

Instituto de Computação



Serviços de Recomendação e a Plataforma RecS-DL

Orientado: Daniel Carlos Guimarães Pedronette

Orientador: Ricardo da Silva Torres

Agenda

- Motivações
- Serviço de Recomendação
 - Técnicas colaborativas
 - Técnicas baseadas em conteúdo
 - Técnicas híbridas
- Plataforma RecS-DL
 - Objetivos e Contribuições
 - Arquitetura Geral
 - Serviço de Recomendação
 - Serviço de Configuração
 - Protótipo e Ferramentas

Motivações

- Crescimento de Bibliotecas Digitais
 - Quantidade
 - Abrangência de Domínios
 - Conteúdo
- Necessidades decorrentes do crescimento
 - Adaptação ao usuário
 - Ferramentas para auxílio à escolha de itens e atualização diante de novidades
 - *Serviços personalizados*

Motivações

- Tradicionalmente:
 - softwares não distribuídos
 - complexidade de gerenciamento representa grandes dificuldades
- Tendência:
 - divisão modular em conjuntos de serviços de menor complexidade

Motivações

- Sistemas de Recomendação para Bibliotecas Digitais:
 - Grande maioria dos projetos são voltados para domínios específicos
 - Pouco adaptáveis diante de bibliotecas digitais distintas
 - Restritos em relação a técnicas de recomendação
 - Não descritos segundo metodologias formais

Serviços de Recomendação

- Serviços de Recomendação:

Dada uma grande coleção de objetos disponíveis, sugerir um subconjunto dessa coleção para um determinado usuário, de acordo com seus interesses

- Elementos Importantes:

- Dados de histórico: informações que o sistema dispõe antes do início do processo de recomendação
- Dados de entrada: informação fornecida no momento do pedido de recomendação
- Algoritmo: combina os dados de entrada e de histórico produzindo recomendações

Técnicas de Recomendação

- Prover sugestões personalizadas de acordo com históricos de comportamento
- Redução da sobrecarga de informação
- Técnicas
 - *Collaborative filtering*
 - *Content-based filtering*
 - Técnicas Híbridas

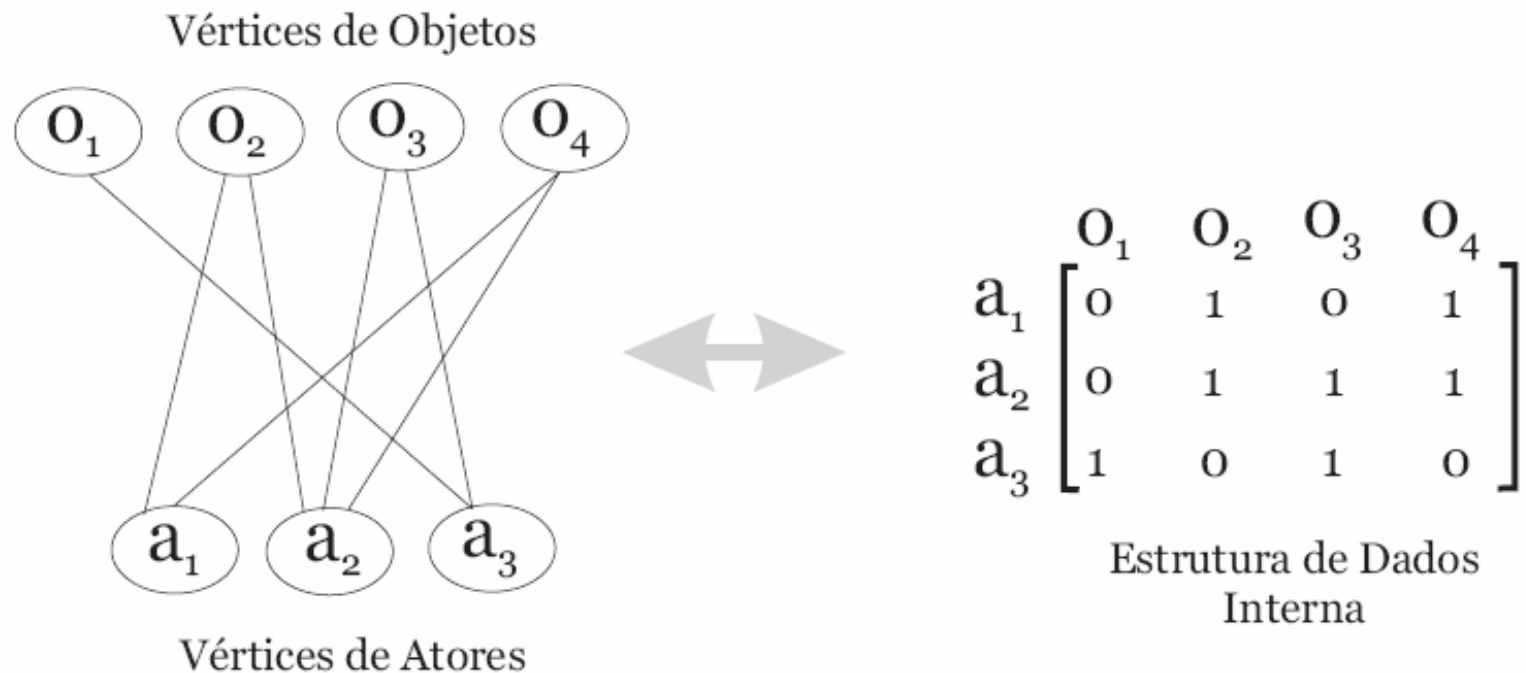
Content-Based Filtering

- Baseado na similaridade de conteúdo entre de itens já “experimentados” pelo usuário.
- Algoritmo mais comum: TF-IDF
- Limitações
 - Conteúdo de dados não textuais é difícil de ser analisado

Collaborative Filtering

- Baseado na colaboração entre usuários
- Semelhança entre perfis e preferências de escolha
- Limitações:
 - Problema da esparsidade
 - Problema do primeiro item

Collaborative Filtering



Técnicas Híbridas

- **Weighted:** as pontuações obtidas de diferentes técnicas são ponderadas e combinadas visando produzir uma única recomendação
- **Switching:** dependendo da situação o sistema escolhe a técnica a ser utilizada
- **Mixed:** recomendações obtidas de diferentes técnicas são apresentadas simultaneamente

Técnicas Híbridas

- **Cascade:** uma técnica refina as recomendações dadas por outra
- **Feature Combination:** características de diferentes fontes de recomendação são implementadas em conjunto, em um único algoritmo de recomendação
- **Meta-level:** o modelo aprendido por uma técnica de recomendação é usado como entrada para outra

Metodologias de Avaliação

- Principais características:
 - Identificação de Metas
 - Seleção de um conjunto de dados
 - Métricas de Precisão
- Experimentos on-line e off-line
- Geralmente:
 - Tentativa de mensurar capacidade do algoritmo em prever escolhas dos usuários

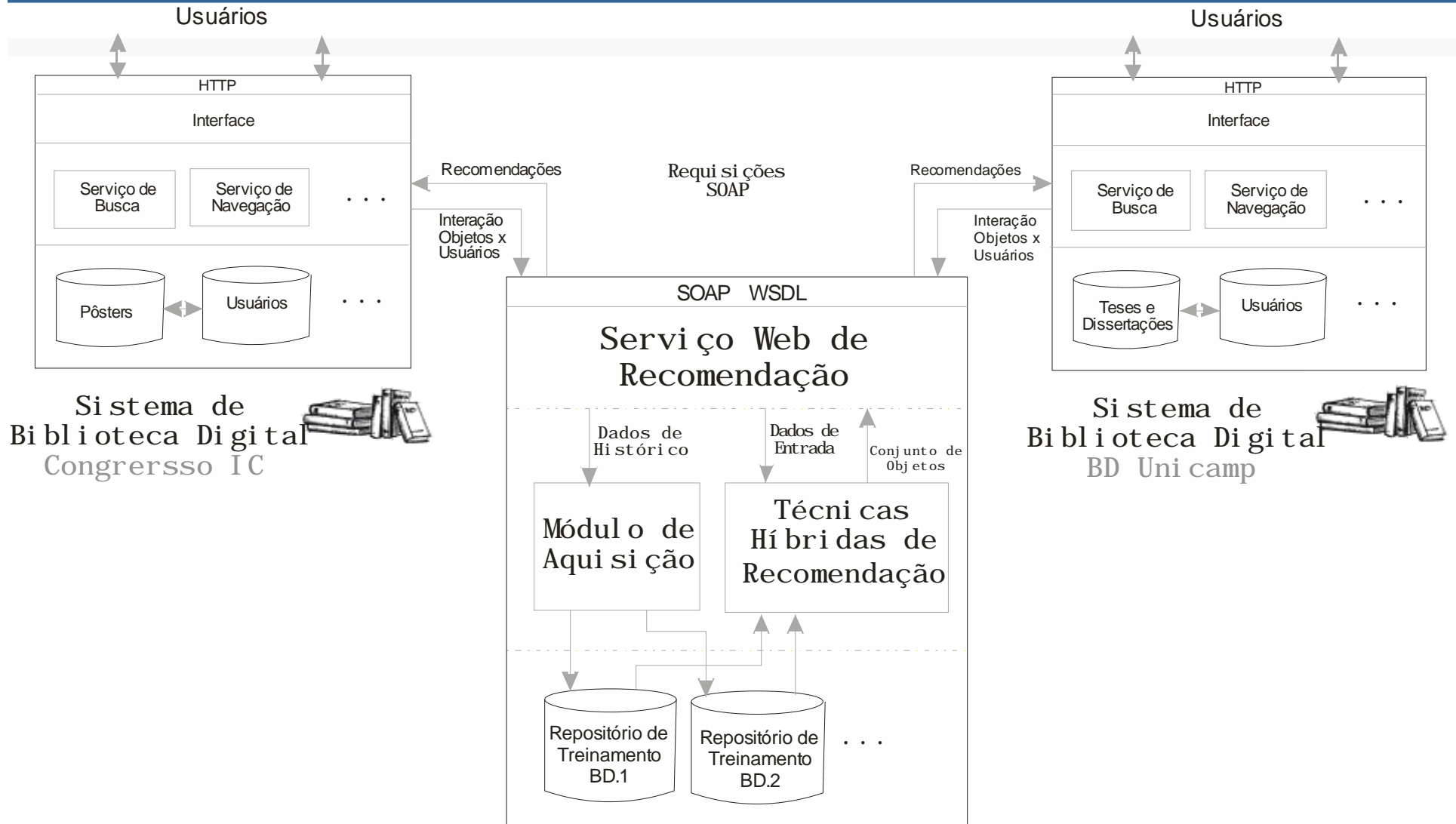
Plataforma RecS-DL

- Objetivos e Contribuições
- Arquitetura Geral
- Serviço de Recomendação
- Serviço de Configuração
- Protótipo
- Ferramentas

Objetivos e Contribuições

- Especificação e Implementação de uma Plataforma de Serviços de Recomendação:
 - Passível de ser utilizado por diferentes sistemas de bibliotecas digitais
 - Independência:
 - Domínio de aplicação
 - Tecnologia
 - Técnicas de Recomendação
 - Acessível através de padrões de mercado
 - Descrito segundo arcabouços formais

Arquitetura Geral



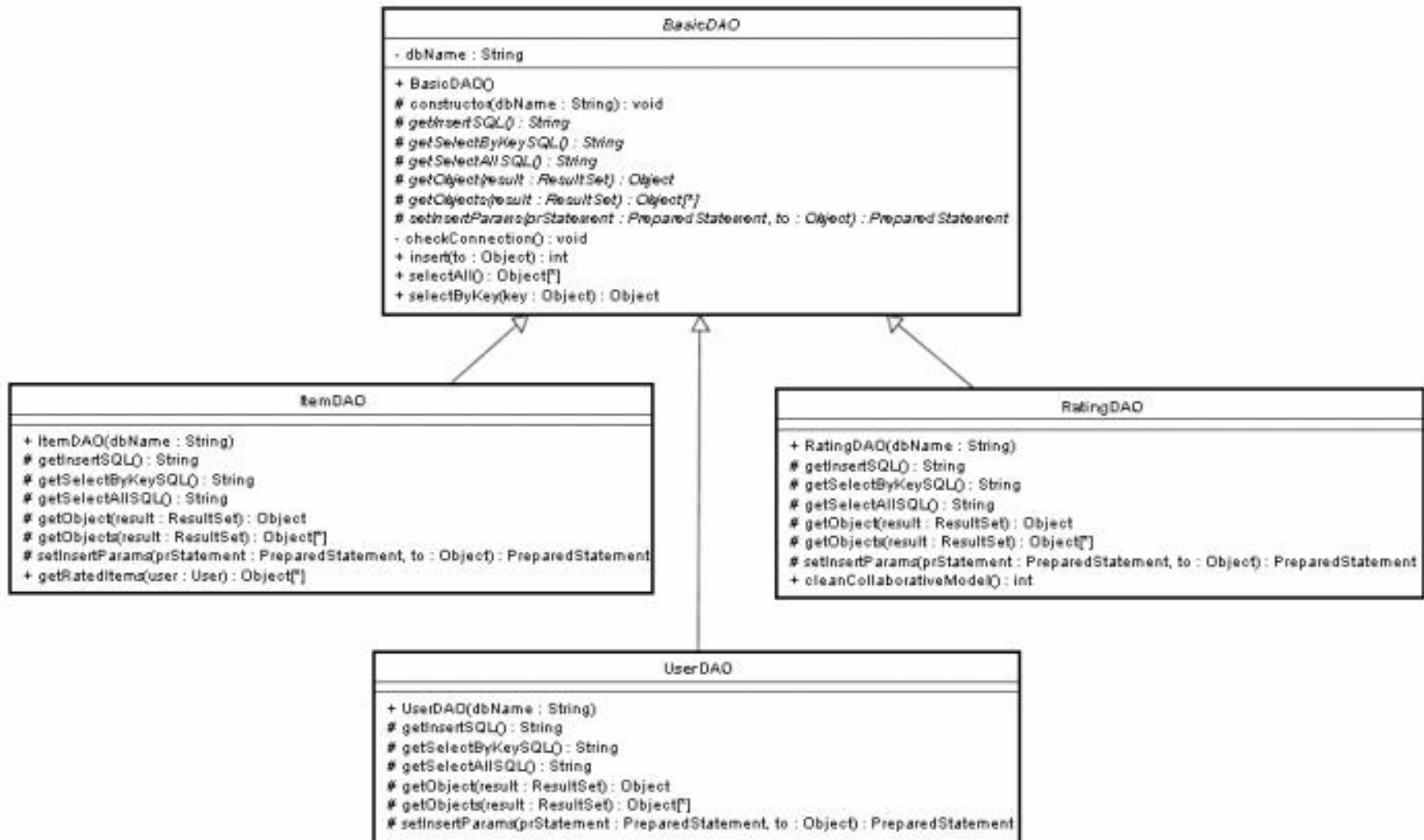
Serviços Web

- Sistema de software projetado para prover interoperabilidade entre máquinas em uma rede.
- Composição de sistemas através de serviços:
 - Arquitetura Orientada a Serviços (SOA)
- Aplicação em Bibliotecas Digitais

Serviço de Recomendação

- **Aquisição de dados:**
 - Inserção
 - Importação OAI
- **Recomendação:**
 - Atualização de modelos.
 - Recomendação por Item
 - Recomendação por Usuário

Módulo de Aquisição



Módulo OAI

Biblioteca Digital ativa: bibdig1

Bib. Digital

Ativar

Recomendação

Recomendação por Item:

Item Id:

Nenhum engine ativo!

Recomendação por Usuário:

User Id:

Nenhum engine ativo!

Módulo de Aquisição

Consultar Item:

Item Id:

Consultar Usuário:

Usuário Id:

- [Inserir Item](#)
- [Inserir Usuário](#)
- [Inserir Rating](#)

- [Listar todos os Itens](#)
- [Listar todos os Usuários](#)

Metadados OAI

OAI: Importar Registro

URL:

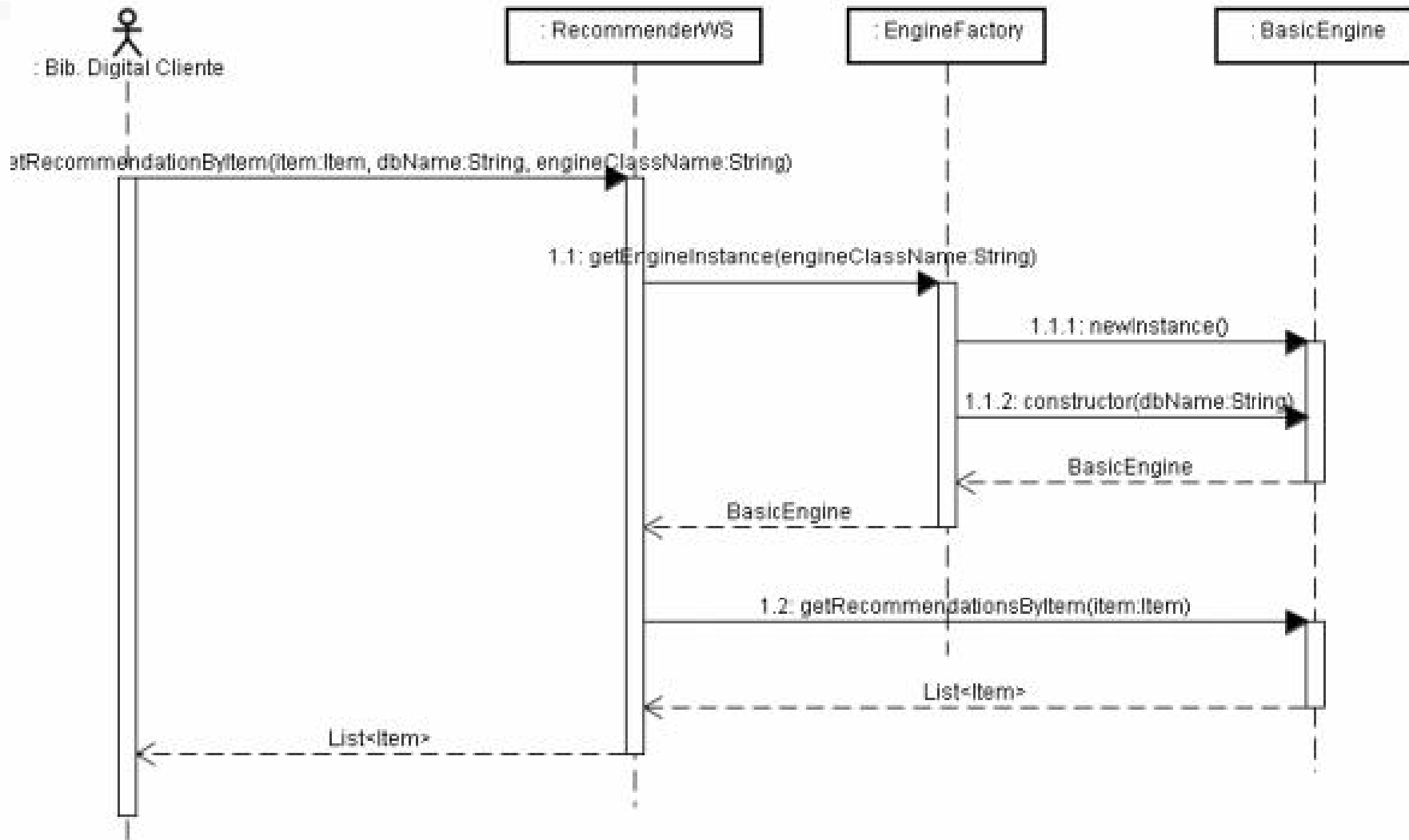
Identificador:

OAI: Importar Lista de Registros

URL:



Recomendação



Protótipo: Recomendação por Item

Consulta Item

Item:

Item Id: 29

Identificador: oai_dc:29

Título: Batman Forever (1995)

Autor:

Descrição:

The Dark Knight of Gotham City confronts a dastardly duo: Two-Face and the Riddler. Formerly District Attorney Harvey Dent, Two-Face incorrectly believes Batman caused the courtroom accident which left him disfigured on one side; he has unleashed a reign of terror on the good people of Gotham. Edward Nygma, computer-genius and former employee of millionaire Bruce Wayne, is out to get the philanthropist; as The Riddler he perfects a device for draining information from all the brains in Gotham, including Bruce Wayne's knowledge of his other identity. Batman/Wayne is/are the love focus of Dr. Chase Meridan. Former circus acrobat Dick Grayson, his family killed by Two-Face, becomes Wayne's ward and Batman's new partner Robin the Boy Wonder.

Assunto:

Relações:

Linguagem:



Recomendações:

Item Id	Identificador	Título	Autor	Detalhes
231	oai_dc:231	Batman Returns (1992)		Detalhes
403	oai_dc:403	Batman (1989)		Detalhes
254	oai_dc:254	Batman & Robin (1997)		Detalhes
1184	oai_dc:1184	Endless Summer 2, The (1994)		Detalhes
392	oai_dc:392	Man Without a Face, The (1993)		Detalhes
1393	oai_dc:1393	Stag (1997)		Detalhes

Recomendação por Item:

ICMultilensRecommender

ICLuceneRecommender

Item Id:

Protótipo: Recomendação por Usuário

Items em Ordem de Rating (UserId: 100)

Item Id	Identificador	Título	Autor	Detalhes
315	oai_dc:315	Apt Pupil (1998)		Detalhes
269	oai_dc:269	Full Monty, The (1997)		Detalhes
258	oai_dc:258	Contact (1997)		Detalhes
272	oai_dc:272	Good Will Hunting (1997)		Detalhes
294	oai_dc:294	Liar Liar (1997)		Detalhes
300	oai_dc:300	Air Force One (1997)		Detalhes
328	oai_dc:328	Conspiracy Theory (1997)		Detalhes
347	oai_dc:347	Wag the Dog (1997)		Detalhes
690	oai_dc:690	Seven Years in Tibet (1997)		Detalhes
751	oai_dc:751	Tomorrow Never Dies (1997)		Detalhes
879	oai_dc:879	Peacemaker, The (1997)		Detalhes
898	oai_dc:898	Postman, The (1997)		Detalhes
900	oai_dc:900	Kundun (1997)		Detalhes
1235	oai_dc:1235	Big Bang Theory, The (1994)		Detalhes
349	oai_dc:349	Hard Rain (1998)		Detalhes
678	oai_dc:678	Volcano (1997)		Detalhes
1007	oai_dc:1007	Twisted (1997)		Detalhes



Recomendação por Usuário:

- ICMultilensRecommender
 ICLuceneRecommender

User Id:

Recomendações:

Item Id	Identificador	Título	Autor	Detalhes
28	oai_dc:28	Apollo 13 (1995)		Detalhes
126	oai_dc:126	Spitfire Grill, The (1996)		Detalhes
597	oai_dc:597	Eraser (1996)		Detalhes
1226	oai_dc:1226	Night Falls on Manhattan (1997)		Detalhes
301	oai_dc:301	In & Out (1997)		Detalhes

Serviço de Configuração

- **Contas de Bibliotecas Digitais:** permite a criação ou configuração de contas de Bib. Digitais.
- **Instalação de Engines:** possibilita que um engine seja instalado por meio de um Serviço Web de configuração.
- **Ativação de Engines e Ajuste de Parâmetros:** esse método permite a ativação e configuração de engines para cada Bib. Digital

Configuração de Bibliotecas Digitais

Instalação & Configuração de Engines no Webservice

Instalar novo Engine:

Nome do Engine:

Arquivo de Biblioteca (.jar):

Descritor (.xml):

Engines Instalados:

ICHybridRecommender	Detalhes	Parâmetros
ICLuceneRecommender	Detalhes	Parâmetros
ICMultilensRecommender	Detalhes	Parâmetros

Configuração de Engines nas Bibliotecas Digitais

Ativar & Configurar Engine em Bib. Digital:

Nome do Biblioteca Digital:

Nome do Engine:

Configurar Engines em Bib. Digital:

Nome do Biblioteca Digital:

Criação e Configuração de Bibliotecas Digitais

Criar Biblioteca Digital: (utilizando BD padrão)

Nome da Biblioteca Digital:

Configurar Biblioteca Digital: (utilizando BD existente)

Nome da Biblioteca:

Driver de BD:

URL JDBC:

Usuário:

Senha:



Engines de Recomendação

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<engine>
  <!-- ===== -->
  <!-- Informacoes gerais do Engine de Recomendacao -->
  <!-- ===== -->
  <general xmlns:dc="http://purl.org/dc/elements/1.1/">
    <!-- Nome simplificado/Identificador do Engine -->
    <dc:identifier>ICLuceneRecommender</dc:identifier>

    <!-- Descricao do Engine -->
    <dc:description>Engine de Recomendacao baseado em conteudo que
      utiliza biblioteca Apache Lucene</dc:description>

    <!-- Tipo/Tecnica utilizada: COLLABORATIVE,CONTENTBASED, HYBRID -->
    <dc:type>CONTENTBASED</dc:type>

    <!-- Autor -->
    <dc:creator>Daniel Pedronette</dc:creator>

    <!-- Nome completo da classe de implementacao do Engine -->
    <class-name>br.unicamp.ic.recommender.lucene.ICLuceneRecommender</class-name>

    <!-- Versao -->
    <version>1.0</version>
  </general>
```

A

```
<!-- ===== -->
<!-- Parametros de configuracao do Engine -->
<!-- ===== -->
<parameters>
  <parameter>
    <!-- Identificador do parametro -->
    <name>INDEXES_LOCAL</name>

    <!-- Descricao da funcao do parametro -->
    <description>Local de armazenamento de indices (Lucene)</description>

    <!-- Valor padrao do parametro -->
    <default-value>LuceneIndex</default-value>

    <!-- Tipo: KEY_VALUE(parametro tipo chave-valor),DDL_DB(representa uma
      configuracao de Banco de Dados,SYSTEM_FILE(representa um caminho
      no sistema de arquivos -->
    <type>SYSTEM_FILE</type>
  </parameter>

  <parameter>
    <name>MAX_SIZE_CONTENTBASED_LIST</name>
```

B

ICLuceneRecommender



ICLuceneRecommender.xml
Documento XML



ICLuceneRecommender.jar
Executable Jar File

C

```
public abstract class BasicEngine {
  (...)

  public abstract void makeModel();

  public abstract List getRecommendationsByItem
    (br.unicamp.ic.recommender.entity.Item item);

  public abstract List getRecommendationsByUser
    (br.unicamp.ic.recommender.entity.User user);
}
```

Configuração de Engines

Biblioteca Digital: bibdig1; **Engine:** ICLuceneRecommender

Parâmetros

Parâmetro: INDEXES_LOCAL
Descrição:
Tipo: SYSTEM_FILE
Valor:

Parâmetro: MAX_ITEMS_RATED_BY_USER
Descrição:
Tipo: KEY_VALUE
Valor:

Parâmetro: MAX_SIZE_CONTENTBASED_LIST
Descrição:
Tipo: KEY_VALUE
Valor:

Parâmetro: MIN_DOCUMENT_FREQ_CONTENT_BY_ITEM
Descrição:
Tipo: KEY_VALUE
Valor:

...

Ferramentas utilizadas

WebService

RecommenderWS
RecommenderWSConfig

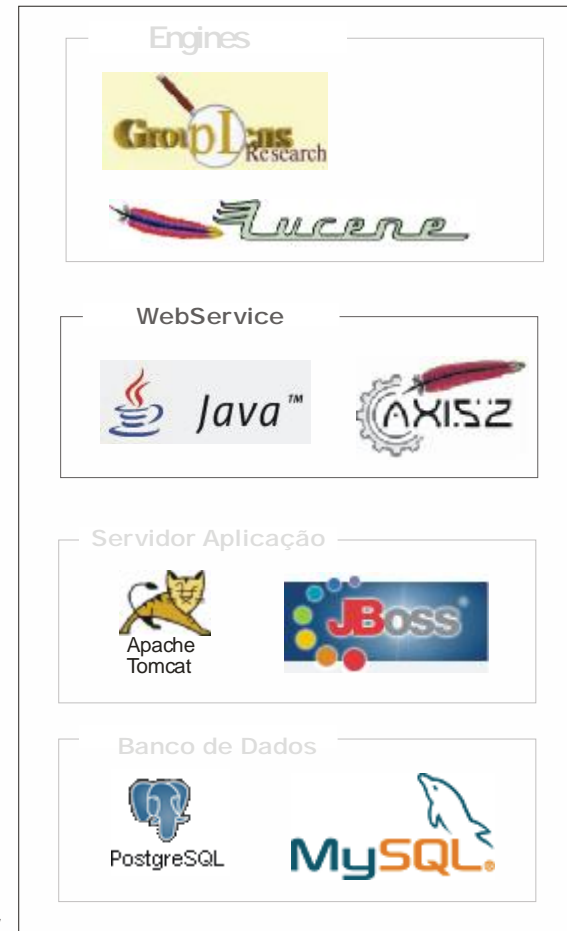
Cliente

RecommenderWSProxy
RecommenderWSConfigProxy

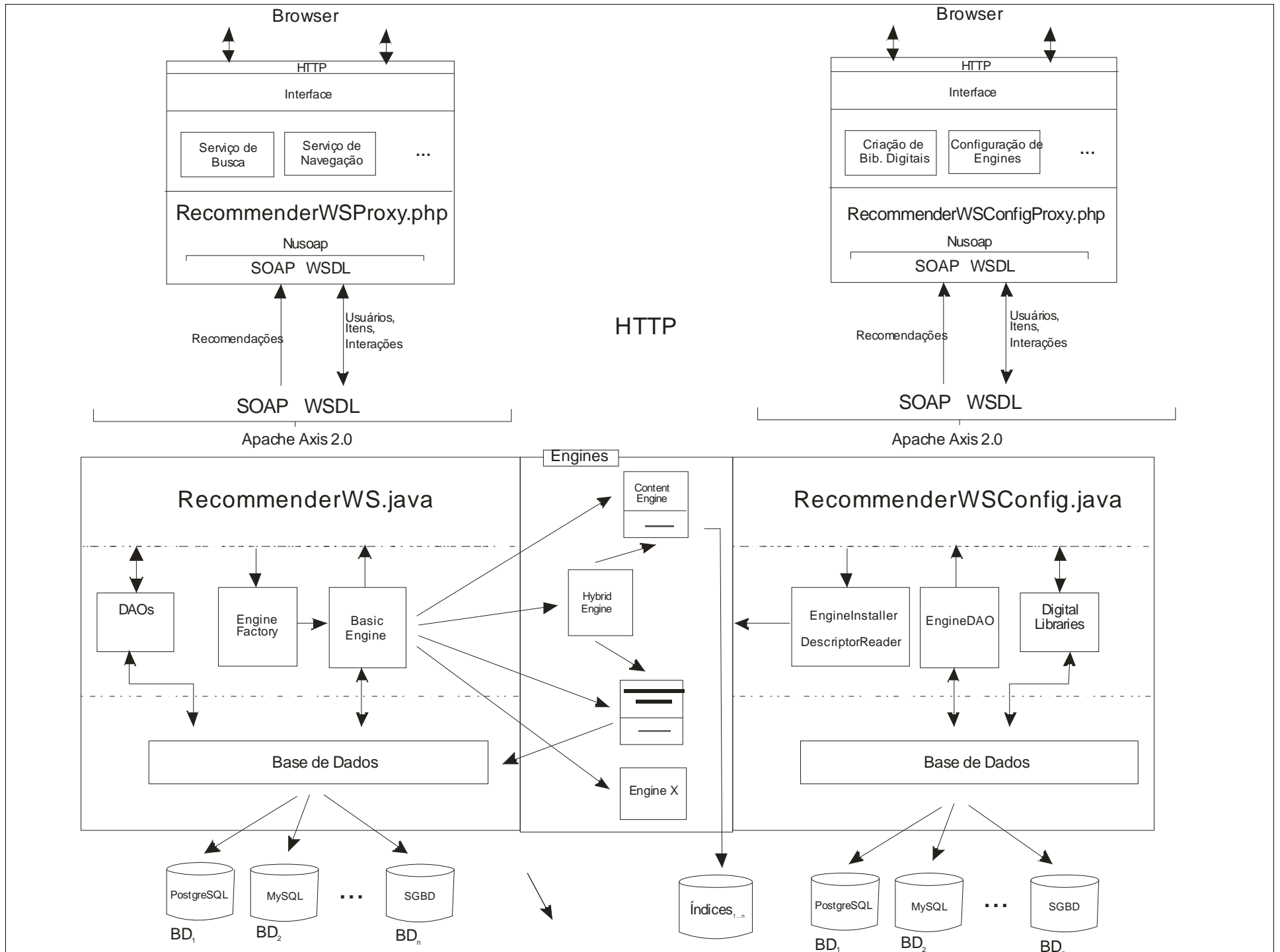


Apache 2.2.3, PHP 5.2, Nusoap 0.6.5

↔
HTTP/SOAP



Apache Lucene 2.0.0
Apache Axis 2.0
JBoss 4.0.5 / Tomcat 5.5.7
PostgreSQL 8.2 / MySQL 5.0.27



Dúvidas

- Dúvidas?