

## Chapter

# Personalização de Páginas Web através dos Sistemas de Recomendação

Eliseo Berni Reategui, Sílvio César Cazella e Fernando Santos Osório

### *Abstract*

*Most of the interfaces developed nowadays are designed for a standard user. Therefore, such interfaces end up by neglecting the interests and needs of each person. Through personalization methods it is possible to create a different interface for each user, changing its structure or its contents according to the profile of each one. One of the techniques employed for the personalization of interfaces are recommender systems, conceived initially to enable users to receive personalized contents through the sharing of information. By monitoring the user's actions, these systems are able to identify contents, items or actions to be recommended in a personalized way. An appropriate book recommendation, for example, can make the difference between conquering a customer or losing him/her. Due to this need to "seduce" the users, personalization techniques are perceived as an important factor in the moment a person has to be "captivated". This chapter gives an overview about recommender systems used in personalization tasks. It describes the most popular techniques and strategies, presents examples of applications that employ recommender systems, discussing certain implications in interface design. At the end of the chapter, new trends in the area are highlighted, such as community formation and the semantic web.*

### *Resumo*

*A maior parte das interfaces dos sistemas desenvolvidos hoje em dia é concebida para atender as necessidades de um usuário padrão. Deste modo, tais interfaces acabam negligenciando necessidades e interesses particulares de cada indivíduo. Através de métodos de personalização é possível criar uma interface diferente para cada usuário, modificando sua estrutura ou seu conteúdo de acordo com o perfil de cada um. Uma das técnicas empregadas na personalização de interfaces são os sistemas de recomendação, criados inicialmente para permitir que usuários pudessem receber conteúdo personalizado através do compartilhamento de informações. Através do monitoramento das ações dos usuários, estes sistemas são capazes de identificar conteúdo, itens ou ações a serem recomendados de forma personalizada. A recomendação adequada de um livro, por exemplo, pode fazer a diferença entre conquistar o usuário ou perdê-lo. Devido a esta necessidade de conquista, a personalização tem se apresentado como um fator facilitador no momento de "cativar" o usuário. Este capítulo dá uma visão geral sobre os sistemas de recomendação utilizados nas tarefas de personalização. Descreve as técnicas e estratégias de recomendação mais utilizadas, apresenta exemplos de aplicações que empregam sistemas de recomendação, discutindo suas implicações em aspectos relacionados à concepção de interfaces. Por fim, o capítulo aborda tópicos que apontam para novas tendências na área, como a formação de comunidades e a web semântica.*

## 1. Introdução à personalização

Os princípios fundamentais de uma interface inteligente estão no aumento de sua flexibilidade, na melhoria de sua usabilidade e na ampliação do poder de interação para todos os usuários [Maybury 2001]. Colocar em prática tais princípios exige o tratamento de alguns problemas: a exposição do usuário a uma quantidade exacerbada de informação; o fornecimento de ajuda na resolução de programas complexos; a execução de algumas tarefas que antes deveriam ser executadas pelo usuário; a criação de sistemas personalizados [Ehlert 2003]. Este capítulo trata do último problema, ou seja, como considerar diferenças entre usuários e fornecer métodos personalizados de interação, tópico bastante investigado na área de sistemas hipermídia adaptativos. Nestes, apresentações personalizadas são construídas de acordo com dados coletados através da interação com os usuários. Tais sistemas têm sido utilizados em muitas aplicações, como em sistemas de informação, *help desks*, sistemas de recuperação de informação e sistemas educativos [De Bra 2000]. Três funções são executadas por um sistema hipermídia adaptativo:

- Enquanto o usuário navega no sistema, toda sua interação é monitorada. Com base nas informações coletadas, mantém-se um modelo do usuário: dados demográficos, itens visualizados, interesses, preferências, entre outros.
- A apresentação do documento pode ser modificada de modo a sugerir ao usuário os próximos passos. *Links* podem ser adicionados, modificados, removidos, reorganizados ou comentados.
- O sistema pode mostrar, esconder, enfatizar ou esmaecer fragmentos de uma página, assegurando que seu conteúdo inclua a informação apropriada, em um nível adequado de dificuldade ou detalhe.

Estes conceitos são conceitos de personalização de interface que podem trazer várias vantagens. Por exemplo, através da constante seleção de conteúdos relacionados aos interesses do usuário, um sistema personalizado pode reduzir o tempo que estes levam para encontrar informações relevantes. Além disso, identificando relacionamentos entre itens (por exemplo: quem compra item X também compra item Y), um sistema personalizado é capaz de identificar em tempo real itens de interesse do usuário, “bombardeando-o” visualmente com conteúdo ou produtos que lhe interessam. Deste modo, a probabilidade de que um usuário acesse ou adquira um item é bem maior do que em sistemas não personalizados. Por fim, através da oferta sistemática de serviços/produtos/conteúdos mais interessantes para o usuário, o sistema consegue fazer com que ele se torne um visitante/cliente fiel.

Apesar da pesquisa em personalização ter trazido resultados significativos, tais interfaces receberam muitas críticas já que a adaptação em tempo real frequentemente contradiz o princípio da estabilidade [Tsandilas 2004]: como é possível o sistema se comportar de forma previsível e compreensível se mudanças são feitas constantemente em sua interface? Algumas diretivas, no entanto, podem tornar menos prejudicial esse efeito:

- Deixar o mais claro possível o processo de personalização da interface: a transparência do sistema e o controle do usuário são princípios de usabilidade que devem ser obedecidos por uma interface personalizada [Nielsen 2006]. Ou

seja, deve-se buscar explicar os processos de adaptação da interface para permitir que o usuário mantenha a percepção de que controla o sistema (ao invés de ser controlado por ele). Quando as lojas de comércio eletrônico informam o usuário que “usuários que compram X também compram Y”, estão explicando, de forma simplificada, como os resultados apresentados foram encontrados (naturalmente todo o processo para se chegar a tal resultado não poderia ser detalhado sob o risco de deixar o usuário ainda mais confuso).

- Personalizar sempre os mesmos componentes ou áreas da tela: para que as próprias modificações da interface passem a formar um padrão, é importante que o conteúdo personalizado apareça sempre da mesma maneira, muitas vezes em uma área fixa da tela. Nesta área, uma resenha indicando a presença de conteúdo personalizado cria certa estabilidade no sistema, além de ressaltar ao usuário um serviço adicional que está sendo prestado.
- Não distrair a interação do usuário: uma interface pró-ativa pode acabar atraindo a atenção desnecessariamente, tornando o ato de personalização mais importante do que a própria tarefa sendo executada pelo usuário. Por isso, sempre deve ser possível ao usuário ignorar a pró-atividade da interface. Deve-se buscar sugerir ações/conteúdos ao invés de executar estas ações ou dirigir o usuário a estes conteúdos de maneira automática.

Na Internet, personalizar a seleção de produtos, itens ou informações apresentadas a um usuário tornou-se um grande desafio. A recomendação adequada de um livro, por exemplo, pode fazer a diferença entre conquistar o usuário ou perdê-lo. Devido a esta necessidade de conquista, a personalização tem-se apresentado como um fator facilitador no momento de "cativar" o usuário.

Este capítulo apresenta a personalização através dos sistemas de recomendação, descrevendo as técnicas mais utilizadas e mostrando como podem ser aplicadas em exemplos práticos. Iniciaremos na seção 2 com uma introdução a personalização através dos sistemas de recomendação, apresentando em seguida na seção 3 questões referentes a formação de perfis de usuários. Na seção 4, questões relativas às estratégias de recomendação são apresentadas, e na seção 5 questões relativas às técnicas. A seção 6 apresenta exemplos de aplicações de sistemas de recomendação. Na seção 7, tendências e desafios na área são descritos, tais como formação de comunidades virtuais, recomendação em ambientes virtuais e web semântica. No final do capítulo são apresentadas conclusões destacando a relevância do tema nos dias de hoje bem como no futuro.

## **2. A Personalização através dos sistemas de recomendação**

Com a quantidade de informações e com a disponibilidade facilitada das mesmas pelo uso da Internet, as pessoas se deparam com uma diversidade muito grande de opções. Muitas vezes um indivíduo possui muito pouca ou quase nenhuma experiência pessoal para realizar escolhas entre as várias alternativas que lhe são apresentadas. A questão relevante neste momento refere-se a como proceder nestes casos? Para minimizar as dúvidas e necessidades que temos frente à escolha entre alternativas, geralmente confiamos nas recomendações que são passadas por outras pessoas, as quais podem chegar de forma direta (*word of mouth*) [Shardanand e Maes 1995], cartas de

recomendação, opiniões de revisores de filmes e livros, impressos de jornais, entre outros.

Os sistemas de recomendação auxiliam no aumento da capacidade e eficácia deste processo de indicação já bastante conhecida na relação social entre seres humanos [Resnick e Varian 1997]. Em um sistema típico as pessoas fornecem recomendações como entradas e o sistema direciona estas informações para os indivíduos considerados interessados potenciais. Um dos grandes desafios deste tipo de sistema é realizar o casamento correto entre os que estão recomendando e aqueles que estão recebendo a recomendação, ou seja, definir e descobrir este relacionamento de interesses.

A maior parte da pesquisa na área de recomendação é relacionada à definição de algoritmos de busca de informação personalizada [Breese 1998]; [Geyer-Schulz 2002]; [Sarwar 2000]. Entretanto, a simples utilização de um sistema de recomendação introduz questões relacionadas à área de interface humano-computador (IHC), na medida em que a composição de uma página personalizada envolve a modificação de sua estrutura e apresentação de conteúdo diferente para usuários distintos. As recomendações podem ser misturadas a outros itens em uma página, ou podem ocupar uma área específica da tela para enfatizar a seleção personalizada de itens. Dependendo da área e do tipo de aplicação envolvida, uma ou outra política pode ser mais apropriada.

Os proponentes do primeiro sistema de recomendação, denominado Tapestry [Goldberg et al. 1992][Resnick e Varian 1997], criaram a expressão *filtragem colaborativa*, visando designar um tipo de sistema específico no qual a filtragem de informação era realizada com o auxílio humano, ou seja, pela colaboração entre os grupos de interessados. Esta técnica tem sido utilizada com sucesso em diversos projetos de pesquisa bem como em *websites* comerciais [Schafer et al. 2000]. O algoritmo por trás da filtragem colaborativa é baseado na idéia de que o usuário ativo é mais propenso a se interessar itens que usuários semelhantes preferem. Para dar suporte a isso, um escore de similaridade entre o usuário ativo e todos os outros usuários é calculado. As recomendações são geradas através da seleção de itens de usuários com maior grau de similaridade.

Os websites de comércio eletrônico são atualmente o maior foco de utilização dos sistemas de recomendação, empregando diferentes técnicas para encontrar os produtos mais adequados para seus clientes e aumentar deste modo sua lucratividade. Introduzido em julho de 1996 o My Yahoo foi o primeiro *website* a utilizar a personalização em grandes proporções, utilizando a estratégia de customização onde o usuário indicava explicitamente vários critérios de preferência de exibição de sua página [Manber et al. 2000]. Hoje em dia, um grande número de *websites* emprega os sistemas de recomendação para levar aos usuários diferentes tipos de sugestões personalizadas, tais como Amazon.com, Drugstore.com, BarnesAndNoble.com, e no Brasil Submarino.com.br, Siciliano.com.br. As estratégias de personalização mais utilizadas trazem ao usuário ofertas casadas ("clientes que compraram item X também compraram item Y"), itens de sua preferência, itens mais vendidos em suas categorias favoritas, entre outras. Para possibilitar quaisquer destas formas de personalização, é necessário coletar e armazenar informações sobre os usuários. A próxima seção aborda este tema,

mostrando como fazer a identificação e coletar dados de um usuário, de forma implícita ou explícita.

### 3. Formação de perfis de usuários

Um dos pontos fundamentais para trabalhar com sistemas de recomendação, está na correta definição do perfil dos usuários do sistema. Como em sistemas de recuperação não se está lidando com recuperação pura de informações, mas sim com recomendação baseada nas preferências apresentadas de forma explícita ou implícita pelo usuário, faz-se necessário à definição do perfil do usuário. Segundo Montaner [Montaner et. al 2003], a geração e manutenção destes perfis requerem a tomada de cinco decisões de projeto de sistemas deste tipo: técnica de representação do perfil, técnica para gerar o perfil inicial, feedback da relevância do item, técnica de aprendizado de perfil e a técnica para adaptação do perfil.

A representação do perfil constitui-se no primeiro passo de um projeto de sistema de recomendação pois as técnicas que serão aplicadas dependerão desta decisão. Um sistema de recomendação não pode iniciar as atividades sem que o perfil do usuário tenha sido previamente definido, pois para iniciar suas atividades o sistema deve saber o máximo possível sobre os usuários. No momento da geração do perfil inicial é importantíssimo ter-se a maior quantidade de informação possível sobre o real interesse e necessidade do usuário.

Após o início das recomendações terem sido feitas é importante que o próprio usuário apresente um *feedback* sobre a relevância da recomendação realizada, visando permitir que o sistema aprenda mais sobre os gostos, interesses e preferências gerais do usuário. O *feedback* pode ser fornecido de forma explícita ou implícita pelo usuário, tornando-se elemento fundamental para o aumento da acurácia do sistema. Este *feedback* e a análise do comportamento do usuário são profundamente importantes, uma vez que os gostos do mesmo ou necessidades podem se alterar com o tempo, sendo assim faz-se necessária à aplicação de técnica adequada para adaptar o perfil do usuário para novos interesses e análise da utilidade dos interesses antigos.

Quanto ao aproveitamento e exploração dos perfis Montaner explica que três critérios são essenciais para caracterizar os sistemas de recomendação: o método de filtragem de informação (demográfica, baseada em conteúdo, colaborativa ou híbrida), o casamento do item-perfil (quando baseado em conteúdo) e técnicas de casamento de perfil de usuários que apresentam similaridades de gostos (quando colaborativo). A tabela 3.1 apresenta a proposta completa dos critérios de análise.

Tabela 3.1: Critérios de análise de perfil

<b>Geração e manutenção do perfil</b>	<b>Exploração do perfil</b>
Representação do perfil do usuário	Técnica de adaptação de perfil
Geração do perfil inicial	Técnica de casamento de perfil do usuário-item
Técnica de aprendizado de perfil	Técnica de casamento de perfil de usuário
<i>Feedback</i> da relevância do item	Método de filtragem de Informação

### 3.1. Geração e Manutenção de Perfil de Usuário

Este tópico engloba quatro critérios de análise: representação do perfil, geração do perfil inicial, técnica de aprendizado de perfil e *feedback* da relevância do item.

#### 3.1.1. Representação do perfil do usuário

A construção de perfis adequadamente definidos constitui-se em uma das tarefas fundamentais e o sistema dependerá do resultado da mesma para obter sucesso. Esta tarefa é de grande valor quando se trabalha com conteúdo e para a questão colaborativa, uma vez que no primeiro busca-se a relação com o conteúdo do item e no segundo busca-se a relação entre perfis. Várias abordagens têm sido aplicadas para auxiliar à consecução desta tarefa, algumas delas estão descritas a seguir:

- **Histórico:**

Alguns sistemas procuram preservar uma lista de vendas realizadas, outros o histórico de navegação na Web procurando melhor definir o perfil do usuário. Junto a estas informações é comum e necessário preservar o valor de relevância de um item para o usuário (fornecido por *feedback*). Constitui-se em uma abordagem típica de *e-commerce* onde é preservada a lista de venda de produtos e pontuações fornecidas pelos usuários junto ao perfil do mesmo. Como exemplo pode se citar a Amazon.com e CDNOW.com. Uma outra abordagem similar é aplicada no sistema WebSell, na qual o perfil é definido com a aplicação de duas listas, uma com pontuação de produtos vendidos e ditos interessantes e outra com alguns ditos não interessantes, constituindo exemplos positivos e negativos respectivamente.

- **Vector Space Model:**

Nesta abordagem os itens são representados com o auxílio de um vetor de características dos mesmos, normalmente são palavras ou conceitos com um valor associado. Este valor pode ser booleano ou real, o primeiro representa a presença ou não de uma dada característica em um dado item, e o segundo a intensidade (frequência) da ocorrência do mesmo. Por exemplo, o sistema Webmate utiliza uma representação de vetores de características múltiplas. A idéia básica é representar cada um dos documentos como um vetor, sendo assim documentos com conteúdo similar possuem vetores similares, sendo considerados produtos similares para serem recomendados ao usuário.

- **Classificadores:**

Alguns sistemas utilizam a classificação como uma técnica para aprender o perfil do usuário. Para a construção destes classificadores são aplicáveis, por exemplo, Redes Neurais, Classificadores Bayesianos e Árvores de Decisão.

- **Matriz de avaliações Usuário-Item:**

Alguns sistemas de filtragem colaborativa mantêm uma matriz de avaliações de usuário-item, como um perfil do usuário, pois nesta matriz é possível verificar o que é mais ou menos interessante para o usuário. Cada célula da matriz contém uma avaliação representando a avaliação do usuário para um determinado item, caso exista um vazio, significa que o usuário não avaliou o referido item.

### 3.1.2. Geração do perfil inicial

Para que um sistema de recomendação funcione de forma adequada é importante a aquisição do máximo de informação possível sobre o usuário, para tanto o contato com o mesmo é essencial. A questão é que usualmente os usuários não estão dispostos a despendar muito tempo para definir seus interesses para criar um perfil. Para que ocorra a aquisição de informações para o perfil do usuário pode ser implementada uma página Web onde o usuário pode manualmente fornecer seus interesses, pode ser implementado o reconhecimento automático de padrões pelo sistema de recomendação (observação do comportamento do usuário) ou pode-se implementar procedimentos semi-automáticos como o proporcionado pelos conjuntos de treinamento.

As técnicas abordadas são descritas abaixo:

#### a) Vazio:

São sistemas que não se preocupam com o perfil inicial e inicia com uma estrutura de perfil vazia. A estrutura de perfil começa a ser preenchida através de um método dito automático de reconhecimento quando o usuário começa a interagir com o sistema.

#### b) Conjunto de Treinamento:

Constitui-se em um conjunto de exemplos de interação de usuário com o sistema, o qual é usado para inferir o perfil do usuário. Uma maneira adequada e fácil para obter este conjunto de treinamento é pedir que o usuário avalie um conjunto de exemplos concretos como sendo relevantes ou irrelevantes para os interesses do mesmo (Sorensen e McElligot apud [Montaner et al. 2003]). Esta técnica apresenta a vantagem de obtenção fácil de informação sobre os gostos do usuário, e como desvantagem pode se citar que o usuário será submetido a avaliação de alguns exemplos que podem não ser de interesse do mesmo fazendo com que os resultados acabe sendo pouco significativos.

#### c) Manual:

Nesta técnica o usuário é solicitado a registrar seus interesses em um formulário com palavras chaves e tópicos. Uma vantagem da aplicação desta técnica é a transparência do comportamento do sistema ao usuário, pois desta forma quando um item é recomendado ao usuário, o sistema pode facilmente inferir sobre o porque da recomendação. Um dos pontos negativos neste tipo de sistema refere-se ao grande esforço que deve ser despendido pelo usuário, outro se refere à dificuldade do usuário determinar no que este está realmente interessado.

### 3.1.3. Técnicas de aprendizado de perfil

Alguns sistemas apresentam uma fase de aprendizado *off-line* durante a qual os sistemas aprendem e criam um modelo de comportamento do usuário. Após este aprendizado o modelo de comportamento é aplicado em tempo real. Alguns sistemas não necessitam de técnicas de aprendizado de perfil, por exemplo: sistemas que adquirem informação de perfil diretamente de BD (relação produtos vendidos); sistemas de filtragem colaborativa, pois estes guardam a matriz de pontuação user-item como perfil.

Uma outra técnica aplicada para o aprendizado de perfil refere-se ao agrupamento que propõe a criação de grupos de usuários similares baseado nos dados obtidos sobre o mesmo. Tradicionalmente sistemas de filtragem colaborativa são freqüentemente baseados no casamento do perfil do usuário atual com agrupamento de perfis similares obtidos pelo sistema.

Outra técnica aplicada é a classificação (explicado na Seção 5.2.1). Para construir um classificador em um sistema de recomendação aplicam-se as informações do item e perfil do usuário como entrada e obtém como saída a categoria que representa o quanto significativo a recomendação será para o usuário.

As árvores de decisão são aplicadas para funções de valores discretos no qual o aprendizado obtido é apresentado na forma de um AD. Estas árvores podem ser representadas como um conjunto de IF-THEN, podendo ser representadas por regras. A AD aprende pelo particionamento do conjunto de treinamento obtendo classes de itens que podem ser classificados de interesse ou não [Quinlan 1993].

### **3.1.4. Retorno da relevância de item (*Relevance Feedback*)**

Os interesses inicialmente esboçados por um indivíduo podem sofrer grandes alterações durante o tempo em que este interage com o sistema de recomendação. Como exemplo poderia ser citado um pesquisador que inicia sua pesquisa em uma área e com o tempo e evolução de suas pesquisas e projetos, começa a mudar o foco de interesse. Para que o sistema de recomendação tenha êxito será essencial a percepção desta mudança de interesse do usuário para que não ocorra a oferta de itens que não casam com os interesses atuais do usuário. O *feedback* constitui-se em um excelente indicador desta mudança.

Os sistemas de recomendação trabalham com informações que permitem a verificação de *feedback*. Por exemplo, a informação positiva que se constitui em itens que foram adquiridos pelo usuário (houve interesse), e a informação negativa que se constitui em itens em que o usuário não apresentou interesse (não houve interesse). A informação positiva pode ser observada como um sucesso de recomendação, ou seja, a recomendação é realizada e o usuário realmente adquire o item ou pelo *feedback* numérico onde o usuário aponta o seu alto grau de interesse em itens deste tipo. A informação negativa pode ser trabalhada nas informações de itens recomendados e não adquiridos, frente à determinação de características comuns destes itens ou pela pontuação explícita fornecida pelo usuário, neste caso pontuação baixa (Holte apud [Montaner et al. 2003]). As duas maneiras para obter este *feedback* positivo ou negativo, baseiam-se na observação implícita da interação do usuário com o sistema ou no uso de informação explícita fornecida pelo usuário. Alguns sistemas não trabalham com a atualização do perfil do usuário, sendo assim não necessitam do *feedback*, mas permitem que o próprio usuário atualize o próprio perfil.

#### **a) *Feedback* explícito:**

Este tipo de *feedback* ocorre em sistemas que requisitam do usuário um retorno sobre um dado item. Uma forma de implementar este tipo de *feedback* pode ser observado no *site* para avaliação de câmeras digitais <http://www.dpreview.com/reviews>.

Segundo Montaner existem três abordagens fundamentais para trabalhar o *feedback* explícito: gostar/não gostar, avaliação (*ratings*) e comentários via texto não-estruturado. Em sistemas que aplicam a avaliação de gostar / não gostar os usuários são requisitados a explicitar julgamentos de itens em escalas ditas binárias, ou seja, colocar a sua opinião sobre um dado item (há interesse ou não existe nenhum interesse). Em abordagens que aplicam escalas numéricas discretas, o usuário deve pontuar o item (exemplo, Amazon.com). Alguns *sites*, por sua vez, encorajam a inclusão de comentários textuais de seus usuários. Estes comentários são aplicados para auxiliar o usuário no momento de tomar a decisão de obter um determinado item ou não, a questão é que o usuário deve despende tempo lendo e avaliando o conteúdo descritivo para verificar se o comentário realmente é positivo ou não. Caso esta quantidade de comentários tome volume considerável, pode-se levar o usuário a enfrentar o problema inicial que era a sobre carga de informação. O uso do *feedback* explícito apresenta algumas limitações básicas:

- 1) A relevância ou não de um item constitui-se em algo extremamente particular uma vez que está ligada aos interesses do momento. Caso o usuário seja apresentado a um número N de itens e solicitado a pontuar, a tendência é que este avalie bem os primeiros que são apresentados e lhe suprem as necessidades e avalie com menor atenção os demais.
- 2) A escala numérica pode não ser totalmente adequada para reportar percepções humanas.
- 3) Os usuários não fornecem *feedback* suficiente mesmo quando são encorajados a fazê-lo. O usuário procura realizar as atividades de interação com o sistema até que este esteja com suas necessidades satisfeitas, pontuar item não faz parte das atividades a que este se dispõe.

#### **b) *Feedback* implícito:**

Quando um sistema aplica um *feedback* implícito isto significa que alguma técnica automática infere as preferências do usuário através do monitoramento contínuo das ações dos mesmos. Como a obtenção de *feedback* explícito é algo extremamente difícil, busca-se obter informações relevantes relativas ao usuário a partir de questões implícitas. A maioria dos métodos implícitos que são aplicados aos sistemas de recomendação obtém *feedback* relevante pela análise de *links* que foram seguidos pelo usuário, pelo histórico de compras do mesmo, pelo histórico de navegação e pelo tempo despendido pelo usuário observando uma determinada página (quanto maior o tempo deve ser inferido um maior interesse).

Assim como o *feedback* explícito o implícito apresenta algumas dificuldades como lidar com a incerteza do porquê um dado usuário está realizando tal ação, por exemplo, o tempo de permanência em uma página pode ter como motivo outros fatores além de interesse, como realização de uma atividade em paralelo.

#### **c) Abordagem Híbrida:**

O fornecimento do *feedback* implícito ajuda a minimizar os esforços do usuário enquanto que o explícito permite uma melhor inferência sobre o perfil do mesmo.

Nichols (Nichols apud [Montaner et al. 2003]) comenta que a combinação entre estes dois tipos de feedback permite que o implícito atue como uma forma de averiguar o explícito, ou seja, através de verificação dos comportamentos dos usuários é possível inferir se o que o mesmo explicita está condizente com seus atos. Por exemplo, caso um usuário realize a avaliação de um item (explicitamente) e despenda pouco tempo na avaliação do mesmo (implícito), o valor desta avaliação pode ser decrescido de importância (peso menor).

### **3.2. Exploração do Perfil do Usuário**

Um sistema de recomendação pode trabalhar recomendando itens com o auxílio de agentes (em um sistema inteligente) que devem tomar decisões de acordo com as informações que lhe foram previamente fornecidas ou que este tenha obtido implicitamente. Sendo assim o agente estará explorando o perfil do usuário.

#### **3.2.1. Métodos de Filtragem de Informação**

Segundo Montaner existem, basicamente, três métodos para a realização da filtragem de informação: demográfico, baseado em conteúdo e colaborativo (Seção 5). Procurando superar as dificuldades apresentadas na aplicação destes métodos de forma isolada é proposta uma abordagem híbrida.

#### **3.2.2. Técnica de casamento de perfil do usuário e item**

Nos sistemas de recomendação, geralmente, os perfis de usuários são utilizados para recomendar novos itens considerados relevantes para os mesmos. Os sistemas de filtragem baseados em conteúdo aplicam comparação direta entre perfil do usuário e conteúdo de novos itens, sendo assim uma técnica de casamento de perfil do usuário e item deve ser implementada.

Uma técnica possível refere-se ao casamento entre palavras-chave, onde é realizada uma contagem de termos no conteúdo do item e verificada frequência de ocorrência de cada uma, verificam-se as de maior ocorrência com o perfil do usuário. Caso o usuário venha a adquirir um item, as demais palavras significativas, mas não pertencentes ao perfil podem auxiliar na atualização do perfil deste usuário. Outra técnica amplamente aplicada refere-se à verificação do vizinho mais próximo, estes algoritmos baseiam-se no cálculo da distância do item de interesse para o resto dos outros itens novos ou já existentes. No seu trabalho de pesquisa Montaner destaca a utilização de CBR (*Case Based Reasoning*) em sistemas de recomendação, onde os perfis de usuários são vistos como um conjunto de casos passados, como exemplo existe o sistema WebSell.

A técnica de classificação também é aplicável neste caso, uma vez que é possível construir, através de um conjunto de treinamento baseado em características dos itens, um modelo que indique se o perfil irá se interessar ou não por este tipo de item.

### **3.2.3. Técnica de casamento de perfil de usuário**

Sistemas de filtragem colaborativa casam pessoas com interesses similares e deste ponto em diante realiza as recomendações. Normalmente o processo de filtragem colaborativa inclui encontrar usuários similares, identificar os vizinhos mais próximos e realizar predições baseando-se nas avaliações dos vizinhos. Buscando encontrar usuários similares são aplicados o cálculo do vizinho mais próximo, agrupamento e classificação (Seção 5).

### **3.2.4. Técnica de adaptação de perfil de usuário**

Os interesses dos usuários podem não ser tão estável como se espera, ou seja, com o tempo o interesse pode apresentar alterações. Desta maneira, os *feedbacks* mais recentes podem ser mais significativos do que os históricos. Existe a necessidade de uma técnica que ajudará a adaptar o perfil do usuário aos novos interesses. Uma questão importante para ser respondida é referente a como descobrir em que momento ocorreu ou iniciou esta mudança de interesses.

Uma técnica de adaptação poderia ser a adaptação manual, ou seja, no momento em que o usuário estiver interessado este se reporta ao sistema e atualiza seus interesses. Esta técnica obriga que o usuário despenda mais esforços. A adição de novas informações refere-se a uma técnica possível. Nesta técnica o perfil do usuário é atualizado através da adição de novas informações extraídas frente ao *feedback* do usuário. Neste caso existe a atualização do perfil, mas não há o esquecimento dos interesses passados.

Outra técnica refere-se à função gradual de esquecimento. Nesta técnica baseia-se na idéia de que o esquecimento natural é um processo gradual (Webb apud [Montaner et al. 2003]). A idéia básica é a de fornecer um peso inicial para as observações feitas, e com o passar do tempo este peso irá sofrer descontos até que a observação se torne sem importância e seja esquecida. Desta forma, os novos interesses passam a ter os maiores pesos.

### **3.3. Privacidade em Sistemas de Recomendação**

Para que um Sistema de Recomendação possa recomendar itens apropriadamente para seus usuários, o sistema deve constantemente coletar dados sobre os usuários. Estes dados referem-se aos comportamentos (por exemplo, navegação e compra) e possivelmente os demográficos. Esta coleta é feita muitas vezes de forma implícita, isto é, sem que o usuário necessariamente perceba que informações sobre ele estão sendo armazenadas na medida em que utiliza o sistema. Esta prática traz algumas questões relacionadas à privacidade.

Pesquisas mostram que a maior parte dos usuários está disposta a fornecer informações suas para que possam receber ofertas personalizadas. No entanto, o censo americano mostra que 75% dos usuários daquele país se preocupam com a possível divulgação de dados que fornecem às empresas [Torres 2004]. Os usuários buscam sempre conhecer os objetivos da coleta de dados e se estes dados serão fornecidos a terceiros.

Portanto, a utilização de Sistemas de Recomendação não teria problema algum se as empresas que os empregam tivessem uma coleta de dados associada a uma política de privacidade adequada. Entretanto, não é incomum encontrar empresas, que além de coletar dados para personalizar o relacionamento com seus clientes, vendam estes dados para outras empresas, uma prática que alimenta o aumento de *spam*. Atualmente não existem leis de privacidade *online* muito claras, sendo importante que as políticas de privacidade das empresas sejam precisas.

Uma política de privacidade não é apenas uma exigência legal, mas também a única forma de garantir que uma empresa agirá com honestidade e protegerá as informações dos seus clientes. No momento em que existir um padrão e uma forma simples de identificação dos clientes, as políticas de privacidade serão mais eficientes.

Uma nova alternativa está sendo desenvolvida pela *World Wide Web Consortium* (W3C) e chamada de *Platform for Privacy Preferences* (PP3). A PP3 é um protocolo que tem como objetivo principal implementar métodos para definir políticas de privacidade padronizadas, as quais podem ser compreendidas e processadas por computadores. Quando um cliente visita um *website* a PP3 é executada e pergunta ao cliente quais as informações ele deseja compartilhar com o *site*, criando assim a política de privacidade. Por exemplo, um cliente talvez deseje informar o seu e-mail e número do telefone com a promessa de que o *site* somente utilizará essas informações em suas iterações e com a devida permissão. Algumas empresas estão propondo selos que garantem que a política de privacidade de um *website* é adequada e cumprida com rigor.

#### **4. Estratégias de recomendação**

No universo da Internet, os principais objetivos dos sistemas de recomendação são a fidelidade e o conseqüente aumento da lucratividade das empresas. Diferentes estratégias podem ser empregadas para personalizar ofertas para um usuário, cada uma exigindo um grau de complexidade distinto no tratamento das informações coletadas. As estratégias mais utilizadas são descritas nas subseções a seguir.

##### **4.1. Listas de recomendação**

Esta estratégia consiste em manter listas de itens organizados por tipos de interesses. Não há necessidade de análise mais profunda de dados do usuário para criação destas listas, apenas a observação dos tipos de itens mais populares, e ordenação destes em grupos tais como: "Itens mais vendidos", "Idéias para presentes", entre outros. A figura 4.1 apresenta um exemplo de lista de recomendações da Livraria Cultura. No exemplo é apresentada uma lista dos livros mais vendidos na seção de informática. No menu vertical na esquerda da imagem é possível observar outras listas, como *sugestões*, *novidade* e *pré-vendas*.



FIGURA 4.1: Lista dos livros mais vendidos em informática (livrariacultura.com.br)

A principal vantagem neste tipo de estratégia está na facilidade de implementação. Basta manter-se listas de acordo com as necessidades de marketing, de aumento de lucratividade, de disseminação de informações (no caso de recomendação de conteúdo), etc. A desvantagem é que as recomendações não são dirigidas a cada usuário independentemente, mas sim a todos os usuários sem distinção.

#### 4.2. Avaliações de usuários

Uma das estratégias mais utilizadas em sistemas de recomendação são as avaliações dos usuários. Ou seja, além de comprar um produto o usuário também deixa um comentário sobre o item adquirido. É muito comum vermos opinião de usuários na forma de ícones dispostos ao lado do item visualizado. Tal mecanismo serve como forma de compartilhamento de informações, fornecendo subsídios aos usuários para seleção de determinado item.

Name	Ratings CNET Editors   Users	Date added	Total downloads	
BatchPhoto 1.3.6	User rating: 	08/09/2006	4,339	<input checked="" type="checkbox"/> Download Now
The Print Shop Essentials 21	User rating: 	10/16/2005	32,332	<input checked="" type="checkbox"/> Download Now
Picture Library 1.4 build 85	User rating: 	11/30/2005	10,265	<input checked="" type="checkbox"/> Download Now

FIGURA 4.2: Avaliações de softwares (download.com)

**MP3 Player Memória Portátil Digital c/ LCD iPod Nano 2GB Apple Preto**

**DISPONIBILIDADE**  
Em Estoque: Entrega 1 dia útil para Grande São Paulo\*  
[Outras localidades?](#)

Avaliação dos clientes: ★★★★★

ADICIONAR À LISTA DE DESEJOS

**AMBIENTE 100% SEGURO**  
Clique e veja mais detalhes

De:	R\$ 4.340,00
Submarino:	R\$ 1.190,00
Economize:	R\$ 150,00

**COMPRA**

ou 12X de R\$ 99,17 sem juros no cartão

**SAIBA COMO PAGAR**  
Clique e conheça as formas de pagamento

Nunca algo tão pequeno reuniu tantas possibilidades. Até 500 músicas. Fotos e capas de disco em cores. Podcasts. Audiobooks. Jogos, relógios, contatos e calendários. Tudo em um iPod que pesa menos ... [Leia mais.](#)

FIGURA 4.3: Avaliação de equipamento (submarino.com.br)

As avaliações dos clientes são muito úteis para assegurar outros consumidores da qualidade e utilidade dos produtos comercializados. No entanto, para que um sistema possa funcionar corretamente com base nos comentários do usuário, é preciso que haja veracidade das opiniões fornecidas. Alguns sistemas propõem mecanismos para incentivar os usuários a contribuir com opiniões verídicas, através de brindes, bônus, etc. Do ponto de vista de implementação, este também é um mecanismo fácil de implementar, na medida em que não exige nenhum tipo de dispositivo inteligente. Basta armazenar e disponibilizar as avaliações de usuários sobre os itens tratados, e apresentar estas opiniões no momento apropriado.

### 4.3. Recomendações por associação

Este tipo de recomendação é obtido através de técnicas capazes de encontrar em uma base de dados associações entre itens avaliados por usuários (comprados, lidos,...). É outro tipo de recomendação muito comum em *websites* de comércio eletrônico, como mostra a figura a seguir.

RELACIONAMENTO SICILIANO (11) 3649-4747 ou 0300-789-3649 | A Siciliano | Nossas Lojas | Franquias

CADASTRO ATENDIMENTO ONLINE STATUS DO PEDIDO PEDIDO ATUAL

HOME LIVROS NACIONAIS LIVROS IMPORTADOS PROMOÇÕES

**SICILIANO**

**MEGA BUSCA**

Título

[Pesquisa avançada >>](#)

**SEÇÕES**

- ARTES, ESPORTES E LAZER
- AUTO-AJUDA
- CIENCIAS
- CULINARIA
- DICIONARIOS
- DIREITO
- INFANTO JUVENIS
- INFORMATICA
- LITERATURA
- LIVROS DIDÁTICOS
- LIVROS INFANTIS
- MEDICINA
- NEGOCIOS
- PSICOLOGIA
- RELIGIOES - CRENÇAS
- TURISMO - MAPAS E GUIAS DE VIAGEM

**Faça Um Site Dreamweaver 8**  
Carlos A. J. Oliviero

ISBN: 8536501154  
Editora: ERICA  
Número de páginas: 224  
Encadernação: Brochura  
Edição: 2006

- Leia a sinopse
- Outros títulos do autor

De: **R\$ 59,00** Por: **R\$ 54,28**  
Desc.: **R\$ 4,72**

Prazo de entrega:  
Normalmente de 3 a 6 dias úteis

**QUEM COMPROU ESSE PRODUTO TAMBÉM COMPROU:**

- Adobe Photoshop Cs2**  
Renato Nogueira Perez  
Avila
- Flash 8 Profissional**  
Silvana Tauhata Ynemine

FIGURA 4.4: Recomendação por associação (siciliano.com.br)

No exemplo ilustrado na figura 4.4, a partir da seleção de um livro sobre construção de *websites*, outras obras relacionadas são apresentadas. Este tipo de recomendação, chamado aqui de recomendação por associação, é a forma mais complexa de recomendação. Ela exige uma análise mais profunda dos hábitos do usuário para a identificação de padrões e recomendação de itens com base nestes padrões.

#### 4.4. Associação por conteúdo

Também é possível fazer recomendações com base no conteúdo de determinado item, por exemplo um autor, um compositor, um editor, etc. Para possibilitar este tipo de recomendação, é necessário que se encontrem associações num escopo mais restrito. Por exemplo: os livros A e B de determinado autor são frequentemente vendidos em conjunto. A figura abaixo mostra este tipo de recomendação em uma livraria virtual.



**Word of Mouse: The Market Power of Collaborative Filtering**  
by [John Riedl](#), [Joseph A. Konstan](#), [Eric Vrooman](#)

**List Price:** \$40.95  
**Price:** \$8.76  
**You Save:** \$2.19 (20%)  
**Availability:** Available for download now

**READY TO BUY?**  
**Add to Shopping Cart**

**Customers who bought titles by John Riedl also bought titles by this author:**

- [Soumen Chakrabarti](#)

▶ [Explore similar authors](#)

FIGURA 4.5: Recomendação por associação de conteúdo (amazon.com)

No exemplo, o sistema recomenda o autor Soumen Chakrabarti para um usuário que selecionou uma obra do autor John Riedl. Diferentes técnicas podem ser utilizadas para se obter este tipo de recomendação.

#### 4.5. Análise de seqüências de ações

Outra forma de fazer a seleção de recomendações é através da análise de seqüências de ações do usuário, como a sucessão de cliques efetuados por ele durante uma visita. Este tipo de avaliação permite determinar, por exemplo, quais são os itens adquiridos pela maior parte dos usuários depois de consultarem determinado item. No exemplo abaixo, o sistema indica que 65% dos usuários que visualizam o livro *Prioritizing Web Usability* adquirem o próprio livro; 14% compram a obra *Don't Make Me Think: A Common Sense Approach to Web Usability*; 10% compram *Designing Web Usability: The Practice of Simplicity*; e assim sucessivamente.

**Prioritizing Web Usability (Paperback)**  
by Jakob Nielsen, Hoa Loranger

---

**What do customers ultimately buy after viewing items like this?**

- 65% buy the item featured on this page: [Prioritizing Web Usability](#) by Jakob Nielsen ★★★★★ \$31.50
- 14% buy [Don't Make Me Think: A Common Sense Approach to Web Usability](#) by Steve Krug ★★★★★ \$35.00
- 10% buy [Designing Web Usability : The Practice of Simplicity](#) by Jakob Nielsen ★★★★★ \$29.70
- 6% buy [Designing Interfaces](#) by Jenifer Tidwell ★★★★★ \$32.97
- 5% buy [Homepage Usability: 50 Websites Deconstructed](#) by Jakob Nielsen ★★★★★ \$25.19

[Compare these items](#) [Explore Similar Items](#)

FIGURA 4.6: Análise de seqüências de ações (amazon.com)

Esta é mais uma forma de apontar relacionamentos entre itens através de uma busca automatizada em um cadastro de compras/consultas.

#### 4.6. Suas Recomendações

Este tipo de seção é inteiramente dedicado a sugestões feitas para o usuário, trazendo recomendações a partir da utilização de diferentes técnicas. A figura a seguir apresenta um exemplo de página de recomendação onde são trazidas sugestões a partir de dados coletados de maneira implícita e explícita.

No exemplo apresentado, há uma lista de produtos recomendados. Na figura, vemos quatro itens, e alguns *links* de seções de interesse. O sistema salienta, ao alto da página, que se trata de uma área personalizada, para busca de itens já adquiridos, recomendações, entre outros. Os itens apresentados foram selecionados de acordo com informações inferidas a partir dos dados de navegação e de compra do usuário.

The screenshot shows the drugstore.com website interface. At the top, there's a navigation bar with categories like home, prescriptions, health, personal care, hair care, beauty, skin care, men's grooming, vitamins, GNC, diet & fitness, food, and SALE. A search bar is present. Below the navigation, there's a 'your shopping bag' section with a 'Banana Box' item. The main content area is titled 'Welcome to your store, eliseo!' and includes a 'your recommendations' section. This section lists several products with their prices and 'Buy' buttons: 'Good Home Co. Surface Cleaner Beach Days - 34 fl oz' for \$14.00, 'Mrs. Meyer's Carry-All Cleaning Kit Lemon Verbena - 1 kit' for \$19.99, 'Mrs. Meyer's Counter Top Spray Geranium - 16 fl oz' for \$4.99, and 'Caldrea Counter Top Cleanser Green Tea Patchouli - 8 fl oz' for \$3.20. To the right of the recommendations, there's a 'Shop Your Store' section with links to various departments like natural bath & body, maxi pads, men's hair color, natural lip balm, face cloths, bar soaps, cleaning solutions, natural laundry, bar soap cleansers, and diapers. There are also links to 'customers' favorite brands' and 'customers' favorite departments'.

FIGURA 4.7: Página de recomendações

## **5. Técnicas de recomendação**

Várias tecnologias têm surgido visando a identificação de padrões de comportamento (consumo, pesquisa, etc.) e utilização destes padrões na personalização do relacionamento com os usuários. Estas técnicas fundamentam o funcionamento dos sistemas de recomendação e são apresentadas nas subseções a seguir.

### **5.1. Filtragem de informação**

A demanda por tecnologias de filtragem de informação não é algo novo [Loeb e Terry 1992]. Peter Denning já escrevia em um volume da *Communications of ACM* do ano de 1982 (Peter Denning apud [Loeb e Terry 1992]) sobre a preocupação no que se refere à quantidade de informação que estava sendo gerada pelos diversos tipos de sistema e recebidas pelos usuários. Ele destacava que toda a atenção estava concentrada na geração da informação para suprir as necessidades do usuário. Destacava também que era importante se preocupar com o recebimento da informação, com o controle de processo, de recuperação e filtragem da informação para que esta alcançasse a pessoa que deveria utilizá-la.

A questão portanto é: Como deve ser entendida a expressão filtragem de informação? Para Belvin [Belvin e Croft 1992] filtragem de informação é o nome utilizado para descrever uma variedade de processos que envolvem a entrega de informação para as pessoas que realmente necessitam delas. Com o tempo este termo tornou-se muito popular e artigos técnicos foram escritos descrevendo o emprego de filtragem em diversas aplicações, como e-mail, documentos eletrônicos em escritórios, entre outros. No entanto, a distinção entre filtragem e processos relacionados com recuperação de informações não era sempre clara. As expressões “recuperação de informação” e “filtragem de informação” descrevem soluções que procuram auxiliar na solução de problemas referentes a sobrecarga de informações. A expressão “recuperação de informação” tradicionalmente envolve armazenamento, índices, e tecnologia para recuperação de documentos textuais. Para que a recuperação realmente ocorra o usuário deve descrever sua necessidade de informação em um formulário, com a consulta (*query*) do que necessita. A partir deste momento, o sistema busca casar a necessidade explicitada com os documentos armazenados. Este tipo de abordagem tende a manter uma base com característica mais estática no armazenamento das informações, e a início da interação é provocada pelo usuário.

Diferentemente da recuperação [Foltz e Dumais 1992] a filtragem propõe uma abordagem distinta. Esta abordagem geralmente mantém um perfil dos interesses do usuário. A idéia é que esta abordagem tenha como característica um maior tempo de duração no que tange ao casamento de interesses. Ou seja, a filtragem não se refere ao momento, mas sim a preferências, enquanto a recuperação baseia-se na percepção de uma necessidade do momento. Esta filtragem deve ser aplicada a cada novo item adicionado procurando verificar se este atende ao usuário. A seguir são descritas técnicas de filtragem aplicáveis a sistemas de recomendação.

#### **5.1.1. Filtragem Baseada em Conteúdo**

Segundo Herlocker [Herlocker 2000], por muitos anos os cientistas têm direcionado seus esforços para aliviar o problema ocasionado com a sobrecarga de informações

através de projetos que integram tecnologias que automaticamente reconhecem e categorizam as informações. Alguns *softwares* têm como objetivo gerar de forma automática descrições dos conteúdos dos itens e comparar estas descrições com os interesses dos usuários visando verificar se o item é ou não relevante para cada um [Balabanovic e Shoham 1997]. Esta técnica é chamada de filtragem baseada em conteúdo [Herlocker 2000][Ansari et al. 2000] por realizar uma seleção baseada na análise de conteúdo do itens e no perfil do usuário.

A descrição de interesses do usuário é obtida através de informações fornecidas por ele próprio ou através de ações, como seleção e aquisição de itens. Muitas ferramentas que esta abordagem aplicam técnicas como indexação de frequência de termos (Salton apud [Herlocker 2000]). Neste tipo de indexação, informações dos documentos e necessidades dos usuários são descritas por vetores com uma dimensão para cada palavra que ocorre na base de dados. Cada componente do vetor é a frequência que uma respectiva palavra ocorre em um documento ou na consulta do usuário. Claramente, os vetores dos documentos que estão próximos aos vetores da consulta do usuário são considerados os mais relevantes para ele.

Outros exemplos de tecnologias aplicadas para filtragem baseada em conteúdo são índices de busca booleana, onde a consulta constitui-se em um conjunto de palavras-chave unidas por operadores booleanos (Cleverdon apud [Herlocker 2000]); sistemas de filtragem probabilística, onde raciocínio probabilístico é aplicado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário (Fuhr, Robertson, Wong apud [Herlocker 2000]); e interfaces de consultas com linguagem natural, onde segundo o autor as consultas são colocadas em sentenças naturais (Jacobs, Lewis, Strzalkowski apud [Herlocker 2000]).

Uma maneira de trabalhar com a filtragem baseada em conteúdo é através de uma solicitação de análise de itens feita ao próprio usuário, onde este deve avaliar alguns itens indicando se estes são de interesse ou não. Uma vez realizada a avaliação, o sistema busca itens que "casam" em conteúdo com o que foi classificado como de interesse, e desconsidera os que "casam" em conteúdo como o que foi classificado de não interesse.

Sistemas deste tipo apresentam limitações como: o conteúdo de dados pouco estruturados é difícil de ser analisado (e.g. vídeo e som); o entendimento do conteúdo do texto pode ser prejudicado devido a uso de sinônimos; pode ocorrer a super especialização, pois o sistema procura se basear em avaliações positivas e negativas feitas pelo usuário, não apresentando conteúdos que não fechem com o perfil.

### **5.1.2.Filtragem Colaborativa**

A abordagem da filtragem colaborativa (FC) foi proposta para atender pontos que estavam em aberto na filtragem baseada em conteúdo [Herlocker 2000] [Ansari et al. 2000]. A Filtragem Colaborativa não se constitui em algo novo e apresenta uma proposta diferenciada dos sistemas de filtragem uma vez que não requerem compreensão ou conhecimento de conteúdo dos itens.

Nos sistemas colaborativos a essência está na troca de experiências entre as pessoas que possuem interesses comuns. Nestes sistemas os itens são filtrados baseado na avaliação feita pelos usuários daqueles itens, ao invés do conteúdo dos itens.

Sistemas de Recomendação Colaborativa procuram prever a pontuação de um item para um consumidor em particular baseado em como outros usuários com gostos semelhantes ao usuário alvo previamente pontuaram o mesmo item. Mais formalmente, a pontuação  $R(u,i)$  do item  $i$  para o usuário  $u$  é estimada baseada na pontuação  $R(u',i)$  atribuída ao mesmo item  $i$  pelos usuários  $u'$  que são similares ao usuário em questão. O valor de uma pontuação desconhecida  $r_{u,i}$  para o usuário  $u$  e item  $i$  é normalmente computado como uma agregação de pontuações de outros usuários (por exemplo, os  $N$  mais similares) para o mesmo item  $i$  [Adomavicius e Tuzhilin 2005]:

$$r_{u,i} = \underset{u' \in \hat{U}}{\text{aggr}} r_{u',i} \quad (1)$$

Onde  $\hat{U}$  denota o conjunto de  $N$  usuários  $u'$  que são mais similares ao usuário  $u$  e que pontuaram o item  $i$ .

Uma questão importante em sistemas colaborativos refere-se a como coletar as preferências dos usuários, como definir no que o mesmo encontra-se interessado. Basicamente existem duas abordagens propostas: uma refere-se à obtenção destas preferências de forma explícita, onde o usuário apresenta explicitamente o que lhe é importante; outra propõe que se aprenda de forma implícita sobre o usuário. A idéia fundamental é aprender sobre o usuário dentro do comportamento que o mesmo apresenta (ações que este realiza).

A FC reforça o conceito de comunidade, na qual cada usuário contribui com suas avaliações para o desempenho geral do sistema. A entrada tradicional de um algoritmo de FC é uma matriz na qual as linhas representam usuários e as colunas os itens. Cada dado na matriz corresponde à avaliação (*rating*) dada por um indivíduo para determinado item.

A técnica de Filtragem Colaborativa pode ser descrita em três fases [Herlocker 2000], explicada do ponto de vista das recomendações realizadas para um usuário, também chamado de usuário alvo (*target*):

1. Calcular o peso de cada usuário do sistema em relação à similaridade com o usuário alvo (métrica de similaridade).
2. Selecionar um subconjunto de usuários com maiores similaridades (vizinhos) para considerar na predição.
3. Normalizar as avaliações e computar as predições ponderando as avaliações dos vizinhos com seus pesos.

Esta técnica também é chamada de “*k-nearest-neighbor*” ou “*user-based*” [Herlocker 2000]. No primeiro passo, a definição da similaridade pode ser realizada através de diversos coeficientes, sendo mais comumente aplicado o coeficiente de correlação de Pearson [Herlocker 2000] [Schardanan e Maes 1995].

(2)

$$w_{a,u} = \frac{\sum_{i=1}^m [(r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)]}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 * \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Na equação (2),  $w_{a,u}$  é a correlação do usuário alvo  $a$  com um determinado usuário  $u$ ,  $r_{a,i}$  é a avaliação que o usuário alvo deu para o item  $i$ , e  $\bar{r}_a$  constitui-se na média de todas as avaliações comuns do usuário alvo  $a$  e usuário  $u$ ; e  $r_{u,i}$  é a avaliação que o usuário  $u$  deu para o item  $i$ , e  $\bar{r}_u$  constitui-se na média de todas as avaliações comuns do usuário  $u$  e usuário alvo  $a$ . Observa-se a necessidade de mais de uma avaliação em comum para que a correlação seja viável, e os resultados variam entre 1 para similaridade total, e -1 para total dissimilaridade. No passo três, o cálculo da predição pode ser efetuado através da equação (3).

(3)

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n |w_{a,u}|}$$

O valor da predição  $p_{a,i}$  do item  $i$  para o usuário alvo  $a$  é a média ponderada das avaliações dadas ao item  $i$  pelos  $n$  vizinhos  $u$  do usuário alvo  $a$ . A quantidade  $n$  de melhores vizinhos (com maiores correlações) é uma escolha de cada sistema que utiliza a FC. Por exemplo, de acordo com a tabela 5.1 o usuário U2 avaliou os itens de A até J com as seguintes pontuações [5;1;2;2;1;7;1;-;6;5] (em uma escala de avaliação de [1-7] referindo-se a itens menos interessantes (1) a itens mais interessantes (7), conforme sua avaliação). Como se pode verificar, o item H não foi pontuado (avaliado) por este usuário.

Tabela 5.1: Usuários X Itens

Usuários	Itens									
	A	B	C	D	E	F	G	H	I	J
U1	1	-	-	2	1	7	-	-	2	3
U2	5	1	2	2	1	7	1	?	6	5
U3	1	5	6	4	3	1	2	-	-	2
U4	6	5	4	1	2	-	-	5	-	3
U5	1	-	2	5	6	7	-	1	-	4

Usuários	Itens									
	A	B	C	D	E	F	G	H	I	J
U6	-	2	-	3	-	1	-	2	-	-
U7	4	2	2	-	-	6	1	6	-	6
U8	-	1	-	3	7	6	-	3	-	1
U9	6	-	-	6	6	2	-	-	5	6

De acordo com a matriz apresentada, o usuário U2 não gostou do item B, avaliando-o com o valor 1. Por outro lado, este usuário gostou do item I, avaliando-o com valor 6. Pode-se observar também que o usuário U2 tende a concordar com U7 e discordar de U3 de acordo com o valor de pontuação apresentado.

Portanto, a filtragem colaborativa recomenda itens de acordo com a similaridade de avaliações fornecidas pelos usuários para os mesmos itens. Para medir essa similaridade, pode-se utilizar o coeficiente de Pearson [Shardanand e Maes 1995].

Tendo novamente como exemplo os usuários U2 e U7 da tabela 5.1, as avaliações dos itens em comum seriam as pontuações [5;1;2;7;1;5] e [4;2;2;6;1;6] respectivamente, gerando uma média de 3,5 para ambos usuários. Aplicando a equação (2), obtêm-se o valor de 0,94. Como se pode observar, os usuários U2 e U7 são bastante similares, tendo em vista que eles concordam em suas avaliações. Tendo como usuário alvo U2 e aplicando o algoritmo do coeficiente de Pearson em todos usuários, têm-se os seguintes resultados apresentados na tabela 5.2:

Tabela 5.2: Similaridade através do coeficiente de Pearson

Usuário	Pearson (U2)
U1	0,63
U3	-0,66
U4	0,34
U5	0,09
U6	-0,78
U7	0,94
U8	0,11
U9	-0,66

De acordo com os resultados da tabela 5.2 observa-se que o usuário U2 é bastante similar a U7, relativamente similar a U4 e não é similar a U3, U6 e U9 em gostos. Uma vez, descobertos quais são os usuários similares ao usuário alvo, pode-se então gerar a predição de novos itens para este usuário.

Aplicando-se a equação (3) para prever a nota que o usuário U2 atribuiria ao item H, levando em consideração todos os itens que os usuários vizinhos, com limiar superior a 0,7 (limite assumido), pontuaram em comum com o usuário alvo, tem-se o resultado de 5,22. Este resultado significa que caso o usuário U2 consumisse o item H este daria como feedback uma nota igual a 5,22 (predição) para este item (tendo como base a avaliação dos vizinhos mais próximos).

O autor Herlocker apresenta uma revisão completa dos objetivos, base de dados, e algoritmos referentes a filtragem colaborativa em [Herlocker 2004].

### **5.1.3. Filtragem Híbrida**

A abordagem da filtragem híbrida procura combinar os pontos fortes da filtragem colaborativa e filtragem baseada em conteúdo visando criar um sistema que possa melhor atender as necessidades do usuário [Herlocker 2000] [Ansari et al. 2000]. Essa abordagem é constituída de vantagens proporcionadas pela filtragem baseada em conteúdo e pela filtragem colaborativa, unindo o melhor das duas técnicas e eliminando as fraquezas de cada uma.

## **5.2. Descoberta de Conhecimento em Base de Dados (DCBD)**

Uma forma de realizar a extração de conhecimento dos bancos de dados dos usuários constitui-se na aplicação de técnicas e ferramentas de mineração de dados. Uma das definições de DCBD encontrada na bibliografia é a que segue:

*“...um processo de extração não trivial de informações potencialmente úteis, as quais não são previamente conhecidas e encontram-se implícitas em grandes coleções de dados” [Zaiane 2000].*

As próximas seções apresentam diferentes tipos de descoberta de conhecimento aplicáveis nos sistemas de recomendação.

### **5.2.1. Tipos de descoberta**

O desenvolvimento de sistemas de DCBD está relacionado com diversos domínios de aplicações, como por exemplo: análises corporativas, medicina, biologia, mercado, etc [Mccarthy 2000]. De acordo com a aplicação algumas técnicas de mineração podem ser mais eficientes do que outras.

#### **a) Regra de associação**

Neste tipo de função cada tupla da base de dados consiste em um conjunto de atributos denominados itens [Cabena 1997] [Agrawal 1993]. Cada tupla corresponde a uma transação, e um item pode assumir um valor verdadeiro ou falso (binário). Uma regra de associação constitui-se em um relacionamento  $X \Rightarrow Y$ , onde  $X$  e  $Y$  são conjuntos de itens e a intersecção entre os mesmos constitui-se em um conjunto vazio. Cada regra de associação é associada a um fator de suporte (FSup), e a um fator de confiança (FConf). O FSup constitui-se na razão do número de tuplas que satisfazem  $X$  e  $Y$ , sobre o número total de tuplas ( $FSup = |X \cup Y| / N$ ). O FConf constitui-se na razão do número de tuplas que satisfazem  $X$  e  $Y$  sobre o número de tuplas que satisfazem  $X$  ( $FConf = |X \cup Y| / |X|$ ).

A função de extração de regras a partir de um banco de dados consiste em encontrar todas as regras possíveis com FSup e FConf maiores ou iguais a um FSup e FConf especificada pelo usuário. A tabela 5.3 apresenta um exemplo de conjunto de dados que pode ser minerado para identificar se determinado item tem ou não probabilidade de ser adquirido. A primeira coluna apresenta um identificador (ID) da

transação, e as outras colunas indicam se um determinado item foi ou não adquirido na transação correspondente. Suponha que o usuário especificou os parâmetros  $FSup = 0.3$  e  $FConf = 0.8$ .

Tabela 5.3: Entrada de dados para a descoberta de regras de associação

ID	Artigo1	Artigo2	Artigo3	Artigo4	Artigo5	Artigo6	Artigo7
1	Não	Sim	Não	Sim	Sim	Não	Não
2	Sim	Não	Sim	Sim	Sim	Não	Não
3	Não	Sim	Não	Sim	Sim	Não	Não
4	Sim	Sim	Não	Sim	Sim	Não	Não
5	Não	Não	Sim	Não	Não	Não	Não
6	Não	Não	Não	Não	Sim	Não	Não
7	Não	Não	Não	Sim	Não	Não	Não
8	Não	Não	Não	Não	Não	Não	Sim
9	Não	Não	Não	Não	Não	Sim	Sim
10	Não	Não	Não	Não	Não	Sim	Não

As regras de associação descobertas a partir dos dados da tabela 5.3 são apresentadas a seguir, aplicando-se os valores de  $FSup$  e  $FConf$  maiores ou iguais aos especificados pelo usuário.

Conjunto de itens frequentes: Artigo2, Artigo4.  $FSup = 0.3$   
 Regra: Se (Artigo2) então (Artigo4).  $FConf = 1$ .

Conjunto de itens frequentes: Artigo4, Artigo5.  $FSup = 0.4$   
 Regra: Se (Artigo4) então (Artigo5).  $FConf = 0.8$ .  
 Regra: Se (Artigo5) então (Artigo4).  $FConf = 0.8$ .

Estas regras descobertas poderiam servir como indicativo de artigos que devam ser oferecidos de forma casada, uma vez que são acessados ou adquiridos juntamente (segundo o que revelou a amostra da base de transações). Através das regras de associação torna-se possível identificar associações entre diferentes itens que participaram ou não de um conjunto significativo de transações.

## b) Classificação

Nessa função cada uma das tuplas pertence a uma das classes entre um conjunto pré-definido de classes [Cabena 1997] [Quinlan 1993]. A classe de uma tupla é indicada por um valor especificado pelo usuário em um atributo meta, ou atributo objetivo. As tuplas consistem de atributos preditivos e um atributo objetivo, o último indicando a qual classe a tupla pertence. O princípio maior da classificação consiste em descobrir algum relacionamento entre os atributos preditivos e o atributo objetivo, para revelar um conhecimento que possa ser utilizado para prever classes de uma tupla desconhecida, ou seja, que não possui uma classificação.

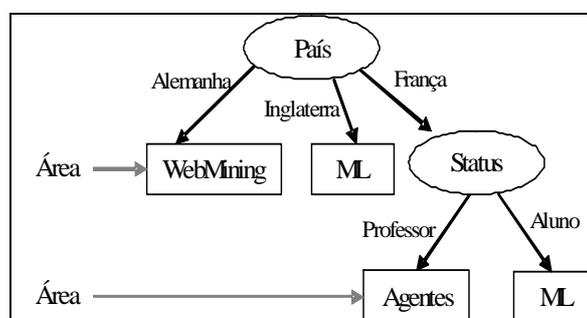
Um exemplo possível para explicar o uso de classificação é o que segue: supondo que um sistema de recomendação guarde algumas informações sobre os seus usuários (por exemplo, *status* na Universidade, país e área de interesse), seria interessante utilizar esses dados para prever que tipo de usuário estaria mais interessado em qual área de pesquisa.

O sistema poderia então concentrar os esforços de recomendação para classificar os usuários. Para prever se o usuário estará ou não interessado em determinada área pode ser feito o uso de uma árvore de decisão [Holte e Yan 1996].

Considerando apenas os atributos *status*, país e área de interesse dos usuários como relevantes para a previsão, são apresentados na figura 5.1 (a) os valores dos atributos selecionados, junto com o valor do atributo objetivo e área de interesse. A árvore de decisão gerada com a aplicação de classificação para o exemplo acima poderia apresentar o formato apresentado na figura 5.1 (b).

Status	País	Área de Interesse
Professor	França	Agentes
Pesquisador	Inglaterra	ML
Aluno	França	ML
Pesquisador	Inglaterra	ML
Professor	França	Agentes
Aluno	Alemanha	WebMining
Aluno	Alemanha	WebMining
Pesquisador	Inglaterra	ML
Professor	França	WebMining
Professor	França	ML
Aluno	França	ML

(a)



(b)

FIGURA 5.1: (a) Dados preditivos e objetivo; (b) Árvore de Decisão

O conhecimento descoberto é frequentemente representado na forma de regras do tipo SE-ENTÃO. Abaixo são descritas as regras obtidas através da classificação:

- Se (PAÍS = Alemanha) então Área = WebMining
- Se (PAÍS = Inglaterra) então Área = ML
- Se (PAÍS = França e Status = Professor) então Área = Agentes
- Se (PAÍS = França e Status = Aluno) então Área = ML

Observa-se através destes resultados que fica mais fácil para um sistema de recomendação decidir o que recomendar no caso de um novo usuário cadastrado.

A classificação é um tipo de método classificado como aprendizado supervisionado, uma vez que definimos qual é o atributo objetivo e os atributos preditivos [Cabena 1997]. Em um sistema de recomendação poderíamos utilizar a classificação para identificar a classe de itens novos em categorias pré-definidas ou mesmo para classificar usuários pelos seus perfis.

### c) Agrupamento

De uma forma natural as pessoas procuram visualizar os dados segmentados em grupos discretos, onde o que permite a união destes objetos em um mesmo grupo é a similaridade existente entre os mesmos [Cabena 1997].

Nesta função de DCBD, o algoritmo de agrupamento automaticamente deve criar grupos através de partições de um banco de dados em conjunto de tuplas similares. A partição acontece de forma que as tuplas de valores e atributos semelhantes sejam reunidas nos mesmos grupos. Um agrupamento de qualidade surge onde a similaridade intra-classe é alta e a inter-classe é baixa. O agrupamento é um tipo de aprendizado não-supervisionado uma vez que fica a cargo do algoritmo a definição de quais atributos e seus valores serão utilizados para montar agrupamentos de dados similares.

Um dos métodos de agrupamento mais utilizado é o *k-means*. O primeiro passo no processo é encontrar  $k$  casos para servirem de “semente” para o processo, por exemplo os  $k$  primeiros casos de um banco de dados.

O segundo passo no processo é encontrar no banco de dados, para cada um dos registros armazenados, a semente mais próxima dele. A figura a seguir ilustra este processo, onde os pontos escuros representam as sementes, os pequenos círculos transparentes representam os outros registros do banco de dados, e os grandes círculos tracejados representam os grupos.

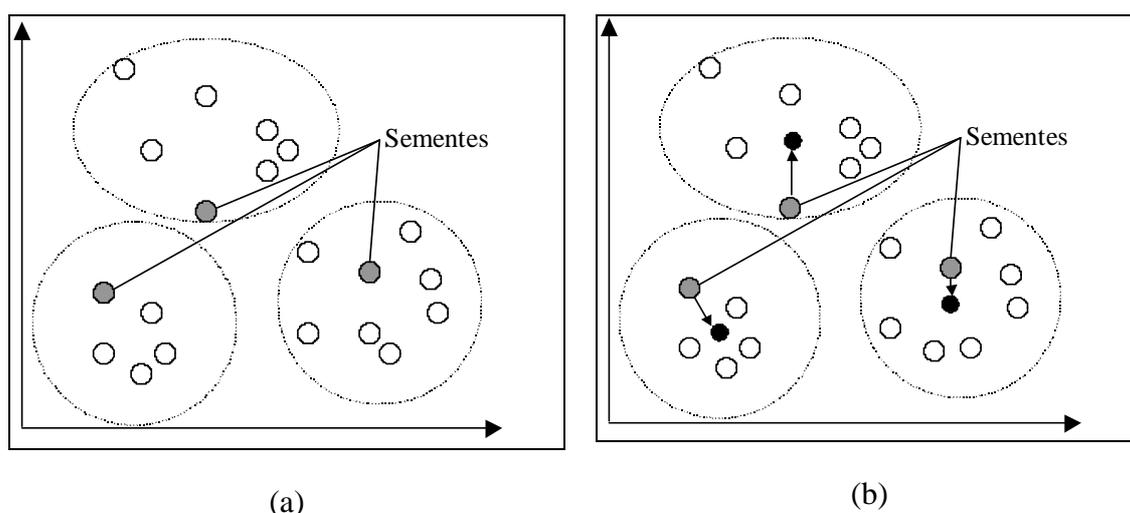


Figura 5.2: (a) Formação inicial de grupos; (b) Recálculo dos centrúides

Neste momento, todos os registros já foram associados a cada um dos grupos. O próximo passo do algoritmo é determinar os novos centróides a partir de todos os elementos existentes em cada cluster, como exemplifica a figura a seguir.

Este passo é então repetido até que os centróides não precisem ser recalculados. Uma utilidade para agrupamentos poderia ser a reunião de usuários por características comuns ou similares. Um sistema de recomendação pode empregar estes agrupamentos para oferecer itens apropriados de acordo com as características de cada grupo.

## **6. Sistemas de recomendação: exemplos de aplicações**

Nesta seção serão descritos alguns exemplos de Sistemas de Recomendação, tanto acadêmicos como comerciais. Deve-se destacar que nenhum dos sistemas descritos a seguir propõe soluções para incluir a relevância da opinião dos usuários nas recomendações geradas, sendo que esta tese de doutorado apresenta preocupação com esta questão e propõe uma maneira de incluir a mesma em Sistemas de Recomendação.

A seguir são descritos alguns Sistemas de Recomendação. Uma listagem completa de Sistemas de Recomendação e comparação entre os mesmos pode ser encontrada em [Adomavicius e Tuzhilin 2005] e [Montaner et al. 2003].

### **6.1. Projetos Acadêmicos**

Abaixo são descritos alguns Sistemas de Recomendação desenvolvidos em projetos acadêmicos:

#### **6.1.1. Referral Web**

Este projeto procurava possibilitar que através de um sistema fosse possível encontrar e visualizar redes sociais, que de acordo com o autor seriam grupos de pessoas ligadas por atividades profissionais [Kautz et al. 1997].

Para criar esta rede o sistema utiliza-se de textos já existentes na Web. Uma vez que o usuário entra com seus dados é feita uma varredura na Web buscando documentos que mencionem o usuário (aplica um motor de busca e recuperação de informações). A partir dos documentos recuperados e relacionados com o usuário, são extraídos nomes de outros indivíduos citados. No final é construída uma rede global de citações representada na forma de um grafo. Neste sistema, a formação da rede está baseada somente na relação entre usuário, documento e citações, a importância e relevâncias dos trabalhos ou indivíduos não são trabalhadas.

#### **6.1.2. RINGO**

O RINGO é um sistema desenvolvido para recomendação personalizada de música e foi desenvolvido no *Massachusetts Institute of Technology* [Shardanand e Maes 1995].

Este trabalho explorava similaridades entre os gostos dos diferentes usuários para recomendar itens. Isto se baseia no fato de que os gostos das pessoas apresentam tendências gerais e padrões entre os gostos das pessoas e entre os grupos de pessoas.

Neste sistema os usuários descreviam suas preferências musicais para o sistema através da avaliação de algumas músicas. Estas avaliações constituíam o perfil dos usuários. O sistema usava estes perfis para gerar as recomendações para usuários individuais.

O Ringo comparava os perfis dos usuários para determinar quais os usuários que apresentavam gostos similares (gostavam dos mesmos álbuns e/ou não gostavam dos mesmos, por exemplo). Para o seu funcionamento, primeiramente usuários similares eram identificados, a partir desta identificação e comparação de perfis, o sistema podia prever o quanto o usuário gostaria de um álbum / artista que ainda não foi avaliado pelo mesmo. Quando o usuário acessava pela primeira vez o Ringo era apresentado a uma lista de 125 artistas. O usuário avaliava os artistas de acordo com o quanto gostava de ouvi-los. Caso o usuário não estivesse familiarizado com o artista ou não possuísse uma forte opinião sobre o mesmo, este era solicitado a não avaliar para não ocorrerem distorções. Os usuários eram aconselhados a pontuar os artistas única e exclusivamente pelo fato de gostarem ou não dos mesmos. Para esta pontuação era aplicada uma escala de avaliação de 7 pontos, sendo o valor 1 (não gosta), 4 (indiferente) e 7 (adora).

A lista de artistas enviada para os usuários era selecionada em duas partes. Uma parte da lista era gerada pelos artistas mais pontuados (Ranqueamento), isto assegurava que um novo usuário tivesse a oportunidade de pontuar artistas que outros já tivessem pontuado, então existiria algo em comum entre os perfis das pessoas. A outra parte da lista era gerada através de uma seleção randômica a partir de um banco de dados sobre os artistas.

Uma vez que o perfil inicial do usuário tivesse sido submetido, o usuário podia pedir ao Ringo por predições, ou seja, uma pessoa podia pedir ao Ringo para: 1) sugerir novos artistas / álbuns que o usuário gostaria de obter ou ouvir; 2) listar artistas / álbuns que o usuário não gostaria; 3) realizar uma predição sobre um artista / álbum específico.

O retorno dado pelo Ringo aos usuários não incluía nenhuma informação em particular sobre a identidade dos outros usuários que contribuíram com as recomendações. Deve-se destacar esta última informação, em sistemas colaborativos ou de informação social, a identidade de quem avaliou deve ser mantida em segredo. O sistema Ringo também permitia a escrita de comentários (como *feedback* explícito) sobre o produto recomendado. Os próprios usuários podiam incluir novos artistas e álbuns no banco de dados. O usuário podia receber uma mensagem sobre os novos artigos, bem como, sobre as novidades do sistema.

### **6.1.3. GroupLens**

O sistema GroupLens [Konstan et al. 1997] foi um projeto implantado usando Filtragem Colaborativa para recomendar leituras na Usenet News<sup>1</sup>. Desenvolvido por pesquisadores da Universidade de Minnesota, considera que o grande volume de

---

<sup>1</sup> *Usenet News* constitui-se em uma idéia inicial de grupos de discussão na Internet. Os tópicos são separados em categorias de grupos, por exemplo, “*alt.rec.humor*” para assuntos de humor e “*comp.lang.java*” com relação à linguagem de programação Java.

informações produzidas nesse meio, somadas às diferenças existentes de perfis de usuários, formam um candidato ideal para uso da FC.

O Sistema coleta a avaliação dos usuários referente aos artigos que estes leram (escala de [1-5] pontos) e utilizava esta avaliação para identificar os vizinhos mais próximos com avaliações semelhantes e predizer se os usuários iriam gostar dos artigos novos baseando-se nas avaliações dos usuários vizinhos.

#### 6.1.4. Fab

O Fab constitui-se em um sistema que procura combinar sistemas de filtragem colaborativa e filtragem baseada em conteúdo (abordagem híbrida), procurando eliminar muitas das fraquezas encontradas em cada uma das abordagens. Foi desenvolvido pela Universidade de Stanford que procurava recomendar páginas Web (documentos) para usuários [Balabanovic e Shoham 1997].

Objetivando a construção de uma arquitetura híbrida, são mantidos perfis de usuários baseado na análise de conteúdo, e estes perfis são comparados diretamente para determinar usuários similares para recomendação colaborativa. O sistema constitui-se em uma implementação distribuída e foi construído como parte do projeto da biblioteca digital da Universidade de Stanford.

O sistema implementava dois agentes, um responsável pela coleta dos documentos na Web (*Collection Agent*) e outro responsável pela seleção de páginas (*Selection Agent*) que seriam apresentadas aos usuários como recomendação, conforme a Figura 6.1 Após a recomendação o usuário poderia avaliar o item aplicando uma escala de avaliação de [1-7] pontos.

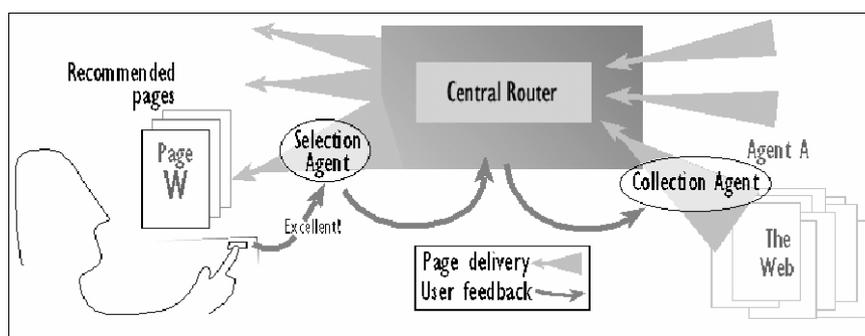


Figura 6.1: Visão geral da arquitetura do Fab [Balabanovic e Shoham 1997]

Buscando criar uma arquitetura híbrida, os perfis dos usuários eram criados usando a análise do conteúdo dos documentos que seriam aplicados para descobrir quais são os usuários semelhantes. Uma vez que os usuários semelhantes fossem obtidos aplicava-se a filtragem colaborativa para recomendar. Sendo assim o usuário recebe recomendação de algo que feche com o seu perfil e com o perfil de seus usuários similares. Com esta arquitetura híbrida o sistema minimizava os problemas existentes nas duas abordagens e conseguia as seguintes vantagens:

- aplicando a abordagem de filtragem colaborativa as experiências dos usuários são levadas em consideração;

- aplicando a abordagem de filtragem baseada em conteúdo torna-se possível levar consideração itens não vistos por outros usuários e caso não existam usuários com gostos semelhantes ao usuário, mesmo assim é possível fazer recomendações.

### 6.1.5.MovieLens

O projeto MovieLens<sup>2</sup> foi desenvolvido por pesquisadores da Universidade de Minnesota constituindo-se em um projeto que visa realizar recomendações de filmes. O projeto é aberto ao público, que pode se cadastrar, efetuar algumas avaliações para formar um perfil inicial para receber as recomendações [Riedl et al. 1999].

Os usuários podem solicitar recomendações geradas por predições, por exemplo, quando se quer saber qual seria a avaliação do usuário para um determinado filme ainda não visto pelo mesmo. A Figura 6.2 apresenta um exemplo de recomendações realizadas pelo MovieLens: nota-se que o filme “Fugitive, The” não foi ainda avaliado, e portanto uma predição está sendo realizada de quatro e meio estrelas.

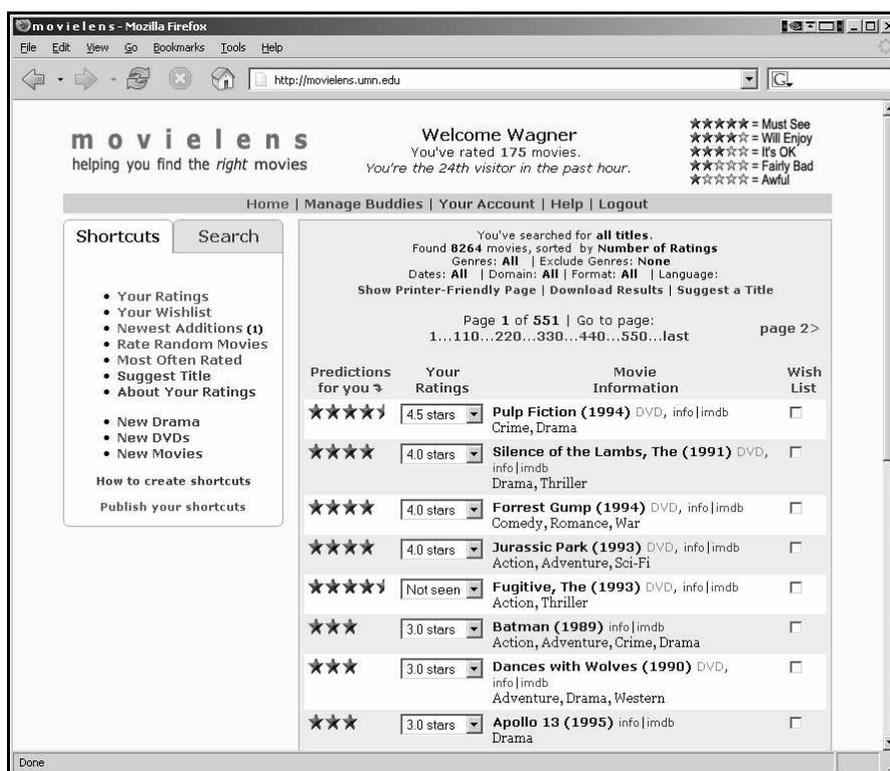


Figura 6.2: Interface de recomendações do MovieLens

Uma das grandes contribuições deste projeto para a comunidade de Sistemas de Recomendação é disponibilizar gratuitamente uma base de dados<sup>3</sup> com cerca de 100 mil

<sup>2</sup> <http://movielens.umn.edu>

<sup>3</sup> <http://www.grouplens.org>

avaliações dadas por seus usuários. Esta base de dados tem sido utilizada por diversos projetos no momento da validação e experimentação de suas propostas.

### 6.1.6.W-RECMAS

O sistema W-RECMAS (*Web-RECommender system based on Multi-Agent System for academic paper recommendation*) constitui-se em um Sistema de Recomendação, baseado em Sistemas Multiagentes e mineração de dados, projetado para recomendar artigos científicos e indivíduos para pesquisadores[Cazella 2005] [Cazella 2006]. Através deste sistema, os pesquisadores recebem recomendações de acordo com seu perfil e terão conhecimento de qual a relevância da opinião do grupo ou do indivíduo que participou da recomendação. Espera-se que esta combinação da métrica de relevância e predição de itens diminua a sobrecarga de informação, pois esta deverá servir de parâmetro para auxiliar o usuário alvo na escolha de qual artigo (item) acessar primeiro.

Para o cálculo do Ranking do Recomendador foram utilizadas informações disponibilizadas pelo servidor do CNPQ<sup>4</sup> e do servidor da Capes<sup>5</sup>, mais precisamente informações do sistema CV-Lattes e do sistema Qualis. A Figura 6.3 apresenta a interface de recomendação de artigos do protótipo. Nesta interface pode-se observar que o usuário está recebendo uma recomendação baseada em um grupo de recomendadores com relevância de opinião média (entre 4-6) na área de interesse da qual o usuário possui interesse. Pode-se observar que o sistema apresenta a predição calculada de opinião do usuário pelo artigo, sendo que o sistema “acredita” que o usuário irá considerar este artigo Bom (4 estrelas).

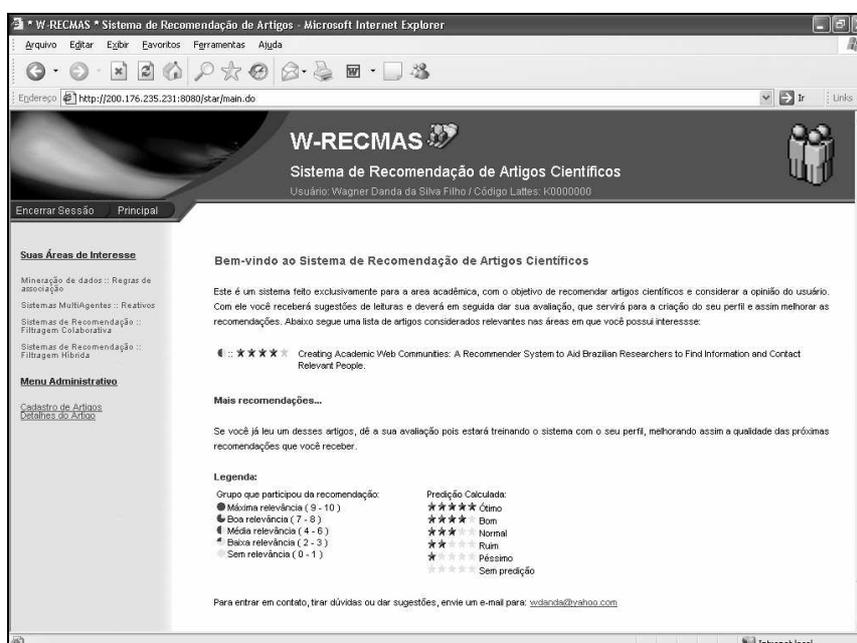


Figura 6.3: Interface de recomendação de artigos no protótipo W-RECMAS

<sup>4</sup> <http://www.cnpq.br>

<sup>5</sup> <http://www.capes.gov.br>

A Figura 6.4 apresenta as legendas utilizadas na interface de recomendação. Em (a) pode-se observar a classificação da relevância da recomendação (Ranqueamento do Recomendador) do grupo que ajudou na formação da recomendação. Em (b) observa-se a classificação das possíveis predições de opinião do usuário alvo da recomendação quanto ao artigo recomendado. Nesta mesma interface o usuário pode fornecer o *feedback* quanto ao artigo recomendado clicando sobre as estrelas.

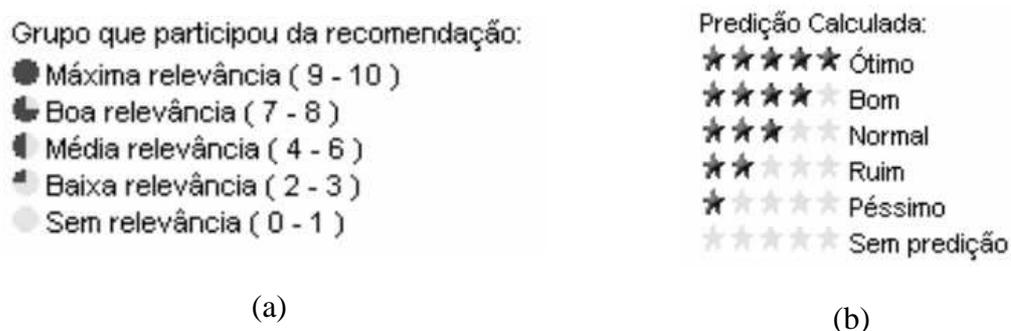


Figura 6.4: (a) Legenda da relevância dos recomendadores; (b) Legenda da predição calculada para recomendação

Maiores informações sobre o sistema e as publicações relacionadas poderão ser encontradas em: <http://www.inf.unisinos.br/~cazella/phd.html>.

## 6.2. Sites Comerciais

Abaixo são apresentados alguns *sites* comerciais que aplicam os conceitos referentes a Sistemas de Recomendação.

### 6.2.1. Amazon.com<sup>TM</sup>

O *site* mundialmente conhecido Amazon ([www.amazon.com](http://www.amazon.com)) é um portal de vendas on-line com diversas categorias de produtos. Este realiza vendas de livros entre outros itens. Os exemplos abaixo são focados na Seção de compras de livros [Schafer et al. 2000].

- **Clientes que compraram:** A Amazon.com utiliza a técnica “Clientes que compraram” que consiste em recomendar livros freqüentemente comprados por clientes que compraram esse mesmo livro em questão. A outra lista de recomendações que o *site* utiliza é recomendar autores cujos livros são freqüentemente comprados por clientes que compraram trabalhos do autor do livro em questão.
- **Amazon.com entregas:** Os clientes selecionam a partir de listas de categorias os seus interesses pessoais para que periodicamente lhes sejam enviados por e-mail as últimas recomendações e novidades nas categorias previamente selecionadas.
- **Classificação de livros:** Essa técnica permite que os clientes façam suas classificações aos livros comprados. Eles podem classificá-los em uma escala que vai do “gostar” ao “detestar”. Aqui se tem um retorno direto do cliente sobre os produtos negociados. Depois de classificar um produto o cliente pode solicitar outras recomendações que serão trazidas conforme o gosto do cliente.

- **Comentários dos clientes:** Essa característica permite que os clientes recebam recomendações textuais sobre o produto pesquisado, baseado na opinião de pessoas que já utilizaram o produto. Na página de cada produto encontra-se a classificação dele em uma escala de avaliação de [1-5] e os comentários escritos dos clientes que compraram. Neste caso valeria ressaltar que não existe nenhuma referência à relevância das opiniões emitidas pelos clientes que avaliaram um produto.

### 6.2.2. eBay

Assim como no site da Amazon.com<sup>TM</sup>, o *site* de leilões *on-line* eBay.com<sup>TM</sup>, acessível no endereço [www.ebay.com](http://www.ebay.com), possui mais de uma variação de estratégias de recomendação, também apresentado por Konstan [Konstan et al. 1997]. Segue a explicação de cada uma destas estratégias [Schafer et al. 2000]:

- **Direito de resposta:** Permite aos compradores e vendedores avaliar o seu parceiro de negócio de acordo com o grau de satisfação da transação de compra. Isto gera uma pontuação que demonstra a confiabilidade de cada negociante (reputação). Quanto mais positiva a pontuação, mais confiável é o negociante, e quanto mais negativa a pontuação, menos confiável é o negociante.
- **Comprador pessoal:** Permite aos clientes indicar os itens que eles têm interesse em comprar, para que, em uma periodicidade definida, o *site* possa enviar os resultados da busca por esses itens.

### 6.2.3. CD NOW

O site CD NOW ([www.cdnow.com](http://www.cdnow.com)) é um portal de venda *on-line* de cd (*compact discs*), vídeos (VHS) e dvd (*digital video disc*). O portal possui vários serviços de recomendação listados abaixo [Schafer et al. 2000]:

- **Recomendador de Álbuns:** O recomendador de álbuns chamado *Album Advisor*, do *site* CDNOW<sup>TM</sup> ([www.cdnow.com](http://www.cdnow.com)) trabalha de dois modos diferentes: no modo de visualização por álbum, o *site* disponibiliza uma página com informações de cada álbum e recomenda também 10 outros álbuns que podem despertar interesse do usuário e na visualização por múltiplos artistas, os consumidores podem selecionar até três artistas e o sistema recomenda 10 álbuns relacionados aos artistas em questão.
- **My CDNOW:** A Seção My CDNOW permite aos consumidores criar sua própria “biblioteca” de músicas, baseada em álbuns e artistas de sua preferência. Os clientes indicam quais são os álbuns e artistas favoritos. Ao comprar do *site*, os artistas e álbuns são incluídos automaticamente na sua lista de adquiridos. Ao entrar nessa lista, o cliente pode, classificar essas informações como “adquirido e preferido” ou “adquirido, porém não preferido”.

Quando os clientes solicitam recomendações, o sistema faz predições de 6 álbuns que o consumidor poderia gostar, baseado na lista de adquiridos e preferidos. A lista de sugestões contém campos para que o usuário preencha e com isso registre um feedback sobre a qualidade da recomendação. Cada álbum sugerido pode ser classificado em “Eu compro”, “Adicionar a lista de desejados” ou “Não é para mim”.

## 7. Tendências e desafios

A pesquisa em sistemas de recomendação apresenta-se como algo promissor e bastante difundida, tendo perdido o caráter totalmente acadêmico uma vez que a aplicação da tecnologia gerada já é aplicada na prática em inúmeros *sites* comerciais. Uma ampla discussão sobre tendências e desafios na área são apresentadas em [Adomavicius, G. e Tuzhilin, A. 2005] e no material disponibilizado em <http://www.grouplens.org/beyond2005/>, referente ao *workshop beyond personalization*. Nesta Seção serão descritas algumas tendências de pesquisas na área.

### 7.1. Formação de Comunidades Virtuais

Com o desenvolvimento das ferramentas tecnológicas, principalmente aquelas promovidas pelo advento da Internet, emergem na sociedade novas formas de relação, comunicação e organização das atividades humanas, entre elas, merecem destaque o estudo de redes sociais virtuais. Estas redes apoiadas por computadores utilizam-se de diferentes recursos, entre eles: e-mails, fóruns, listas de discussão, sistemas de boletins eletrônicos (BBSs), grupos de notícias, Chats, Softwares Sociais como Orkut.com.

Segundo o autor Case [Case et al. 2001], pode-se definir o termo Comunidade como sendo:

*“...um local onde pessoas com interesses em comum reúnem-se para trocar experiências, responder perguntas, promover colaboração, obter conhecimentos e compartilhar casos de sucessos e insucessos”.*

As primeiras formas de Comunidades Virtuais foram criadas baseadas em usuários que poderiam encontrar-se e discutir temas não interessando a distância que estes encontravam-se [Hauben e Hauben 1997] [Figallo 1998]. Conforme Rheingold [Rheingold 2000], uma Comunidade Virtual representa o surgimento de comunidades socialmente motivadas pelo interesse mútuo na Internet. O mesmo autor descreve comunidades virtuais como sendo:

*“...comunidade virtuais são agregações sociais que emergem da rede, quando um número suficiente de pessoas envolvem-se nesta discussão pública por um determinado período, visando formar redes de relacionamento utilizando o cyberspace”.*

Uma comunidade torna-se uma excelente fonte de conhecimento e informação, e talvez o mais importante, esta auxilia na diminuição da duplicação de esforço, uma vez que se pode basear na experiência que os membros tiveram no passado. Por exemplo, pode-se imaginar uma situação hipotética onde um estudante da graduação necessita iniciar uma pesquisa para realizar seu trabalho de conclusão de curso, este estudante poderá perder muito tempo nesta busca por material, passando por caminhos que levarão ao sucesso ou não, caminhos talvez trilhados anteriormente por outros estudantes, ou seja, esforço já despendido por outros. Entretanto, caso este estudante estivesse escrito em uma comunidade com indivíduos com interesses em comum, estes poderiam fornecer lhe alguns atalhos nesta busca por informação, talvez diminuindo o tempo perdido pelo aluno na busca por informação [Cazella e Alvares 2005].

Segundo Case [Case et al. 2001], para se criar uma comunidade é necessário enfrentar alguns desafios específicos: a) os membros da comunidade deveriam ser direcionados diretamente para pessoas que pensam de forma semelhante a sua (*like-*

*minded*) e para informações relevantes; b) os membros da comunidade deveriam ter acesso a informação requerida sem sentirem-se sobrecarregados; c) os membros da comunidade deveriam ser informados sobre outros membros que pensam de forma semelhante a sua (*like-minded*); d) os membros deveriam disseminar informação para as pessoas apropriadas.

Algumas comunidades virtuais são amplamente divulgadas na Internet, possuindo um número significativo de membros segundo apresentado na Tabela 7.1, proposta por Golbek [Golbek 2005]:

Tabela 7.1: Número de membros em comunidades virtuais

Website	URL	Número de membros
Adult Friend Finder	http://adultfriendfinder.com	15.700.000
Alt.com	http://alt.com	2.600.000
Amigos	http://amigo.com	3.500.000
Asia Friend Finder	http://asiafriendfinder.com	6.000.000
Black Planet	http://blackplanet.com	14.000.000
Friend Finder	http://friendfinder.com	3.600.000
LiveJournal	http://livejournal.com	5.700.000
My Space	http://myspace.com	6.000.000
Orkut	http://orkut.com	3.000.000
Tickle	http://tickle.com	18.000.000

Fonte: [Golbeck 2005], p. 16.

Atualmente existem várias pesquisas que fazem uso de comunidades virtuais para viabilizar o funcionamento dos sistemas de recomendação. Estas pesquisas focam o trabalho na identificação de relações de *trust* entre usuários [Massa et al. 2004] [Golbeck 2005]. A abordagem aplicada nestas pesquisas concentra-se na idéia de que recomendações feitas por pessoas nas quais um indivíduo apresenta uma relação de confiança possuem maiores chances de serem aceitas e terem melhores *ratings*. O grande desafio aqui se concentra em como identificar estas relações de confiança. Esta relação pode ser explicitada pelo usuário (ver [Golbeck 2005]) e a partir desta informação é possível personalizar com base nas pessoas em que o usuário confia, ou pode ser inferida quanto a “proximidade” destas relações e um peso pode ser atribuído as mesmas (ver [Massa et al. 2004]).

## 7.2. Personalização de Ambientes Virtuais

Os avanços tecnológicos permitiram uma evolução importante nas interfaces humano-computador, onde a Internet teve um papel muito importante, inicialmente através da simples troca de textos e arquivos, passando pelos documentos e hiperdocumentos, e mais recentemente com os conteúdos multimídia (textos, áudios, imagens e vídeos integrados). Atualmente, com o barateamento e a popularização das placas de hardware aceleradoras para o processamento gráfico 3D, houve uma grande expansão da quantidade de aplicações gráficas disponíveis, destacando-se principalmente os jogos e os ambientes gráficos 3D. A interação humano-computador nestes ambientes tridimensionais ocorre de uma forma mais natural, uma vez que aproximamos o “nosso mundo real” do “mundo virtual”, que é apresentado pelo computador. As interfaces orientadas a menus, as aplicações convencionais e mesmo o uso tradicional da Internet, requerem do usuário um certo domínio prévio de conhecimentos relativos aos procedimentos de interação com estes ambientes, onde o acesso a ambientes virtuais

tridimensionais, pode usualmente ser feito por “leigos”, e até mesmo por crianças ainda não alfabetizadas, com um mínimo de conhecimentos prévios.

Por outro lado, podemos reaproveitar muitos dos conceitos e técnicas que já vem sendo usados nos ambientes tradicionais baseados em textos e documentos, adaptando-os para os ambientes virtuais 3D. Portanto, a personalização de páginas e de conteúdos disponibilizados na Web pode tomar uma forma mais sofisticada se for considerada a interação e adaptação de Ambientes Virtuais em 3D [Osório et al. 2004]. A personalização de ambientes virtuais 3D é um tema bastante recente de pesquisa, onde entretanto, podemos encontrar trabalhos que propõem formas diferentes de personalização do ambiente virtual e de seu uso em problemas práticos que abrangem aplicações de: e-commerce (lojas virtuais), ensino à distância (ambientes virtuais de aprendizagem) e jogos [Chittaro e Ranon 2000, 2002][Santos 2004][Santos e Osório 2004a, 2005]. Por trás da idéia de adaptação de ambientes virtuais temos o fato de que é relativamente fácil traçar o perfil comportamental do usuário em um ambiente tridimensional, uma vez que podemos “rastrear” precisamente por onde ele passou, para onde olhou (e.g. métodos do tipo *line-of-sight*) e as interações que teve com os elementos do ambiente virtual. Além disto, outras técnicas de coleta de perfil, como por exemplo, a coleta explícita de dados, o monitoramento do uso de ferramentas de busca, a aquisição de itens, podem também continuar a ser usadas para completar o perfil do usuário em um ambiente virtual 3D. De posse das informações sobre o perfil do usuário, e sobre os itens (categorias, posição espacial, etc), podemos recomendar uma reorganização espacial do ambiente, de modo a reposicionar os objetos neste ambiente.

Esta técnica de reorganizar os produtos em uma loja (no mundo real) é bastante usada no comércio tradicional, onde em alguns supermercados os produtos são expostos de modo a destacar promoções, incentivar a venda de um produto específico colocando ele em uma posição mais visível/acessível (notamos muitas vezes uma diferença de preço entre produtos colocados diretamente a nossa frente e outros produtos em prateleiras mais altas ou baixas, de acesso mais difícil), ou mesmo, colocando produtos próximos uns dos outros de modo a estimular que o cliente leve ambos (usando inclusive técnicas de mineração e análise da “cesta de compras” para determinar quais produtos devem ficar lado a lado). Em um mundo virtual o reposicionamento de objetos é uma tarefa trivial, mais fácil do que no mundo real, e pode envolver não apenas os objetos dispostos em prateleiras ou seções, mas podemos inclusive reposicionar uma loja inteira dentro de um shopping virtual, onde por exemplo, é possível mudar uma loja inteira de eletrônicos trocando de posição com a loja de perfumaria e cosméticos. A figura 7.1 apresenta um exemplo de loja virtual que se adaptam aos seus clientes.

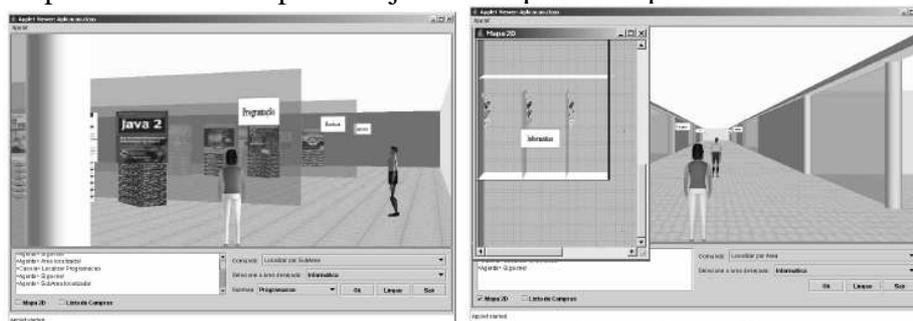


FIGURA 7.1: Loja virtual - Livraria Personalizada (Santos e Osório 2004a)

Para realizar esta adaptação do ambiente virtual 3D, são necessários os perfis dos usuários, assim como informações de categorias e de descrição de itens. Técnicas de recomendação de novos itens também podem usadas para adicionar novos itens que possam ser de interesse dos usuários. Dentre os principais trabalhos que propõem o uso de modelos de usuários na adaptação de um ambiente virtual podem ser citados os desenvolvidos por [Chittaro e Ranon 2000, 2002] e o modelo AdapTIVE desenvolvido por [Santos 2004]. Uma descrição mais detalhada do funcionamento destes sistemas pode ser encontrada junto as referências citadas nesta seção e em [Osório et al. 2004].

### 7.3. Web Semântica

Um dos grandes desafios atuais em relação aos conteúdos disponibilizados através da Internet, é transformar estas informações representadas usualmente apenas por um conjunto de dados (semi)estruturados em verdadeiros conhecimentos. A idéia por trás da Web Semântica [Librelotto et al. 2005] é a introdução de anotações e a associação de significados aos dados disponibilizados na Web. Sendo assim, pretende-se dar um significado, uma interpretação para estas informações, e mais além, deseja-se possibilitar também a manipulação e compartilhamento destas informações de modo inteligente e automatizado [Berners-Lee et al. 2001]. Desta forma, o significado dos dados passa a ser “entendido” tanto pelo homem, quanto pela máquina.

Na seção anterior discutimos sobre os ambientes virtuais, onde os objetos possuem uma descrição gráfica 3D, entretanto apenas esta descrição geométrica não permite que se associe um significado aos dados, como por exemplo, uma lixeira (cesta de lixo) virtual é um lugar onde colocamos coisas que não nos interessam mais e uma outra cesta (cesta de compras) é um lugar onde colocamos os objetos que nos interessam. A geometria pode ser a mesma, mas o sentido associado ao seu uso é bastante diferente. Este mesmo problema ocorre nas páginas Web, que apesar de apresentar conteúdos sob a forma de textos, muitas vezes não possuem associado um contexto ou significado. Isto pode levar aos problemas típicos das atuais ferramentas de busca, onde ao buscar informações sobre “*Recomendação e Privacidade*” (relativas à personalização na Web) obtemos como resposta a “*recomendação* de um restaurante com um ambiente que oferece uma total *privacidade* para reuniões”; a busca por palavras-chave pode até ter funcionado corretamente, mas o contexto da resposta obtida não se enquadra no contexto desejado. Para resolver problemas como estes, tem sido largamente usada na Web a adição de anotações que descrevem de forma mais padronizada (passível de interpretação e manipulação por ferramentas computacionais) os dados. O padrão recomendado pelo W3C (*World Wide Web Consortium*) para ser adotado em anotações na Web é o XML [Bray et al. 2000].

A anotações em XML permitem uma melhor descrição dos dados, entretanto, é necessário um nível maior de abstração, que pode ser obtido através da utilização de uma linguagem voltada para a representação de conhecimentos. As linguagens usadas este fim vem sendo pesquisadas junto a Inteligência Artificial, como por exemplo: Redes Semânticas, Frames, Taxonomias, Ontologias, entre outros formalismos. O uso de Ontologias [Gruber 1993][Ushold e Gruninger 1996] permite que seja feita uma especificação explícita de uma conceitualização, tornando explícita a ligação entre a forma e o conteúdo da informação. Uma Ontologia descrevendo um domínio de conhecimentos irá permitir e facilitar o acesso aos dados e conhecimentos, a interopera-

bilidade entre sistemas e a comunicação pela troca de conhecimentos e não apenas de dados brutos. Destaca-se aqui a OWL (*Web Ontology Language*) e ferramentas como o *Protege* [Gennari et al. 2003], para a modelagem e manipulação de Ontologias.

A importância e os impactos da Web Semântica, através do uso de anotações e de Ontologias, junto aos sistemas de recuperação de informações e de personalização na Web é descrito através dos seguintes exemplos: (i) Em Sistemas de Busca e Recuperação de Informações com Filtragem de Conteúdo, podemos ter uma melhor descrição do perfil do usuário e dos conteúdos das páginas Web, recuperando as informações de acordo com o contexto da busca, e descartando páginas consideradas fora do contexto da busca; (ii) Em Sistemas de Personalização e Recomendação: uma melhor descrição, classificação e organização dos itens, assim como do perfil do usuário, permitirá uma melhor personalização/recomendação, além disto, podemos também pensar em sistemas que possam buscar informações complementares (interoperabilidade e troca de conhecimentos) a fim de melhor aconselhar o usuário, inclusive realizando atividades de alto nível através de complexas análises de dados (e.g. recuperar e processar informações presentes nas páginas pessoais de usuários na Web ou em seu Curriculum Vitae; buscar informações complementares sobre produtos na Web ou sobre hábitos de consumo destes produtos). Podemos concluir esta seção, com a certeza de que a Web Semântica será de grande relevância em um futuro próximo e trará importantes contribuições para os sistemas de personalização e de recomendação.

## **8. Conclusões e comentários finais**

Vimos neste capítulo como os sistemas de recomendação podem ser utilizados na personalização de páginas Web. Foram apresentadas diversas técnicas neste capítulo, onde foi visto que podemos melhorar significativamente a recuperação, apresentação e sugestão de informações (produtos e itens) aos usuários. A personalização da interface com o usuário, considerando as informações sobre o perfil e comportamento deste usuário, sobre as características e categorias dos itens a serem recomendados, e inclusive sobre os perfis e comportamentos de usuários similares, levam a criação de sistemas que podem aumentar de modo significativo a satisfação dos usuários, uma vez que estes lhe ofereçam boas recomendações e informações de seu interesse. As diversas aplicações e tendências nesta área de personalização e recomendação de conteúdos demonstram claramente a grande importância e o futuro promissor desta área de estudos.

É importante destacar que a maior parte dos exemplos apresentados neste capítulo se referiam à personalização das interfaces de websites. No entanto, outras aplicações dos sistemas de recomendação são possíveis também fora do ambiente puramente computacional, onde encontram cada vez mais mercado. Uma destas aplicações é a de *call-centers*. A maior parte das empresas que lida diretamente com os clientes possui hoje em dia um serviço de *call-center*. Além da infra-estrutura pesada, grande parte do custo de um *call-center* está associado ao pessoal que nele trabalha. Portanto, otimizar este serviço e rentabilizá-lo é algo muito importante. Os sistemas de recomendação permitem, por exemplo, que se possa recomendar a um usuário que está comprando por telefone, produtos que ele provavelmente estará interessado de acordo com as mercadorias já escolhidas. Concluimos este capítulo destacando que

independente do tipo de interface usada (web, 3D, call-center), a personalização é uma poderosa e importante ferramenta para seduzir e cativar usuários na era da informação.

## Referências

- Adomavicius, G.; Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, New York, v. 17, n. 6, p. 734-749, June.
- Agrawal, R. et al. (1993). Mining Association Rules between sets of items in large databases., In: *International Conference Management of Data (SIGMOD-93).Proceedings...*, p.207-216.
- Ansari, A. et al. (2000). Internet Recommendation Systems. *Journal of Marketing Research*, v.37, n.3, p. 363-375, Aug.
- Balabanovic, M.; Shoham, Y. (1997). Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM*, New York, v.40, n.3, p. 66-72, Mar.
- Belvin, N. J.; Croft, W. B.(1992). Information Filtering and Information Retrieval: two sides of the same coin?. *Communications of the ACM*, New York, v.35, n.12, p. 29, Dec.
- Berners-Lee, Tim; Hendler, James and Lassila, Ora (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., and Maler, E. (2000). Extensible Markup Language (XML). World Wide Web Consortium. <http://www.w3.org/TR/REC-xml>.
- Breese, J., Heckerman, D. & Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, July, Morgan Kaufmann.
- Cabena, P. et al.(1997). *Discovering data mining: from concept to implementation*, Prentice-Hall.
- Case, S. et al. (2001). Enhancing E- Communities with agent- based systems. *Computer*, Los Alamitos, v.34, n.7, p. 64- 69.
- Cazella, S. C., Alvares, L. O. C. (2005). Creating Academic Web Communities: A Recommender System to Aid Brazilian Researchers to Find Information and Contact Relevant People. In: *Annual Global Information Technology Management World Conference, GITMA, 2005. Proceedings...* [S.l.: s.n.], 2005. p. 45-48.
- Cazella, S. C., Alvares, L. O. C. (2006). An Architecture based on multi-agent system and data mining for recommending research papers and researchers. In: *International Conference on Software Engineering and Knowledge Engineering, 2006. Proceedings...* [S.l.: s.n.], 2006. p. 67-72.
- Chittaro L., Ranon R. (2000). Adding Adaptive Features to Virtual Reality Interfaces for ECommerce. *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, Lecture Notes in Computer Science 1892*, Springer-Verlag, Berlin, August.
- Chittaro, L. , Ranon, R. (2002). Dynamic Generation of Personalized VRML Content: a General Approach and its Application to 3D E-Commerce. *Proceedings of the 7th International Conference on 3D Web Technology*, ACM Press, New York, February.
- Cook, D. J., Holder, L. B. (2000). Graph-based data mining. *IEEE Intelligent Systems*, Los Alamitos, v.15, n.2, p. 32-41, Mar./Apr.
- De Bra, P. et al. (2000). Making General-Purpose Adaptive Hypermedia Work. *WebNet 2000*, p. 117-123.
- Ehlert, P.A.M. (2003). Intelligent user interfaces: introduction and survey, Research Report DKS03-01 / ICE 01, Delft University of Technology, The Netherlands.

- Figallo, C. (1998). *Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and maintaining a Competitive Edge*. New York: Wiley Computer.
- Foltz, P. W., Dumais, S. T. (1992). Personalized Information Delivery: An Analysis of Information Filtering Methods. *Communications of the ACM*, New York, v.35, n.12, p. 51-60, Dec.
- Gennari, John, Musen, Mark et al. (2003). The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, Vol.58, Issue 1. January, p.89-123
- Geyer-Schulz, A., Hahsler, M. (2002). Evaluation of Recommender Algorithms for an Internet Information Broker based on Simple Association Rules and on the Repeat-Buying Theory. *Proceedings WEBKDD'2002*, Eds. B. Masand, M. Spiliopoulou, J. Srivastava and O. Załane. Edmonton, Canada, pp.100–114. July 2002.
- Golbeck, J. (2005). *Computing and applying trust in web-based social networks*. 2005. 184 f. Tese ( Doutorado em Ciência da Computação ) – Computer Science Department, University of Maryland, EUA.
- Goldberg, D., Nichols, D., Oki, B. M. e Terry, D. (1992). Using collaborative filtering to weave an information Tapestry. *Communications of the ACM*, New York, v.35, n.12, p. 61-70, Dec.
- Gruber, T. R. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N. and Poli, R., editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands. Kluwer Academic Publishers. p.1-23.
- Hauben, M.; Hauben, R. (1997). *Netizens: On the History and Impact of Usenet and the Internet*. Los Alamitos, CA: IEEE Computer Society.
- Herlocker, J. L. (2000). *Understanding and Improving Automated Collaborative Filtering Systems*, Tese de Doutorado (Doutorado em Ciência da Computação), University of Minnesota, Minnesota.
- Herlocker, J. L. et al. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, New York, v.22, n.1, p. 5-53. Jan.
- Holte, R. C. ; Yan, J. Y. (1996). “Inferring What a User is Not Interested”, In: *Advances in Artificial Intelligence, Proceedings...*, p. 159-171.
- Kautz, H. et al. (1997). Referral Web: combining social networks and Collaborative filtering. *Communications of the ACM*, New York, v.40, n.3, p. 63-65, Mar.
- Konstan, J. A. et al.(1997). GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, New York, v.40, n.3, p. 77-87, Mar.
- Librelotto, Gilberto R. et al. (2005). Representação de Conhecimento na Semantic Web. XXIV JAI – Jornadas de Atualização em Informática, XXV Congresso da SBC, Unisinos – São Leopoldo. p.1210-1261, Julho.
- Loeb, S.; Terry, D. (1992). Information Filtering. *Comunications of ACM*, New York, v.35, n.12, p.26, Dec.
- Manber, Udi et al. (2000). Experience with Personalization on Yahoo! *Communication of the ACM*, New York.
- Massa, P. et al. (2004). Trust-aware Collaborative Filtering for Recommender System. In: *International Conference on Cooperative Information System, 2004*, Agia Napa, Cyprus. *Proceedings...* Berlin: Springer-Verlag, 2004. p. 492-? (Lecture Notes in Artificial Intelligence, v. 3290 ).
- Maybury, M.T. (2001). Intelligent user interfaces for all., in *User interfaces for all: concepts, methods, and tools.*, Stephanidis, C. (editor), Lawrence Erlbaum Associates, Inc., Publishers, Maway, NJ, USA.
- Mccarthy, J. (2000). Phenomenal Data Mining. *Communications of the ACM*, v.43, n.8, p. 75-

79, Aug.

- Montaner, M et al. (2003). A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*. Netherlands : Kluwer Academic Publishers, pp. 285-330, Aug.
- Nielsen, J. Ten Usability Heuristics. Conteúdo acessado em 1 de setembro de 2006 e disponível em: [http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html).
- Osório, F. S.; Musse S. R.; Santos, C. T.; Heinen, F.; Braun, A. e Silva A.T. (2004). Ambientes Virtuais Interativos e Inteligentes: Fundamentos, Implementação e Aplicações Práticas. XXIII JAI – Jornadas de Atualização em Informática, XXIV Congresso da SBC, Salvador. p.239-288, Julho.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*, San Mateo: Morgan Kaufmann.
- Resnick, P. e Varian, H. R. (1997). Recommender Systems. *Communications of the ACM*, New York, v.40, n.3, p. 55-58, Mar.
- Rheingold, H. (2000). *The virtual community: homesteading on the electronic frontier*. Massachusetts, USA: MIT Press.
- Riedl, J. et al. (1999). Combining Collaborative Filtering with Personal Agents for Better Recommendations, In: *Proceedings of AAAI, Proceedings...*, vol. 35, p.439-446. Press.
- Santos, C. (2004). Um Ambiente Virtual Inteligente e Adaptativo Baseado em Modelos de Usuário e Conteúdo. Dissertação de Mestrado, PIPCA – UNISINOS, São Leopoldo.
- Santos, C.; Osório, F. S. (2004a). AdapTIVE: An Intelligent Virtual Environment and its Application in E-Commerce. In: *COMPSAC 2004 / IEEE*, Hong-Kong: IEEE Computer Society. v. 1, p. 468-473.
- Santos, C.; Osório, F. S. (2005b). Potential Applications of the AdapTIVE Model: Distance Learning, E-Commerce and Games. In: *Virtual Concept 2005*, Biarritz, França. Springer-Verlag. v. 1, p. 1-10.
- Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2000). Analysis of recommender algorithms for e-commerce. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, p.158-167, 2000.
- Schafer, J.B. et al. (2000). Recommender Systems in E-Commerce. In: *ACM CONFERENCE ON E-COMMERCE, 2., 2000*, Minneapolis. *Proceedings...* New York: ACM Press, 2000.
- Shardanand, U; Maes, P. (1995). Social information filtering: Algorithms for automating "word of mouth", In: *Human Factors in Computing Systems. Proceedings...*, 1995, p. 210-217.
- Torres, R. (2004). *Personalização na Internet*. [S.l.]: Novatec, 2004.
- Tsandilas, T. & Schraefel, M. C. (2004). Usable Adaptive Hypermedia Systems". *Hypermedia*, V10, Issue 1, p.5-29. June.
- Uschold, Mike; Gruninger, Michael (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*. Vol. 11, no. 2, pp. 93-136.
- Zaiane, O. R. (2000). Web Mining: Concepts, Practices and Research, In: *Simpósio Brasileiro de Banco de Dados, SBBD, 15., João Pessoa*. Tutorial... João Pessoa: CEFET-PB; Porto Alegre: PUCRS, v.43, n.8., pp. 410-474.