

# Uma plataforma de serviços de recomendação para bibliotecas digitais

Daniel Carlos Guimarães Pedronette<sup>1</sup>, Ricardo da Silva Torres<sup>1</sup>

<sup>1</sup>Instituto de Computação – Caixa Postal 6176  
Universidade Estadual de Campinas – UNICAMP  
CEP 13083-970, Campinas, SP, Brasil

daniel.pedronette@students.ic.unicamp.br, rtorres@ic.unicamp.br

**Abstract.** *This paper presents a platform for recommendation services, called RecS-DL, to support the use of recommendation tools by digital library applications. The proposed RecS-DL platform is independent of application domain, technology, and recommendation techniques. The recommendation services offered by the platform can be easily incorporated into digital libraries systems. Furthermore, new recommendation engines can also be plugged into the platform in a dynamic way. We present the results obtained from experiments conducted with real digital libraries and from evaluations made by potential users. Experimental results show that the platform facilitates the interoperability of recommendation tools in digital libraries systems.*

**Resumo.** *Este artigo apresenta uma plataforma de serviços de recomendação, chamada RecS-DL, visando ampliar as possibilidades de utilização das ferramentas de recomendação por bibliotecas digitais. A Plataforma RecS-DL apresentada é independente de domínio de aplicação, de tecnologias e técnicas de recomendação. O serviço de recomendação oferecido pode ser facilmente agregado às bibliotecas digitais clientes, assim como novos mecanismos de recomendação podem ser acoplados à plataforma de maneira dinâmica. São apresentados também os resultados de experimentos conduzidos, envolvendo bibliotecas digitais reais e avaliações por potenciais usuários. Resultados experimentais mostram que a plataforma facilita a interoperabilidade de ferramentas de recomendação em bibliotecas digitais.*

## 1. Introdução

Uma biblioteca digital pode ser definida, de maneira bastante ampla, como uma coleção organizada de objetos digitais associada a um conjunto de serviços acessíveis em ambientes distribuídos, visando atender às necessidades de comunidades de usuários [Gonçalves 2004]. Há anos as bibliotecas digitais apresentam um crescimento vertiginoso, seja em quantidade de informação, seja em abrangência de domínios de utilização.

Neste cenário, a localização e escolha do conteúdo desejado, assim como a tarefa de manter-se atualizado diante de grandes volumes de informação tornam-se um grande desafio. Dessa forma, o processo de recomendação, já bastante difundido em relações sociais humanas, passa a ter um importante papel nas bibliotecas digitais.

De maneira geral, uma recomendação consiste em, a partir de uma grande coleção de itens disponíveis, sugerir um conjunto desses itens para um usuário de acordo com seus interesses. Atualmente, os sistemas de recomendação estão presentes nos mais diversos segmentos, tais como recomendação de vídeos, músicas, livros, entre outros, havendo diversas ferramentas de recomendação descritas na literatura e em utilização em diversas aplicações de mercado. Entretanto, vários aspectos estruturais das ferramentas existentes geralmente limitam sua aplicação e reuso. Entre esses aspectos podem-se destacar:

- **Domínio de aplicação:** as ferramentas são projetadas tendo em vista uma Biblioteca Digital ou domínio específico, impossibilitando sua reutilização em outros cenários;

- **Técnicas de recomendação:** os pacotes disponíveis são, em sua maioria, sistemas monolíticos. Os serviços oferecidos não são extensíveis e geralmente baseiam-se em uma única técnica de recomendação;

- **Tecnologia:** a implementação das ferramentas é realizada em linguagem, plataforma ou interface específica, limitando o acesso de aplicações clientes.

Este artigo apresenta uma plataforma de serviços de recomendação, chamada RecS-DL, visando reduzir as limitações das ferramentas existentes e ampliar a utilização das ferramentas de recomendação por bibliotecas digitais. A abordagem proposta consiste em criar uma plataforma flexível o bastante para tornar-se independente de domínio de aplicação, de tecnologias utilizadas e extensível sob o ponto de vista de técnicas de recomendação. A contribuição central deste trabalho consiste na especificação, modelagem, implementação e validação experimental da Plataforma RecS-DL.

## 2. Trabalhos relacionados

### 2.1. Técnicas de recomendação

Os sistemas de recomendação vêm sendo bastante explorados na literatura desde o início da década de 90 [Goldberg et al. 1992]. As técnicas existentes podem ser divididas em duas grandes categorias: filtragem colaborativa e filtragem baseada em conteúdo.

A filtragem colaborativa baseia-se principalmente no compartilhamento de experiências entre indivíduos que apresentam interesses semelhantes. Assim, as ações e análises de um usuário, considerando uma porção particular de informação, são armazenadas para benefício de toda uma comunidade [Herlocker 2000]. As experiências podem ser registradas por meio de avaliações, histórico de comportamento, entre outros.

A filtragem baseada em conteúdo é caracterizada pela recomendação de objetos, baseando-se na correlação entre o conteúdo de itens e preferências dos usuários em relação a estes itens. As técnicas baseadas em conteúdo apresentam uma importante vantagem em relação às técnicas colaborativas: não apresentam o problema de esparsidade, ou seja, não necessitam de um grande histórico de avaliações para um bom desempenho. Entretanto, sistemas desse tipo apresentam limitações, dado que o conteúdo de dados não textuais é difícil de ser analisado (imagens, vídeos) e o entendimento do conteúdo do texto pode ser prejudicado devido ao uso de sinônimos ou termos muito específicos.

A filtragem colaborativa e baseada em conteúdo são complementares: uma linha de pesquisa comum em sistemas de recomendação converge para métodos e algoritmos que sejam capazes de combinar as recomendações de técnicas diferentes.

## 2.2. Infra-estrutura das ferramentas de recomendação

A difusão das diversas técnicas de recomendação originou outros desafios de pesquisa: como proporcionar que essas técnicas sejam utilizadas de forma comum por diversas aplicações clientes, implementadas em tecnologias e para domínios diferentes, como torná-los configuráveis a cada aplicação e acessíveis por meio de *interfaces* padronizadas.

Alguns pacotes e ferramentas têm sido propostos visando prover acesso às técnicas de recomendação. As ferramentas CoFE - Collaborative Filtering Engine [CoFE 2004] e Multilens [Miller 2003], por exemplo, podem ser utilizadas e acopladas às aplicações que necessitem de técnicas de recomendação. São independentes de domínio de aplicação e utilizam banco de dados relacionais para armazenamento dos repositórios de treinamento. Possibilitam a extensão de mecanismos de recomendação, todavia, baseando-se apenas no modelo colaborativo. Não permitem a utilização de outras técnicas de recomendação e têm limitações quanto à *interface* de acesso, baseada exclusivamente na tecnologia Java.

O pacote Duine [van Setten et al. 2002] consiste em uma ferramenta que suporta várias técnicas de recomendação e configuração de parâmetros para os algoritmos disponíveis. A ferramenta também é independente de domínio e possibilita o armazenamento das informações em banco de dados. Todavia, a ferramenta apresenta limitações relacionadas ao uso de linguagem específica pelas aplicações clientes. Já a abordagem apresentada neste artigo apresenta soluções que a tornam independente, tanto de técnicas de recomendações, quanto da linguagem da aplicação cliente.

O projeto EasyUtil [EasyUtil 2006] oferece um serviço de recomendação colaborativo operando sobre o protocolo HTTP. As requisições são realizadas por meio de parâmetros enviados a uma determinada URL e os resultados são codificados em formato XML. A ferramenta suporta aplicações em linguagens e domínios diversos. Entretanto, apresenta desvantagens, dado que se limita às técnicas colaborativas e não oferecem interfaces formalmente definidas, baseando-se apenas na URL do serviço.

O pacote Taste [Owen 2005] consiste em uma ferramenta de recomendação baseada no modelo colaborativo. A ferramenta pode ser acoplada a aplicações clientes ou pode ser acessada a partir de interface de Serviços Web. Embora portátil tecnologicamente, a ferramenta limita-se às técnicas colaborativas de recomendação.

O projeto *The MobLife Recommender* [Petteri Nurmi 2006] também oferece um serviço de recomendação baseado em Serviços Web e interfaces formalmente definidas em WSDL. A utilização de Serviços Web possibilita boa portabilidade, todavia, a ferramenta também não permite a utilização de outras técnicas de recomendação.

A Plataforma RecS-DL proposta neste artigo unifica as vantagens das ferramentas existentes, integrando Serviços Web e flexibilidade quanto ao uso de diversas técnicas de recomendação.

## 3. Plataforma RecS-DL

### 3.1. Problemas considerados

A definição da arquitetura da plataforma foi realizada tendo como principal objetivo o atendimento dos requisitos de flexibilidade propostos. O atendimento desses requisitos,

por sua vez, resultou em problemas a serem considerados durante o projeto da plataforma. Os principais desafios enfrentados e as soluções propostas são listados a seguir:

- **Como possibilitar a utilização de várias técnicas e ferramentas de recomendação utilizando uma mesma interface de acesso?** A solução proposta consistiu na definição de *interfaces* comuns e na criação de *engines* de recomendação, detalhados na Seção 3.3.2. Os *engines* podem ser instalados de maneira uniforme na plataforma, possibilitando que diversas técnicas ou ferramentas de recomendação existentes possam ser encapsuladas.

- **Como garantir a independência de domínio de aplicação?** A solução proposta consistiu na utilização do padrão Dublin Core [Dublin Core Metadata Initiative 2008] na modelagem dos repositórios de dados mantidos pela Plataforma RecS-DL. Dado que a plataforma permite também o armazenamento de dados binários, um *item* da biblioteca digital pode ser um livro, uma tese, um filme, uma música. A recomendação por conteúdo para cada tipo de informação pode ser realizada por meio de *engines* específicos para cada formato.

- **Qual tecnologia utilizar para garantir que os serviços de recomendação sejam portáteis e acessíveis entre ambientes tecnológicos distintos?** A plataforma utilizou Serviços Web para suas *interfaces* permitindo que os serviços oferecidos possam ser utilizadas por *softwares* implementados em linguagens e sistemas operacionais diversos.

- **Qual o conjunto de funcionalidades que deveria ser oferecido às bibliotecas digitais clientes?** A definição das funcionalidades de recomendação desejadas para a plataforma demandou uma análise das ferramentas e técnicas existentes. O fruto desse trabalho permitiu a definição das *interfaces* da Plataforma RecS-DL. Uma descrição formal da plataforma pode ser encontrada em [Pedronette 2008].

- **Qual a estrutura dos engines de recomendação?** Para que o conceito de *engine* proposto se tornasse factível, foi necessário criar um padrão para a sua construção. Esse padrão é basicamente composto por diretrizes para implementação em tecnologia Java e na definição de um conjunto de metadados para descrição do *engine*.

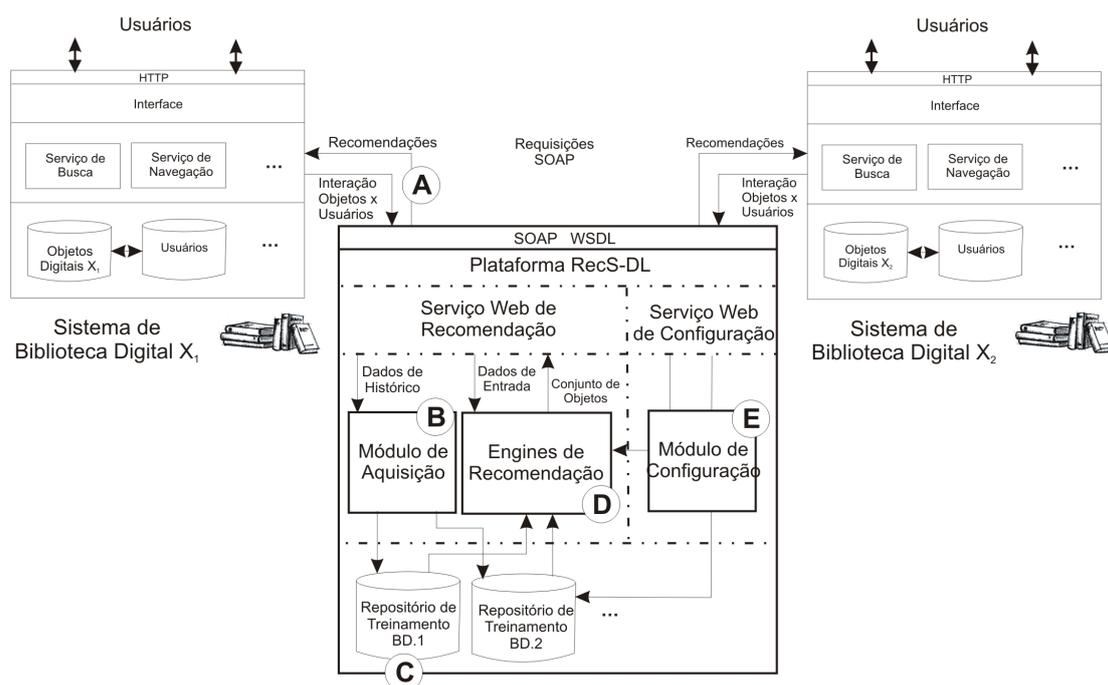
- **Como viabilizar o acesso da plataforma aos dados das bibliotecas digitais clientes?** A Plataforma RecS-DL necessita acessar os dados das bibliotecas digitais clientes como entrada para execução das técnicas de recomendação. Dada essa necessidade, foi definido o Módulo de Aquisição da plataforma, discutido em detalhes na Seção 3.3.1.

- **Como garantir flexibilidade para a criação e configuração de bibliotecas digitais na plataforma?** A plataforma pode ser utilizada por várias bibliotecas digitais clientes que podem apresentar características e configurações bastante distintas. De modo a garantir a flexibilidade da plataforma, foi criado um Serviço de Configuração, que será apresentado em detalhes na Seção 3.4.

## 3.2. Arquitetura da Plataforma RecS-DL

A Figura 1 ilustra a arquitetura geral da plataforma [Pedronette 2008].

São representadas na figura duas bibliotecas digitais realizando requisições de recomendação à Plataforma RecS-DL. Como é possível observar, essas aplicações estão isoladas da plataforma, mantendo características e tecnologias próprias e comunicando-se



**Figura 1. Arquitetura Geral da Plataforma RecS-DL.**

por meio de uma *interface* de requisições de serviços.

A *interface* entre a plataforma e as bibliotecas digitais clientes utiliza a tecnologia de Serviços Web, de forma que os métodos de acesso aos serviços da plataforma são definidos segundo o padrão WSDL e a chamada dos serviços por meio de requisições SOAP, ambos baseados em XML e utilizando o protocolo HTTP. A parte (A) da Figura 1 ilustra essa interação.

Haja vista que as aplicações das bibliotecas digitais clientes estão isoladas, a Plataforma RecS-DL deve prover mecanismos de acesso às informações dessas aplicações. Isso é realizado por meio do Módulo de Aquisição, ilustrado na parte (B) da Figura 1. As informações obtidas, por sua vez, são armazenadas em repositórios distintos para cada biblioteca digital cliente, como ilustrado na parte (C) da figura.

Utilizando as informações obtidas pelo Módulo de Aquisição, técnicas de recomendação podem ser aplicadas para oferecimento de recomendações às bibliotecas digitais clientes. Isso é realizado pelos *Engines* de Recomendação, ilustrados na parte (D) da Figura 1.

Tendo em vista a possibilidade de utilização da Plataforma RecS-DL por diversas bibliotecas digitais, um grande número de configurações podem ser personalizadas. Essas ações são realizadas no Módulo de Configuração, ilustrado na parte (E) da Figura 1. As funcionalidades oferecidas pela Plataforma RecS-DL são divididas em dois serviços: o Serviço de Recomendação e o Serviço de Configuração, apresentados a seguir.

### 3.3. Serviço de recomendação da Plataforma RecS-DL

Este serviço é responsável por todos os procedimentos compreendidos no processo de recomendação: a aquisição de dados pela plataforma, a construção de modelos de

recomendação e, por fim, a requisição de um conjunto de recomendações. É composto pelo Módulo de Aquisição e pelos *Engines* de Recomendação.

### 3.3.1. Módulo de aquisição

Dado que as técnicas de recomendação baseiam-se principalmente em dados de histórico de comportamento dos usuários, a plataforma dispõe de um módulo de aquisição de dados para treinamento. O objetivo desse módulo é prover mecanismos para acesso aos dados das bibliotecas digitais clientes.

As informações gerenciadas pelo Módulo de Aquisição são armazenadas em um SGBD, para posterior utilização pelos *engines* e respectivas técnicas de recomendação. O conjunto de dados gerenciado pela plataforma é armazenado da mesma maneira, mas em um repositório distinto para cada biblioteca digital cliente da plataforma. Dessa forma, são gerenciados dados de Usuários, Itens, *Ratings*, *Engines* e parâmetros de configuração de cada biblioteca digital.

O Módulo de Aquisição oferece três maneiras diferentes para importação de dados, com o objetivo de atender necessidades diversas de várias bibliotecas digitais:

- **Serviços Web:** são disponibilizados vários métodos de Serviços Web para inclusão e consulta das principais entidades da plataforma (Usuário, Item e *Rating*).
- **Visões:** quando o repositório da biblioteca digital cliente for implementado utilizando um banco de dados relacional, a Plataforma RecS-DL pode ser configurada para acessar as informações diretamente no banco de dados da biblioteca digital por meio de visões.
- **OAI:** é possível realizar a importação de metadados por meio do protocolo *Open Archives Initiative* (OAI) [Lagoze and de Sompel 2001] que permite o compartilhamento de metadados entre bibliotecas digitais.

### 3.3.2. Engines de recomendação

Os *engines* consistem em componentes que implementam técnicas de recomendação ou encapsulam ferramentas de recomendação de terceiros, desenvolvidos segundo padrões e *interfaces* definidas pela Plataforma RecS-DL. O funcionamento desses elementos assemelha-se ao conceito de *plugin*, já bastante difundido em computação, nos quais módulos de software podem ser desenvolvidos e facilmente instalados sobre uma aplicação principal.

Os *engines* de recomendação realizam operações como: (i) acesso às informações das bibliotecas digitais clientes utilizando o Módulo de Aquisição; (ii) consulta aos parâmetros de configuração; (iii) recuperação de conexões para acesso ao repositório de treinamento; (iv) criação de lista de recomendações.

Os *engines* de recomendação são compostos basicamente por dois elementos [Pedronette 2008]: um pacote de implementação utilizando tecnologia Java e um conjunto de metadados XML baseados no padrão Dublin Core. A estrutura para implementação de *engines* definida pela plataforma concentra-se na implementação de

uma classe principal. Essa classe deve estender a classe abstrata *BasicEngine*, definida pela plataforma, implementando os métodos de acesso às técnicas de recomendação. O conjunto de metadados consiste em arquivos XML contendo as principais informações de criação e configuração dos *engines* divididas em dois grandes grupos: Informações Gerais e Parâmetros de Configuração.

### 3.4. Serviço de configuração da Plataforma RecS-DL

Dada a grande versatilidade do conceito de *engines*, verificou-se a necessidade da disponibilização de serviços de configuração da plataforma. Dessa forma, foi criado um Serviço Web complementar, contendo métodos específicos de configuração. Este serviço é de vital importância para a flexibilidade da plataforma, garantindo que ela possa ser adaptada e configurada de acordo com as necessidades de cada biblioteca digital.

Os seguintes métodos são oferecidos pelo Serviço de Configuração:

- **Instalação de *engines*:** o método *installEngine* possibilita que um engine implementado seja instalado na plataforma. O método de instalação de *engines* recebe três parâmetros de entrada: a biblioteca de implementação, os metadados XML e o nome (identificador) do engine, que também consta no arquivo XML.
- **Criação de contas de bibliotecas digitais:** a Plataforma RecS-DL pode ser acessada/utilizada por várias bibliotecas digitais clientes de forma simultânea. Isso é permitido a partir do uso do conceito de contas para bibliotecas digitais clientes.
- **Ativação de *engines* e ajuste de parâmetros:** Dado que a plataforma pode operar com várias bibliotecas digitais distintas e com características particulares, é possível que um *engine* de recomendação apresente melhores resultados em uma determinada biblioteca do que em outra. Dessa forma, é relevante que seja possível escolher e configurar os *engines* de recomendação que serão utilizados por cada biblioteca digital. Assim, a plataforma proposta permite a ativação de *engines* e ajuste de parâmetros.

## 4. Protótipo

Um protótipo foi construído visando implementar os requisitos e projeto previamente discutidos nas seções anteriores. O desenvolvimento do protótipo foi dividido em quatro etapas:

- **Plataforma RecS-DL:** implementação da Plataforma RecS-DL propriamente dita, dada pelos Serviços de Recomendação e Configuração;
- ***Engines* de Recomendação:** implementação de *engines* de recomendação para possibilidade de teste efetivo dos métodos da plataforma e da estrutura de *engines* proposta;
- **Aplicação Cliente:** construção de uma aplicação cliente com o objetivo de simular o comportamento de uma biblioteca digital cliente.
- **Ferramenta de instalação para a plataforma:** construção de uma aplicação responsável por instalar a Plataforma RecS-DL, a Aplicação Cliente e os softwares servidores necessários a execução dessas aplicações.

## 4.1. Engines de recomendação

O protótipo procurou validar a independência de técnicas de recomendação. Dessa forma, foram desenvolvidos e instalados na plataforma três *engines* de recomendação, cada um deles baseado numa técnica de recomendação distinta. São eles: ICMultilensRecommender: *engine* de recomendação baseado em técnicas colaborativas; ICLuceneRecommender: *engine* de recomendação que implementa técnicas baseadas em conteúdo e ICHybridRecommender: que implementa uma técnica híbrida de recomendação baseada em pesos para as técnicas combinadas. Mais detalhes sobre os engines implementados podem ser encontrados em [Pedronette 2008].

## 4.2. Aplicação cliente

Para a aplicação cliente foram construídas *interfaces* Web para todos os métodos dos serviços oferecidos pela plataforma, visando prover meios para testes efetivos de todos os métodos. Dessa forma, tornou-se possível a simulação de chamadas de bibliotecas digitais clientes à Plataforma RecS-DL. A Figura 2 ilustra a *interface* da aplicação cliente construída.

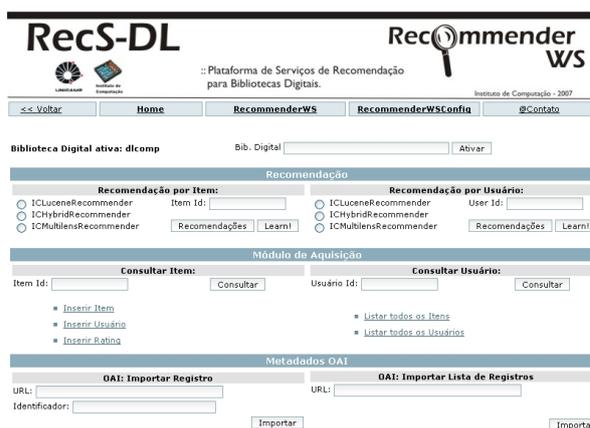


Figura 2. Aplicação Web cliente.



Figura 4. Ferramenta de controle dos softwares servidores.



Figura 3. Ferramenta de instalação.



Figura 5. Software Nou-Rau integrado à Plataforma RecS-DL.

## 4.3. Ferramenta de instalação

A Plataforma RecS-DL requer a instalação e configuração de vários *softwares* servidores para seu completo funcionamento. Outra dificuldade identificada consiste no controle

de execução desses *softwares*. A solução encontrada foi a implementação de uma ferramenta de instalação responsável por instalar e configurar todos os *softwares* necessários, configurar a Plataforma RecS-DL e a aplicação cliente e, por fim, criar uma ferramenta de controle de execução na máquina onde a plataforma foi instalada. A Figura 3 ilustra a ferramenta de instalação e a Figura 4 a ferramenta de controle de *softwares* servidores, ambas para ambiente Windows.

## 5. Validação

Esta seção tem como objetivo descrever os experimentos realizados e respectivos resultados obtidos. Foram realizados estudos de caso, descritos em [Pedronette 2008], e experimentos envolvendo potenciais usuários e uma aplicação real, a Biblioteca Digital da Unicamp, apresentados a seguir.

### 5.1. Biblioteca Digital da Unicamp

Os experimentos realizados nessa etapa consistem na incorporação de funcionalidades de recomendação à Biblioteca Digital da Unicamp utilizando a Plataforma RecS-DL. O objetivo consiste em utilizar todo o repositório e integrar a Plataforma RecS-DL ao *software* Nou-Rau, que gerencia a Biblioteca Digital da Unicamp.

A primeira etapa do experimento consistiu na criação de um ambiente de simulação da Biblioteca Digital da Unicamp. Foi utilizado um microcomputador Pentium IV, 3.0GHz, 2 GB de memória RAM DDR2 e HD de 300 GB, com sistema operacional Linux Debian. Esse microcomputador recebeu uma cópia fiel daquele que hospedava em ambiente de produção a Biblioteca Digital da Unicamp em Julho/2007. Dados aferidos em Maio/2007 indicam um total de 13.620 teses e dissertações armazenadas.

A próxima etapa do experimento consistiu na instalação e configuração da Plataforma RecS-DL e dos *softwares* que são pré-requisito para o seu funcionamento. Dada a instalação da plataforma, o último conjunto de procedimentos consistiu na integração da plataforma ao *software* Nou-Rau que gerencia a Biblioteca Digital da Unicamp.

A Figura 5 ilustra uma tela do *software* Nou-Rau exibindo uma consulta a uma tese da Biblioteca Digital da Unicamp, na parte (A). A seta ilustra a integração com a Plataforma RecS-DL, na parte (B), responsável por gerar recomendações por conteúdo para o item consultado.

### 5.2. Experimentos com potenciais usuários

Os experimentos consistem em avaliações realizadas por potenciais usuários da Plataforma RecS-DL. Potenciais usuários seriam administradores ou desenvolvedores de bibliotecas digitais que tenham interesse em utilizar técnicas de recomendação. Os usuários participantes dos experimentos consistem em estudantes de graduação e pós-graduação do Instituto de Computação da Unicamp, cursando a disciplina de Bibliotecas Digitais do segundo semestre de 2007. Participaram dos experimentos 11 estudantes, divididos em 5 duplas e um indivíduo. Estes estudantes foram convidados a participar dos experimentos pois criaram e configuraram bibliotecas digitais durante o curso. Sabe-se também que um dos estudantes é administrador de bibliotecas digitais reais.

Foi preenchido também um questionário visando identificar o perfil dos participantes, conhecimentos sobre tecnologias utilizadas e formação acadêmica. Verificou-se

que aproximadamente metade, totalizando 47% dos participantes, possuem nível superior. A outra metade divide-se entre alunos de graduação, com 27%, e mestrandos e doutorandos, com outros 26%. A área de formação dos participantes é predominantemente em Computação. Os participantes relataram conhecimentos medianos em Serviços Web, bons conhecimentos em Banco de Dados e pouco conhecimento em Serviços de Recomendação.

Os experimentos e avaliações realizadas buscaram avaliar a possibilidade de integração e utilização da Plataforma RecS-DL por bibliotecas digitais. Dessa forma, embora os *engines* de recomendação incluam a implementação de técnicas de recomendação, o objetivo desses experimentos com potenciais usuários não é avaliar as técnicas ou a qualidade das recomendações, mas sim verificar se a Plataforma RecS-DL atinge seu objetivo de aperfeiçoar a interoperabilidade dos sistemas de recomendação. A metodologia utilizada e atividades realizadas para efetivação dos experimentos foi composta por 5 etapas, descritas a seguir:

1. **Criação de uma biblioteca digital:** os participantes dos experimentos, divididos em grupos, executaram a tarefa de criar uma biblioteca digital para os Relatórios Técnicos do Instituto de Computação da Unicamp. Foram utilizados nessa tarefa *softwares* de grandes projetos mundiais da área: Greenstone, Fedora e D-Space. O Greenstone [Witten et al. 2005] é um *software* de biblioteca digital dos mais populares, desenvolvido pela University of Waikato e financiado pela UNESCO. O Fedora [Payette and Lagoze 1998] é mantido em esforço conjunto pela Cornell University Information Science e pela University of Virginia Library. O D-Space [Tansley et al. 2003] é fruto de uma colaboração entre MIT Libraries e Hewlett-Packard.

2. **Palestra sobre a Plataforma RecS-DL:** foram realizadas duas palestras de 35 a 40 minutos para os participantes. Foram apresentados conceitos de serviços de recomendação e as principais características da Plataforma RecS-DL.

3. **Disponibilização e leitura da documentação da plataforma:** um site contendo material de documentação Plataforma RecS-DL foi disponibilizado para consulta.

4. **Execução das tarefas estipuladas:** foram estipulados dois conjuntos de tarefas a serem executadas. As tarefas foram descritas em formulários apresentados aos participantes e executadas em grupos de duas pessoas.

5. **Preenchimento dos questionários:** para cada conjunto de tarefas executados, foram elaborados relatórios e preenchidos questionários pelos participantes. Os relatórios descreveram detalhes das atividades executadas. Os questionários buscaram avaliar a plataforma sob diversos aspectos a partir das tarefas executadas. Os questionários foram divididos em dois blocos:

- **Avaliação Quantitativa:** respondidas individualmente e direcionadas para respostas numeradas em escalas de 1 a 5 indicando grau de satisfação, complexidade, etc.
- **Avaliação Qualitativa:** respondida pelos grupos e composta de questões dissertativas abertas, permitindo ampla discussão a respeito dos funcionalidades da plataforma. Para interpretação dos resultados foram analisadas todas as repostadas e identificadas as recorrências de temas abordados.

Os resultados dos experimentos são apresentados nas seções seguintes.

### 5.2.1. Experimento I: Instalação da plataforma

O objetivo deste experimento é avaliar o procedimento de instalação da Plataforma RecS-DL sob dois aspectos: o primeiro deles consiste em verificar se a instalação é um procedimento passível de execução por usuários potenciais, sem a necessidade de conhecimentos prévios sobre a plataforma; o segundo aspecto consiste em avaliar se a ferramenta de instalação desenvolvida auxilia no procedimento de instalação. As tarefas descritas para execução do experimento consistem nos procedimentos necessários para que a Plataforma esteja instalada e efetivamente em funcionamento.

#### Questionário Quantitativo

A Figura 6 apresenta os resultados da avaliação do procedimento de instalação quanto à complexidade. Como já era previsto, dada a necessidade de instalação de vários softwares servidores, a avaliação da instalação teve os seguintes resultados: enquanto 37% dos participantes classificaram como “Simples”, 43% classificaram como “Complexo” ou “Muito Complexo”.

Outra hipótese a ser verificada era a de que a ferramenta de instalação desenvolvida poderia facilitar o procedimento de instalação. A Figura 7 apresenta um gráfico que exhibe resultados bastante positivos da avaliação da ferramenta de instalação. A avaliação foi realizada em relação à simplicidade na manipulação da ferramenta, 80% dos participantes classificaram como “Simples” ou “Muito Simples”. Tais resultados apresentam bons indícios de que o objetivo de simplificar o processo de instalação foi alcançado.

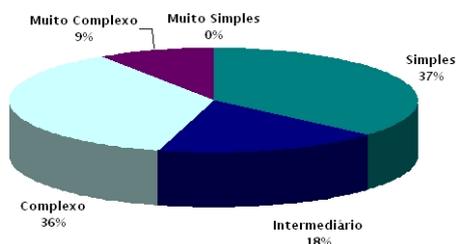
Além da simplicidade da ferramenta de instalação, pretendia-se avaliar também a efetividade da ferramenta. A Figura 8 apresenta os resultados da avaliação de eficácia, ou seja, se a ferramenta atende às necessidades. Embora os resultados gerais da avaliação sejam bons, com classificações positivas de aproximadamente 67% dos participantes, ainda há 9% e 27% dos participantes que avaliaram que a ferramenta “Não Atende” ou “Atende Parcialmente”. As avaliações negativas foram atribuídas a *bugs* de implementação.

Vale lembrar também que, independente dos resultados das avaliações, todos os participantes conseguiram executar completamente os procedimentos estipulados pelo experimento. Esse fato, por si só, já confirma a possibilidade de execução da instalação da plataforma por potenciais usuários.

#### Questionário Qualitativo

Vários participantes apontaram como aspecto positivo a simplicidade da instalação utilizando a ferramenta oferecida pela plataforma, integrando os *softwares* servidores em um único instalador e isolando o usuário de detalhes específicos dos servidores utilizados. Outros pontos positivos citados foram o tempo necessário para o procedimento de instalação, a ferramenta de controle dos servidores e em relação à documentação o Guia Rápido oferecido.

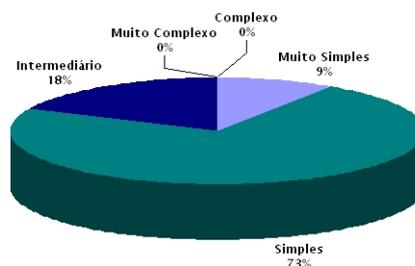
Alguns *bugs* de implementação foram identificados durante os experimentos. A ferramenta de instalação apresentava problemas para instalação em pastas cujos nomes eram extensos e não exibia um relatório de possíveis erros para o usuário ao término da



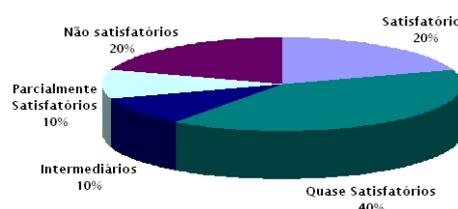
**Figura 6. Avaliação do procedimento geral de instalação.**



**Figura 8. Avaliação: A ferramenta de instalação atende às necessidades?**



**Figura 7. Avaliação da manipulação da ferramenta de instalação.**



**Figura 9. Avaliação dos resultados obtidos pelo Módulo de Aquisição.**

instalação. Tais problemas demandaram correções e geraram novas funcionalidades para as ferramentas de instalação e controle. As correções foram efetuadas e os participantes puderam experimentar uma nova versão nos experimentos seguintes.

Algumas possíveis melhorias também foram apontadas nas respostas dos questionários, indicando também possíveis trabalhos futuros. Entre os aspectos citados estão a necessidade de um desinstalador, e a portabilidade da ferramenta de instalação para ambiente Linux e a necessidade de maior documentação.

### 5.2.2. Experimento II: Aquisição de dados e utilização da plataforma

Este experimento tem como objetivo avaliar a Plataforma RecS-DL em relação à importação de dados e à integração a outros softwares de bibliotecas digitais. A primeira parte do experimento consiste em verificar a possibilidade da importação de dados de bibliotecas digitais pela Plataforma RecS-DL. A segunda parte do experimento tem como objetivo verificar a possibilidade de integração dessas bibliotecas à plataforma, de modo a oferecer serviços de recomendação.

#### Questionário Quantitativo

A primeira avaliação, de importação de dados, que é executada pelo Módulo de Aquisição da plataforma é apresentada na Figura 9. Os resultados são ligeiramente positivos, apresentando 60% das avaliações entre “Quase satisfatórios” e “Satisfatórios”. Entretanto, 40% das avaliações entre intermediário e negativo indicam possíveis problemas neste módulo. Uma provável causa do índice de insatisfação apresentado pelo Módulo de Aquisição consiste na complexidade relatada por alguns participantes nos procedimentos de importação. Embora todos os participantes tenham concluído completamente o procedimento de importação de dados, a avaliação quanto a complexidade foi predominantemente negativa. Tal avaliação pode ter sido causada pelo pouco conhecimento das

tecnologias utilizadas pelos participantes, como foi identificado no questionário de perfis.

A próxima avaliação refere-se ao procedimento de integração da Plataforma RecS-DL a outros *softwares* de bibliotecas digitais. Esse procedimento era o maior candidato a apresentar dificuldades, já que demanda esforços de criação ou alteração de aplicações capazes de interagir entre a Plataforma RecS-DL e o *software* escolhido.

Apesar das dificuldades, todos os participantes concluíram completamente as atividades dos experimentos. Tais resultados confirmam que a Plataforma RecS-DL pode oferecer um serviço de recomendação a ser incorporado em outros *softwares* de bibliotecas digitais. Os relatórios construídos pelos participantes apresentam telas onde os softwares utilizados foram integrados à Plataforma RecS-DL. A Figura 10 exhibe o *software* Greenstone, a Figura 11 o *software* Fedora, a Figura 12 o *software* DSpace, e por fim, a Figura 13 ilustra tela da Plataforma RecS-DL apresentando as recomendações em resposta à requisição do *software* DSpace.

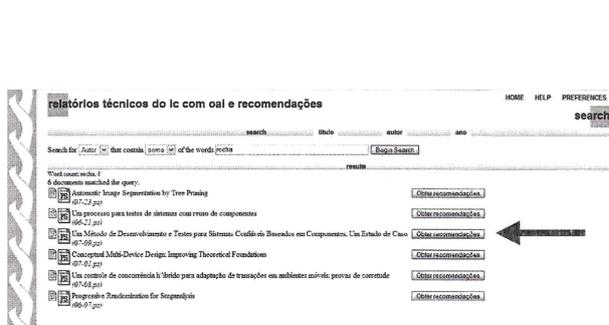


Figura 10. Integração com o *software* Greenstone.

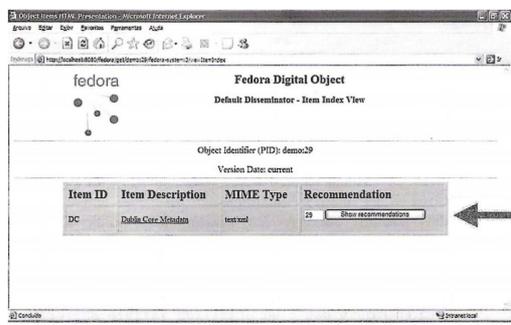


Figura 11. Integração com o *software* Fedora.

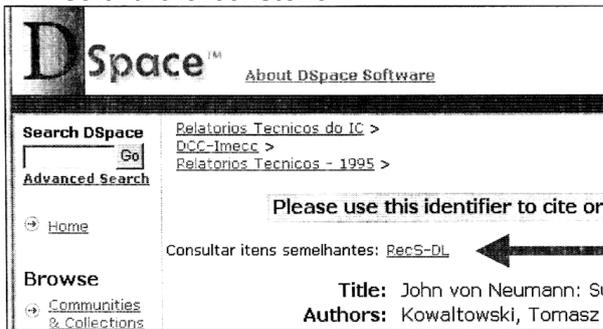


Figura 12. Integração com o *software* DSpace.

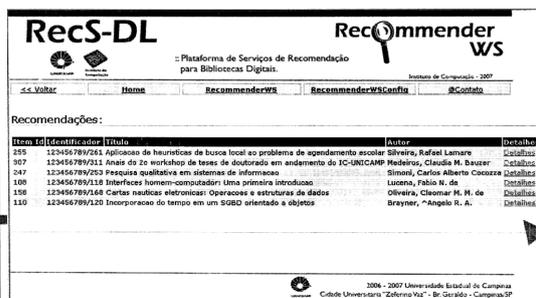
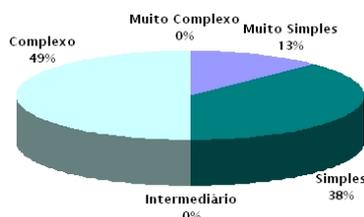


Figura 13. Plataforma RecS-DL integrada ao *software* DSpace.

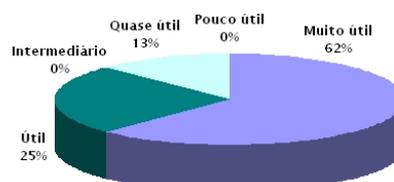
Foi realizada também pelos participantes uma avaliação quanto a complexidade do procedimento de integração, cujos resultados estão apresentados na Figura 14. Dadas as dificuldades previstas para os procedimentos, podem-se considerar os resultados positivos, já que 51% dos participantes classificaram as atividades de “Simples” ou “Muito Simples”. A variação significativa dos resultados pode ser provavelmente atribuída à utilização de vários *softwares* de bibliotecas digitais, criando cenários e níveis de dificuldades distintos.

Por fim, a última avaliação quantitativa realizada refere-se ao potencial de utilidade da Plataforma RecS-DL, como provedora de serviços de recomendação. Os parti-

participantes dos experimentos, após a instalação e utilização da plataforma realizaram uma avaliação do potencial de utilidade da plataforma. Os resultados, ilustrados na Figura 15, são muito positivos: 87% dos participantes classificaram a Plataforma RecS-DL como “Útil” ou “Muito útil”. A avaliação apresenta bons indícios de que as funcionalidades oferecidas estão adequadas.



**Figura 14. Avaliação do processo de integração da Plataforma RecS-DL.**



**Figura 15. Avaliação da utilidade da Plataforma RecS-DL.**

### Questionário Qualitativo

Todos os participantes concluíram as atividades propostas pelo experimento, configurando e integrando a Plataforma RecS-DL às bibliotecas digitais. Todavia, os procedimentos de importação de dados e integração a outros softwares de bibliotecas digitais tiveram mais pontos de dificuldades apontados do que o procedimento de instalação.

O Módulo de Aquisição de Dados foi o que concentrou as maiores dificuldades encontradas. Aspectos negativos apontados consistiram na necessidade de conhecimento do *software* e do modelo de dados do *software* gerenciador da biblioteca digital para a realização da integração.

Uma conclusão quase que consensual entre os participantes é de que a Plataforma RecS-DL representa uma ferramenta inovadora, de grande utilidade e funcional, haja vista a efetiva conclusão dos experimentos, mas que algumas melhorias ainda seriam desejáveis para o oferecimento da Plataforma RecS-DL como produto final.

## 6. Conclusões

Este artigo apresentou a Plataforma RecS-DL, que visa ampliar a utilização das ferramentas de recomendação combinando Serviços Web, independência de domínio de aplicação, de técnica de recomendação e de linguagem das aplicações clientes.

Experimentos com potenciais usuários e com uma biblioteca digital real foram conduzidos e os resultados obtidos demonstraram que as funcionalidades oferecidas pela Plataforma RecS-DL são úteis e efetivamente aplicáveis a bibliotecas digitais reais. Os experimentos comprovaram que os procedimentos de instalação, importação de dados e integração a outros softwares de bibliotecas digitais são factíveis por potenciais usuários sem experiência prévia com a plataforma.

Em relação às novas funcionalidades há várias possibilidades de trabalhos futuros e melhorias de módulos existentes, como a implementação de novos *engines*, utilizando outras técnicas de recomendação e métodos para inclusão de informações de *relevance feedback*. Questões relacionadas à segurança de acesso também constituem alvo de trabalhos futuros.

## 7. Agradecimentos

Os autores agradecem a FAPESP, CAPES, CNPQ, FAEPEX, Microsoft eScience and Tablet PC Technology and Higher Education Projects pelo apoio financeiro. Os autores também são gratos aos colaboradores da Biblioteca Central da Unicamp, CCUEC e DGA.

## Referências

- CoFE (2004). CoFE - Collaborative Filtering Engine. Disponível em <http://eecs.oregonstate.edu/iis/cofe/>. Acessado em 07 de maio de 2008.
- Dublin Core Metadata Initiative (2008). Dublin Core Metadata Initiative. Disponível em <http://dublincore.org>. Acessado em 07 de maio de 2008.
- EasyUtil (2006). EasyUtil Recommendation Service. Disponível em <http://easyutil.com/>. Acessado em 07 de maio de 2008.
- Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70.
- Gonçalves, M. A. (2004). *Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications*. PhD thesis, Virginia Polytechnic Institute and State University.
- Herlocker, J. L. (2000). *Understanding and improving automated collaborative filtering systems*. PhD thesis, University of Minnesota.
- Lagoze, C. and de Sompel, H. V. (2001). The open archives initiative: building a low-barrier interoperability framework. In *JCDL '01: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, pages 54–62, New York, NY, USA.
- Miller, B. N. (2003). *Toward a Personal Recommender System*. PhD thesis, University of Minnesota. Adviser-John Riedl, University of Minesota.
- Owen, S. (2005). Taste Documentation. Disponível online em <http://sourceforge.net/projects/taste/>. Acessado em 07 de maio de 2008.
- Payette, S. and Lagoze, C. (1998). Flexible and extensible digital object and repository architecture. In *Second European Conference on Research and Advanced Technology for Digital Libraries*, pages 21–31. Springer.
- Pedronette, D. C. G. (2008). Uma plataforma de serviços de recomendação para bibliotecas digitais. Master's thesis, Unicamp.
- Petteri Nurmi, Jukka Suomela, E. L. (2006). The MobiLife Recommender. Disponível em <http://www.cs.helsinki.fi/group/acs/mobilife/>. Acessado em 07 de maio de 2008.
- Tansley, R., Bass, M., Stuve, D., Branschovsky, M., Chudnov, D., McClellan, G., and Smith, M. (2003). The DSpace institutional digital repository system: Current functionality. In *JCDL '03: Proceedings of the 3th ACM/IEEE-CS joint conference on Digital libraries*, pages 87–97.
- van Setten, M., Veenstra, M., and Nijholt, A. (2002). Prediction strategies: Combining prediction techniques to optimize personalization. In *Personalization in Future TV'02 at the Adaptive Hypermedia 2002 conference*, pages 78–91, Malaga, Spain.
- Witten, H., Ian, Bainbridge, and David (2005). Creating digital library collections with greenstone. *Library Hi Tech*, 23(4):541–560.