de autores visitados, por exemplo, e pode-se constatar uma sobrecarga de informações, justamente o que o RIA-BD procura atenuar.

Uma solução para resolver o problema pode ser limitar a quantidade máxima de resultados para os visitados que serão apresentados aos usuários. Nesta situação serão apresentadas prioritariamente as obras cujos autores tiverem um maior número de acesso. Ao resultado dos visitados deve ser adicionada a explicação que existem mais itens de menor relevância e que estão ocultos, oferecendo um *link* para que eles possam ser apresentados.

Com a implementação do RIA-BD na LBC e disponibilizada a comunidade espera-se ampliar os testes, incluindo estudantes e professores da área de comunicações, onde se pode obter inclusive estatísticas de melhora em relação ao sistema convencional.

6.5 Resumo

O capítulo apresentou a integração do sistema RIA-BD à LBC, as novas funcionalidades implementadas e os resultados iniciais dos testes.

Apesar dos testes iniciais apresentarem resultados satisfatórios verifica-se a necessidade dos mesmos serem realizados por uma comunidade real de usuários de literatura, pois os mesmos podem verificar detalhes e necessidades que porventura tenham sido omitidos inicialmente.

Capítulo 7. Conclusões e Trabalhos Futuros

O desenvolvimento da tecnologia está ocorrendo nos últimos tempos de forma muito rápida, gerando uma integração muito maior dos sistemas computacionais com áreas como a de comunicação. Uma grande beneficiada com esta integração é o campo da Internet, onde as tecnologias se desenvolvem, surgem e se renovam quase diariamente.

As bibliotecas digitais surgem neste cenário como uma ferramenta capaz de suprir uma grande demanda por conhecimento que sempre existiu na humanidade, com a própria Internet sendo uma grande biblioteca digital. É natural que, após passado um primeiro momento de euforia, onde vários serviços puderam ser transportados para um meio digital "virtual ", esses mesmos serviços se voltem agora para um aprimoramento de suas técnicas.

O grande conhecimento adquirido pelos serviços de comunicação como indexação de documentos, armazenamento ordenado para uma posterior recuperação, e até mesmo a busca por conhecer os seus usuários, como já tratava o livro de Nocetti (1980) sobre disseminação da informação, serviram como base para uma nova era do conhecimento.

Este trabalho se posiciona exatamente neste ponto da evolução deste sistema, onde se procura verificar os meios de facilitar o acesso das pessoas comuns aos sistemas de informação computacional, como as bibliotecas digitais.

Com este objetivo foram levantadas informações sobre as bibliotecas digitais, como foi o seu desenvolvimento, suas arquiteturas mais clássicas, os sistemas de gerenciamento da informação, sistemas que utilizam modelagem de usuários para facilitar a interação homem-máquina, como as pesquisas em Hiperdocumentos Adaptativos e Sistemas de Recuperação de Informação.

Pode-se identificar como pontos fracos da técnica:

- Definição do peso de relevância deve ser aprimorado para uso em bibliotecas digitais;

- Conversão do metadado dos visitados para os preferidos deve ser melhor investigado.

Como pontos fortes:

- Apresentação dos resultados facilita o usuário encontrar rapidamente a informação que procura;
- A técnica é simples de ser implementada, não querendo dizer contudo que não possa ser tomada como base para trabalhos mais complexos, como por exemplo, podem ser adicionadas técnicas de IA para prever resultados esperados pelos usuários, técnicas de *Data Minning* , etc.

Como parte do trabalho, foi implementado um protótipo para validar a técnica apresentada. Devido ao pouco tempo para o desenvolvimento, os testes foram realizados na fase de desenvolvimento do mesmo, pelo autor, devendo posteriormente, ser incorporado a biblioteca original onde serão realizados testes com vários níveis de usuários e com prazos mais longos.

Os testes iniciais foram satisfatórios, melhorando a apresentação dos resultados, com a informação buscada sendo mais rapidamente acessada, melhorando a interação com o sistema da biblioteca de uma forma geral.

Como trabalhos futuros, espera-se implementar integralmente a técnica na biblioteca digital de literatura do projeto SIDIE. A investigação do peso de relevância é um outro item que deve ser investigado, através de aplicação de outras técnicas de recuperação de informações como visto no capítulo 4 e de investigações com a RIA-BD.

Uma outra proposta seria a formalização de modelagem de usuários padrões para bibliotecas digitais. Para tal, seria necessária a adoção de padrões de representação dos metadados que definem o perfil e protocolos permitindo o acesso a tais perfis.

Capítulo 8. Referências

(Aldenderfer, 1984) Aldenderfer, M.S.; Blashfied, R. K. Cluster Analysis. Beverly Hills, CA: Sage, 1984. 88p.

(Allgayer, 1989) Allgayer, J.,Harbusch, K., Kobsa, A., Reddig, C.,Reithinger, N., Schmauks, D. XTRA: A Natural-Language Access System to Expert Systems. International Journal of Man-Machine Studies 31, 161-195.

(Amato, 2005) Amato, Giuseppe; Straccia, Umberto. User Profile Modeling and Applications to Digital Libraries. Istituto di Elaborazione dell'Informazione – C.N.R., Pisa, Italy. Disponível em: http://faure.iei.pi.cnr.it/~{amato,straccia}. Acesso em 2005.

(Antunes, 2003) Antunes, Fernando Carlos Ponce de Leon Antunes. Tradutor de consultas XML/Dublin Core para Sentenças SQL para recuperação de Recursos de Bibliotecas Digitais Baseadas em Banco de Dados Relacionais. Dissertação de Mestrado, CPGCC/UFSC. 2003.

(Arms, 1997) Arms, William Y.; Blanchi, Christopher; Overly, Edward A.; An Architecture for Informationin Digital Libraries. Corporation for National Research Initiatives. Reston, Virginia. D-Lib Magazine, Febrary, 1997. Disponível em: http://www.dlib.org/dlib/february97/cnri/02arms1.html. Acesso em 2005.

(Arms, 2005) Arms, William. The online edition of Digital Libraries. MIT Press, updated with additional material by the author. Disponível em http://www.cs.cornell.edu/wya/DigLib/new/Chapter1.html. Acesso novembro, 2004.

(Baeza-Yates, 1999) Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier de. Modern Information Retrieval. New York: ACM Press, 1999.

(Barreto, 2001) Barreto, Jorge Muniz. Inteligência Artificial no Limiar do Século XXI, Abordagem Híbrida simbólica Conexionista e Evolutiva. 3ª ed. Duplic, 2001.

(Brusilovsky, 1996) Brusilovsky, Peter. Methods and Techniques of Adaptive Hypermedia. User Modeling and User Adapted Interaction. 1996, v 6, n 2-3, pp 87-129.

(Brusilovsky, 1998) _. Efficient Techiniques for Adaptive Hypermedia, Human Computer Interaction Institute School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213 USA, 1998.

(Brusilovsky, 2001) _.Adaptive Hypermedia. Kluwer Academic Publishers. User Modeling and User-Adapted Interaction. 11:87-110.2001.

(CEN, 2004) CEN. European Committee for Standardization. Disponível em http://www.cenorm.be/cenorm/index.htm.Acesso em 2004.

(Chakrabarti, 2003) Chakrabarti, Soumen.Mining The Web Discovering Knowledge from Hypertext Data . Morgan Kaufmann Publishers. By Elsevier Science (USA) 2003. 345 p.

(Chen, 1998) Chen, L., Sycara, K. WebMate: Personal Agent for Browsing and Searching. In PTOC of the 2nd Int'l Conf on Autonomous Agents, 132-139. 1998.

(Chen, 2002) Chen, M., LaPaugh, A.S., Singh, J.P. Predicting Category Accesses for a User in a Structured Information Space. SIGIR'2, August 11-15, 2002, Tampere, Finland.

(Chen, 2002) Chen, Mao, Paugh, Andrea La; Singh, Jaswinder Pal. Categorizing Information Objects from User Access Patterns. CIKM'02, November 4-9, MCLean, Virginia, USA. ACM 2002.

(CIMI, 2004) CIMI, Consortium for the Computer Interchange of Museum Information. Disponível em : http://www.cimi.org/. Acesso em 2004.

(Conallen, 2003) Conallen, Jim. Desenvolvendo Aplicações Web com UML. Editora Campus.2003.

(Crestani, 1999) Crestani, Fabio; Rijsbergen, Cornelis J. Van. A Model for Adaptive Information Retrieval. Kluwer Academic Publishers, Boston. 1-30. 1999.

(Chaves, 2002) Chaves, Marcirio Silveira. Padrões em Bibliotecas Digitais. PUCRS, PPGCC. 2002. Disponível em: http://www.inf.pucrs.br/~mchaves/pg_portugues/master/trabalhos/ti1.pdf. Acesso em 2005.

(Cunha, 1999) Cunha, Murilo Bastos da. Desafios na Construção de uma Biblioteca Digital. Revista eletrônica Ciência da Informação. Vol. 28, nº 1999. Disponível em http://www.ibict.br/cienciadainformacao/archive.php. Acesso em 2005.

(DCMI, 2005) DCMI, Libraries Working Group. DC- Library Application Profile. Disponível em: http://dublincore.org/documents/library-application-profile/. Acesso em 2005

(Deniman, 2003) Deniman, Dave; Sumner, Tamara; et al. Merging Metadata and Content-Based Retrieval. Journal of Digital Information, volume 4 issue 3, article nº 231, 2003-11-06. Disponível em: http://jodi.ecs.soton.ac.uk/Articles/v04/i03/Deniman/. Acesso em 2005.

(Dizaj, 2003) Dizaj, S. Maleki; Z.A. Othman; H. O. Nyongesa, J. Siddiqi. Evolutionary Reinforcement of User Models in An Adaptive Search Engine. IEEE Computer Society http://csdl.computer.org/comp/proceedings/wi/2003/1932/00/19320706abs.htm. Acesso em 2005.

(DLF, 1998) DLF, Digital Library Federation.Waters, Donald J. What are Digital Libraries ? Clir issues number 4, july/august 1998. Disponível em: http://www.clir.org/pubs/issues/issues04.html. Acesso em 2005.

(Fernandes, 2003) Fernandes, Anita Maria da Rocha. Inteligência Artificial Noções Gerais. Visual Books, 2003.

(Ferneda, 2004) Ferneda, Edberto. Recuperação de Informação: Análise sobre a Contribuição da Ciência da computação para a Ciência da Informação. Tese de doutorado USP. Disponível em: http://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230/. Acesso em 2005.

(Ferreira, 1997) Ferreira, José Ricon. A biblioteca digital. Revista USP nº 35, Informática/Internet, Setembro/Outubro/Novembro de 1997. Disponível em http://www.ime.usp.br/~is/infousp/rincon/rincon.htm. Acesso 2005.

(Fox, 1995) Fox, Edward A. Digital Libraries. Communications of the ACM, 38(4), april/1995, pp.23-8.

(Geisler, 2002) Geisler, Gary; et.al. Creating Virtual Collection in Digital Libraries: Beneficts and Implementation Issues. Procediing in ACM, july 2002.

(IBICT, 2001) IBICT, Instituto Brasileiro de Informação em Ciência e Tecnologia. Projeto Técnico da Biblioteca Digital Brasileira - BDB. Disponível em http://www.ibict.br/secao.php?cat=BDB. Acesso 2005.

(ISO, 2003) ISO. Standard 15836-2003 (February 2003).Disponível em: http://www.niso.org/international/SC4/n515.pdf.

(Kass, 1988) Kass, R. Acquiring a Model of the User's Beliefs from a Coopertive Advisory Dialog. Ph.D. Thesis. Depto. Of Information and Computer Science, University of Pennsylvania, Philadelphia, PA.1988.

(Kobsa, 1985) Kobsa, A. Benutzermodellierung in Dialogsystemen. Springer Verlag, Berlin, Heidelberg. 1985.

(kobsa, 1990) _. Modeling the User's Conceptual Knowledge in BGP-MS, a User Modeling Shell System. Computational Intelligence 6, 193-208. 1990.

(Kobsa, 1995) _; Pohl, W. The BGP-MS User Modeling System. User Modeling and User-Adapted Interaction .4(2), 59-106. 1995.

(Kobsa, 2001) _. Generic User Modeling Systems. User Modeling and User-Adapted Interaction 11: 49-63. Kluwer Academic Publishers. 2001.

(Lagoze, 1995) Lagoze, Carl; James R. Davil. Dienst: An Architecture for Distributed Document Libraries. Communications of The ACM, April 1995, vol.38 nº 4.

(Lee, 1997) Lee, Dik L., et. al. Document Ranking and the Vector-Space Model. IEEE Computer Society. Disponível em: http://csdl.computer.org/comp/mags/so/1997/02/s2067abs.htm. Data 1997.

(Letícia, 2003) Letícia, Teresinha Letícia da Silva. Uma Arquitetura de Recuperação de Informações Usando Recursos de Hiperdocumentos Adaptativos Aplicada a Bibliotecas Digitais. Dissertação de Mestrado, CPGCC/UFSC. 2003.

(Levacov, 1997) Levacov, Marilia. Bibliotecas Virtuais: Problemas, Paradoxos, Controvérsias. Intexto, PPGCOM/UFRGS, 1997.

(Lucier, 1995) Lucier, Richard E..Buindilg Digital Library for The Healt Science: Information Space Complementign Information Place. Bulletin of The Medical Library Association, 83 (3) jul/1995, pp. 346-50.

(Liu, 2001) Liu, Duen-Ren; Lin, Yuh-Jaan; et.al. A Framework for Personalized E-Catalogs: an Integration of XML-based Metadata, User Models and Agents. Proceedings of the 34 th Hawaii International conference on System Sciences – IEEE. 2001.

(MARC, 2005) MARC. Understanding MARC Bibliographic: Machine Readable Cataloging. Disponível em: http://www.loc.gov/marc/umb/. Acesso 2005.

(Marcondes, 2002) Marcondes, Carlos Henrique; Luis Fernando Sayão. Documentos digitais e Novas Formas de Cooperação entre Sistemas de Informação em C& T. Ciência da Informação, vol. 31, nº 3, IBICT, 2002.

(Nallapati, 2004) Nallapati, Ramesh. Discriminative Models for Information Retrieval. ACM Press, New York, NY, USA. Pages 64-71, 2004.

(NASA, 2004) NASA, National Aeronautics and Space Administration. Directory Interchange Format (DIF), Writers'S Guide, Version 9. May 2004. Disponível em: http://gcmd.gsfc.nasa.gov/User/difguide/difman.html. Acesso em 2005.

(NCSA , 2004) NCSA. Disponível em: http://www.ncsa.uiuc.edu/. Acesso em 2004.

(NISO, 2001) NISO. Standard Z39.85-2001 (Setember 2001). Disponível em: http://niso.org/standards/resources/Z39-85.pdf.

(Nocetti, 1980) Nocetti, Milton A..Disseminação Seletiva da Informação: Teoria e Prática. ABDF – Associação dos Bibliotecários do Distrito Federal, 62p. 1980.

(NDLTD , 2004) NDLTD. Disponível em: Networked digital Library of Theses and Dissertations. Acesso em 2004.

(OAI, 2004) OAI. The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Document Version 2004/1012T15:31:00Z. Disponível em: http://www.openarchives.org/OAI/openarchivesprotocol.html. Acesso em 2005.

(Ogle, 1996) Ogle, Virginia; Wilensky, Robert. Testbed Development for The Berkeley Digital Library Project. D-Lib Magazine. University of California, Berkeley, july/august 1996.

(OLC, 2004) OLC.Online Computer Library Center. Disponível em: http://www.oclc.org/. Acesso em 2004.

(Palazzo, 2002) Palazzo, Luiz Antônio Moro. Anais do XXII Congresso da Sociedade Brasileira de Computação.Convergências Tecnológicas Redesenhando as Fronteiras da Ciência e da Educação. Cap.7. pg. 286-325. SBC- 15 a 19 de julho de 2002.

(Park, 1998) Park ,Youn-Woo; Lee, Eon-Seok. A New Generation Method of An User Profile for Information Filtering on the Internet. 13 th International Conference on Information Networking (ICOIN'98), January, 21 -23 1998. Tokyo, Japan.

(Pistori, 2000) Pistori, Jeferson. Arquitetura de Implementação de Uma Biblioteca Digital Multimídia Distribuida. Dissertação de Mestrado, CPGCC/UFSC. 2000.

(Reza, 1994) Reza, Fazlollah M.. An Introduction to Information Theory. Dover Publications, Inc. New York,1994

(Rossetto, 2005) Rossetto, Marcia; Adriana Hypolito Nogueira. Aplicações de Elementos Metadados Dublin Core para Descrição de Dados Bibliográficos On-Line da Biblioteca Digital de Teses da USP. Universidade de São Paulo – USP, Sistema Integrado de Bibliotecas – Departamento Técnico. Disponivel em: http://www.usp.br/sibi. Acesso em 2005.

(Sdorra, 1999) Sdorra, P. Bollmann,;V. V. Raghavan,; H. Sever. Term Preference Weight. *Proceedings of the 14th International Symposium on Computer and Information Sciences. (ISCIS'99)* ,1999, pp. 360-369.

(SIDIE, 2001) SIDIE, Projeto CNPq, Sistema de Disponibilização de Informações para o Ensino.2001. Disponível em: http://www.sidie.nurcad.ufsc.br/. Acesso em 2005.

(Silva, 2003) Silva, Edilene Cristiana da. Adaptando padrões, atividades e serviços das Bibliotecas Tradicionais para as Bibliotecas Digitais. Dissertação de Mestrado - CPGCC- UFSC, 2003

(Sleeman, 1985) Sleeman, D. UMFE: A User Modelling Front-end Subsyustem. International Journal of Man-Machine Studies 23, 71-88. 1985.

(Smith, 1996) Smith, Terence R. The Meta-Information Environment of Digital Libraries. D-Lib Magazine, July/August 1996. Disponível em: http://www.dlib.org/metrics/public/papers/dig-lib-scope.html. Acesso 2005.

(Sumner, 1997) Sumner, Robert G.Jr.; Shaw, W.M. Jr..An Investigation of Relevance Feedback Using Adaptive Linear and Probabilistic Models. CITESEER, 1997.

(USP, 2004) USP, Biblioteca Digital de Teses e Dissertações. Disponível em: http://www.teses.usp.br/. Acesso em 2004.

(Vosgrau, 2005) Vosgrau, Sonia Regina Casselhas. Et.al. Formato MARC 21 Holdings para publicações seriadas. Disponível em: http://libdigi.unicamp.br/document/?down=1202. Acesso em 2005.

(Weitzel, 2003) Weitzel, Simone da Rocha; Sueli Mara S.P. Ferreira. Comunicação Científica e o Protocolo OAI: Uma Proposta na Área de Ciências da Comunicação. INTERCOM, Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 2003.

(Widyantoro, 1999) Widyantoro, Dwi H.; Thomas R. Loerger; John Yen. An Adaptive Algorithm for Learning Changes in User Interests. CIKM'99 - Kansas City, MO, USA. ACM- 1999.

(Wives, 2004) Wives, Leandro Krug. Utilizando Conceitos como Descritores de Texto para o Processo de Identificação de conglomerados (clustering) de Documentos. Disponível em: http://www.inf.ufrgs.br/~palazzo/alunos/defendidas.htm. Março de 2004.

(Wu, 2001) Wu, Hongjing; Erik de Kort, Paul de Bra. Design Issues for General Purpose Adaptive Hypermedia Systems. HT'01, AARHUS, Denmark. ACM - 2001