

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIAS DA COMPUTAÇÃO – BACHARELADO

FERRAMENTA DE INDEXAÇÃO E RECUPERAÇÃO DE
DOCUMENTOS ELETRÔNICOS UTILIZANDO ORACLE
INTERMEDIA

EMERSON IMMIAOVSKY

BLUMENAU
2007

2007/1-10

EMERSON IMMIAOVSKY

**FERRAMENTA DE INDEXAÇÃO E RECUPERAÇÃO DE
DOCUMENTOS ELETRÔNICOS UTILIZANDO ORACLE
INTERMEDIA**

Trabalho de Conclusão de Curso submetido à
Universidade Regional de Blumenau para a
obtenção dos créditos na disciplina de
Trabalho de Conclusão de Curso II do curso de
Ciências da Computação — Bacharelado.

Prof. Alexander Roberto Valdameri – Orientador

**BLUMENAU
2007**

2007/1-10

**FERRAMENTA DE INDEXAÇÃO E RECUPERAÇÃO DE
DOCUMENTOS ELETRÔNICOS UTILIZANDO ORACLE
INTERMEDIA**

Por

EMERSON IMMIAOVSKY

Trabalho aprovado para obtenção dos créditos
na disciplina de Trabalho de Conclusão de
Curso II, pela banca examinadora formada
por:

Presidente: _____
Prof. Alexander Roberto Valdameri, Mestre – Orientador, FURB

Membro: _____
Prof. Paulo Roberto Dias, Mestre – FURB

Membro: _____
Prof. Wilson Pedro Carli, Mestre – FURB

Blumenau, 29 de Junho de 2007

Dedico este trabalho a todos os amigos,
especialmente aqueles que me ajudaram
diretamente na realização deste.

AGRADECIMENTOS

À Deus, pelo seu imenso amor e graça.

À minha família, que mesmo longe, sempre esteve presente.

Aos meus amigos, pelos empurrões e cobranças.

À minha esposa Marcia e filha Nicole que em tantas horas dedicadas à realização deste trabalho se mostraram compreensivas pela minha ausência.

Ao meu orientador, Alexander Roberto Valdameri, por ter acreditado na conclusão deste trabalho.

Aceite com sabedoria o fato de que o caminho está cheio de contradições. Há momentos de alegria e desespero, confiança e falta de fé, mas vale a pena seguir adiante.

Paulo Coelho

RESUMO

A evolução tecnológica está direcionada para uma era onde qualquer informação pode estar digitalmente armazenada. Livros, manuais, contratos, imagens, vídeos, áudios e tantos outros tipos de mídias podem ser encontrados nos sistemas de bibliotecas digitais. Mas, para que todo este conteúdo digital esteja disponível de forma ágil e eficaz, é necessário que se tenha meios de localizar e recuperar as informações relevantes ao contexto da pesquisa. Esse trabalho apresenta uma solução para esse problema, através do desenvolvimento de um sistema de pesquisa por conteúdo dos documentos textos. O sistema tem como principal função, auxiliar os usuários no processo de pesquisa e obtenção de documentos eletrônicos através do conteúdo contidos em bibliotecas digitais e na Internet.

Palavras-chave: Biblioteca digital. Oracle interMedia. Recuperação de informação. Oracle Text. Documentos eletrônicos. Indexação por conteúdo.

ABSTRACT

The technological evolution is directed for an age where any information can be digitally stored. Books, manuals, contracts, images, videos, voice and as much other types of medias can be found in the systems of digital libraries. But, so that all this digital content is available of agile and efficient form, it is necessary that if it has ways to locate and to recoup the excellent information to the context of the research. This work presents a solution for this problem, through the development of a system of research for content of documents texts. The system have as main function, to assist the users in the research process and electronic document attainment through the content contained in digital libraries and the Internet.

Key-words: Digital library. interMedia Oracle. Recovery of information. Oracle Text. Electronic documents. Indexation for content.

LISTA DE ILUSTRAÇÕES

Figura 1 – Telas de pesquisa básica e avançada ao acervo eletrônico da UNICAMP	15
Figura 2 – Telas de pesquisa básica e avançada da biblioteca digital da USP	15
Figura 3 – Tela de busca ao acervo digital da FURB	16
Figura 4 – Estrutura de uma aplicação utilizando os recursos do Oracle Text	18
Figura 5 – Interface de busca da biblioteca digital da UNICAMP	19
Figura 6 - Interface de busca da biblioteca digital da FURB	20
Figura 7 – Funcionamento geral do sistema SIRDE	23
Figura 8 – Funcionamento geral do sistema durante a pesquisa	24
Figura 9 – Diagrama de casos de uso do SIRDE.....	26
Figura 10 – Diagrama de classes do SIRDE	27
Figura 11 – Diagrama de atividades	28
Figura 12 – Organização do banco de dados	30
Figura 13 – Modelo de dados	32
Figura 14 – Estrutura das aplicações publicadas no servidor <i>Tomcat</i>	33
Quadro 1 – Trecho de código de pesquisa básica.....	34
Quadro 2 – Trecho de código Java de criação da conexão com o banco de dados	35
Quadro 3 – Trecho de código com a consulta à documentos utilizando o InterMedia	35
Quadro 4 – Trecho de código de apresentação dos resultados	36
Quadro 5 – Localização dos documentos nos repositórios locais	37
Quadro 6 – Trecho de código da carga dos documentos Web	38
Quadro 7 – Trecho de código da indexação dos arquivos pelo InterMedia	38
Quadro 8 – Sincronização do índice já criado pelo InterMedia	39
Quadro 9 – Trecho de código para otimização do índice InterMedia	39
Quadro 10 – Exemplo de índice fragmentado e após a otimização rápida.....	40
Quadro 11 – Exemplo de otimização completa do índice InterMedia	40
Quadro 12 – Exemplo de otimização do índice utilizando o método recriar	40
Figura 15 – Tela principal do SIRDE.....	41
Figura 16 – Tela de <i>login</i> do administrador	41
Figura 17 – Tela de parâmetros gerais do SIRDE.....	42
Figura 18 – Lista de repositórios	43
Figura 19 – Tela para manter repositórios locais	43

Figura 20 – Tela para manter repositórios Web	44
Figura 21 – Lista de restrições.....	45
Figura 22 – Tela de indexação e otimização	46
Figura 23 – Tela de consulta e limpeza de <i>logs</i>	47
Figura 24 – Tela para pesquisa básica	47
Quadro 13 – Operadores lógicos do InterMedia	48
Figura 25 – Tela para pesquisa avançada	48
Figura 26 – Tela de resultados da busca e pesquisa nos resultados	49

LISTA DE SIGLAS

ANSI – *American National Standards Institute*

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

EA – *Enterprise Architect*

FURB – Universidade Regional de Blumenau

IHC – Interfaces Humano-Computador

MACI – Mecanismo Automático de Carga e Indexação

PDF – *Portable Document Format*

RTF – *Rich Text Format*

SBU – Sistema de Biblioteca da Unicamp

SIRDE – Sistema de Indexação e Recuperação de Documentos Eletrônicos

SQL – *Structured Query Language*

SQL/MM – SQL Multimídia

SRI – Sistema de Recuperação de Informação

UML – *Unified Modeling Language*

UNICAMP – Universidade de Campinas

USP – Universidade de São Paulo

XML – *eXtensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 OBJETIVOS DO TRABALHO	13
1.2 ESTRUTURA DO TRABALHO	13
2 FUNDAMENTAÇÃO TEÓRICA	14
2.1 BIBLIOTECAS DIGITAIS	14
2.2 ORACLE INTERMEDIA	16
2.3 INTERFACE PARA BIBLIOTECA DIGITAL	18
2.4 TRABALHOS CORRELATOS	20
3 DESENVOLVIMENTO DO SISTEMA.....	22
3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO.....	22
3.2 VISÃO GERAL.....	23
3.3 ESPECIFICAÇÃO	25
3.3.1 Diagrama de casos de uso	25
3.3.2 Diagrama de classes	26
3.3.3 Diagrama de atividades	28
3.4 MODELAGEM DE DADOS	29
3.5 IMPLEMENTAÇÃO	33
3.5.1 Técnicas e ferramentas utilizadas.....	33
3.5.2 Operacionalidade da implementação	40
3.6 RESULTADOS E DISCUSSÃO	49
4 CONCLUSÕES.....	51
4.1 EXTENSÕES	52
REFERÊNCIAS BIBLIOGRÁFICAS	53

1 INTRODUÇÃO

A busca pelo conhecimento é uma necessidade constante em qualquer fase da vida, começando já na infância onde as pessoas são instigadas à leitura. Durante todo o período de aprendizado, é muito importante que se tenha uma base para se orientar e buscar o conhecimento necessário para a realização dos estudos e da formação profissional.

Devido ao impulso científico e tecnológico, o homem presenciou o surgimento de uma sociedade baseada no conhecimento. O conhecimento já era usado a certo tempo nas organizações, mas só recentemente passou a ter o seu reconhecimento como recurso essencial (SANTOS; PASSOS, 2000, p. 26).

Historicamente, as bibliotecas tradicionais eram as únicas grandes bases de informação, nas quais o estudante podia pesquisar e aperfeiçoar seu aprendizado. Porém, com o avanço tecnológico, houve uma mudança na forma de armazenar e disponibilizar as coleções de obras de informação. Surgiram então as chamadas bibliotecas digitais.

Atualmente as bibliotecas possuem sistemas de catalogação e buscas de obras para facilitar a localização dos livros de interesse do acadêmico. Porém, se a procura não for por um livro, autor ou palavra-chave específica, a localização torna-se mais demorada e, em muitos casos, não se encontra o material adequado ou, se encontrado, pode não ser o que se espera.

O grande volume de informações disponíveis na sociedade e nas organizações baseadas em conhecimento, aliado à falta de associação entre conteúdos oferecidos e demanda, problemas de forma de apresentação e restrições de busca e de acesso existentes nos Sistemas de Recuperação de Informação (SRI) dificultam o processo de filtrar e obter a informação necessária dentro dos crescentes espaços virtuais de informação disponibilizados para os usuários. (MARCHIORI, 1997, p. 58).

Devido a essa dificuldade na localização do material de interesse, muitos estudantes deixam de utilizar os benefícios disponibilizados pelas bibliotecas. O que se percebe cada vez mais é a preferência por pesquisas na Internet. Este fato se dá pela facilidade e precisão nas respostas da busca oferecida por alguns *sites*, citando como exemplo, o Google. (HOHAGEM, 2005).

Procurando trazer uma maior eficiência às buscas feitas na biblioteca da Universidade Regional de Blumenau (FURB), este trabalho apresenta os métodos para a construção de uma ferramenta de indexação e recuperação de conteúdo digital, usando o módulo InterMedia do banco de dados Oracle, que é capaz de indexar arquivos multimídias. Assim, as buscas aos documentos digitais da referida biblioteca poderão ser feitas também pelo seu conteúdo,

proporcionando resultados melhores e mais confiáveis num menor espaço de tempo.

1.1 OBJETIVOS DO TRABALHO

O objetivo deste trabalho é a construção de uma ferramenta para indexação e recuperação, por conteúdo, de documentos eletrônicos utilizando o módulo InterMedia do banco de dados Oracle.

Os objetivos específicos do trabalho são:

- a) localizar os documentos digitais armazenados no servidor da biblioteca da FURB e de outras instituições, indexando-os pelo seu conteúdo;
- b) indexar os documentos eletrônicos do acervo digital da biblioteca da FURB e de outras instituições disponíveis publicamente na Internet através de seu endereço eletrônico, ampliando a oferta de conteúdo;
- c) oferecer mecanismos alternativos de busca e manipulação dos resultados, como *download* e pesquisa nos resultados;
- d) disponibilizar uma interface de administração do sistema permitindo a manutenção das configurações básicas e a inclusão de repositórios a serem indexados pela ferramenta;
- e) permitir que o administrador do sistema possa consultar as mensagens geradas pelos processamentos e erros ocorridos durante a indexação.

1.2 ESTRUTURA DO TRABALHO

Esse trabalho descreve no capítulo 1 a introdução ao conteúdo desse trabalho apresentando os objetivos, no capítulo 2 são apresentadas as áreas relacionadas com a fundamentação teórica, sendo elas: bibliotecas digitais, Oracle InterMedia e interface para bibliotecas digitais, no capítulo 3 é apresentado o processo de desenvolvimento da ferramenta, incluindo os requisitos, algoritmos, técnicas, diagramas e trechos de código e por fim, no capítulo 4 são apresentadas conclusões, análise dos resultados obtidos e sugestões para extensão do presente trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados aspectos teóricos relacionados ao trabalho, tais como: bibliotecas digitais, Oracle InterMedia, interface para bibliotecas digitais e sistema atual da biblioteca da FURB. Por fim, são apresentados alguns trabalhos correlatos.

2.1 BIBLIOTECAS DIGITAIS

Segundo (LEINER, 1998) bibliotecas digitais são coleções organizadas de documentos, onde cada fonte de informação possui dois atributos relacionados: os relativos ao seu conteúdo e os que identificam de forma descritiva o documento.

O conceito representa um processo gradual e evolutivo como resultante da utilização do computador nas últimas décadas (CUNHA, 1999). Elas surgem como uma resposta ao fenômeno da explosão da quantidade de informação, disponíveis na rede Internet, sendo este um dos maiores fenômenos em termos de processamento da informação no século XX.

As bibliotecas digitais estão em crescente fase de pesquisas e já começam a aparecer alguns projetos como o portal da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) que disponibiliza vários periódicos de todas as áreas do saber em meio digital. A possibilidade do surgimento de projetos como este está intimamente ligada com a evolução tecnológica nas redes de computadores e nos novos formatos digitais que surgiram (IUNES, 2004).

Uma biblioteca digital é constituída por documentos primários que são digitalizados na forma material (disquetes, CD-ROM, DVD), ou na forma eletrônica disponibilizada via sistemas ou Internet (BIBLIOTECAS, 2007). Permite o acesso aos documentos eletrônicos remotamente através de um computador ligado a uma rede bem como a utilização simultânea por vários usuários.

O Sistema de Bibliotecas da UNICAMP (SBU), através da web, possibilita o acesso à informação gerada pela Universidade e pela comunidade científica do país e exterior. O acesso e a consulta ao material catalogado é livre e irrestrito ao público em geral. É subdividida em monografias e periódicos, periódicos eletrônicos, biblioteca digital, base de dados Perie, Edubase e Hemered (UNICAMP, 2004). A pesquisa por documentos é feita

através de tópicos ou informando uma expressão de busca normal ou avançada, e os resultados podem ser ordenados por relevância, título ou data. A figura 1 mostra as telas do sistema.

<p>Procurar: todos os tópicos</p> <p>Para encontrar documentos contendo uma ou mais palavras, preencha o seguinte campo:</p> <p>Procurar por: <input type="text"/> <input type="button" value="Procurar"/> Procura avançada</p> <p>Para acessar um documento dado o seu código, basta indicá-lo abaixo</p> <p>Código: <input type="text"/> <input type="button" value="Acessar"/></p>	<p>Procurar: todos os tópicos</p> <p>Para encontrar documentos contendo uma ou mais palavras, preencha os seguintes campos:</p> <p>Seleção: <input type="text" value="todas as palavras"/> Ordenação: <input type="text" value="relevância"/> Resultados por página: <input type="text" value="10"/></p> <p>Procurar por: <input type="text"/> <input type="button" value="Procurar"/> Procura normal</p> <p>Para acessar um documento dado o seu código, basta indicá-lo abaixo</p> <p>Código: <input type="text"/> <input type="button" value="Acessar"/></p>
--	--

Fonte: adaptado de UNICAMP (2004).

Figura 1 – Telas de pesquisa básica e avançada ao acervo eletrônico da UNICAMP

Na Universidade de São Paulo (USP), através do portal do conhecimento estão disponíveis digitalmente as teses e dissertações defendidas na USP para consulta e download (USP, 2005). Conforme apresentado na figura 2, a pesquisa no acervo pode ser feita informando uma expressão nas formas básica ou avançada, sendo que na avançada a busca é realizada por título, autor ou resumo.

<p>Pesquisa Simples em Acervo</p> <p>Pesquisar por: <input type="text"/></p> <p>Para fazer uma pesquisa no acervo, digite a palavra no formulário e pressione a imagem </p> <p></p> <p>Copyright © 2001-2005 by Universidade de São Paulo. Todos os direitos reservados.</p>	<p>Pesquisa Avançada em Acervo</p> <p>Unidades: <input type="text" value="Todas"/></p> <p>Ano de defesa: <input type="text" value="Todos"/></p> <p>Área: <input type="text" value="Todas"/></p> <p>Título: <input type="text"/></p> <p>Autor: <input type="text"/></p> <p>Resumo: <input type="text"/></p> <p>Documento: <input type="text" value="Teses e Dissertações"/></p> <p></p> <p></p> <p>Copyright © 2001-2005 by Universidade de São Paulo. Todos os direitos reservados.</p>
---	--

Fonte: adaptado de USP (2005).

Figura 2 – Telas de pesquisa básica e avançada da biblioteca digital da USP

Na biblioteca digital da FURB estão disponibilizadas as dissertações, teses, monografias, trabalhos de conclusão de curso, artigos, fotografias, ilustrações, obras de arte, revistas, vídeos e outros documentos de interesse ao desenvolvimento científico, tecnológico e artístico (FURB, 2005). Utiliza as mesmas ferramentas do Sistema Integrado de Bibliotecas

para o registro, busca e a recuperação da informação. Uma vez localizado o documento desejado, o texto completo pode ser acessado a partir do resultado apresentado de cada registro e a visualização é mostrada no formato *Portable Document Format* (PDF) (FURB, 2005).

O sistema atual da biblioteca digital da FURB possui indexação somente por palavras chaves e por um resumo que são cadastrados no momento da catalogação. Conforme apresentado na figura 3, a pesquisa ao acervo é realizada através de uma expressão de busca por autor, assunto ou título.

A interface de busca rápida apresenta um cabeçalho azul com o texto "BUSCA RÁPIDA". Abaixo dele, há um campo de texto rotulado "Expressão de Busca:". Seguem as opções de pesquisa: "Pesquisa:" com radio buttons para "Livre" (selecionado), "Autor", "Assunto" e "Título". Abaixo, "Listar por:" com radio buttons para "Título" (selecionado) e "Referência Bibliográfica". Na base, há dois botões: "Procurar" e "Limpar".

Fonte: (FURB, 2005).

Figura 3 – Tela de busca ao acervo digital da FURB

2.2 ORACLE INTERMEDIA

O módulo Oracle InterMedia é um gerenciador de conteúdo multimídia disponível desde a versão 9i do Oracle Database. Sendo uma parte integrante do Oracle Database Server, o módulo possui todos os seus benefícios tais como: velocidade, escalabilidade, segurança e poder (PELSKI, 2005b).

O InterMedia é capaz de operar com qualquer tipo de mídia, sendo projetado com especial interesse para aplicações que rodam em cima da plataforma Internet. A adição de campos com conteúdo multimídia é transparente e oferece suporte para os mais populares servidores web e ferramentas de autoria.

O acesso ao InterMedia pode ser feito com a *Structured Query Language* (SQL) padrão *American National Standards Institute* (ANSI), SQL orientado a objeto ou através de

SQL Multimedia (SQL/MM). Aplicações escritas em Java ou em C++ já possuem bibliotecas de acesso otimizadas facilitando assim a integração desse novo módulo a projetos já existentes. As informações extraídas no metadados pode ser passada à aplicação pelo formato *eXtensible Markup Language* (XML) (PELSKI, 2005a).

O Oracle InterMedia permite a pesquisa pelo conteúdo de documentos textos em vários formatos, tais como PDF e *Rich Text Format* (RTF). Assim, uma pesquisa pode ter melhores resultados com um número menor de palavras-chave num menor período de tempo. O resultado da pesquisa pode ser classificado por ordem de relevância, ou seja, pela quantidade de vezes que a palavra ou expressão pesquisada aparece no conteúdo. Desta forma, o usuário pode, por exemplo, pesquisar por uma função em particular da linguagem Java, ao invés de ter que listar todos os livros de Java disponíveis (PELSKI, 2005b).

Independente do local onde estão armazenados os documentos eletrônicos, podendo ser em uma tabela de banco de dados ou em um servidor de arquivos separados do servidor do banco de dados, o InterMedia pode indexar os documentos sem dificuldade alguma desde que seja disponibilizado acesso via *link* de Internet ou caminho de rede. Isso aumenta em muito as possibilidades de integração com outras bibliotecas sem acarretar num aumento no espaço consumido para armazenar a mídia (PELSKI, 2005b).

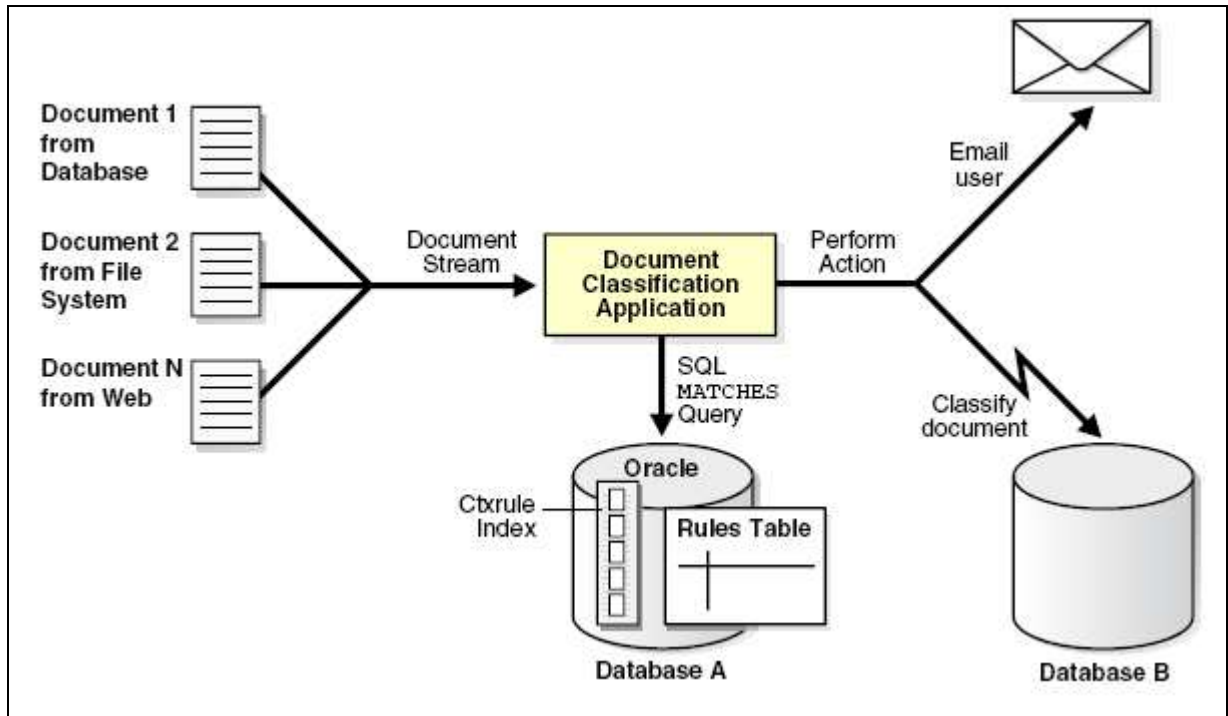
A instalação do módulo InterMedia vem junto ao pacote de módulos do banco de dados, podendo ser instalado junto com o banco ou em momentos isolados. Todo o módulo compreende um conjunto de funções e objetos para processar diferentes tipos de dados multimídia dentre eles, estão os áudios, vídeos, imagens, e documentos binários.

Para documentos textos sendo eles binários ou não, o InterMedia possui um submódulo chamado Oracle Text que possibilita transformar um banco de dados Oracle numa poderosa ferramenta de busca, permitindo que sejam realizadas pesquisas por textos em colunas de tipos normais como *varchar* e *clob*, ou do tipo binário como *blob*, bem como referencias à documentos externos ao banco de dados (ORACLE, 2005a).

Para utilização do Oracle Text, é necessário ter uma tabela contendo uma coluna onde será armazenado o próprio documento ou uma referencia ao documento externo podendo ser um caminho de rede ou de Internet. Uma vez tendo o documento armazenado ou referenciado na coluna da tabela, utilizam-se as funções do Oracle Text criar parâmetros e indexar todo o conteúdo do arquivo. E ao criar o índice com todas as palavras encontradas, o InterMedia cria tabelas de controles internos para otimizar a localização do registro sem armazenar um volume muito grande de dados (ORACLE, 2005b).

A figura 4 mostra como pode ser estruturada uma aplicação utilizando os recursos do

InterMedia especificamente voltadas ao Oracle Text.



Fonte: (ORACLE, 2005b).

Figura 4 – Estrutura de uma aplicação utilizando os recursos do Oracle Text

2.3 INTERFACE PARA BIBLIOTECA DIGITAL

A Internet abriga uma quantidade gigantesca de informação em ciência e tecnologia de natureza variada. Apesar das incontáveis ferramentas de recuperação de informação existentes na rede, como os *search engines*, catálogos e guias, essas informações não apresentam uma organização que permita aos pesquisadores consultá-las com facilidade, obtendo, em tempo hábil, resultados capazes de atender às suas necessidades (OLIVEIRA; CARDOSO, 2003).

As bibliotecas digitais apresentam o mesmo problema citado anteriormente. Oferecem basicamente duas interfaces para o usuário: uma simples e outra avançada. Na interface simples é oferecido um campo para se digitar alguma palavra-chave que se relacione com o item a ser pesquisado. Caso ele possua um domínio sobre as opções oferecidas, ele pode partir para a interface de busca avançada que possui várias opções a fim de filtrar e obter melhores resultados. Porém, o usuário não é incentivado a usar a interface avançada e não há nenhuma ligação entre elas (CARDOSO, 2000, p. 14).

A interface de uma biblioteca digital deve possibilitar ao usuário opções de busca

necessárias e adequadas ao contexto da pesquisa de maneira a oferecer incentivo ao descobrimento de novas formas de se realizar a consulta, recuperação e visualização do documento.

Para construção de interfaces para bibliotecas digitais, deve-se avaliar o grau de satisfação dos utilizadores referentes à usabilidade e critérios ergonômicos de Interfaces Humano-Computador (IHC), tanto para usuários com pouca experiência na utilização da Internet, como para os considerados experientes. Cada utilizador tem uma visão, uma idéia diferente da interface analisada (RAABE et al., 2002).

Para que o usuário possa ser incentivado a pesquisar em seu acervo, a biblioteca deve oferecer uma interface simples, porém poderosa, que esteja bem documentada para oferecer auxílio aos que não estejam bem familiarizados com o seu funcionamento. Assim, o usuário poderá encontrar o que deseja em um período de tempo menor e mais facilmente.

A seguir são apresentadas duas figuras referentes as interfaces de pesquisa de duas grandes universidades. A figura 5 mostra a tela de busca da biblioteca digital da UNICAMP e a figura 6 mostra a interface da FURB.

Fonte: (UNICAMP, 2004).

Figura 5 – Interface de busca da biblioteca digital da UNICAMP

Fonte: (FURB, 2005).

Figura 6 - Interface de busca da biblioteca digital da FURB

2.4 TRABALHOS CORRELATOS

Esta sessão apresenta trabalhos correlatos que descrevem soluções de indexação e recuperação de conteúdos de documentos eletrônicos. Porém, não utilizam a tecnologia do Oracle InterMedia como recurso principal, mas somente como um componente agregado aos projetos.

ITGis Softwares e Equipamentos (1999), desenvolveu um sistema chamado *iSMART Geospatial Suite*, que permite aos usuários finais publicar e pesquisar dados espaciais para construção de mapas em ambiente internet/intranet sem dificuldades e com muita rapidez. Para que esta facilidade e agilidade pudessem atingir níveis aceitáveis e satisfatórios, foi utilizado o recurso InterMedia da Oracle para o armazenamento e recuperação das informações do sistema.

SOUZA, AMARO e CASTRO (2005), descrevem um sistema de gerenciamento geográfico para monitoramento ambiental dos estuários e áreas adjacentes com o objetivo de controlar as alterações na paisagem natural ocasionadas por processos erosivos de origem natural bem como de origem social decorrentes dos interesses econômicos e políticos. Para o tratamento e processamento dos dados do sistema, foi utilizado o Oracle InterMedia para

processar, armazenar e consultar informações em arquivos de imagens, áudio, vídeo e de outros formatos, oferecendo, por exemplo, um ambiente confiável para dados multimídia.

Martins (2005) comenta sobre soluções aplicadas utilizando produtos Oracle e cita um projeto realizado em Portugal chamado Viseu Digital. Trata-se de um portal de acesso a temas como notícias, informação ao cidadão e às empresas, promoções turísticas, educação e cultura e ainda serviços do governo eletrônico. Este projeto emprega muitas tecnologias inovadoras dentre elas o Oracle InterMedia para o armazenamento e recuperação das informações.

O Google Inc. (2006), lançou recentemente um novo recurso em pesquisas na internet denominado Google Acadêmico que permite aos usuários pesquisarem por várias disciplinas e fontes em um só lugar: artigos revisados por especialistas, teses, livros, resumos e artigos de editoras acadêmicas, organizações profissionais, bibliotecas de pré-publicações, universidades e outras entidades acadêmicas. O Google Acadêmico classifica os resultados de pesquisa segundo a relevância. Como na pesquisa da web, as referências mais úteis são exibidas no começo da página. A tecnologia de classificação do Google leva em conta o texto integral de cada artigo, o autor, a publicação em que o artigo saiu e a frequência com que foi citado em outras publicações acadêmicas (GOOGLE, 2006?).

3 DESENVOLVIMENTO DO SISTEMA

Neste capítulo está descrito todo o processo de desenvolvimento do Sistema de Indexação e Recuperação de Documentos Eletrônicos (SIRDE) e divide-se em cinco seções. A primeira delas relaciona os requisitos funcionais e não funcionais. A segunda seção mostra uma visão geral do funcionamento do sistema. A terceira seção apresenta a especificação através de diagramas que o representam logicamente. A quarta seção, aborda alguns aspectos da implementação. A quinta seção apresenta os resultados e discussões do trabalho.

3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Os requisitos funcionais do sistema são:

- a) o sistema deve ser capaz de efetuar buscas de palavras ou expressões pelo conteúdo de documentos eletrônico no acervo da biblioteca digital da FURB e de outras instituições;
- b) permitir cadastrar *links* de documentos eletrônicos armazenados em bibliotecas digitais de outras instituições;
- c) apresentar uma lista dos resultados da busca ao usuário por ordem de relevância;
- d) possibilitar que os usuários possam pesquisar nos resultados de uma busca anterior;
- e) oferecer mecanismo de visualização do documento eletrônico através do acesso direto ao documento ou a tela de visualização dos dados do documento quando existir;
- f) permitir que o sistema possa ser administrado para inclusão de repositórios locais ou remotos;
- g) possibilidade do administrador do sistema poder executar a indexação total ou rápida e de utilizar recursos de otimização dos índices criados.

Os requisitos não funcionais do sistema são:

- a) ser desenvolvido em Java com a ferramenta NetBeans 5.5;
- b) utilizar servidor de aplicação para Internet Apache Tomcat 4.5 ou superior;
- c) armazenar os índices de pesquisa em banco de dados Oracle 10G;

- d) utilizar o módulo Oracle InterMedia versão 10G instalado no banco de dados Oracle 10G;
- e) construir os diagramas da UML do sistema na ferramenta Enterprise Architect (EA);
- f) ter uma área reservada para administração do sistema com uma única conta.

3.2 VISÃO GERAL

A figura 7 mostra o funcionamento geral do sistema SIRDE e na seqüência está descrito cada uma das ações.

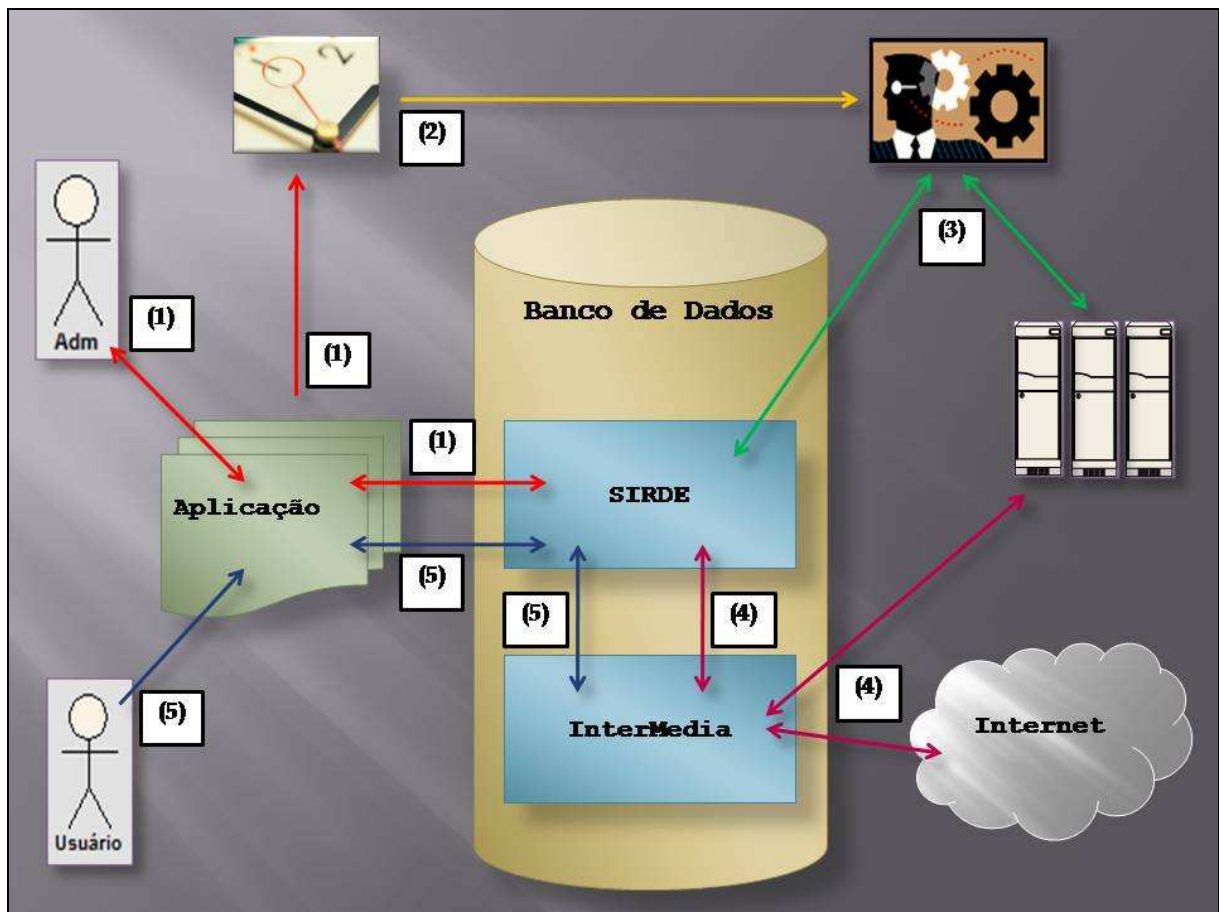


Figura 7 – Funcionamento geral do sistema SIRDE

A descrição da seqüência das ações acima apresentadas segue da seguinte forma:

- a) ação (1): o administrador utiliza a aplicação para cadastrar repositórios, restrições e parâmetros, verificar *log*, agendar o dia e hora do agente de busca;
- b) ação (2): o mecanismo de agendamento de tarefas aciona o agente de busca e

- indexação;
- c) ação (3): o agente realiza conexão com o banco de dados do sistema e executa consulta dos repositórios cadastrados para identificar quais servidores e endereços físicos deverá iniciar a busca por arquivos. E para cada um, acessa o servidor de arquivos na estrutura mencionada e recupera o nome dos arquivos encontrados cadastrando-os em uma tabela no banco de dados. Após finalização da busca por arquivos, o agente aciona o mecanismo do InterMedia para indexação dos documentos através do endereço de internet de cada um;
 - d) ação (4): o InterMedia recupera cada arquivo da tabela de documentos a indexar e através do endereço de internet de cada um, ele acessa via internet o documento e cria um índice com todas as palavras de todos os arquivos de forma otimizada;
 - e) ação (5): o usuário utiliza a aplicação para efetuar a busca por uma expressão, onde é acionado funções do InterMedia para utilização do índice previamente criado e recuperar todos os arquivos contendo a expressão solicitada pelo usuário.

A figura 8 representa o funcionamento geral do sistema quando o usuário efetua uma pesquisa por um determinado assunto.

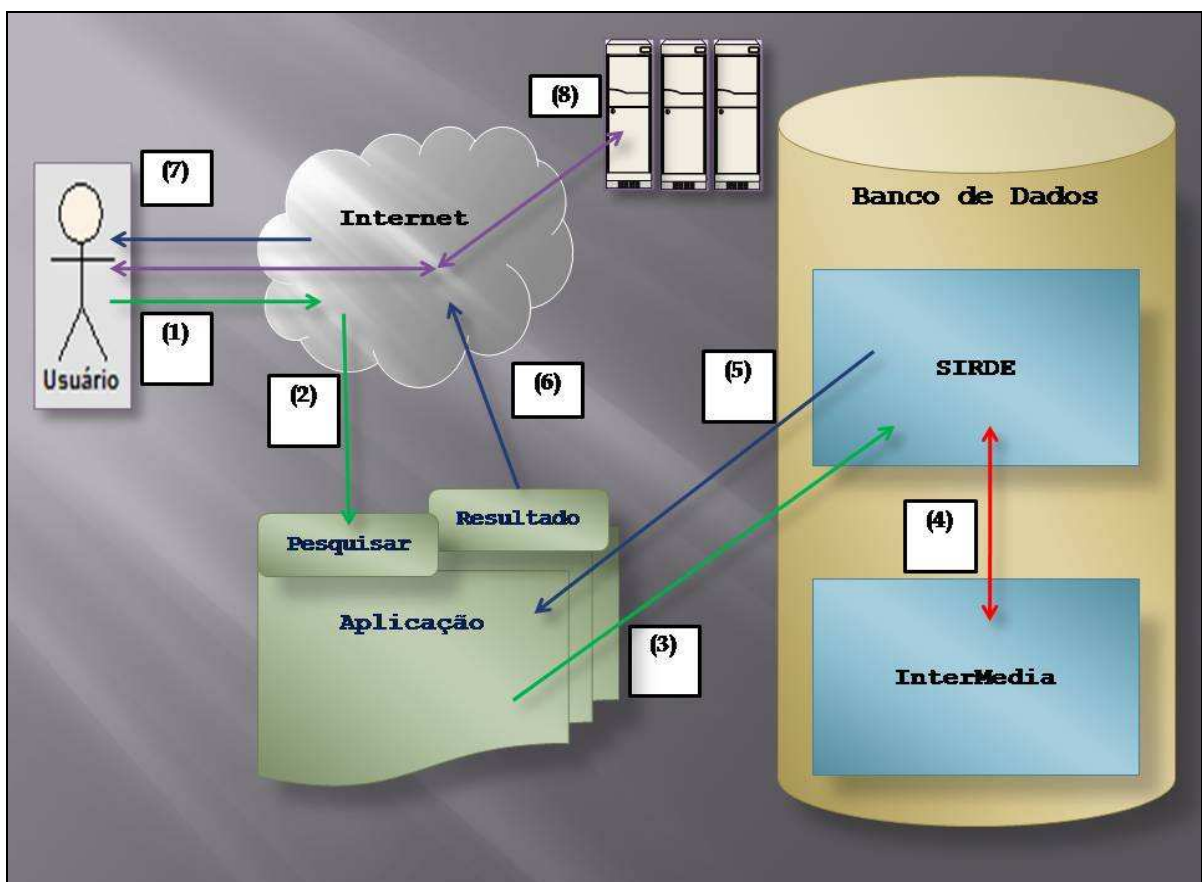


Figura 8 – Funcionamento geral do sistema durante a pesquisa

A descrição da seqüência das ações acima apresentadas segue da seguinte forma:

- a) ação (1): o usuário acessa o SIRDE através da internet;
- b) ação (2): o usuário informa a expressão de busca na tela de pesquisa básica ou avançada;
- c) ação (3): a aplicação acessa o banco de dados e executa as funções de consulta passando os parâmetros informados pelo usuário;
- d) ação (4): o banco de dados SIRDE utilizam funções de manipulação do índice para localizar os registros de documentos conforme expressão do usuário, gravando suas referencias em uma tabela de resultados;
- e) ação (5): a aplicação recebe retorno do fim da busca do banco de dados através de um identificador de resultados;
- f) ação (6): a aplicação monta página de resultados;
- g) ação (7): a aplicação envia resposta para o navegador do usuário que é direcionado para a página de resultados;
- h) ação (7): o sistema carrega o documento eletrônico para visualização;
- i) ação (8): o usuário localiza e clica no *link* do documento encontrado e a aplicação abre a página Web da instituição inicia o download do documento.

3.3 ESPECIFICAÇÃO

A especificação do sistema apresenta-se através dos diagramas da UML, utilizando para tal os diagramas de casos de uso, classes e atividades, e por último o modelo físico de dados.

3.3.1 Diagrama de casos de uso

“Um caso de uso é uma descrição de um conjunto de seqüências de ações, inclusive variantes, que um sistema executa para produzir um resultado de valor observável por um ator” (BOOCH; RUMBAUGH; JACOBSON, 2000, p. 220). Sendo assim, essa modelagem é usada para descrever o que o sistema deve fazer, ou seja, os diagramas de casos de uso descrevem os requisitos funcionais do sistema. A figura 9 mostra o diagrama de casos de uso

do SIRDE.

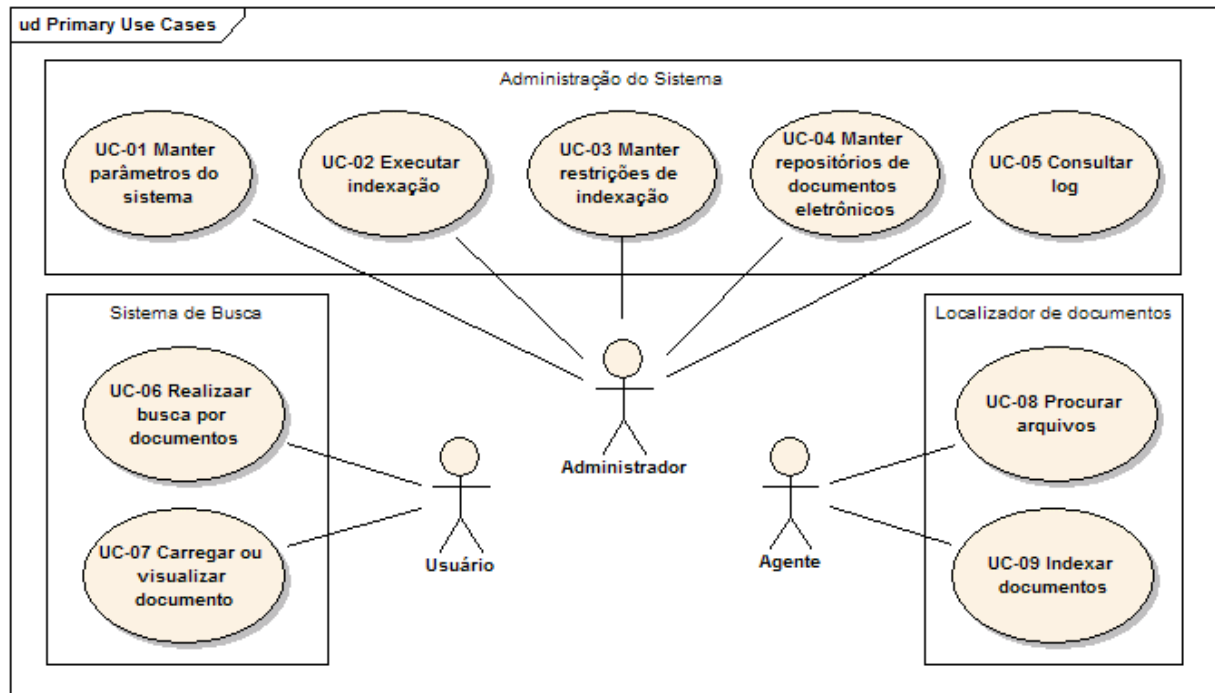


Figura 9 – Diagrama de casos de uso do SIRDE

O diagrama de casos de uso acima apresentado pode ser descrito da seguinte maneira:

- o administrador utiliza o sistema para manter parâmetros, restrições e repositórios, executar indexação e otimizar os índices, e consultar os *logs* da aplicação.
- o agente localiza os documentos eletrônicos no servidor e grava suas referências no banco de dados;
- o agente aciona as funções do InterMedia para iniciar a indexação dos documentos previamente localizados;
- o usuário informa uma expressão para pesquisa e o sistema iniciará o processo de pesquisa no banco de dados já indexado e após o processo de pesquisa é apresentado os resultados obtidos ao usuário;
- o usuário seleciona um documento para visualização ou *download*.

3.3.2 Diagrama de classes

“Um diagrama de classe é um diagrama que mostra um conjunto de classes, interfaces e colaborações e seus relacionamentos” (BOOCH; RUMBAUGH; JACOBSON, 2000, p. 105). Dessa forma, os diagramas de classes apresentam as classes que compõem um sistema bem como os relacionamentos entre elas.

A Figura 10 apresenta o diagrama de classes do SIRDE.

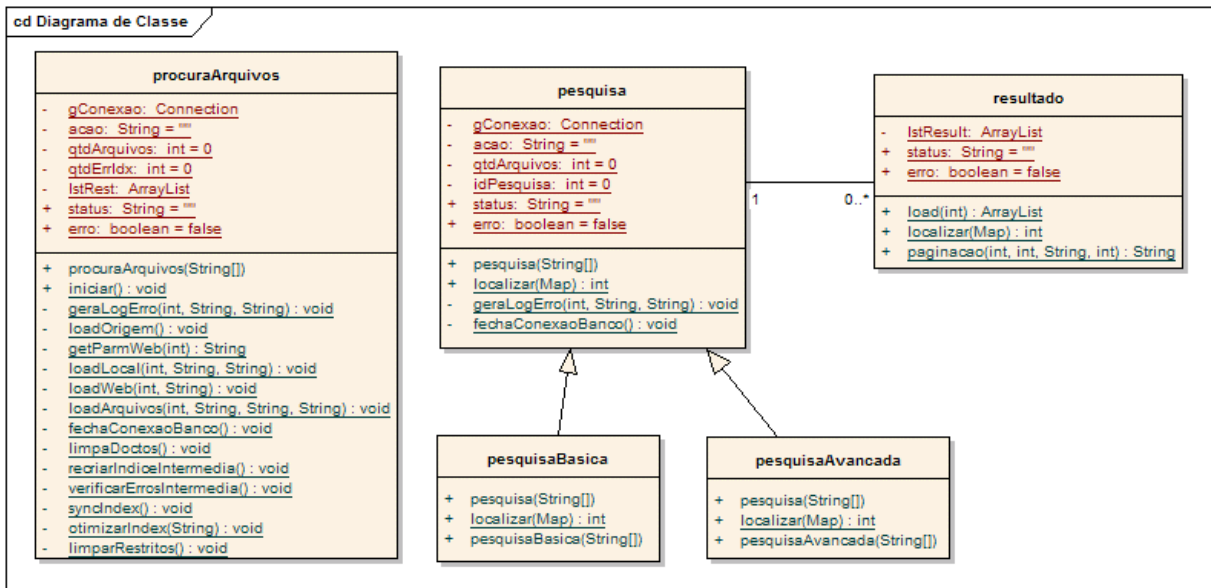


Figura 10 – Diagrama de classes do SIRDE

A descrição do diagrama de classes do sistema é apresentada da seguinte maneira:

- a classe *procucaArquivos* contém os métodos necessários para recuperar informações do banco de dados e percorrer os diretórios e subdiretórios dos servidores de arquivo, assim como a indexação dos documentos encontrados. É responsável também pela indexação dos documentos de outras instituições cadastrados no sistema;
- a classe *pesquisa* é uma super classe contendo atributos e métodos que são comuns às classes de pesquisa básica ou avançada. É utilizada no momento em que o usuário efetua uma pesquisa por documentos eletrônicos no sistema;
- a classe *pesquisaBásica* contém os métodos necessários para realização de pesquisa da expressão de busca informada pelos usuários;
- a classe *pesquisaAvançada* contém os métodos e atributos para realizar pesquisa da expressão de busca de acordo com as características avançadas informadas pelos usuários;
- a classe *resultados* contém os métodos e atributos para apresentação das informações referente aos documentos eletrônicos encontrados bem como controles para realizar nova busca em cima dos resultados.

3.3.3 Diagrama de atividades

O diagrama de atividades descreve a seqüência das atividades de um sistema, elas apresentam comportamento condicional e por meio dessas condições apresentarão mudanças de estado (BOOCH; RUMBAUGH; JACOBSON, 2000, p. 257).

A Figura 11 apresenta o diagrama de atividades do SIRDE, permitindo a visualização sequencial da execução do sistema.

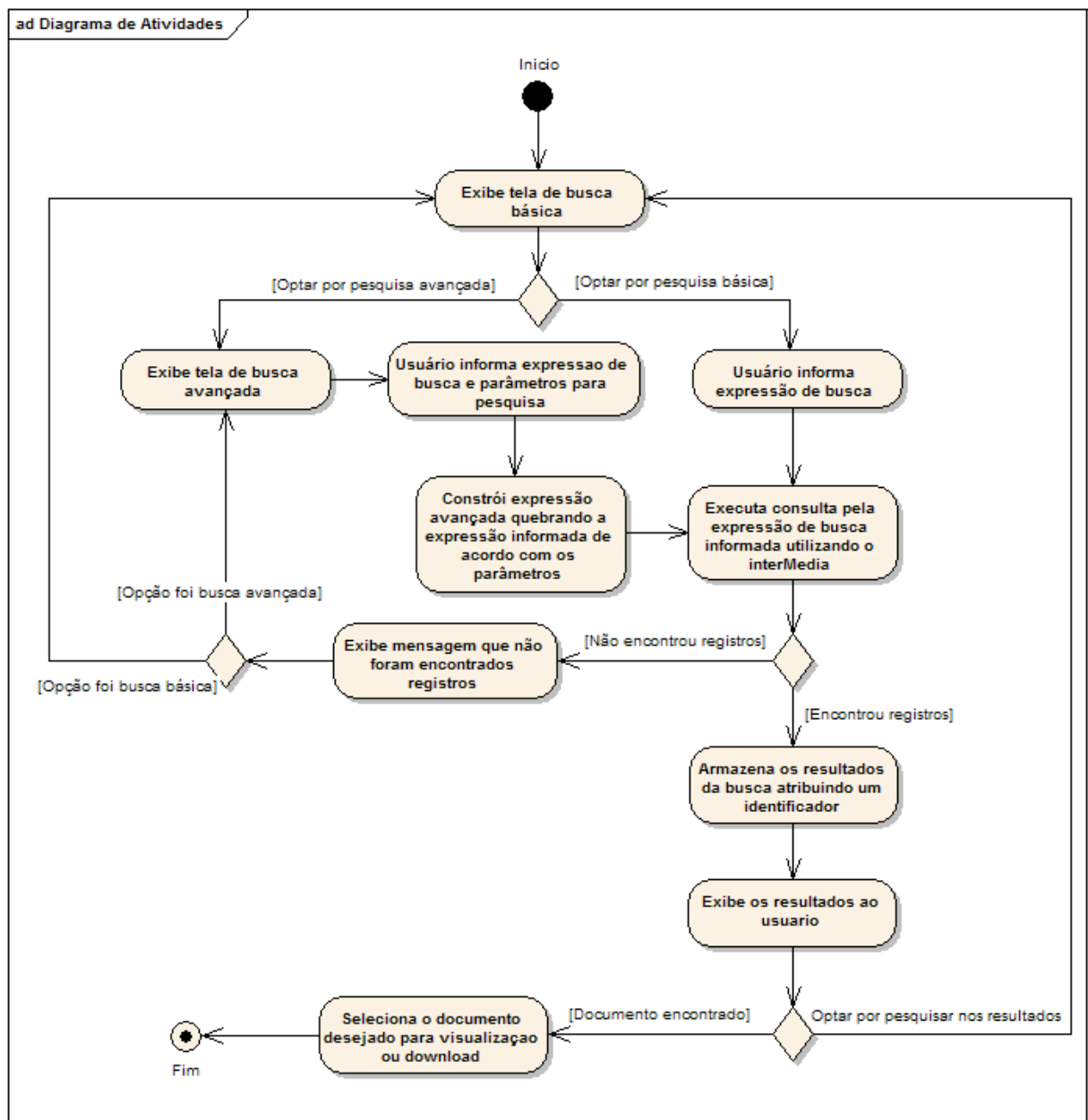


Figura 11 – Diagrama de atividades

A descrição do diagrama de atividades apresenta-se da seguinte maneira:

- o sistema apresenta a tela de busca básica;

- b) se o usuário optar por pesquisa básica, informa a expressão de busca, e o sistema executa a consulta no banco de dados utilizando o InterMedia;
- c) se o usuário optar por pesquisa avançada, o sistema exibe a tela com parâmetros avançados para pesquisa;
- d) o usuário informa a expressão de busca e os parâmetros avançados;
- e) o sistema constrói a expressão de busca avançada quebrando a expressão informada de acordo com os parâmetros e executa a consulta no banco de dados utilizando o InterMedia;
- f) se o sistema não encontrou registros, exibe mensagem ao usuário;
- g) se a tentativa de consulta realizada foi básica, o sistema exibe a tela de busca básica novamente;
- h) se a tentativa de consulta realizada foi avançada, o sistema exibe a tela de busca avançada novamente;
- i) se o sistema encontrou registros, armazena-os em uma tabela do banco identificando-os caso o usuário precise pesquisar nos resultados;
- j) o sistema exibe a tela com os resultados da pesquisa;
- k) se o usuário optar por pesquisar nos resultados, o sistema exibe a tela de busca básica passando o identificador dos resultados para que a pesquisa seja realizada somente nos registros encontrados;
- l) se o usuário optar por visualizar ou carregar um documento encontrado, o sistema irá direcionar para outra janela do navegador acessando o seu endereço de internet.

3.4 MODELAGEM DE DADOS

O armazenamento das informações é feito no banco de dados Oracle, uma vez que é necessário o recurso InterMedia que está diretamente ligado ao esquema do sistema.

A Figura 12 apresenta como está organizado o banco de dados com os esquemas do SIRDE e do InterMedia.

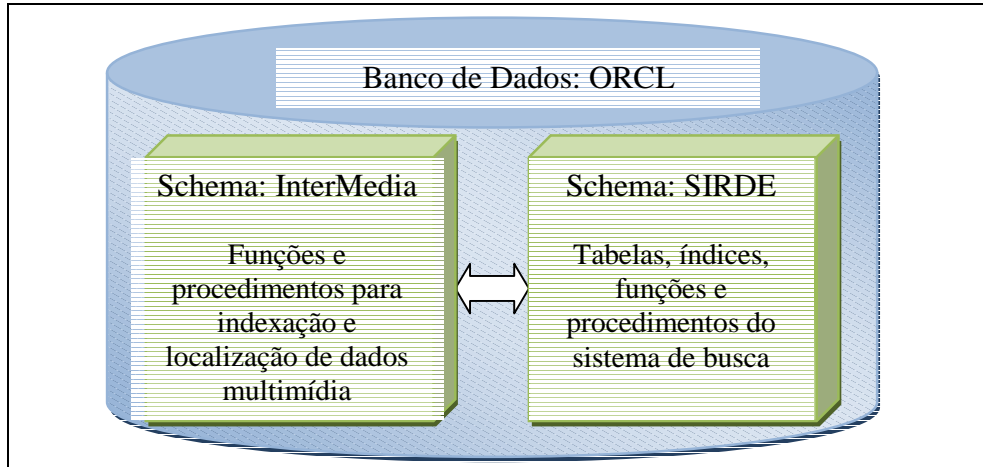


Figura 12 – Organização do banco de dados

Na seqüência apresenta-se a descrição do modelo físico de dados do sistema e em seguida a representação gráfica pela figura 13, construída com os recursos da ferramenta EA.

O modelo de dados está organizado da seguinte maneira:

- a) na tabela *PARM_SIRDE* é armazenado um único registro contendo informações de relevância geral do sistema. Esta tabela não possui relacionamento com outras, pois trata-se de uma tabela de configurações gerais. Os atributos desta tabela são: *SIG_SISTEM*, *NOM_SISTEM*, *DIA_SEMAN_AGENTE*, *HOR_EXEC_AGENTE* e *PSW_ADMIN*;
- b) na tabela *LOG_SIRDE* são gerados vários registros referentes ao coportamento do sistema bem como mensagens de erros ocorridos durante os processamentos diversos sendo da utilização dos usuários ou do agente de busca e indexação. Também é armazenado as mensagens de alerta e erros gerados pelo InterMedia. Esta tabela não possui relacionamento pois os registros de *log* não estão ligados a qualquer outro registro e sim ao comportamento do sistema. Os atributos desta tabela são: *NUM_SEQ_LOG*, *COD_ERRO*, *DSC_ERRO*, *DSC_LOCAL_ERRO* e *DAT_OCORR*;
- c) na tabela *RESTRICAO* são mantidos informações das restrições que o sistema deverá considerar para restringir determinados arquivos, como por exemplo extensões específicas, partes de nomes ou ainda por faixa de tamanho. Não possui relacionamentos, pois uma restrição deve ser considerada para qualquer repositório. Os atributos desta tabela são: *COD_RSTRCAO*, *DSC_RSTRCAO*, *IND_TIPO_RSTRCAO*, *REG_RSTRCAO*, *MIN_KBYTES* e *MAX_KBYTES*;
- d) na tabela *ORIGEM* são mantidas informações de origens de documentos que serão utilizados pelo agente para localizar documentos nos servidores locais ou na

internet. Os atributos desta tabela são: *NUM_SEQ_ORIGEM*, *NOM_DOMINIO_WEB*, *IND_ORIGEM*, *NOM_HOST*, *URL_VIEW_WEB* e *DSC_DETALHES*;

- e) na tabela *PARM_LINK_VIEW_WEB* são mantidas informações sobre os parâmetros necessários para enviar no endereço de internet no momento em que o usuário selecionar a opção de visualização do documento. Esta tabela possui relacionamento com a tabela *ORIGEM*, pois os parâmetros são específicos para cada origem de documentos. Os atributos desta tabela são: *NUM_SEQ_ORIGEM*, *NOM_PARM*, *VAL_FIXO_PARM* e *IND_PARM_VARVEL*;
- f) na tabela *REPOSITORIO_WEB* são mantidas informações sobre documentos de outras instituições em que o agente não consegue recuperar. Esta tabela possui relacionamento com a tabela *ORIGEM*, pois uma origem pode conter vários documentos sendo que cada documento é específico de uma origem. Os atributos desta tabela são: *NUM_SEQ_ORIGEM*, *COD_DOCTO_WEB*, *NOM_DOCTO_WEB*, *URL_DOCTO_WEB* e *DSC_RESUMO_DOCTO_WEB*;
- g) na tabela *REPOSITORIO_LOCAL* são mantidas informações sobre o local em que estão armazenados os documentos eletrônicos em servidores onde o agente poderá localizá-los através de uma rede local. Esta tabela possui relacionamento com a tabela *ORIGEM* pois uma origem poderá ter vários diretórios de documentos. Os atributos desta tabela são: *NUM_SEQ_ORIGEM*, *END_ESTRU_RAIZ*, *END_WEB_RAIZ* e *FLG_PESQ_SUBDIRTRO*;
- h) na tabela *DOCTO_INDEX* são armazenados os registros localizados nos repositórios de arquivos eletrônicos locais pelo agente e os documentos de outras instituições que foram cadastrados no repositório Web. Esta tabela possui relacionamento com a tabela *ORIGEM* para manter a ligação com cada origem do documento encontrado e indexado. Optou-se pela desnormalização repetindo algumas colunas de tabelas relacionadas para ganho de desempenho no momento da pesquisa pelo usuário. Os atributos desta tabela são: *NUM_SEQ_ORIGEM*, *COD_DOCTO*, *NOM_DOCTO*, *URL_IDX_DOCTO*, *URL_VIEW_DOCTO* e *DSC_RESUMO_DOCTO*;
- i) na tabela *RESULT_PESQ* são armazenadas todos os documentos encontrados quando o usuário efetua um busca. Cada registro é gravado com um identificador da pesquisa para que o sistema mostre ao usuário somente os registros referentes à sua pesquisa e também para o caso da necessidade de refinar a pesquisa anterior.

Esta tabela possui relacionamento com a tabela de documentos indexados para evitar redundância de dados. Os atributos desta tabela são: *NUM_SEQ_PESQ*, *NUM_SEQ_ORIGEM*, *COD_DOCTO* e *NUM_RELVNCIA*.

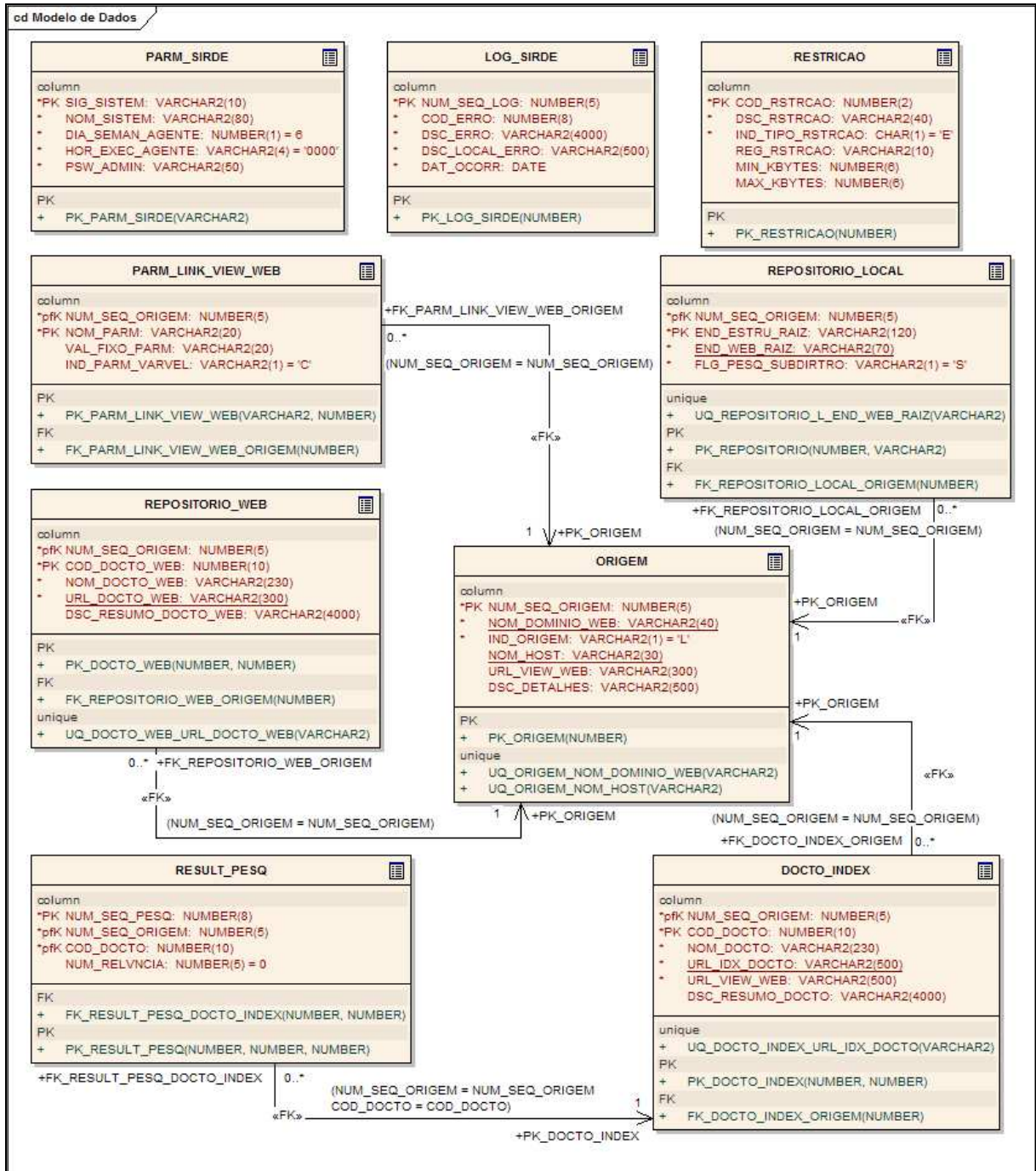


Figura 13 – Modelo de dados

3.5 IMPLEMENTAÇÃO

A implementação do sistema está dividida em duas seções. Na primeira seção, apresentam-se algumas considerações sobre as técnicas e ferramentas utilizadas para o seu desenvolvimento. Na segunda seção, apresenta-se a operacionalidade ou o funcionamento da implementação.

3.5.1 Técnicas e ferramentas utilizadas

Para o desenvolvimento deste trabalho foram utilizadas técnicas e ferramentas como a utilização do servidor de aplicação java Web TomCat 5.5 para executar o SIRDE, banco de dados Oracle 10G para armazenamento de dados, os recursos do InterMedia instalado e configurado no banco de dados e a ferramenta de desenvolvimento IDE NetBeans 5.5.

Para a publicação do sistema, é importante que se tenha um pacote de distribuição que normalmente possui a extensão *war* gerado pela ferramenta de desenvolvimento e em seguida utilizar a própria aplicação do *Tomcat* para efetuar a publicação do sistema desenvolvido. A figura 14 mostra o servidor configurado e as aplicações liberadas com destaque à aplicação SIRDE.



Figura 14 – Estrutura das aplicações publicadas no servidor *Tomcat*

As informações do sistema e a geração dos índices estão armazenadas em um banco de dados Oracle 10g por ser um banco confiável, de alto desempenho e especialmente porque possui o recurso InterMedia que possibilitou a construção do SIRDE. Como o modelo de dados já apresentado na figura 14, foi totalmente feito na ferramenta EA, utilizou-se uma função da própria ferramenta para a geração do *script* de criação das tabelas. A execução

deste script foi feita na interface SQL do Oracle para o banco de dados.

Para disponibilizar os recursos do InterMedia no banco, deve-se selecionar a opção na instalação por instalar também os pacotes do InterMedia. Não há necessidade de qualquer configuração do InterMedia neste momento.

O desenvolvimento do sistema na IDE NetBeans 5.5. trouxe algumas facilidades pois já possui um padrão para projetos Web. Através de páginas dinâmicas escritas em *JavaServer Pages* (JSP), foram construídas interações com os usuários. O quadro 1 mostra um trecho do código no momento em que o usuário submete uma pesquisa.

```

1  <%@ page import="sirdeJava.pesquisa" %>
2  <%@ page import="java.util.*" %>
3  <%@ page import="java.sql.*" %>
4  <%
5      String erros = "";
6      String acao = (request.getParameter("acao") == null ? "" : request.getParameter("acao"));
7      String expr = (request.getParameter("expr") == null ? "" : request.getParameter("expr"));
8      Map parametros = request.getParameterMap();
9
10     if("basica".equalsIgnoreCase(acao)){
11         String[] str = banco.getConexao();
12         str[5] = acao;
13
14         pesquisa psq = new pesquisa(str);
15         if (!psq.erro){
16             int idPsq = psq.localizar(parametros);
17             if (!psq.erro){
18                 response.sendRedirect("./resultados.jsp?pID="+idPsq+"&expr="+expr);
19             }
20         }
21         erros = psq.status;
22     }
23 %>
24 <html>
25 <head>
26     ....
27 </head>
28 <body onLoad="setFocus()">
29     ....
30 </body>
31 </html>

```

Quadro 1 – Trecho de código de pesquisa básica

Todo o processo de acesso ao banco de dados é realizado no momento em que a classe do objeto em questão é instanciada através de seu construtor. O quadro 2 apresenta o trecho do código em que é criada a conexão com o banco de dados.

```

1 package sirdeJava;
2
3 import ...
4
5
6
7
8
9 public class pesquisa {
10     public pesquisa(String args[]) {
11         try{
12             // Inicializacao dos atributos
13             ...
14             // Criacao da conexao com o banco de dados
15             Locale l = new Locale("en", "US");
16             Locale.setDefault(l);
17             Class.forName("oracle.jdbc.driver.OracleDriver");
18             gConexao = DriverManager.getConnection("jdbc:oracle:thin:@" + pHostname + ":"
19                                                     + pPortServer + ":"
20                                                     + pBanco
21                                                     ,pUsername
22                                                     ,pPassword);
23             status = "Conexão realizada!";
24         } catch (Exception e) {
25             e.printStackTrace();
26             status += "Erro não esperado:<br>" + e.toString();
27             erro = true;
28         }
29     }
30     ...
31 }

```

Quadro 2 – Trecho de código Java de criação da conexão com o banco de dados

A utilização do recurso InterMedia ocorre em alguns pontos do sistema e um deles é no momento em que o usuário efetua a pesquisa. O sistema possui uma tabela de resultados da pesquisa onde são gravados todos os registros correspondentes a pesquisa do usuário. O quadro 3 apresenta o trecho de código com a função de localização dos registros onde é acrescentado ao comando *select* a função *contains* do InterMedia para somente carregar os documentos cujo índice possui a expressão solicitada.

```

118 private static void pesquisaBasica(String pExpr){
119     try {
120         String sql = "insert into result_pesq (num_seq_pesq, num_seq_origem, cod_docto, num_relvncia) ";
121         sql += "select ?,num_seq_origem, cod_docto, score(1) ";
122         sql += "from docto_index ";
123         sql += "where contains(url_idx_docto, ?, 1) > 0";
124         sql += "order by score(1) desc";
125
126         PreparedStatement ps = gConexao.prepareStatement(sql);
127         ps.setInt(1, idPesquisa);
128         ps.setString(2, pExpr);
129
130         ps.execute();
131         qtdArquivos = ps.getUpdateCount();
132
133         ps.close();
134         gConexao.commit();
135     } catch (Exception e) {
136         e.printStackTrace();
137         geraLogErro(8002,e.toString(),"Ao efetuar busca básica!");
138         if((e.toString().indexOf("DRG-10599"))>0){
139             status = "O sistema está sendo reindexando. Por favor tente mais tarde!";
140         }else{
141             status = "Ocorreu um erro não esperado!<br>" + e.toString();
142         }
143         erro = true;
144     }
145 }

```

Quadro 3 – Trecho de código com a consulta à documentos utilizando o InterMedia

Após o término da função de pesquisar, o sistema terá a tabela de resultados povoada com os registros encontrados e é realizado um redirecionamento para a pagina de resultados.

Nesta tela, o sistema efetua uma consulta aos registros cujo identificador é igual ao retorno da função de pesquisa. O quadro 4 mostra o trecho de código que carrega a pagina web com os resultados.

```

1  <%@ page import="sirdeJava.pesquisa" %>
2  <%@ page import="sirdeJava.resultado" %>
3  <%
4  Map parametros = request.getParameterMap();
5
6  idPsq = psq.localizar(parametros);
7
8  ArrayList lstResult = null;
9  if(idPsq > 0){
10     lstResult = resultado.load(idPsq);
11 }
12 %>
13 <html>
14 <head>
15     ...
16 </head>
17 <body marginwidth="0">
18     <%
19     if(lstResult.size() > 0){ %>
20         <table border="0" width="100%">
21             <%
22                 HashMap has = null;
23                 for(int i=(regDe-1); i <= (regAte-1); i++){
24                     has = (HashMap)lstResult.get(i);
25                     String codDoc = (String)has.get("cod_docto");
26                     String nomDoc = (String)has.get("nom_docto");
27                     String urlDoc = (String)has.get("url_view_web");
28                     String dscDoc = (String)has.get("dsc_resumo_docto");
29                     String prcRev = (String)has.get("num_relvncia");
30                     %>
31                     <tr>
32                         <td>
33                             <p> <a href="<%=out.print(urlDoc);%>" target="new">
34                                 
35                             </a>
36                             <a href="<%=out.print(urlDoc);%>" target="new">
37                                 <%=out.print(nomDoc);%>
38                             </a>
39                             <br>
40                             <small>... <%=out.print(dscDoc);%>&nbsp;&nbsp;...<br>
41                                 <font class="emphasis"><%=out.print(codDoc);%></font> &raquo;
42                                 <font class="emphasis">27-01-2005 11:19</font> &raquo;
43                                 <font class="emphasis"><%=out.print(prcRev);%></font>
44                             </small>
45                         </td>
46                     </tr>
47                 <%
48                 }
49             %>
50         </table>
51     <%=out.print(prcRev);%>
52     ...
53 </body>
54 </html>

```

Quadro 4 – Trecho de código de apresentação dos resultados

Para a localização dos documentos eletrônicos, foram criados dois tipos de carga, sendo uma para localização de documentos nos repositórios locais através de acesso via rede

local e outra para referencias a documentos em outras instituições.

O processo de carga de documentos locais é feito através de leitura dos diretórios e para cada arquivo encontrado, é gravado na tabela de documentos o nome e sua referência por um endereço Web. O quadro 5 mostra o trecho de código da função de carga da tabela de documentos locais à indexar.

```

489 private static void loadArquivos(int pNumSeqOrigem, String pDirBase, String pUrlBase, String pPesqSubdir){
490
491     try{
492
493         File wrkDirOrigem = new File(pDirBase);
494         File wrkArquivosDir[] = wrkDirOrigem.listFiles();
495
496         for (int x = 0; x < wrkArquivosDir.length; x++) {
497
498             String wrkArq = wrkArquivosDir[x].getCanonicalPath().substring(pDirBase.length());
499             wrkArq = wrkArq.substring(1,wrkArq.length());
500
501             if (wrkArquivosDir[x].isDirectory() && "S".equalsIgnoreCase(pPesqSubdir)) {
502
503                 String novoDirBase = pDirBase + "\\\" + wrkArq;
504                 String novaUrlBase = pUrlBase + "/" + wrkArq;
505                 loadArquivos(pNumSeqOrigem, novoDirBase, novaUrlBase, pPesqSubdir);
506
507             } else {
508
509                 qtdArquivos ++;
510             }
511         }
512     } catch (Exception e) {
513         e.printStackTrace();
514         geraLogErro(e.hashCode(), e.toString(), "Ao gravar arquivo na tabela de indexação");
515         erro = true;
516     }
517 }

```

Quadro 5 – Localização dos documentos nos repositórios locais

Para os documentos de outras instituições, foi criado um procedimento mais simples, pois quando é cadastrado um documento no repositório Web, já são informados todos os dados necessários para acessá-lo. Por esta razão, a carga é feita simplesmente pela consulta aos documentos e a inclusão do mesmo na tabela de indexação. O quadro 6 mostra o trecho de código com os procedimentos para esta atividade.

```

246 private static void loadWeb(int pNumSeqOrigem, String pUrlView){
247
248     try {
249         String sql = "insert into docto_index(num_seq_origem,cod_docto,nom_docto,url_idx_docto" +
250             ",dsc_resumo_docto,url_view_web) ";
251         sql += "select ?,COD_DOCTO_WEB,NOM_DOCTO_WEB,URL_DOCTO_WEB,DSC_RESUMO_DOCTO_WEB,";
252         sql += "replace(replace('"+pUrlView+"','[C]','COD_DOCTO_WEB'),'[N]','NOM_DOCTO_WEB) ";
253         sql += "from repositorio_web ";
254         sql += "where num_seq_origem = ? ";
255
256         if("fast".equalsIgnoreCase(acao)){
257             sql += "and not exists(select 1 from docto_index where url_idx_docto = url_docto_web) ";
258         }
259
260         PreparedStatement ps = gConexao.prepareStatement(sql);
261
262         ps.setInt(1, pNumSeqOrigem);
263         ps.setInt(2, pNumSeqOrigem);
264
265         ps.execute();
266         qtdArquivos += ps.getUpdateCount();
267
268         ps.close();
269     } catch (Exception e) {
270         e.printStackTrace();
271         geralLogErro(e.hashCode(), e.toString(), "Ao carregar arquivos Web!");
272         erro = true;
273     }
274 }

```

Quadro 6 – Trecho de código da carga dos documentos Web

Uma vez a tabela de documentos à indexar carregada com os documentos locais e de outras instituições, o sistema chama o procedimento que aciona o InterMedia que carrega cada documento referenciado na coluna indexada e realiza a indexação de todas as palavras do documento. O quadro 7 apresenta o trecho de código com os comandos para indexação.

```

369 private static void recriarIndiceIntermedia() {
370
371     String wrkSql;
372     try {
373         wrkSql = "drop index I_DOCTO_INDEX_01";
374         Statement st = gConexao.createStatement();
375         st.execute(wrkSql);
376         st.close();
377     } catch (SQLException e) {
378         e.printStackTrace();
379         geralLogErro(7777,e.toString(), "Ao excluir indice intermedia");
380         erro = true;
381     }
382     try {
383         CallableStatement proc = gConexao.prepareCall("{call ctx_ddl.drop_preference('COMMON_DIR')}");
384         proc.execute();
385         proc = gConexao.prepareCall("{call ctx_ddl.create_preference('COMMON_DIR','URL_DATASTORE')}");
386         proc.execute();
387         wrkSql = "create index I_DOCTO_INDEX_01";
388         wrkSql = wrkSql + " on";
389         wrkSql = wrkSql + " DOCTO_INDEX (URL_IDX_DOCTO)";
390         wrkSql = wrkSql + " indextype is ctxsys.context";
391         wrkSql = wrkSql + " parameters ('datastore COMMON_DIR filter ctxsys.inso_filter')";
392         Statement st = gConexao.prepareStatement(wrkSql);
393         st.execute(wrkSql);
394         st.close();
395     } catch (SQLException e) {
396         e.printStackTrace();
397         geralLogErro(7777,e.toString(), "Ao recriar indice InterMedia");
398         erro = true;
399     }
400 }

```

Quadro 7 – Trecho de código da indexação dos arquivos pelo InterMedia

Todo esse procedimento de carregar os arquivos dos repositórios locais e remotos é

realizado no momento em que é executado o comando de indexação completa no sistema, sendo iniciado pelo administrador ou pelo mecanismo de agendamento de tarefas do servidor.

Porém, para que não sejam utilizados muito processamento de máquina e também pelo tempo total de demora em que o processo de indexação completa gasta, foi implementado um mecanismo de indexação rápida, onde somente é adicionado novos arquivos sem que seja excluído o índice todo e recriado. Para isto, o InterMedia disponibiliza um recurso de sincronização de índice, onde somente os novos registros é que serão indexados. Conforme o quadro 8, observa-se o código implementado com a sincronização do índice.

```

426 private static void syncIndex(){
427     try {
428         CallableStatement proc = gConexao.prepareCall("{call ctx_ddl.sync_index('I_DOCTO_INDEX_01', 'LOM')}");
429         proc.execute();
430     } catch (SQLException e) {
431         e.printStackTrace();
432         geralLogErro(7779,e.toString(), "Ao sincronizar indice InterMedia");
433         erro = true;
434     }
435 }

```

Quadro 8 – Sincronização do índice já criado pelo InterMedia

Pelo fato de existir a funcionalidade de sincronização do índice, em alguns momentos se faz necessário que os índices sejam otimizados para se ter uma resposta mais rápida da expressão à ser pesquisada. A necessidade da otimização é devido ao índice ficar muito fragmentado e o InterMedia possui alguns comandos que possibilita esta otimização. No quadro 8 está o trecho de código que implementa a otimização dos índices.

```

437 private static void otimizarIndex(String metodo){
438     try {
439         CallableStatement proc;
440         if("fast".equalsIgnoreCase(metodo)){
441             proc = gConexao.prepareCall("{call ctx_ddl.optimize_index('I_DOCTO_INDEX_01', 'FAST')}");
442         }else if("full".equalsIgnoreCase(metodo)){
443             proc = gConexao.prepareCall("{call ctx_ddl.optimize_index('I_DOCTO_INDEX_01', 'FULL')}");
444         }else{
445             proc = gConexao.prepareCall("{call ctx_ddl.optimize_index('I_DOCTO_INDEX_01', 'REBUILD')}");
446         }
447         proc.execute();
448     } catch (SQLException e) {
449         e.printStackTrace();
450         geralLogErro(7778,e.toString(), "Ao otimizar indice InterMedia pelo método "+metodo);
451         erro = true;
452     }
453 }

```

Quadro 9 – Trecho de código para otimização do índice InterMedia

Conforme observado no quadro acima, foram implementados três tipos de otimização sendo eles, *FAST*, *FULL* E *REBUILD*.

No método *FAST*, o InterMedia simplesmente reorganiza o índice já criado por ordem

de palavras sem eliminar índices repetidos. A tabela 1 representa um exemplo de índice fragmentado pela sincronização e após o comando de otimização rápida.

Índice Fragmentado	Índice otimizado
palavra1 doc1 doc2 doc3	palavra1 doc1 doc2 doc3
palavra2 doc6	palavra2 doc6
palavra3 doc1 doc4	palavra2 doc4 doc5
palavra2 doc4 doc5	palavra3 doc1 doc4
palavra3 doc2 doc5	palavra3 doc2 doc5

Quadro 10 – Exemplo de índice fragmentado e após a otimização rápida

No método *FULL*, o InterMedia reorganiza agrupando os índices. As palavras repetidas são eliminadas. A tabela 2 mostra um exemplo de como é realizado a otimização completa.

Índice Fragmentado	Índice otimizado
palavra1 doc1 doc2 doc3	palavra1 doc1 doc2 doc3
palavra2 doc6	palavra2 doc6 doc4 doc5
palavra3 doc1 doc4	palavra3 doc1 doc4 doc2 doc5
palavra2 doc4 doc5	
palavra3 doc2 doc5	

Quadro 11 – Exemplo de otimização completa do índice InterMedia

Por último, o método *REBUILD* recria todo o índice ordenando tanto as palavras fragmentadas como também a referencia ao registro em que a palavra se encontra. A tabela 3 mostra um exemplo do índice fragmentado e após a reconstrução.

Índice Fragmentado	Índice otimizado
palavra1 doc1 doc2 doc3	palavra1 doc1 doc2 doc3
palavra2 doc6	palavra2 doc4 doc5 doc6
palavra3 doc1 doc4	palavra3 doc1 doc2 doc4 doc5
palavra2 doc4 doc5	
palavra3 doc2 doc5	

Quadro 12 – Exemplo de otimização do índice utilizando o método recriar

3.5.2 Operacionalidade da implementação

O sistema possui duas áreas distintas, sendo uma restrita com acesso através de senha para o administrador, e outra aberta para qualquer usuário efetuar as pesquisas por documentos. A figura 15 mostra a tela inicial de acesso aberto do sistema observando que no canto superior direito possui um *link* de nome “administrar” que dá acesso à área restrita.

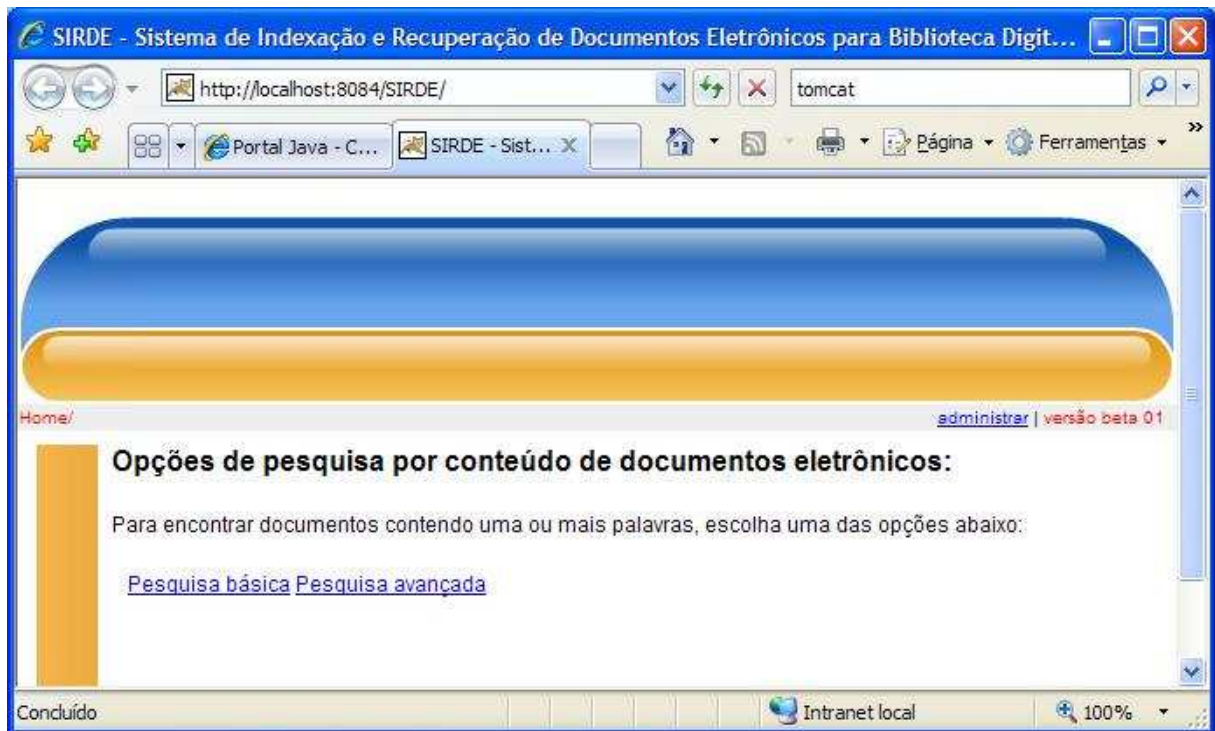


Figura 15 – Tela principal do SIRDE

Inicialmente o administrador do sistema deverá configurar e cadastrar os repositórios e referências de documentos de outras instituições para que os usuários possam efetuar as pesquisas desejadas.

Conforme figura 16, ao clicar no *link* “administrar”, o sistema solicita a senha do administrador que por padrão é “*admin*” e após validar a senha é apresentada a primeira tela da área restrita onde são informados os parâmetros gerais do SIRDE.

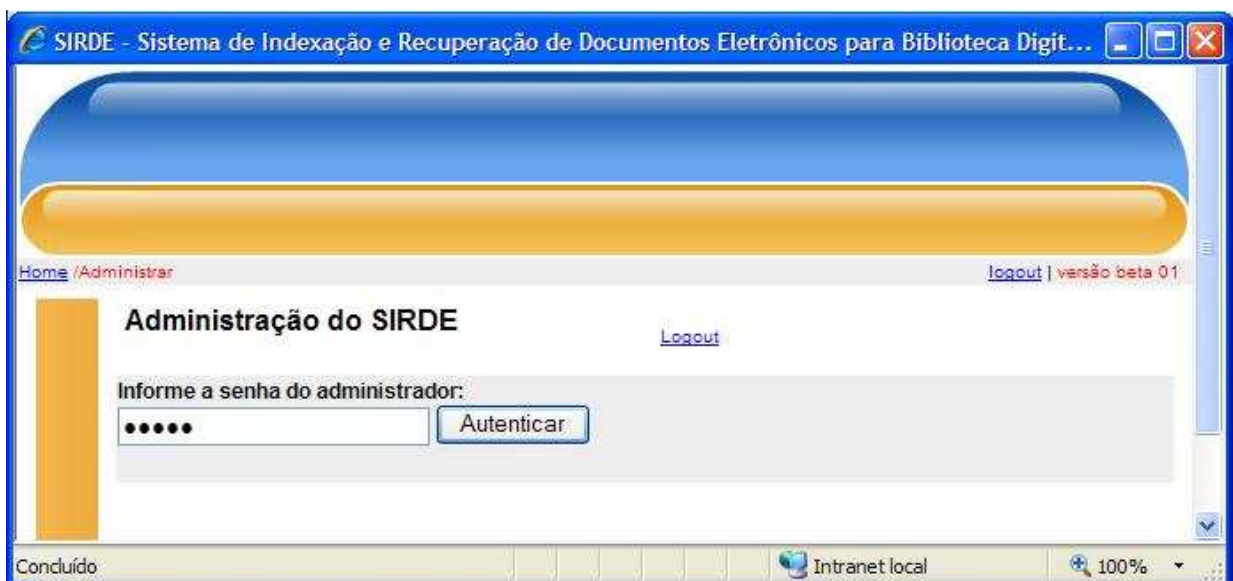


Figura 16 – Tela de *login* do administrador

A figura 17 mostra a primeira tela da área restrita onde o usuário informa os

parâmetros gerais do sistema.

The screenshot displays the 'Administração do SIRDE' interface. At the top, there is a blue and yellow header bar. Below it, a navigation menu includes 'Home / Administrar' and 'logout | versão beta 01'. The main content area is titled 'Administração do SIRDE' and contains a sub-menu: 'Parâmetros Sirde | Repositórios | Restrições | Indexar | LOG | Logout'. The 'Parâmetros do sistema:' section includes the following fields:

- Sigla do Sistema:** Text input field containing 'SIRDE'.
- Nome do Sistema:** Text input field containing 'Sistema de indexação e recuperação de documentos eletrônicos'.
- Dia da Semana do agente:** Dropdown menu set to 'Sexta'.
- Hora do Agente:** Two input fields for hours and minutes, both set to '00'.
- Senha Atual do Administrador:** Text input field with a red note: '*para gravar as alterações é necessário informar a senha atual'.
- Trocar Senha:** Text input field.
- Repetir nova Senha:** Text input field.

At the bottom, there are two buttons: 'Gravar' and 'Desfazer', with a red note '*campos obrigatórios' above them.

Figura 17 – Tela de parâmetros gerais do SIRDE

Nesta tela o administrador pode informar uma sigla e nome para a aplicação, o dia da semana e horário em que o agente irá executar a localização de documentos nos repositórios cadastrados, e efetuar a troca da senha padrão do administrador.

Os repositórios correspondem a locais onde deverão ser localizados documentos a serem indexados e documentos de outras instituições em que o agente não possui acesso via rede e somente acesso via Web. A seguir serão mostradas as telas de cadastro de repositórios e o funcionamento de cada uma delas.

A figura 18 apresenta a tela que lista todas as origens cadastradas. A partir desta tela o administrador poderá incluir uma nova ou alterar uma já existente.

Home / Administrar Logout | versão beta 01

Administração do SIRDE Logout

[Parâmetros Sirde](#) | [Repositórios](#) | [Restrições](#) | [Indexar](#) | [LOG](#) | [Logout](#)

Origem:

...	Dominio	Tipo	Repos.	Host	URL View
...	http://localhost:8084	Local	1	hp-dv6045nr	
...	www.bc.furb.br	Web	4		bdigital/connet/recupera_mfn_xml_dom.php

Figura 18 – Lista de repositórios

Ao selecionar uma origem já existente, o sistema abre a tela de manter origem em que o administrador poderá cadastrar ou alterar as informações da origem. A figura 19 apresenta a tela de manter origens.

Home / Administrar Logout | versão beta 01

Administração do SIRDE Logout

[Parâmetros Sirde](#) | [Repositórios](#) : [Origens](#) | [Restrições](#) | [Indexar](#) | [LOG](#) | [Logout](#)

Manter Origem:

Dominio: Tipo de Repositório:

URL View Web: Host:

Observações:

Repositórios Locais

...	Endereço diretório raiz	URL raiz	Sub-Níveis
...	Publico/docs	SIRDE/docs	Sim

Figura 19 – Tela para manter repositórios locais

Nesta tela é apresentado as informações da origem e os repositórios correspondentes. Se a origem for local, os dados de repositórios são informações referentes ao endereço físico e web, e se a origem for do tipo Web, serão listados os documentos referenciados para o sistema indexar. A figura 20 apresenta a tela de uma origem do tipo Web.

Manter Origem:

Domínio:
 Tipo de Repositório:

URL View Web:
 Host:

Observações:

Parâmetros para visualização dos documentos WEB			
...	Nome	Tipo	Valor
<input type="checkbox"/>	CD_MFN	Variável	Código do documento

Documentos Web			
...	Código	Título	URL
<input type="checkbox"/>	257038	Protótipo de software para troca de dados entre aplicações de comércio eletrônico utilizando o protocolo SOAP	http://www.bc.furb.br/docs/MO/2002/257038_1_1.pdf
<input type="checkbox"/>	140706	Redes neurais aplicadas ao sistema de informacao do jogo de empresas virtual /	http://www.bc.furb.br/docs/MO/1997/140706_1_1.pdf
<input type="checkbox"/>	191597	Empreendedorismo e formação de administradores : uma análise do Curso de Administração da Universidade Regional de Blumenau /	http://www.bc.furb.br/docs/TE/1999/191597_1_1.pdf
<input type="checkbox"/>	212618	Protótipo de um sistema de serviços WAP para a Biblioteca Central da FURB /	http://www.bc.furb.br/docs/MO/2000/212618_1_1.pdf

Figura 20 – Tela para manter repositórios Web

Para o tipo Web, também são informados os parâmetros para visualização do documento. Estes parâmetros são utilizados quando se faz necessário que a visualização do documento seja uma tela do sistema da biblioteca referenciada.

Na seqüência é apresentada a figura 21 que mostra a tela onde são cadastrados as restrições de documentos. São informações que o sistema utiliza para restringir documentos com certas características do tipo extensão, tamanhos ou que o nome contém determinados trechos de nomes.

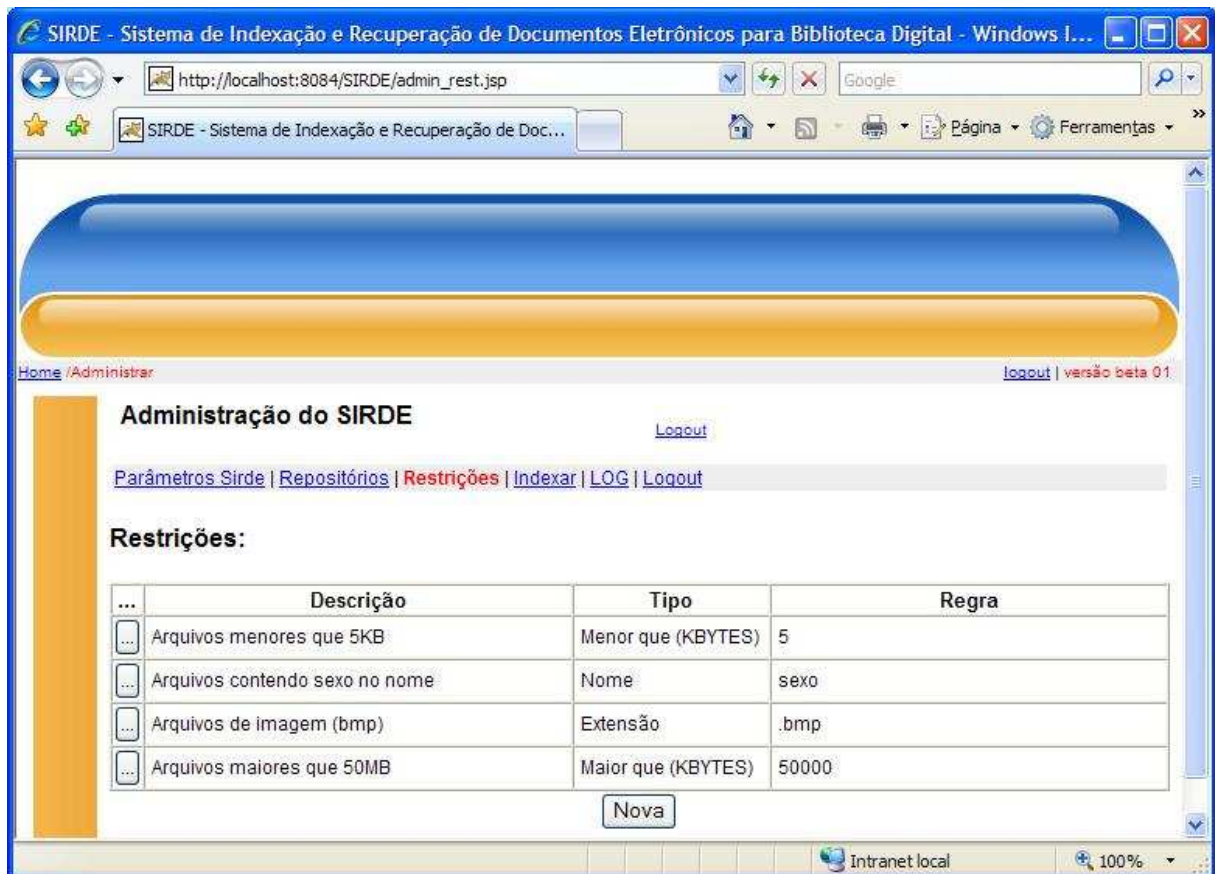


Figura 21 – Lista de restrições

A partir desta tela o administrador poderá incluir novas restrições ou alterar alguma já existente. As regras para restringir documentos poderão ser várias do mesmo tipo, pois o sistema irá executar uma a uma até que todas forem aplicadas.

Uma vez cadastradas as origens e seus repositórios e as restrições, o administrador poderá executar a indexação dos documentos ou aguardar que o mecanismo de agendamento de tarefas acione a indexação.

Caso o administrador opte por executar a indexação, o sistema disponibiliza uma sessão para estas atividades através da opção "indexar". Conforme se observa na figura 22, o administrador tem as opções de indexação rápida, total, otimização total ou otimização recriando o índice.

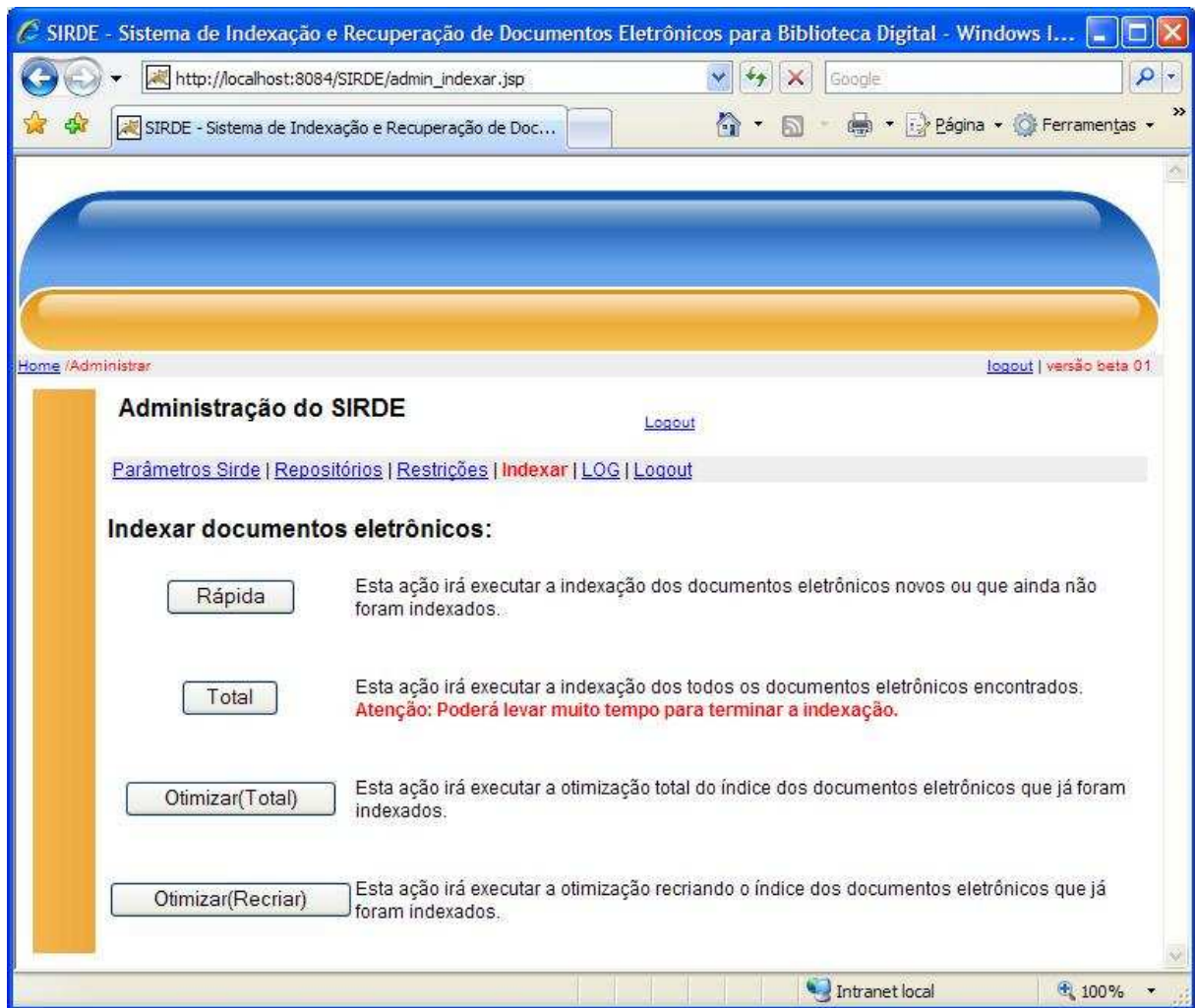


Figura 22 – Tela de indexação e otimização

A indexação rápida permite ao administrador que somente os documentos novos e que ainda não foram indexados seja processados pelo sistema e pelo InterMedia. Já a indexação total faz com que o sistema exclua todos os registros já indexados e processe todas as origens novamente, criando o índice total.

A otimização possui uma característica um pouco diferente pois não incorpora novos documentos ao índice mesmo que existam novos. O sistema simplesmente utiliza os comandos do InterMedia para que este faça a sincronização dos índices existentes.

Um exemplo prático da otimização dos índices foram apresentados na seção 3.5.1 onde foram apresentados exemplos através das tabelas 1, 2 e 3.

Para se ter um controle do comportamento do sistema e das mensagens que o InterMedia gera, o administrador possui a tela de *log* para consultar e ajudá-lo a tomar decisões para que todo o processo execute com eficácia. A figura 23 apresenta a tela de *log*.



Figura 23 – Tela de consulta e limpeza de logs

Nesta tela o administrador possui as opções de excluir os logs já lidos ou redundantes. Para isto, será necessário que sejam selecionado e pressionado o botão de excluir.

Até o momento foi apresentada a área restrita do sistema direcionada ao administrador. A partir de agora, será demonstrado como pode ser utilizado a área aberta onde qualquer usuário possui acesso para efetuar as pesquisas.

A primeira opção é a pesquisa básica em que o usuário informa somente uma expressão de busca. A figura 24 apresenta a tela de pesquisa básica.

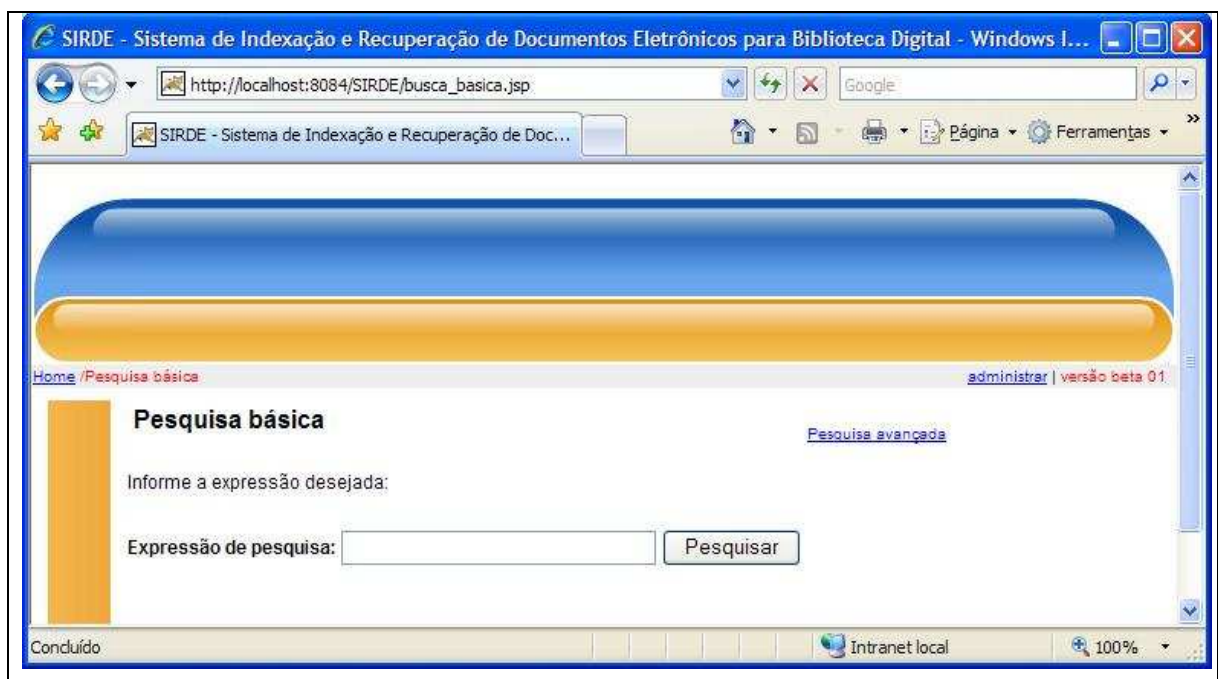


Figura 24 – Tela para pesquisa básica

Nesta opção, embora exista a pesquisa avançada, o usuário poderá utilizar algumas diretrizes para o InterMedia que permite ao usuário efetuar buscas mais refinadas. A tabela 4 apresenta os operadores lógicos permitidos pelo InterMedia.

Operador	Símbolo	Descrição	Exemplo
E	&	Usado para localizar documentos que possuem pelo menos uma ocorrência para cada termo informado	Documentos que tenham as palavras 'cão' e 'gato'. Expressão: 'cão&gato'
Ou		Usado para localizar documentos que possuem pelo menos uma ocorrência de qualquer um dos termos informados	Documentos que tenham qualquer das palavras 'cão' ou 'gato'. Expressão: 'cão gato'
Negação	~	Usado para localizar documentos que possuem um termo e que não possuam outro termo	Documentos que tenham a palavra 'animal' e que não tenha a palavra 'cão'. Expressão: 'animal~cão'
Lista	,	Usado para localizar documentos que possuem uma lista de termos	Documentos que tenham as palavras 'cão', 'gato' e 'rato'. Expressão: 'cão,gato,rato'
Igual	=	Usado para localizar documentos com equivalência de termos	Documentos que tenham a expressão 'cães são animais dóceis' ou 'gatos são animais dóceis'. Expressão: 'cães=gatos são animais dóceis'

Quadro 13 – Operadores lógicos do InterMedia

Os operadores lógicos apresentados acima podem ser utilizados tanto na pesquisa básica quanto na avançada.

A pesquisa avançada possui alguns atributos extras que permitirão ao usuário a pesquisa por documentos específicos, quantidade de registros na tela de resultados e um método de pesquisa que incluem os operadores lógicos para o usuário. A figura 25 apresenta a tela de pesquisa avançada.

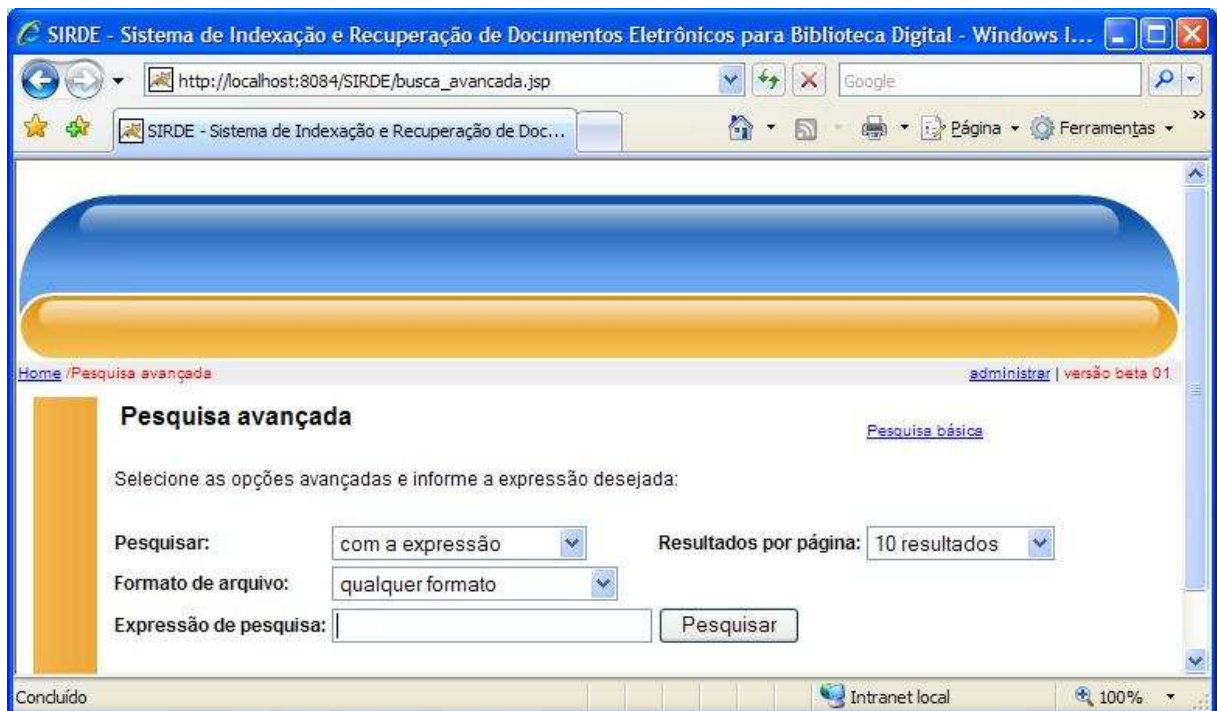


Figura 25 – Tela para pesquisa avançada

Tanto pela tela de pesquisa básica como da tela de pesquisa avançada ao executar a

pesquisa, o sistema redireciona para a tela de resultados. A figura 26 apresenta esta tela.

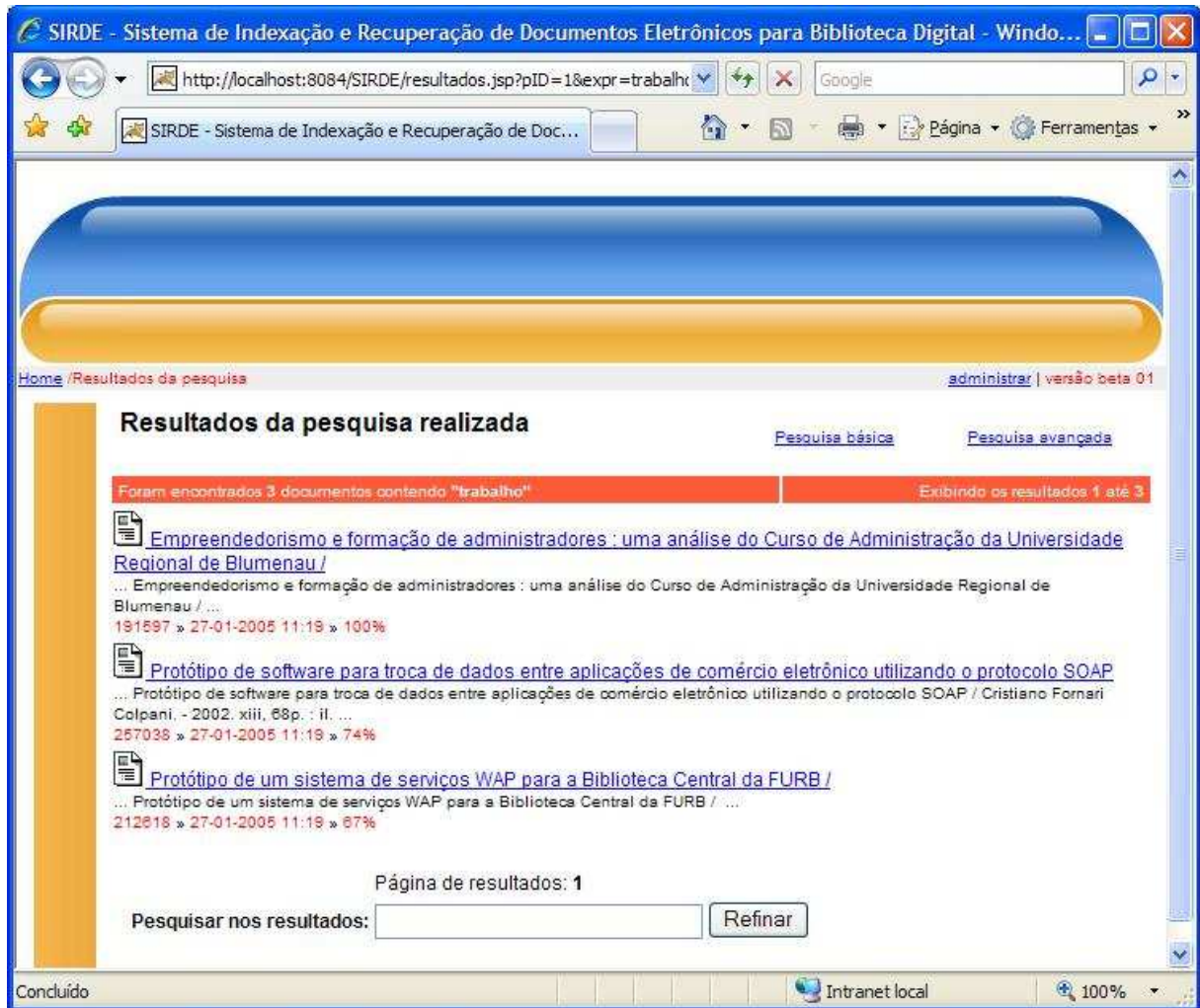


Figura 26 – Tela de resultados da busca e pesquisa nos resultados

Os resultados são apresentados em ordem decrescente de relevância, ou seja, o documento que possui mais ocorrências da palavra ou expressão pesquisada aparecerá no topo da lista.

3.6 RESULTADOS E DISCUSSÃO

Após várias execuções do sistema observou-se excelentes resultados no que diz respeito a localização das informações desejadas. Destacando-se principalmente, a rapidez com que os resultados são apresentados ao usuário e pela relevância dos documentos em relação à expressão de busca. Outro ponto de bons resultados é a forma como o sistema recupera e executa a indexação dos documentos locais ou que estão referenciados na Web. O

InterMedia possibilitou uma fácil implementação de funções que utilizam seus recursos.

Fazendo um comparativo com o Google Acadêmico (GOOGLE, 2006?), percebe-se que ambos apresentam bons resultados justamente pelo fato de recuperarem os documentos pelo contexto e não somente por palavras chaves ou resumos.

Em comparação com o iSMART *Geospatial Suite* (ITGIS 1999), é importante ressaltar que há uma diferença somente no tipo de dados tratado, que no iSMART são dados espaciais para construção e localização de mapas enquanto este trabalho localiza documentos através de seu conteúdo texto. Mesmo assim, ambos obtêm bom desempenho em apresentar os resultados no momento de efetuar as pesquisas desejadas.

Comparando com o sistema de gerenciamento geográfico (SOUZA, AMARO e CASTRO, 2005), ambos possuem métodos de indexação por informações relevantes do conteúdo dos arquivos, porém diferente deste trabalho, trabalha somente com as informações cadastrais das imagens inseridas pelos usuários.

Diferente dos trabalhos citados anteriormente, o Viseu Digital (MARTINS, 2005) pode ser comparado como uma ferramenta mais parecida com este trabalho, pois implementa um sistema de indexação e recuperação de informações textos utilizando o InterMedia da Oracle. A diferença está no conteúdo que é indexado que neste trabalho são documentos eletrônicos de bibliotecas digitais, enquanto que o Viseu Digital indexa documentos do próprio portal.

4 CONCLUSÕES

Este trabalho apresentou um sistema para auxiliar os usuários na busca por material científico e tecnológico de forma rápida e eficaz. O sistema mostrou-se eficaz no atendimento aos objetivos, proporcionando uma excelente opção aos usuários na diminuição do tempo gasto na busca por informações mais relevantes às suas pesquisas.

O InterMedia Text do Oracle viabilizou a construção deste sistema que oferece a busca rápida por palavras ou frases em documentos externos ao banco de dados. Desta forma, contribuiu para o atendimento dos objetivos permitindo que os usuários possam localizar documentos eletrônicos da biblioteca digital da FURB, bem como de outras instituições através da Internet. A possibilidade de o usuário poder visualizar ou até mesmo baixar documentos que são mais relevantes ao critério de pesquisa proporciona uma melhor confiança no sistema de busca contribuindo muito para o aprendizado e estudos dos alunos.

Pesquisas refinadas, como a busca nos resultados desenvolvida neste trabalho, permitem que o usuário encontre mais rapidamente e de forma eficaz o material de interesse. Através do mecanismo de geração de *logs*, o administrador pode acompanhar o comportamento do sistema possibilitando agir para o melhoramento da ferramenta, fazendo com que erros ou situações não tratadas possam ser corrigidas.

Dentre as principais vantagens do sistema apresentado, destaca-se o emprego de técnicas de localização e indexação de documentos baseado em parâmetros indicando a localização do repositório de documentos, e da utilização dos recursos do InterMedia disponíveis nos bancos de dados Oracle. Tais características permitiram o desenvolvimento do trabalho utilizando tecnologias recentes, e que apresentam resultados desejáveis.

Os objetivos previstos foram alcançados visto que o sistema foi desenvolvido utilizando técnicas e recursos permitindo que os usuários façam pesquisas de documentos eletrônicos através do conteúdo e não somente por palavras chaves e que a parametrização dos repositórios realiza a integração com outros sistemas, com visualização e *download* dos arquivos encontrados na biblioteca digital da FURB bem como de outras instituições fazendo uso da Internet.

4.1 EXTENSÕES

Este trabalho pode ser continuado através da implementação de mecanismos de localização de documentos eletrônicos na Web, ou seja, desenvolver rotinas para permitir que o agente identifique através de parâmetros os documentos de outras bibliotecas digitais sem que seja citado o caminho do documento para indexação. Dessa forma o sistema se tornará mais atualizado, pois uma vez que algum documento novo é publicado na biblioteca remota, o agente recupera-o e realiza a indexação automática do mesmo.

REFERÊNCIAS BIBLIOGRÁFICAS

BIBLIOTECAS digitais. In: WIKIPEDIA, a enciclopédia livre. [S.l.]: Wikimedia Foundation, 2007. Disponível em: <http://pt.wikipedia.org/wiki/Biblioteca_digital>. Acesso em: 22 maio 2007.

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **UML: guia do usuário**. Tradução Fábio Freitas. Rio de Janeiro: Campus, 2000.

CARDOSO, Jiani. **Uma proposta de interface de consulta personalizável para bibliotecas digitais**. 2000. 112 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-graduação em Ciência da Computação, Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre.

CUNHA, Murilo. **Desafios na construção de uma biblioteca digital**. Ci. Inf., Brasília, v. 28, n. 3, 1999. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651999000300003&lng=pt&nrm=iso>. Acesso em: 14 abr. 2007. Pré-publicação.

FURB - UNIVERSIDADE REGIONAL DE BLUMENAU. **A biblioteca digital da FURB: apresentação**. Blumenau, [2005?]. Disponível em: <http://www.bc.furb.br/bdigital/conheca_projeto.htm>. Acesso em: 09 set. 2006.

GOOGLE, Inc. **Sobre o Google acadêmico**. São Paulo, [2006?]. Disponível em: <<http://scholar.google.com.br/intl/pt-BR/scholar/about.html>>. Acesso em: 05 abr. 2007.

HOHAGEM, Alexandre (Org.). **Tudo sobre o Google**. [S.l.], [2005?]. Disponível em: <<http://www.google.com.br/intl/pt-br/about.html>>. Acesso em: 9 out. 2006.

ITGIS - Softwares e Equipamentos. **iSMART geospatial suite**. São José do Rio Preto, 1999. Disponível em: <<http://www.itgis.com.br/espacial.asp>>. Acesso em: 09 set. 2006.

IUNES, Ivan. **UNB agencia: acesso mais democrático à informação**. Brasília, 2004. Disponível em: <<http://www.unb.br/acs/unbagenzia/ag0104-05.htm>>. Acesso em: 18 out. 2006.

LEINER, Barry. The NCSTRL Approach to Open Architecture for the Confederated Digital Library. **D-Lib Magazine**, Virginia, Dec, 1998. Disponível em: <<http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/december98/leiner/12leiner.html>>. Acesso em: 25 out. 2006.

MARCHIORI, Patrícia. **Ciberteca ou biblioteca virtual: uma perspectiva de gerenciamento de recursos de informação**. Ci. Inf., Brasília, v. 26, n. 2, 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000200002&lng=en&nrm=iso>. Acesso em: 23 Abr. 2007. Pré-publicação.

MARTINS, Victor. Soluções tecnológicas à medida de cada região. **e-Ciência**, Cartaxo, 2005. Disponível em: <<http://www.cienciapt.net/entrevista/vitormartins.pdf>>. Acesso em: 09 set. 2006.

OLIVEIRA, João; CARDOSO, Jiani. Problemáticas em interfaces de busca de bibliotecas digitais. **Ciência da Computação**, Porto Alegre, 2003. Disponível em: <[http://lsm.dei.uc.pt/ribie/docfiles/txt200372912310Problem%C3%A1ticas%20em%20interfa](http://lsm.dei.uc.pt/ribie/docfiles/txt200372912310Problem%C3%A1ticas%20em%20interfa%C3%A7%C3%B5es%20de%20busca.pdf)>. Acesso em: 03 nov. 2006.

ORACLE, JD Edwards. **Oracle text reference**. Redwood, 2005a. Disponível em: <http://download-east.oracle.com/docs/cd/B19306_01/text.102/b14218/toc.htm>. Acesso em: 22 maio 2007.

ORACLE, JD Edwards. **Oracle text application developer guide**. Redwood, 2005b. Disponível em: <http://download-east.oracle.com/docs/cd/B19306_01/text.102/b14217/toc.htm>. Acesso em: 22 maio 2007.

PELSKI, Sue. **Oracle interMedia user's guide**. Redwood, 2005a. Disponível em: <http://download-east.oracle.com/docs/pdf/B14302_01.pdf>. Acesso em: 13 out. 2006.

PELSKI, Sue. **Oracle interMedia reference**. Redwood, 2005b. Disponível em: <http://download-east.oracle.com/docs/pdf/B14297_01.pdf>. Acesso em: 13 out. 2006.

RAABE, André et al. Avaliação de usabilidade de uma biblioteca digital. **Ciência da Computação e Pós-Graduação em Engenharia de Produção**, Itajaí, 2002. Disponível em: <<http://delivery.acm.org/10.1145/950000/944549/p243-raabe.pdf?key1=944549&key2=6218392611&coll=&dl=acm&CFID=15151515&CFTOKEN=6184618>>. Acesso em: 18 out. 2006.

SANTOS, Gildenir; PASSOS, Rosemary. **Manual de organização de referências e citações bibliográficas para documentos impressos e eletrônicos**. Campinas, SP: Autores Associados; Ed. UNICAMP, 2000.

SOUZA, Clenúbio; AMARO, Venerando; CASTRO, Angélica. Sistema de informação geográfica para o monitoramento ambiental de regiões costeiras e estuarinas do estado do Rio Grande do Norte. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Natal, **Anais...** Natal: INPE, 2005. p. 2383-2388. Disponível em: <<http://marte.dpi.inpe.br/col/ltid.inpe.br/sbst/2004/11.18.18.56/doc/2383.pdf>>. Acesso em: 09 set. 2006.

UNICAMP - UNIVERSIDADE DE CAMPINAS. **Serviços - bibliotecas: acervos eletrônicos**. Campinas, [2004?]. Disponível em: <http://www.unicamp.br/unicamp/servicos/servicos_bibliotecas.html#acervos>. Acesso em: 22 mai. 2007.

USP - UNIVERSIDADE DE SÃO PAULO. **Biblioteca digital de teses e dissertações**. São Paulo, [2005?]. Disponível em: <<http://www.teses.usp.br/biblioteca.html>>. Acesso em: 22 maio 2007.