

The background of the slide is a collage of images. On the right side, there is a computer mouse on a white base. Below it, a hand is shown holding a CD-ROM. The overall color scheme is a mix of light green and white, with a dark green rectangular area on the left containing the title text.

# Ferramenta de indexação e recuperação de documentos eletrônicos utilizando Oracle InterMedia

Acadêmico: Emerson Immianovsky

Orientador: Prof. Alexander R. Valdameri

# Roteiro

- Introdução
- Objetivos
- Fundamentação
- Trabalhos correlatos
- Desenvolvimento
- Operacionalidade
- Conclusões



# Introdução

- Bibliotecas tradicionais
- Bibliotecas digitais
- Sistemas atuais
- Métodos de indexação



# Objetivos

- Sistema de busca por conteúdo
- Indexação biblioteca FURB e outras instituições
- Refinamento das pesquisas
- Interface de administração do sistema
- Verificação de problemas no sistema

# Bibliotecas digitais

- Conceito
- Evolução
- Arquitetura
- Bibliotecas digitais existentes



# Interface biblioteca digital

- Sistema de busca
- Opções de busca
- Ergonomia
- Simplicidade
- Eficiência

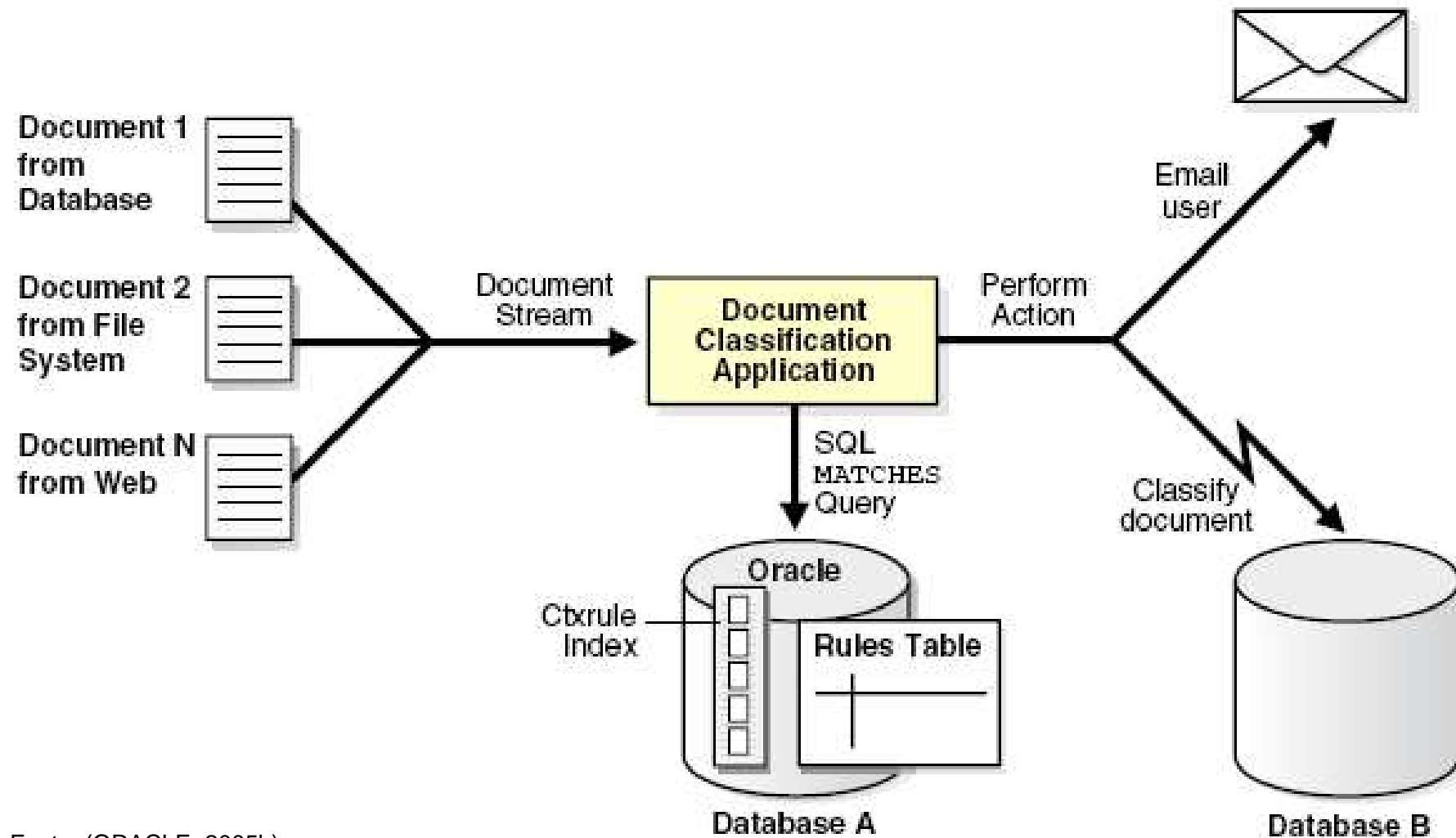


# Oracle InterMedia

- O que é?
- Instalação
- Tipos de indexação
- Documentos suportados
- Funcionamento
- Oracle Text



# Oracle InterMedia



Fonte: (ORACLE, 2005b).



# Interface biblioteca UNICAMP

## Índice

[Página principal](#)

[Documentos](#)

[Usuários](#)

## Procurar: todos os tópicos

Para encontrar documentos contendo uma ou mais palavras, preencha o seguinte campo:

Procurar por:   [Procura avançada](#)

## Ações

[Consultar](#)

[Procurar](#)

[Exibir estatísticas](#)

Para acessar um documento dado o seu código, basta indicá-lo abaixo

Código:

## Procurar por:

[Procura avançada](#)

# Interface biblioteca FURB



Sistema Integrado de Bibliotecas

[Página Principal](#) [Logout](#)

[Voltar](#)

- Conheça o projeto
- Acervo Digital
- Publicação
- Perguntas Frequentes

## BUSCA RÁPIDA

Expressão de Busca:

**Pesquisa:**  Livre  Autor  Assunto  Título

**Listar por:**

Título  Referência Bibliográfica

Procurar

Limpar

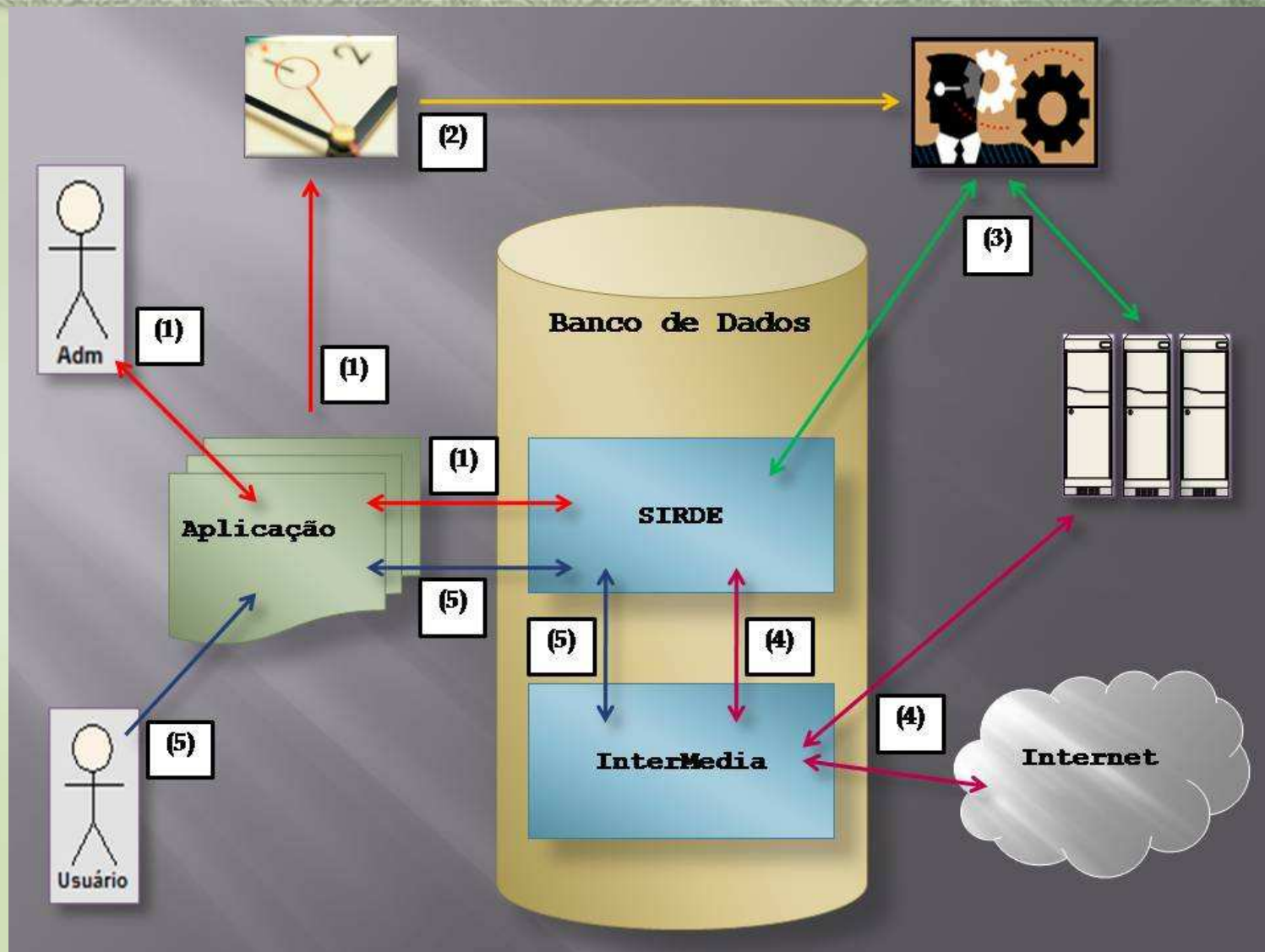
# Trabalhos correlatos

- *iSmart* - ITGis Softwares e Equipamentos (1999)
- Sistema de gerenciamento geográfico - SOUZA, AMARO e CASTRO (2005)
- Viseu Digital – Victor Martins (2005)
- Google Acadêmico - Google Inc. (2006)

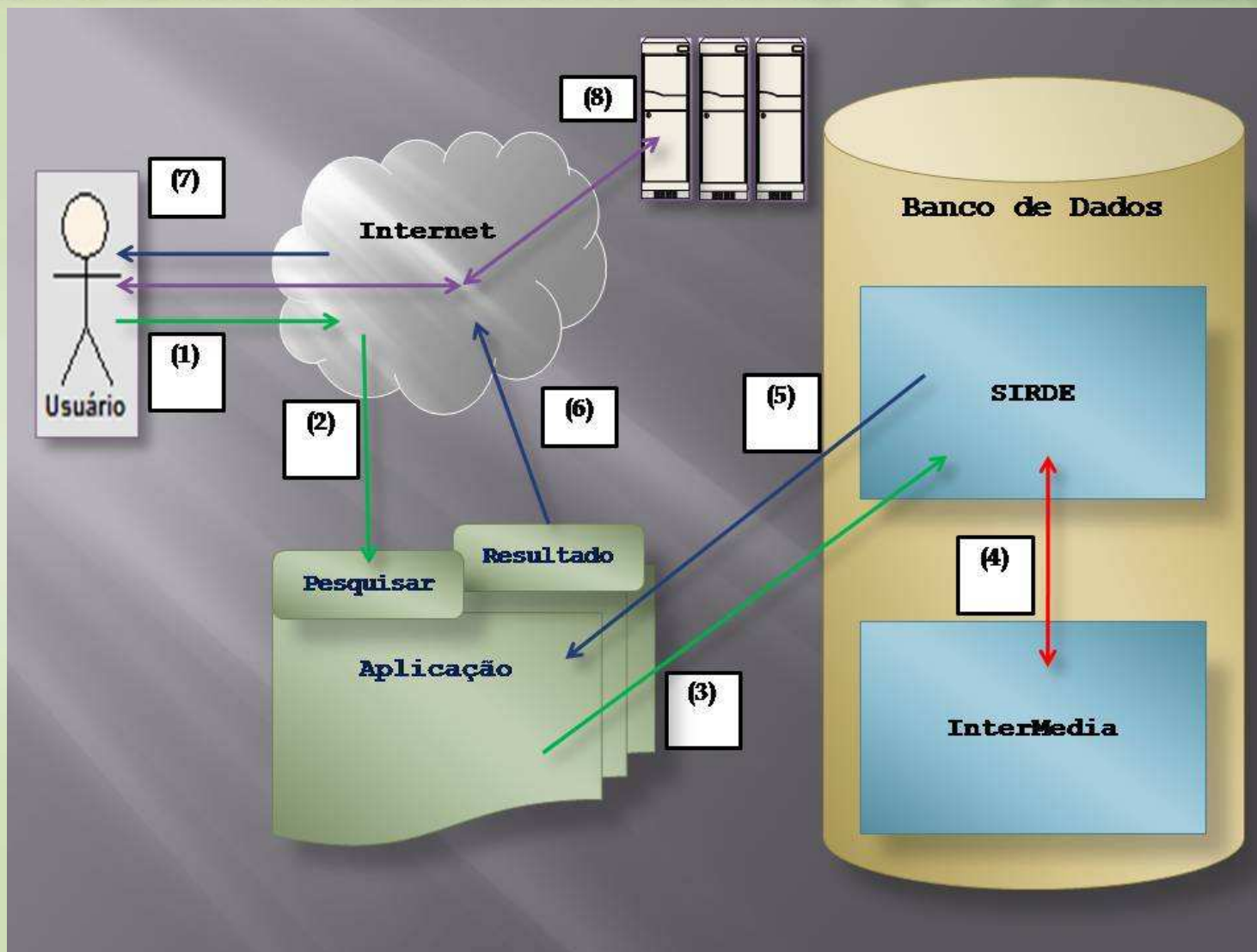
# Requisitos

- Pesquisas por conteúdo
- Integração com biblioteca da FURB e outras instituições através da Internet
- Listar resultados por relevância
- Pesquisar nos resultados
- Visualização e *download*
- Administração
- Indexação automática e manual com otimização

# Visão geral



# Funcionamento da pesquisa

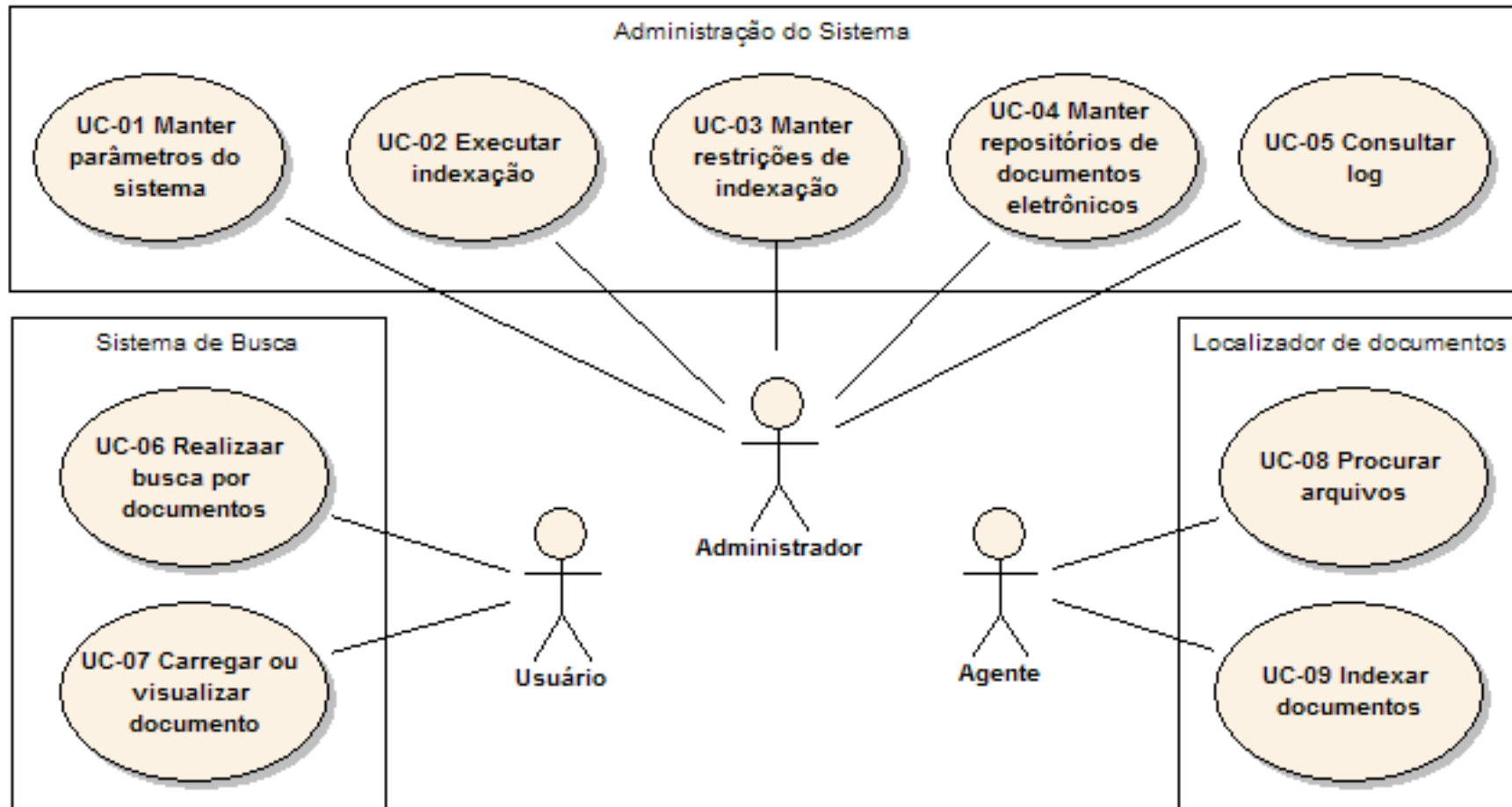


# Especificação

- Diagrama de Caso de Uso
- Diagrama de Classes
- Diagrama de Atividades
- Diagrama de Entidade Relacionamento

# Diagrama de Casos de Uso

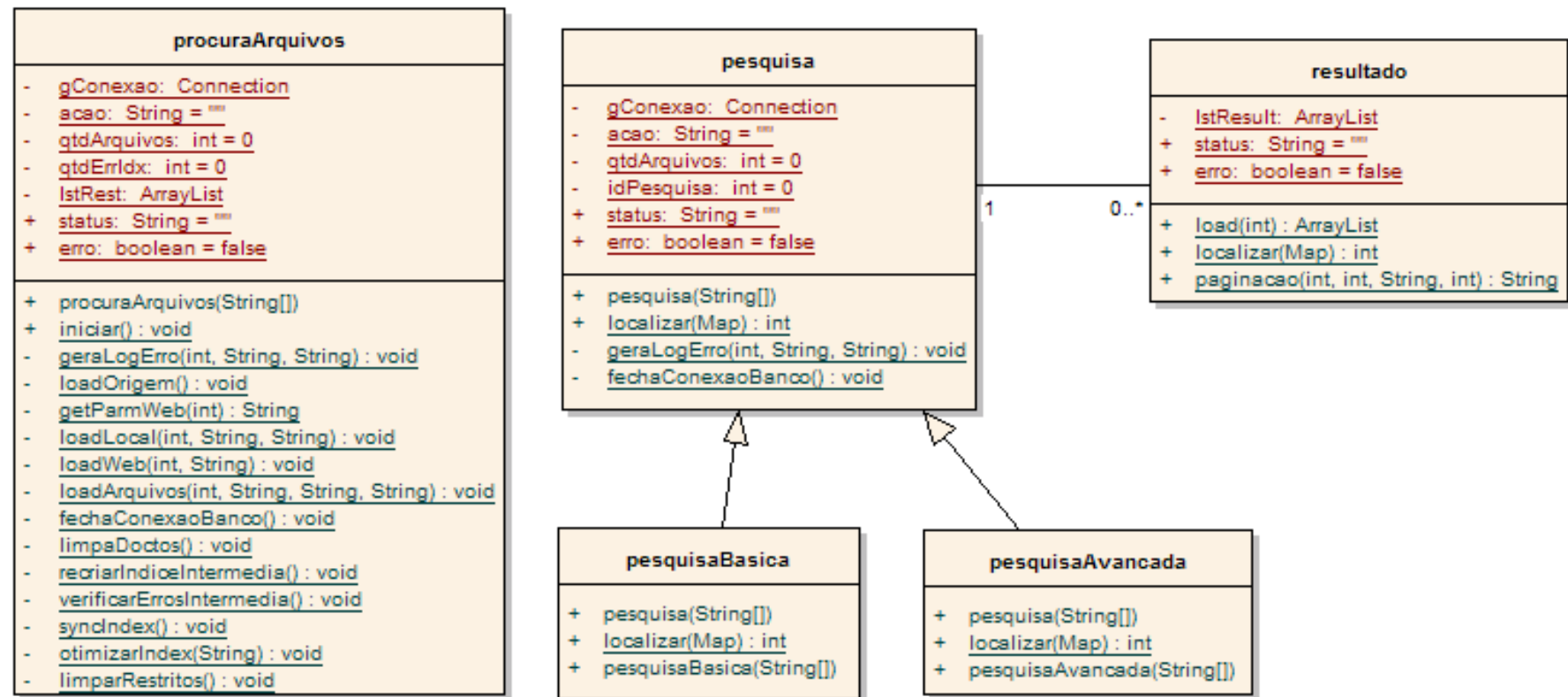
ud Primary Use Cases





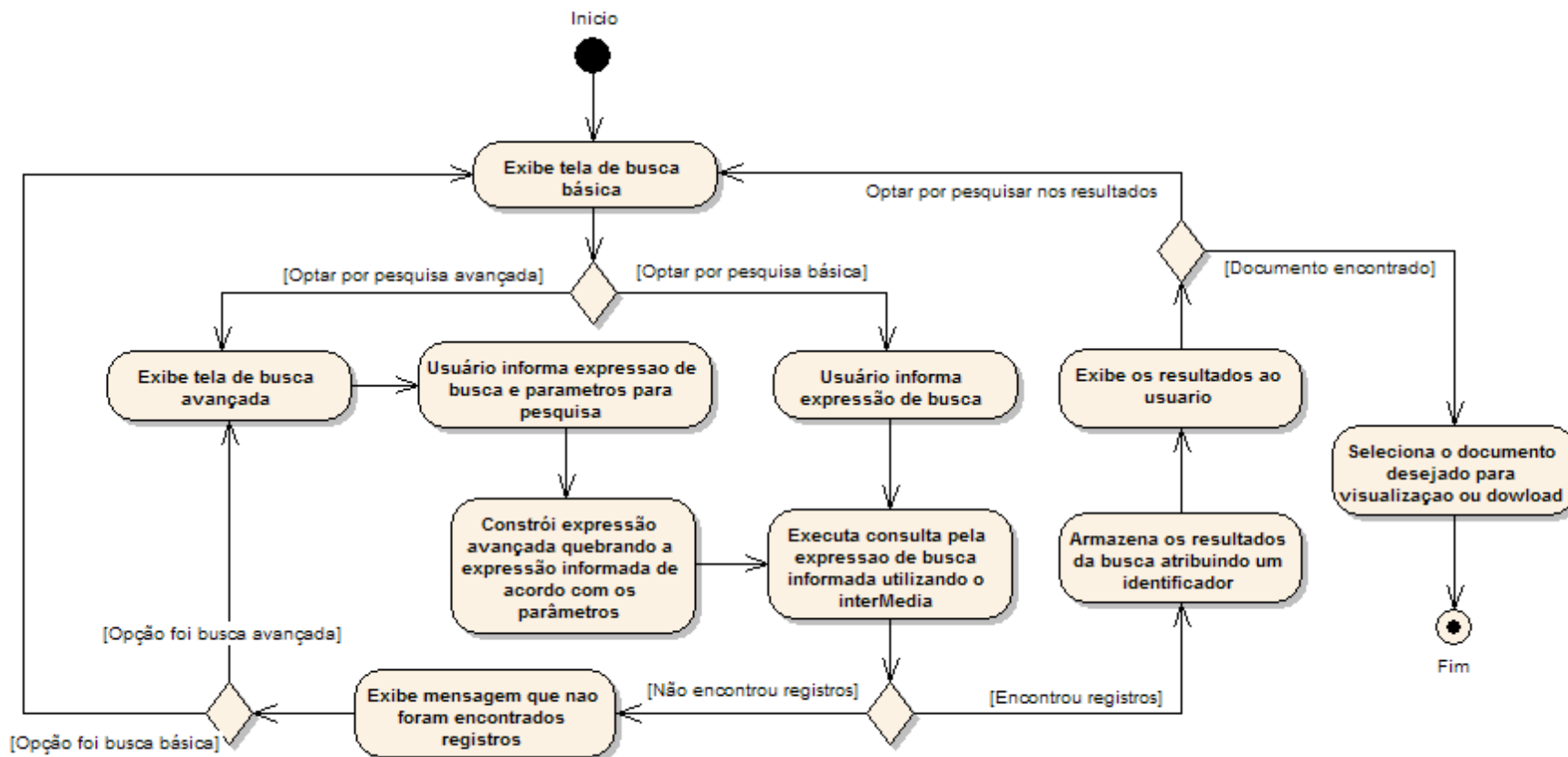
# Diagrama de Classes

UML Diagrama de Classe



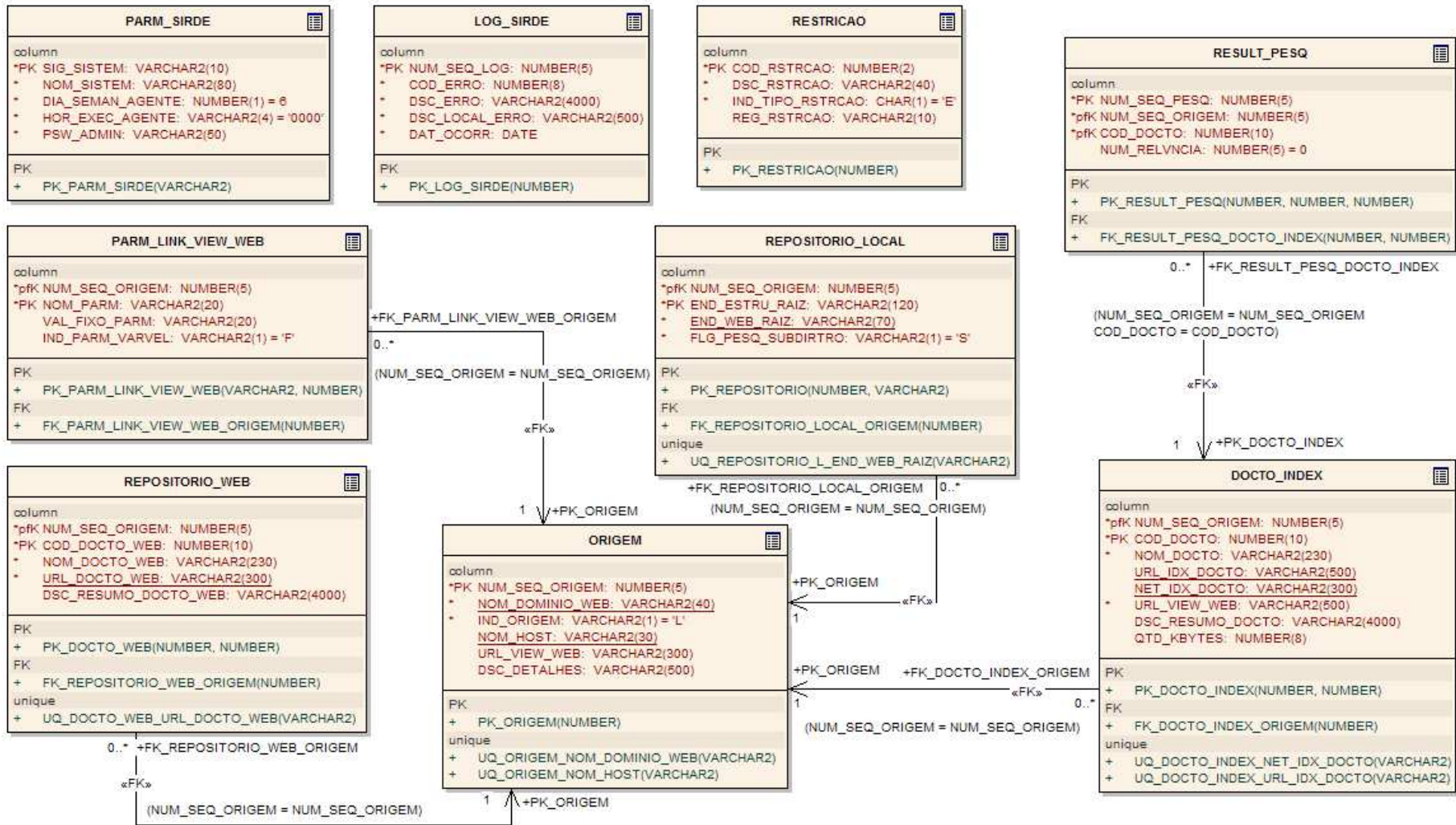
# Diagrama de atividades

ad Diagrama de Atividades



# Diagrama de Dados

cd Modelo de Dados



# Implementação

- Ferramentas utilizadas
  - *Web TomCat 5.5*
  - *SGBD Oracle 10G Express Edition*
  - *Oracle InterMedia 10G*
  - *IDE NetBeans 5.5*
- Linguagem
  - Java 1.5
  - JSP

# Pesquisa com Oracle InterMedia

```
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         geralLogErro(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 }
```

# Indexação pelo Oracle InterMedia

```
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         geraLogErro(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         geraLogErro(7777,e.toString(), "Ao recriar indice InterMedia");
398         erro = true;
399     }
400 }
```

# Exemplos de otimização

## Rápida

Índice Fragmentado	Índice otimizado
palavra1 doc1 doc2 doc3	palavra1 doc1 doc2 doc3
palavra2 doc6	palavra2 doc6
palavra3 doc1 doc6	palavra2 doc4 doc5
palavra2 doc4 doc5	palavra3 doc1 doc6
Palavra3 doc5	palavra3 doc5

## Total

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

# Exemplos de otimização

## Recriação

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



# Operacionalidade

- Acesso irrestrito
  - Principal
  - Pesquisa básica
  - Pesquisa avançada
  - Lista de resultados
  - Pesquisa nos resultados
- Acesso restrito
  - Parâmetros gerais
  - Repositórios
  - Restrições
  - Indexação
  - *LOGs*



# Acesso principal

The screenshot shows a web browser window with the following elements:

- Browser Title Bar:** SIRDE - Sistema de Indexação e Recuperação de Documentos Eletrônicos para Biblioteca Digit...
- Address Bar:** http://localhost:8084/SIRDE/
- Navigation and Tools:** Back, Forward, Stop, Refresh, Home, RSS, Print, Page, and Tools menus.
- Page Header:** Home/ [administrar](#) | versão beta 01
- Main Content:**
  - Opções de pesquisa por conteúdo de documentos eletrônicos:**
  - Para encontrar documentos contendo uma ou mais palavras, escolha uma das opções abaixo:
  - [Pesquisa básica](#) [Pesquisa avançada](#)
- Page Footer:** Concluído, Intranet local, 100%

# Lista de resultados

The screenshot displays the SIRDE (Sistema de Indexação e Recuperação de Documentos Eletrônicos para Biblioteca Digital) web application. The browser window title is "SIRDE - Sistema de Indexação e Recuperação de Documentos Eletrônicos para Biblioteca Digital - Windo...". The address bar shows the URL "http://localhost:8084/SIRDE/resultados.jsp?pID=1&expr=trabalho". The search results page features a blue header with a yellow bar below it. The main content area is titled "Resultados da pesquisa realizada" and includes links for "Pesquisa básica" and "Pesquisa avançada". A red banner indicates "Foram encontrados 3 documentos contendo 'trabalho'" and "Exibindo os resultados 1 até 3". Three search results are listed, each with a document icon, a title, a description, and a relevance score. At the bottom, there is a "Página de resultados: 1" indicator and a "Pesquisar nos resultados:" field with a "Refinar" button. The status bar at the bottom shows "Concluído", "Intranet local", and "100%".

SIRDE - Sistema de Indexação e Recuperação de Documentos Eletrônicos para Biblioteca Digital - Windo...

http://localhost:8084/SIRDE/resultados.jsp?pID=1&expr=trabalho

Google

SIRDE - Sistema de Indexação e Recuperação de Doc...

Home / Resultados da pesquisa administrar | versão beta 01

## Resultados da pesquisa realizada

[Pesquisa básica](#) [Pesquisa avançada](#)

Foram encontrados 3 documentos contendo "trabalho" Exibindo os resultados 1 até 3

- [Empreendedorismo e formação de administradores : uma análise do Curso de Administração da Universidade Regional de Blumenau /](#)  
... Empreendedorismo e formação de administradores : uma análise do Curso de Administração da Universidade Regional de Blumenau / ...  
191597 » 27-01-2005 11:19 » 100%
- [Protótipo de software para troca de dados entre aplicações de comércio eletrônico utilizando o protocolo SOAP](#)  
... Protótipo de software para troca de dados entre aplicações de comércio eletrônico utilizando o protocolo SOAP / Cristiano Fornari Colpani. - 2002. xiii, 68p. : il. ...  
257038 » 27-01-2005 11:19 » 74%
- [Protótipo de um sistema de serviços WAP para a Biblioteca Central da FURB /](#)  
... Protótipo de um sistema de serviços WAP para a Biblioteca Central da FURB / ...  
212618 » 27-01-2005 11:19 » 67%

Página de resultados: 1

Pesquisar nos resultados:

Concluído Intranet local 100%

# Parâmetros gerais

[Home](#) / [Administrar](#) [logout](#) | versão beta 01

## Administração do SIRDE [Logout](#)

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

### Parametros do sistema:

- Sigla do Sistema:

- Nome do Sistema:

- Dia da Semana do agente:

- Hora do Agente:  :

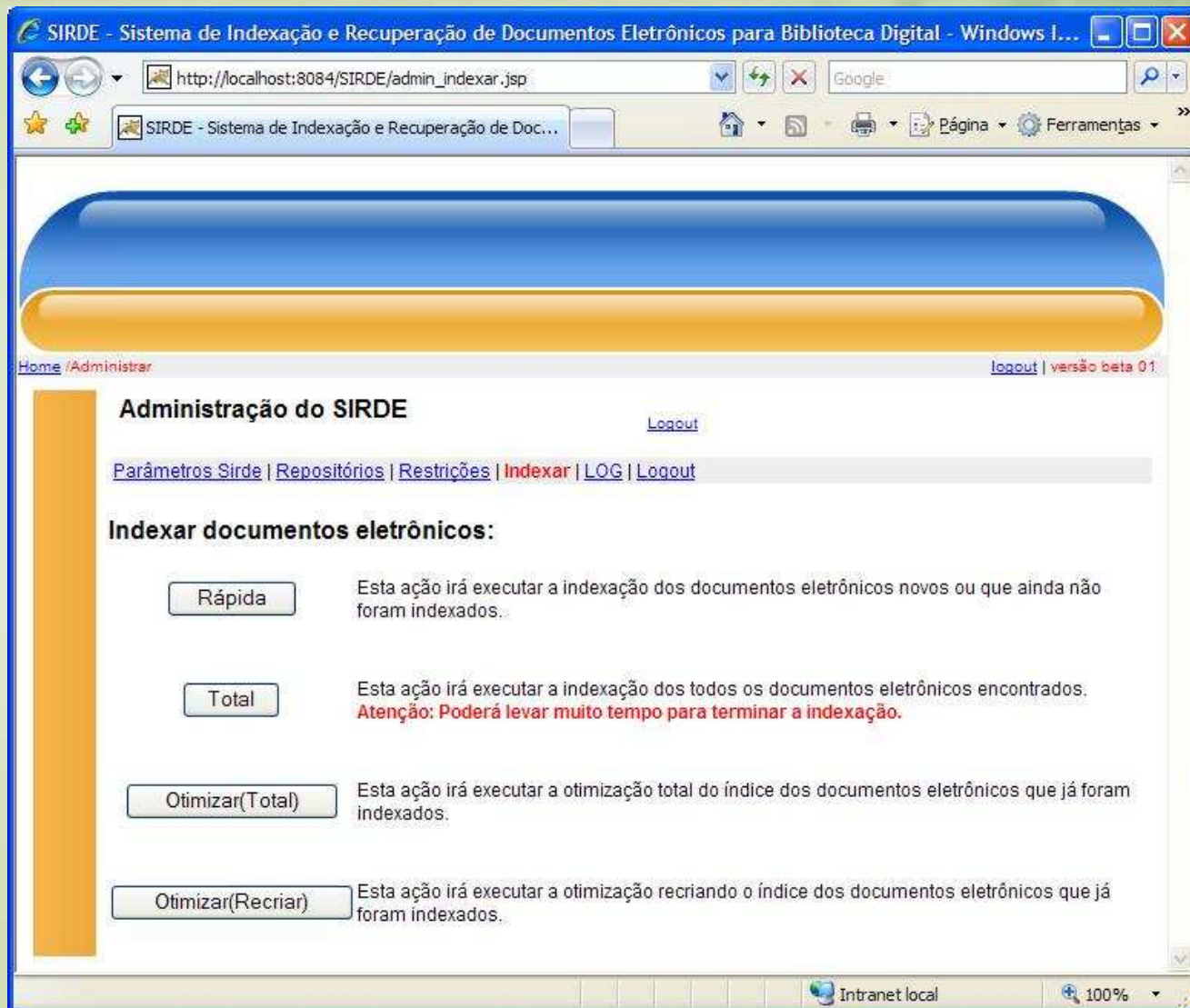
- Senha Atual do Administrador:  \*para gravar as alterações é necessário informar a senha atual

Trocar Senha:

Repetir nova Senha:

\*campos obrigatórios:

# Indexação e Otimização



SIRDE - Sistema de Indexação e Recuperação de Documentos Eletrônicos para Biblioteca Digital - Windows I...

http://localhost:8084/SIRDE/admin\_indexar.jsp

Home / Administrar [logout](#) | versão beta 01

## Administração do SIRDE

[Logout](#)

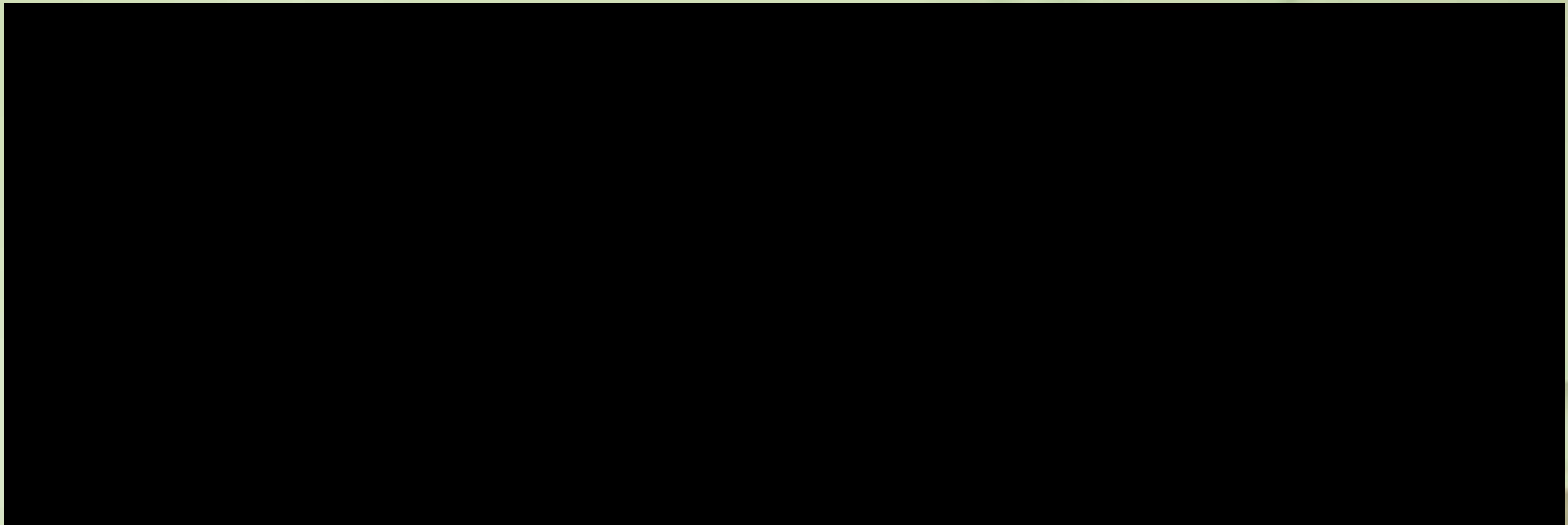
[Parâmetros Sirde](#) | [Repositórios](#) | [Restrições](#) | **Indexar** | [LOG](#) | [Logout](#)

### Indexar documentos eletrônicos:

<input type="button" value="Rápida"/>	Esta ação irá executar a indexação dos documentos eletrônicos novos ou que ainda não foram indexados.
<input type="button" value="Total"/>	Esta ação irá executar a indexação dos todos os documentos eletrônicos encontrados. <b>Atenção: Poderá levar muito tempo para terminar a indexação.</b>
<input type="button" value="Otimizar(Total)"/>	Esta ação irá executar a otimização total do índice dos documentos eletrônicos que já foram indexados.
<input type="button" value="Otimizar(Recriar)"/>	Esta ação irá executar a otimização recriando o índice dos documentos eletrônicos que já foram indexados.

Intranet local 100%

# Operadores de pesquisa



# Resultados e discussões

- Performance
- Relevância
- Integração
- Indexação documentos remotos



# Conclusões

- Excelente ferramenta de pesquisa acadêmica
- Potencial do Oracle InterMedia
- Integração
- Objetivos alcançados





# Extensões

- Construção de agentes de localização e indexação WEB
- Mecanismo de autenticação para sites restritos
- Possibilitar novos tipos de documentos

# Relevância Pessoal

- Aprendizado acadêmico
- Contribuição profissional
- Realização pessoal



# Demonstração do Sistema

**SIRDE**

The background of the slide is a faded, light green image. It depicts a person wearing a white lab coat, likely a scientist or technician. The person is shown from the chest up, holding a petri dish in their left hand and a computer mouse in their right hand. The overall scene suggests a laboratory or research environment where a computer system is being used for data analysis or monitoring.