

# Sistema de Recomendação Acadêmico para Apoio a Aprendizagem

#### Carla Duarte Barcellos - Centro Universitário LaSalle

carla.bo@gmail.com

#### Daniela Leal Musa - Centro Universitário LaSalle

dlmusa@unilasalle.edu.br

#### André Luiz Brandão – Instituto Tecnológico da Aeronáutica

abrandao@ita.br

## Mariusa Warpechowski – Universidade Federal do Rio Grande do Sul

mariusa@inf.ufrgs.br

Resumo. Os sistemas de recomendação surgiram com o objetivo de auxiliar os usuários no processo de busca na Web, indicando informações para o usuário com base no seu perfil. Esses sistemas são de grande valia para o processo de aprendizagem, pois professores e alunos podem realizar buscas por conteúdos educacionais mais apropriados ao seu perfil (nível de formação, áreas de interesse, etc.). Este artigo apresenta um modelo de referência para sistema de recomendação, que dispensa a intervenção do usuário no momento de traçar seu perfil e utiliza técnicas de mineração de dados. O modelo foi implementado no protótipo "Ponto com Filtro" para verificar sua aplicabilidade.

**Palavras-chaves:** Sistemas de Recomendação, Aprendizagem, Perfil de usuário

Abstract. Recommendation Systems arising from the objective to help users on the searching at the Web, suggesting information for the user based on her profile. These systems are important to the learning process, because the teachers and learners would search instructional material according her profile (interest areas, education level). This paper presents a reference model to recommendation systems. In this model is not necessary the user intervention and uses data-mining techniques. The model was implemented in "Ponto com Filtro" prototype to verify its applicability.

**Keywords:** Recommendation Systems, Learning, User Profile



# 1. Introdução

Os sistemas de busca na Web têm sido amplamente utilizados, principalmente para busca de recursos de aprendizagem. Esses sistemas, pelo fato de retornarem muitos dados, geram uma tarefa árdua para os usuários (alunos e professores), pois eles devem realizar uma filtragem nessas informações para encontrar os dados relevantes. Uma forma de auxiliar esse processo é o uso de sistemas de recomendação, que realizam indicações específicas para cada usuário, agilizando o processo de busca na Internet. Nesses sistemas, os usuários são auxiliados na procura de informações, recebendo indicação, por exemplo, de recursos didáticos, que são baseadas no seu perfil (interesses e preferências). Os dados sobre o perfil do usuário não são obtidos facilmente, pois uma das grandes dificuldades nesse processo é a aquisição desses dados. Duas formas de aquisição de dados são normalmente usadas: explícita e implícita. Adquirir dados de forma explícita consiste em coletá-los diretamente do usuário, por exemplo, através de questionários que devem ser respondidos e cujas questões estão relacionadas com os dados pessoais, preferências, metas, etc. Esse método não agrada os usuários, pois o preenchimento de formulários demanda muito tempo, bem como a inexistência de informações sobre a confiabilidade do sistema. Adquirir dados implicitamente resumese em coletar as operações realizadas pelo usuário durante interação com algum site. Técnicas como redes semânticas, bases de regras e lógica fuzzy podem ser aplicadas sobre esses dados de forma a descobrir as preferências do usuário.

O objetivo deste trabalho é integrar sistemas de busca e perfil do usuário para fornecer recomendações mais relevantes. Além disso, pretende-se fornecer uma solução para o problema de esforço do usuário na criação de seu perfil, bem como facilitar o processo de busca tornando-o mais próximo do seu perfil. A solução consiste na criação de um sistema de recomendação acadêmico onde o usuário não necessite participar ativamente, isto é, fornecer dados para traçar seu perfil, nem perder tempo fornecendo informações, através do preenchimento de formulários. O perfil será criado através de informações que o usuário tem disponível na Internet. O usuário, ao realizar uma pesquisa, informará apenas a(s) palavra(s)-chave para a busca, enquanto o sistema encarrega-se de coletar dados para o seu perfil e realizar a busca com base nessas informações (perfil e palavra-chave). Assim, uma das vantagens do sistema é realizar uma busca sobre o assunto desejado de acordo com o perfil do usuário, refinando a pesquisa e apresentando resultados mais específicos e relevantes. Este artigo apresenta o modelo de um sistema de recomendação que utiliza coleta de informação do usuário sem intervenção do mesmo, aplicação de técnicas para classificar os usuários e filtragem nos resultados da pesquisa realizada. Visando avaliar o modelo, foi realizada a sua implementação, que gerou o protótipo do sistema de recomendação batizado de "Ponto com filtro".

Nessa solução, decidiu-se pela criação de um sistema de recomendação de artigos acadêmicos. Esse sistema utiliza-se de dados contidos no currículo Lattes (Lattes) do usuário e de sua página pessoal (caso exista) para a criação do perfil do usuário. As informações do usuário, consideradas necessárias pelo sistema de recomendação são: dados pessoais, formação acadêmica e área de atuação. Nesse protótipo, a pesquisa por artigos acadêmicos é feita com o auxílio da ferramenta de busca Scholar Google (Scholar Google).

Para a avaliação do sistema de recomendação "Ponto com filtro", foram realizados testes com usuários que utilizaram o sistema. Esses usuários responderam a um questionário sobre as recomendações recebidas e funcionalidades da ferramenta. Os resultados dessa avaliação são mostrados no final do artigo. Este artigo está dividido em 5 seções. Os conceitos relacionados são apresentados na seção 2. A seção 3 mostra os trabalhos relacionados. A seção 4 apresenta a solução proposta. A seção 5 descreve o estudo de caso, a implementação, aplicação e análises de avaliação do experimento. Na última seção são apresentadas as conclusões do trabalho.

#### 2. Conceitos relacionados

Os sistemas de recomendação (SR) são utilizados para identificar usuários, armazenar suas preferências e recomendar itens que podem ser produtos, serviços e/ou conteúdos, de acordo com suas necessidades e interesses.

Segundo (Schafer, 2000) a estrutura dos sistemas de recomendação é dividida em quatro processos: identificação do usuário, coleta de informações, estratégias de recomendação e visualização das recomendações. A figura 1 apresenta a estrutura baseada no modelo de Schafer, que é considerada uma das mais completa.

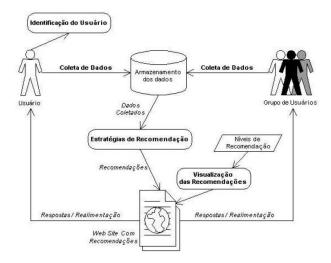


Figura 1 – Estrutura do Sistema de Recomendação (Schafer, 2000)

O processo de identificação do usuário é opcional em um sistema de recomendação. Porém, quando é realizado, esse é o primeiro passo a ser feito, que consiste na coleta dos dados sobre o usuário. Esses dados são armazenados em uma base de dados, para posterior uso na recomendação. Com os dados do usuário, são criadas as recomendações através de uma estratégia de recomendação definida pelo projetista. De posse das recomendações, o usuário visualizá-as (visualização das informações) em um site de fácil de acesso e de compreensão. As seções, a seguir, descrevem cada processo, em detalhes.

#### 2.1.1 Identificação do Usuário

Muitos sistemas realizam a identificação do usuário, pois assim é possível identificar com mais precisão quem está se conectando ao sistema e assim, customizar o sistema de acordo com as características desse usuário.

## 2.1.2 Coleta de Dados

Os dados do usuário podem ser coletados de forma explícita, implícita ou inferência. Na forma explícita, as informações são coletadas do próprio usuário, através do preenchimento de um formulário no site onde o mesmo informa suas preferências. Na forma implícita, as informações são coletadas pelo monitoramento da navegação do usuário com um site, páginas consultadas, histórico de compras, conteúdos selecionados, etc. Os dados coletados são armazenados em um banco de dados e utilizados para a análise das recomendações. É possível aplicar técnicas de mineração de dados nas informações coletadas, visando descobrir relações entre as páginas visitadas, produtos comprados, conteúdos e outros itens oferecidos pelo site (Torres, 2003). A coleta de dados por inferência consiste em descobrir o perfil do usuário com base no comportamento de outros usuários com padrão de comportamento similar ao seu.

### 2.1.3 Estratégia de recomendação

As estratégias de recomendação mais utilizadas são: filtragem de informação, customização, sumarização estatística, individualização, bussiness-rules, correlação e mineração de dados. A técnica de filtragem de informação pode ser baseada em conteúdo, colaborativa ou híbrida. A Filtragem baseada em conteúdo, recomenda itens semelhantes com os que o usuário demonstrou interesse no passado. Segundo (Torres, 2003) os interesses por determinados tipos de conteúdos podem definir o perfil do usuário. Na técnica de Filtragem Colaborativa é permitida a recomendação de produtos e serviços para um determinado cliente com base no que outros clientes com interesses ou preferências semelhantes tenham consumido. Diferente da filtragem de conteúdo que recomenda itens com base no conteúdo dos mesmos, a filtragem colaborativa recomenda itens que foram avaliados. A abordagem híbrida consiste do uso mútuo dessas técnicas.

A técnica de Customização consiste em Adaptar algo para uma determinada pessoa ou grupo de pessoas, conforme suas características. A customização é feita manualmente pelo usuário, que escolhe a maneira como a informação será apresentada. Na customização o usuário possui total controle do site (Garofalakis, 1999).

A Sumarização Estatística é baseada na opinião dos usuários sobre itens de um site, resumos de avaliações da comunidade e médias de popularidade. A porcentagem de usuários que compraram um determinado item ou o número de usuários que recomendam um item específico são exemplos de sumarização estatística.

A individualização é baseada na identificação de comportamentos frequentes dos usuários. Nessa recomendação o site calcula o número de visitas a diferentes páginas e seções para coletar os interesses do usuário de forma implícita (Reategui, 2005) Os dados coletados são referentes ao monitoramento da navegação do usuário pelo site.

Nas recomendações *business-rules* (regras de negócio), o site pode utilizar informações demográficas para direcionar a propaganda de um produto para um segmento do mercado (Schafer, 2000), por exemplo: quando um usuário é de uma determinada região, um certo produto é recomendado para ele, independente de seus interesses e suas necessidades.

A técnica de correlação é baseada na associação de itens que tendem a ocorrer juntamente com uma compra, podendo ser utilizados na identificação de padrões (Garofalakis, 1999). A correlação também pode ser definida através de regras de associação, que demonstra como produtos e serviços se relacionam uns com os outros.

O conhecimento extraído da mineração de dados pode ser utilizado pelos sistemas de recomendação para obter recomendações que sejam mais adequadas às

preferências do usuário. Os principais desafios são a eficiência e a qualidade nas informações recomendadas em tempo real, independente do tamanho do banco de dados. Pode-se aplicar também a mineração dos dados para otimizar o desempenho de um servidor Web, para descobrir que produtos estão sendo comprados simultaneamente, ou para identificar se o site está sendo usado como esperado (Spilliopoulou, 2000).

A solução proposta neste trabalho utiliza filtragem de informação baseada em conteúdo.

# 2.1.4 Visualização das Recomendações

As recomendações aos usuários devem ser apresentadas de maneira que as mesmas possam ser facilmente visualizadas e compreendidas. Além da apresentação das recomendações também é importante identificar os níveis de recomendação.

Segundo (Schafer, 2000), existem três níveis de recomendação: não recomendação, recomendação efêmera e recomendação persistente. Na não-recomendação, as recomendações são iguais para todos os usuários. Como, por exemplo, uma lista com os produtos mais vendidos para todos os seus clientes. Na recomendação efêmera, as recomendações são baseadas inteiramente na navegação de um único usuário e não utiliza informações das navegações anteriores do mesmo. Na recomendação persistente, as recomendações são baseadas no reconhecimento do usuário, e sugere produtos que são do seu interesse, com base nas suas navegações anteriores.

A visualização das recomendações pode ser apresentada através de e-mail, lista de recomendações, navegação, atendimento online e personificação.

#### 3. Trabalhos relacionados

Alguns sistemas de recomendação já têm sido propostos na literatura e visam indicar informações mais relevantes para um usuário, de acordo com o seu perfil.

Um framework para a recomendação de artigos científicos é apresentado por (Freitas, 2005). O sistema considera a opinião dos usuários e utiliza a técnica de filtragem colaborativa. Os usuários recebem recomendações de artigos com base nas informações de seu perfil. Os dados do perfil do usuário são adquiridos de forma explícita, ou seja o mesmo responde um questionário com várias perguntas sobre a sua carreira acadêmica. O sistema realiza uma busca por informações complementares no currículo Lattes do usuário. Umas das vantagens deste trabalho é o uso da opinião dos outros usuários, porém uma das deficiências do sistema é a necessidade que o usuário tem de informar seus dados através de um questionário exaustivo. Além disso, não foi feita nenhuma avaliação junto ao usuário sobre as recomendações recebidas.

Em (Ávila, 2006) é apresentado o SisRecAC, um sistema de metabusca que recomenda artigos científicos indexados pelo Google Scholar. Os artigos localizados e recomendados pela ferramenta de metabusca passam por uma avaliação do próprio usuário, que pode comentá-los e recomendá-los a outras pessoas, criando uma rede de cooperação. Nessa solução não é considerado o perfil do usuário e sim a opinião de um grupo de usuários.

O uso de redes bayesianas para modelar o perfil do usuário em sistema de recomendação é considerado em (Zhang, 2007). O trabalho apenas descreve a eficiência do algoritmo e não a satisfação do usuário após receber uma recomendação com base na técnica sugerida.

Em (Nunes, 2006) é apresentado um Sistema de Recomendação para Apoio à Colaboração. As recomendações são feitas a partir da análise do comportamento dos usuários em um Web Chat e do acesso a uma biblioteca digital. Apenas os dados cadastrais e o grau de interesse do usuário em um determinado assunto são considerados na recomendação. Os dados sobre o perfil do usuário são coletados de forma explícita.

Lopes (2006) propõe um sistema de recomendação acadêmico para bibliotecas digitais cujo perfil do usuário é coletado de forma implícita. O sistema utiliza filtragem por conteúdo e foi testado na Biblioteca Digital Brasileira de Computação (BDBComp, 2005). Os dados do perfil do usuário são coletados do seu currículo Lattes, o qual deve ser fornecido ao sistema pelo próprio usuário.

Conforme observa-se nos trabalhos relacionados, o perfil do usuário é um item muito importante num processo de recomendação, porém os usuários perdem muito tempo informando suas características/preferências ou os sistemas existentes utilizam-se de pouquíssimos dados sobre o usuário de forma a realizar a recomendação. Existe uma necessidade de realizar uma modelagem do usuário de forma mais rápida e eficiente, de modo que o mesmo não necessite informar seus dados.

# 4. Solução Proposta

O objetivo principal deste trabalho é oferecer um sistema de recomendação que gere o perfil do usuário de forma automática, sem o mesmo precisar informar seus dados. As informações sobre o usuário são retiradas da própria Internet, por exemplo, do seu currículo, da sua página Web, etc. Dessa forma evita que o usuário precise preencher formulários visando assim, minimizar o problema de confiabilidade, de tempo, e de informações irrelevantes numa busca. A figura 2 apresenta a solução proposta para o sistema de recomendação acadêmico.



Figura 2 – Sistema de recomendação acadêmico

- 1- O usuário cadastra-se no sistema fornecendo seu nome completo e dados de acesso (login e senha). Assim, é feita uma busca de forma implícita por dados do usuário disponíveis na Internet e assim é feita a criação do seu perfil.
- 2 O usuário informa a palavra-chave (assunto de interesse) que deseja fazer sua busca;

- 3 O sistema de recomendação processa de forma implícita o perfil do usuário e realiza uma seleção dos artigos que são mais adequados para o usuário, enfatizando aqueles que tem referência em mais de uma base de dados de artigos.
- 4 Por fim, o sistema apresenta o resultado dos artigos recomendados.

Nesse sistema, a taxonomia de E/S é por atributo e palavra-chave, com entrega das informações por navegação com apresentação "passiva", onde o usuário não necessita clicar em algum link para ver sua recomendação ou recebê-las por e-mail, ela estará no contexto da página.

O grau de personalização é baseado em atributo e do tipo persistente, onde o perfil do usuário está armazenado e é atualizado periodicamente, e essas informações refletem em sua recomendação.

A técnica mais adequada para esta solução é de mineração de dados baseado em regras associativas, pois os dados devem ser minerados periodicamente e as recomendações criadas por associações de outras fontes.

Visando verificar a aplicabilidade da solução proposta, a mesma foi implementada.

## 5. Sistema Ponto com Filtro

Para comprovar a viabilidade da solução foi implementado um protótipo, que foi batizado de "Ponto com filtro". O objetivo principal do sistema é fornecer ao usuário uma recomendação de artigos mais indicados ao seu perfil. A recomendação leva em consideração a área de interesse do usuário, seu nível de formação, áreas de interesse e idiomas. O perfil do usuário é criado de forma implícita, sendo as informações retiradas da sua página Web e do seu próprio currículo disponibilizado na plataforma Lattes.

O protótipo foi desenvolvido na plataforma PHPDev 4.2.3, com Apache 5, banco de dados MySQL e usa as bibliotecas DomXML e Scholar Google.

O usuário cadastra-se no sistema Ponto com filtro, informa login/senha e submete o endereço do seu site e/ou seu currículo Lattes no formato XML, conforme mostra a Figura 3.



Figura 3 – Tela de cadastro

Os dados relevantes para o perfil do usuário são capturados pelo sistema. Os dados recuperados são: nome completo, e-mail comercial, e-mail residencial, idiomas, áreas de atuação (especialidade e sub-área), nível de formação.

Após cadastro, o usuário pode efetuar pesquisas por artigos acadêmicos. No momento de realização de uma consulta, o sistema de busca do Google Scholar foi utilizado. Cada pesquisa realizada pelo usuário é armazenada no seu histórico, para posterior consulta.

O sistema, para realizar a pesquisa, gera uma "string de pesquisa" que consiste das palavras-chaves informadas pelo usuário, e seu perfil.

A quantidade de resultados é dada pela quantidade de áreas. Por exemplo, para uma área apenas, são mostrados os 100 primeiros resultados, para 2 áreas são mostrados 50 resultados de cada busca por área e assim sucessivamente. O "string de pesquisa" é passado como parâmetro para função "google()", da API do Google, que executa a pesquisa no Scholar Google e gera um arquivo "results.html" com os resultados. Portanto, se existirem seis áreas, o arquivo será gerado e carregado seis vezes.

Para criação do arquivo "results.html" foi preciso utilizar um executável chamado "wget" que salva os resultados da pesquisa do Scholar Google, pois o mesmo bloqueia o acesso de browsers. Assim o programa "wget" muda sua identificação através de um parâmetro permitindo baixar a página de resultados em um arquivo "html". A figura 4 apresenta a tela de resultados do sistema, após uma busca pela palavra-chave "recomendação".



Figura 4 – Tela de resultados

# 6. Avaliação do Ponto Com Filtro

Para a avaliação do sistema de recomendação "Ponto com filtro", foi realizado um estudo de caso, onde 10 usuários, com faixa etária de 29-40 anos de idade, usaram o sistema e posteriormente responderam um questionário sobre as recomendações recebidas, funcionalidades e comparações entre o sistema "Ponto com filtro" e o sistema de busca de domínio público Scholar Google. Desses usuários, apenas um era da área de Arquitetura e os demais da Computação, 72% possuem o título de mestre, 14% o título de doutor e 14% são graduados. Os usuários acessaram o sistema remotamente, ficarem livres para a realização de buscas e tempo de uso. Cada usuário realizou as pesquisas que acharam relevantes e responderam a um questionário com seis perguntas, que envolviam questões sobre a interface do sistema, resultados apresentados e comparação com o Scholar Google. Como resultados do questionário, 100% dos usuários responderam que ficaram satisfeitos com a visibilidade e funcionalidade do sistema e que os dados filtrados da pesquisa corresponderam as informações do currículo. 100% também responderam que tinha ficado claro o objetivo e funcionalidade do sistema. Comparando o sistema "Ponto com Filtro" com o sistema de busca Google scholar, no critério "tempo", os usuários acharam que os dois são similares. Em relação ao critério "resultados de busca apresentados", 57% acharam os resultados apresentados pelo

"Ponto com Filtro" melhores do que os do "Scholar Google". No critério "interface", 47% consideraram a interface do sistema "Ponto com Filtro" mais adequada que a do "Scholar Google"

## 7. CONCLUSÕES

O volume de informações na Internet dificulta o processo de busca e filtragem de informação. Com isso, sistemas de recomendação se tornaram importantes e necessários para auxiliar usuários numa procura. Além disso, a necessidade de personalizar o relacionamento com os usuários está cada vez mais crescente. Visando minimizar o tempo gasto e a resposta de uma busca por informação relevante, este trabalho apresentou um modelo de sistema de recomendação, que dispensa a intervenção do usuário no momento de traçar seu perfil e utiliza técnicas de mineração de dados. O modelo foi implementado para verificar sua aplicabilidade. O protótipo "Ponto com Filtro" utiliza os dados que o usuário possui no seu currículo Lattes para realizar a recomendação. Uma das vantagens é que o usuário não precisa preencher formulários exaustivos para informar o seu perfil. O sistema foi experimentado em um pequeno grupo de usuários, os quais demonstraram bastante satisfeitos com os resultados encontrados. Novos experimentos estão sendo realizados, de forma a avaliar o sistema com um grupo mais abrangente. Como trabalhos futuros pretende-se utilizar de outros dados sobre o usuários, que também seriam retirados do currículo Lattes (artigos publicados) ou até mesmo de suas homepages.

Uma das limitações apresentadas neste trabalho foi o não acesso à base de dados do Lattes, pois o seu acesso é limitada a instituições de ensino, assim necessitou-se solicitar ao usuário seu currículo no formato "xml". Além disso, o uso das API's oferecidas pela ferramenta de busca do Google não funcionam para o Scholar Google, com isso necessitou-se da utilização de um programa que copiasse para um "html" os resultados deste site de busca.

Considerando-se que o objetivo principal deste trabalho era criar um perfil de usuário de forma implícita e gerar respostas relevantes para uma determinada pesquisa, pode-se considerar que o mesmo foi atingido.

# REFERÊNCIAS

Avila, Christiano; Loh, Stanley. Sistema de Recomendação de Artigos Científicos. In: **WebMedia**, 2006, Natal, RN. Anais.

BDBComp: Biblioteca Digital Brasileira da Computação. Disponível em: <a href="https://www.lbd.dcc.ufmg.br/bdbcomp">www.lbd.dcc.ufmg.br/bdbcomp</a>. Acesso em: nov. 2007.

Freitas, Juliana Gonçalves de. **Uma Ferramenta para Clusterização de Perfis de Usuários Baseada em Dados Qualitativos**. 2005. (Graduação em Análise de Sistemas). Universidade do Vale do Rio dos Sinos. Trabalho de Conclusão de Curso.

Garofalakis, Minos N.; RASTOGI, Rajeev; SESHADRI, S.; SHIM, Kyuseok. Data Mining and the Web: Past, Present and Future. In: **WIDM**, 1999. Proceedings...

Lattes. Disponível em: http://lattes.cnpq.br. Acesso em: nov. 2007

Lopes, G; Souto, M. A; Oliveira, JPM. Sistema de Recomendação para Bibliotecas Digitais sob a Perspectiva da Web Semântica. In: **II Workshop de Bibliotecas Digitais**, SBBD 2006, SBC, Florianópolis, 2006. Anais..



Milani, Fábio; Cazella, Sílvio Cezar. Um Modelo para Determinar a Autoridade de Usuários em Sistemas de Recomendação. **III Forum de Inteligência Artificial**, 2005, Canoas, 2005.

Nunes, Marcos Freitas; Borges, Thyago, VINSON, Alexander, LOH, Stanley. Técnica para Analisar a Evolução do Perfil do Usuário com base nas suas Publicações. In: **II Escola Regional de Banco de Dados** 2006, Passo Fundo. RS. Anais..

Reategui, Elisio Berni; CAZELLA, Sílvio Cézar. Sistemas de Recomendação. In: **XXV Congresso da Sociedade Brasileira de Computação**, São Leopoldo, 2005. Anais..

Schafer, J. Ben; Konstan, Joseph; RIEDL, John. Recommender Systems. In: Conference on Electronic Commerce, 2000, Minneapolis. Proceedings...

Scholar Google. Disponível em: http://scholargoogle.com

Spilliopoulou, Myra. Web Usage Analysis and User Profiling. In: International **WEBKDD Workshop**, San Diego, California, USA, 1999, Proceedings..

Torres, Roberto Dias. **Técnicas para Desenvolver Sistemas de Recomendação na Web.**, UFRGS, Porto Alegre, 2003. Dissertação de Mestrado

Zhang, Yi; Koren, Jonathan. Efficient Bayesian Hierarchical User Modeling for Recommendation Systems. In: **SIGIR**'07, July 23–27, 2007, Amsterdam. Proceedings...