

Chapter

1

A Ciência da Opinião:

Estado da arte em Sistemas de Recomendação

Sílvio César Cazella, Maria Augusta S. N. Nunes e Eliseo Berni Reategui

Abstract

Recommending products, information, services or even people to a user on the Internet, presents itself as one of the biggest challenges in the virtual world due to the large amount of data available. The appropriate recommendation of a product, service or person can make the difference between getting or losing a customer. Because of this need for achievement, the systems recommendation and personalization has been presented as a facilitator in time to "captivate" the user. This chapter presents an overview of Recommendation systems used in order to enhance the tasks of the personalization. Describes, the most popular techniques and strategies of recommendation, and brings examples of their use both in the commerce and academy.

Resumo

Recomendar produtos, informações, serviços ou mesmo pessoas a um usuário na Internet apresenta-se como um dos maiores desafios no mundo virtual devido à grande quantidade de conteúdo disponível. A recomendação adequada de um produto, serviço ou pessoa pode fazer a diferença entre conquistar o usuário ou perdê-lo. Devido a esta necessidade de conquista, os Sistemas de Recomendação e personalização tem se apresentado como um instrumento facilitador no momento de "cativar" o usuário. Este capítulo apresenta uma visão geral sobre os Sistemas de Recomendação utilizados no intuito de potencializar as tarefas de personalização. Descreve, ainda, as técnicas e estratégias de recomendação mais popularmente utilizadas, trazendo exemplos de utilização desses tanto na área de Comércio como na Academia.

1. Introdução aos Sistemas de Recomendação

Com a quantidade de informações e com a disponibilidade facilitada das mesmas pelo acesso a Internet, as pessoas se deparam com uma diversidade muito grande de opções. Muitas vezes um indivíduo possui pouca ou quase nenhuma experiência pessoal para realizar escolhas dentre as várias alternativas que lhe são apresentadas. A questão relevante neste momento refere-se a como proceder nestes casos? Para diminuir 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*) [Maes and Shardanand 1995] ou através de textos de recomendação, opiniões de revisores de filmes e livros, impressos de jornais, dentre outros.

Os Sistemas de Recomendação auxiliam no aumento da capacidade e eficácia deste processo de indicação já bastante conhecido na relação social entre seres humanos [Resnick and Varian 1997]. Em um sistema típico, as pessoas fornecem recomendações como entradas que o sistema agrega e direciona para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios deste tipo de sistema é realizar a combinação adequada entre as expectativas dos usuários e os produtos, serviços e pessoas a serem recomendados aos mesmos, ou seja, definir e descobrir este relacionamento de interesses é o grande problema.

Os proponentes do primeiro Sistema de Recomendação denominado *Tapestry* [Goldberg et al. 1992], [Resnick and Varian 1997], cunharam 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. Os autores preferem utilizar a expressão Sistemas de Recomendação, por ser um termo genérico e defendem este posicionamento por dois motivos: primeiro porque os recomendadores podem não explicitar colaboração com os que as recebem, pois um pode não conhecer o outro, e por último os recomendadores podem sugerir itens de interesse particular, incluindo aqueles que poderiam ser desconsiderados. A questão aqui é como verificar o real interesse em colaborar, pois na realidade as recomendações esboçadas pelos recomendadores estão concentradas em seus interesses próprios. Na verdade o grupo a ser formado para receber as recomendações deveria, pelo menos, apresentar indivíduos com interesses comuns (explícitos) ou comportamento comum (implícito). Assume-se que sistemas de filtragem colaborativa e sistemas de filtragem baseada em conteúdo são tipos de Sistemas de Recomendação que aplicam abordagens distintas, mas possuem como finalidade única a recomendação.

Alguns autores, como Montaner [Montaner et al. 2003], destacam que existe um terceiro tipo de filtragem de informação denominada de filtragem demográfica. A filtragem demográfica utiliza a descrição de um indivíduo para aprender o relacionamento entre um item em particular e o tipo de indivíduo que poderia vir a se interessar. Este tipo de abordagem utiliza as descrições das pessoas para conseguir aprender o relacionamento entre um item e o tipo de pessoa que gostaria deste. O perfil do usuário é criado pela classificação dos usuários em estereótipos que representam as características de uma classe de usuários. Dados pessoais são requisitados ao usuário,

geralmente em formulários de registro, e usados como caracterização dos usuários e seus interesses. Como exemplo, Montaner cita o método implantado por Krulwich (Krulwich apud [Montaner et al. 2003]) em *LifeStyle Finder* onde é utilizado um sistema demográfico denominado *PRIZM* da *Claritas Corporation*. Este sistema tem o objetivo de dividir a população americana em 62 agrupamentos demográficos de acordo com seus históricos de compra, características referentes ao tipo de vida e respostas a pesquisas.

Quanto às técnicas de filtragem de informação, Burke [Burke 2002] define, além da conhecida filtragem colaborativa, da baseada em conteúdo e da demográfica, duas outras técnicas. Na primeira delas, intitulada de filtragem baseada em conhecimento, a recomendação dos itens é feita baseada nas inferências das preferências do usuário e suas necessidades através de conhecimento estruturado de forma funcional. Alguns exemplos clássicos dessa abordagem são: *Google* [Brin and Page 1998] e *The Entree* [Burke 2002]. Na segunda técnica, intitulada de filtragem baseada em utilidade, a recomendação é realizada considerando a utilidade dos itens para um determinado usuário, alguns exemplos clássicos são *Tetê-a-Tête* e *PersonaLogic* [Guttman et al. 1998].

Gonzalez [Gonzalez et al. 2007] e Nunes [Nunes 2009] propõem também técnicas de recomendação complementares seguindo a linha proposta por de Burke [Burke 2002]. Essas técnicas envolvem questões psicológicas do usuário combinados as características dos produtos serviços, a mesma intitula-se, segundo Gonzalez, de filtragem baseada em outros contextos. Ele utiliza questões de Inteligência Emocional como inovação para recomendar cursos em um *site* de *e-training* chamado de *emagister.com*. Nunes [Nunes 2009], em uma versão estendida de filtragem baseada em outros contextos, propõe a filtragem baseada em aspectos psicológicos utilizando exemplos de Gonzalez (usa Inteligência Emocional para melhorar a recomendação de cursos). Saari et al [Saari et al. 2004] (recomenda produtos baseado no efeito psicológico que recomendações passadas causaram), Mastoff [Mastoff 2005] (usa satisfação do usuário para prever recomendações satisfatórias a grupos de usuários) e Nunes [Nunes 2009] (usa traços de personalidade para substancialmente tornar mais satisfatória a recomendação de produtos, serviços e pessoas).

Segundo Adomavicius and Tuhilin [Adomavicius and Tuhilin 2005] e Burke [Burke 2002] existem pesquisadores que muitas vezes utilizam mais de uma técnica de filtragem para melhorar a resposta da recomendação. Essa técnica intitula-se de filtragem híbrida. Alguns exemplos clássicos são: o *Fab System* [Balabanovic and Shoham 1997] e o sistema do *Pazzani* [Pazzani 1999].

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 os Sistemas de Recomendação em grandes proporções, utilizando a estratégia de personalização [Manber et al. 2000]. Atualmente, um grande número de *websites* emprega os Sistemas de Recomendação para levar aos usuários diferentes tipos de sugestões, como ofertas casadas ("clientes que compraram item X também compraram item Y"), itens de sua preferência, itens mais vendidos nas suas categorias favoritas, entre outros.

Este capítulo apresenta os Sistemas de Recomendação, descreve as técnicas mais utilizadas e mostra como podem ser aplicados em exemplos práticos. Iniciaremos pela

conceituação de personalização no contexto de Sistemas de Recomendação (Seção 2), seguido pelos mecanismos de coleta de informações geralmente utilizados para formação de perfis de usuários (Seção 3), apresentando a seguir, na Seção 4, questões relativas a estratégias de recomendação e, na Seção 5 questões relativas às técnicas de recomendação. Na Seção 6, serão apresentados exemplos de Sistemas de Recomendação. No final do capítulo, serão apresentados alguns desafios e tendências da área (Seção 7), seguido de uma conclusão que visa destacar a relevância do tema Sistemas de Recomendação.

2. Personalização através de Sistemas de Recomendação

Personalização é o ato de adequar um produto ou serviço para atender as necessidades de um indivíduo [Gyara and Sachdev 2008]. Do ponto de vista computacional, um sistema capaz de tratar cada usuário de maneira individual necessita de um conjunto específico de funções:

- Enquanto o usuário navega no sistema, toda sua interação deve ser monitorada. Com base nas informações coletadas, mantém-se um modelo do usuário: dados demográficos, itens visualizado, 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 apresentar, esconder ou enfatizar fragmentos de uma página, assegurando que seu conteúdo inclua a informação apropriada, em um nível adequado de dificuldade ou detalhe.

Estas são características da função de personalização 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, apresentando-lhe conteúdo ou produtos relevantes.

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. Do ponto de vista do marketing, tal abordagem se contrapõe ao marketing de massa, no qual determinado produto é oferecido a um grande número de pessoas quase indistintamente.

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.

Os Sistemas de Recomendação são grandes aliados da personalização de sistemas computacionais, principalmente na *web*. Os Sistemas de Recomendação são capazes de identificar preferências e sugerir itens relevantes para cada usuário, de acordo com a

análise de seu comportamento de navegação, consulta e/ou compra, preferências, entre outros aspectos. A próxima seção apresenta a formação de perfis de usuários – uma etapa essencial na construção de sistemas personalizados.

3. Formação de Perfis de Usuários

Para que seja possível recomendar produtos, serviços ou pessoas a um usuário é necessário ter-se conhecimento sobre quem é este usuário. Antes mesmo de pensar em capturar e armazenar suas informações pessoais e comportamentais é necessário identificar qual o tipo de informação será relevante para a geração da recomendação visando uma eficiente personalização dos produtos, serviços e pessoas. Para a correta geração da recomendação a definição do perfil do usuário e coleta de informações é imprescindível.

3.1. Identidade do Usuário

Segundo a visão da psicologia clássica, Identidade é definida pela autoconsciência/visão que cada pessoa possui de si mesma, enquanto que na Psicologia Social e Sociologia, Identidade pode ser definida como a forma que cada pessoa é vista sob os olhos da sociedade.

Segundo os pesquisadores de Teoria da Personalidade, o desenvolvimento da Identidade recebe uma importante influência da Personalidade. Boyd [Boyd 2002] descreve dois aspectos diferentes da Identidade: a noção internalizada do “eu” (Identidade Interna) e a versão projetada da internalização do “eu” (Identidade Social). Nessa mesma linha, Erikson [Erikson 1980], por exemplo, acredita que Identidade (EGO) tem uma representação pessoal interna (Identidade Interna) bem como uma representação social (Identidade Social). Giddens [Giddens 1991] concorda que sem experiências sociais o “eu” não pode internalizar evolução. Giddens ainda afirma que a identidade de um indivíduo não é estática, ela pode ser representada em constante evolução, principalmente porque o componente social é dinâmico e esta sempre sendo modificado. Mead [Mead 1934], ainda define “eu” e “mim”, onde “mim” representa o aspecto socializado da pessoa (Identidade Social), enquanto que o “eu” representa como a pessoa se define em relação aos outras pessoas da sociedade (Identidade Individual).

Note que no mundo virtual onde não há presença física e conseqüentemente não há percepção de características sutis da Identidade, várias pistas que possivelmente identificariam dicas de preferências, comportamentos, habilidades sociais, entre outras, são ausentes, ao contrário do que ocorre no mundo real [Donath 1999]. Donath [Donath 2000] afirma que conhecer a Identidade da pessoa é vital para uma adequada personalização de uma ambiente no mundo virtual. Goffman [Goffman 1959] afirma, ainda, que as pessoas se esforçam para se apresentarem como pessoas “aceitáveis” aos olhos da sociedade (em comunidades virtuais, por exemplo).

Considerando a Identidade como um canal importante onde as características objetivas e subjetivas das pessoas emergem, denomina-se de fundamental importância seu uso em Sistemas de Recomendação no intuito de fornecer pistas sobre os futuros comportamentos e necessidades dos usuários em um dado ambiente onde a personalização se faz eficaz, por exemplo.

Tecnicamente, em Ciência da Computação, a tecnologia usada para formalizar a Identidade em um dado ambiente computacional é pelo uso de Perfil/Modelo do Usuário (Identidade Interna) e Reputação do Usuário (Identidade Social).

3.1.1. Perfil de Usuário

Donath [Donath 1999] afirma que para a formação eficiente de uma Identidade Virtual é crucial que o usuário tenha definida sua Identidade Interna e sua Identidade Social. No mundo virtual a Identidade Interna do usuário é definida por ele próprio similar ao mundo real (algumas vezes também é descoberta através de técnicas de *Machine Learning*). Enquanto a Identidade Social é definida pelos outros membros do mundo virtual (elucidada na próxima sessão). Tanto a Identidade Interna, como a Identidade Social são armazenadas no Perfil do Usuário.

Perfis de Usuários são conceitos aproximados, eles refletem o interesse do usuário com relação a vários assuntos em um momento particular. Cada termo que um Perfil de Usuário expressa é, num certo grau, características de um usuário particular [Poo et al 2003] incluindo todas as informações diretamente solicitadas a ele e aprendidas implicitamente durante sua interação na *web* [Carreira et al 2004]. Fisicamente, o Perfil do Usuário pode ser visto como uma base de dados onde a informação sobre o usuário, incluindo seus interesses e preferências, é armazenada e pode ser dinamicamente mantido [Rousseau et al. 2004], [Poo et al. 2003].

Na *web* encontram-se muitos tipos de Perfis de Usuário com diferentes graus de complexidade. Eles são desenvolvidos no contexto de *e-commerce*, *e-learning* e *e-community*, por exemplo. Kobsa [Kobsa 2007] cria uma Modelagem Genérica de Usuário para ser usada como uma *shell* para a criação de categorias de informação sobre o usuário objetivando personalizar as aplicações *web*. O modelo proposto por Kobsa é um dos mais reputados. Paiva [Paiva and Self 1995] também desenvolveu uma *shell* de Modelo de usuário chamado TAGUS, criado para melhor modelar os alunos para atividades de aprendizado.

No *e-commerce*, [Riedl et al. 1999], [Herlocker et al. 2004], [Konstan et al. 1997], [Schafer et al. 1999] e [Schafer et al. 2001], do GroupLens, criaram vários Modelos de Usuário baseado em ranqueamento de filmes, de notícias, entre outros. Esses modelos têm sido usados nos Sistemas de Recomendação criados pelo GroupLens.

Considerando ainda definições de Modelo de usuário, Heckmann [Heckmann et al. 2005], [Heckmann 2005] e [Heckmann and Kruguer 2003] propôs uma Ontologia¹ de um Modelo de usuário Geral (*GUMO*). O *GUMO* é um modelo ubíquo de Modelo de Usuário incluindo muitos aspectos básicos de usuário, partindo desde informação de contato, demográficos, habilidades fisiológicas e psicológicas, estado emocional, estado mental e nutrição. A Ontologia de Heckmann é muito rica e pode ser implementada de acordo com o interesse do projetista de uma *shell* de perfil de usuário. Na figura 3.1 apresentam-se as dimensões básicas propostas por Heckmann no *GUMO*.

¹Uma ontologia é uma especificação de um conceito.

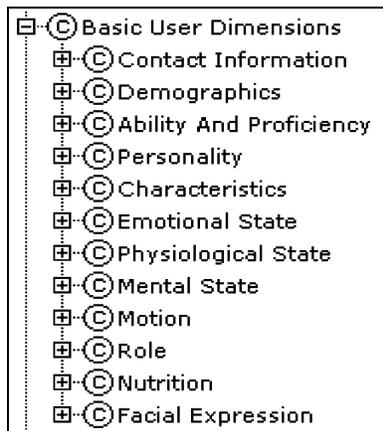


Figura 3.1. Dimensões básicas do usuário segundo modelo GUMO [Heckmann 2005]

Note que os Sistemas de Recomendação para gerar as recomendações e personalizar o ambiente ao usuário necessitam da Identidade Interna do usuário que é definida pelo Perfil de Usuário, como foi apresentado. Porém necessita muitas vezes também da Identidade Social que é definida pela Reputação do usuário, como se apresenta a seguir.

3.1.2. Reputação

Reputação pode ser definida como o retorno social recebido sobre a personalidade de alguém. A Reputação pode ser compatível ou não com a descrição feita no Perfil de Usuário. Josang et al [Josang et al. 2007] descreve Reputação como “a informação normalmente dita ou crível sobre as características de uma pessoa ou coisa e seus estados”.

Resnick [Resnick et al. 2000] define Reputação como a coleção dos *feedbacks* recebidos sobre o comportamento efetuado pelos participantes de uma comunidade. A Reputação ajuda as pessoas escolherem parceiros confiáveis no mundo virtual que são credíveis no mundo real. Geralmente nas redes de Reputação, os usuários encorajam os comportamentos confiáveis discriminando a participação de pessoas desabilitadas moralmente ou desonestas.

Segundo Rein [Rein 2005] a Reputação pode ser também definida como um completo sistema de informações sobre Reputação do usuário que inclui todos os aspectos de um modelo de referência. Esse modelo de referência é baseado em nove aspectos determinantes: conhecimento, experiência, credenciais, endosso, contribuidor, conexões, sinais, *feedback*, contexto e valores sociais. A visão estrutural do modelo é apresentada na figura 3.2.

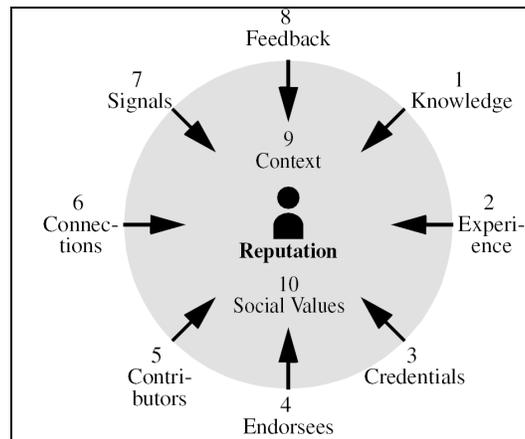


Figura 3.2. Visão estrutural dos nove determinantes da Reputação [Rein 2005]

A visão estrutural de Rein descreve as funcionalidades e comportamentos essenciais do ser humano que são desejáveis e efetivos para possivelmente ser representado através de uma Reputação explícita e fácil de ser medida no usuário.

A Reputação é geralmente aplicada para gerenciar comportamento do usuário durante um processo comercial (*e-commerce*, por exemplo) envolvendo compra e venda de produtos e/ou serviços e também durante processos sociais como combinação social em comunidades virtuais e redes sociais.

Em processos comerciais, como por exemplo, no *eBay* [Resnick et al. 2000] [Resnick et al. 2006] um consumidor compra um certo produto de alguém. Depois disso, ele deixa um *feedback* sobre o produto comprado e/ou o comportamento do vendedor durante o processo de venda.

Em contraste, em situações sociais como, por exemplo, *Orkut*, *IKarma*, *Opinity*, *LinkedIn*, *Mendeley* [Jensen et al. 2002], usuários são membros de comunidades virtuais ou redes sociais. Eles são capazes de coletar gerenciar e promover Reputação de usuário entre seus clientes e contatos da comunidade ou rede. Isto é, usuários (prestadores de serviço) que tem perfil na Rede de Reputação, que é também uma rede social podem ser “tagged” e rankeados pelos seus clientes e/ou contatos. Usuários podem ser encontrados através de *tags* em e-mail ou, também, alguém pode encontrar um contato de um prestador de serviço simplesmente procurando em *tags* na própria rede de reputação.

Aqui, convencionou-se Reputação como uma extensão de um Perfil de Usuário. Ele usa o mesmo tipo de informação armazenada no Perfil de Usuário, porém o conjunto de informações é fornecida por outro alguém (amigo, cliente do usuário, entre outros). Nesse caso, a Identidade é determinada pelos Traços de Personalidade do usuário informados por ele mesmo para Perfil de Usuário e informados por uma outra pessoa para Reputação de Usuário.

Perfil de Usuário e Reputação de Usuário são muito importantes para definir a Identidade do usuário. Dessa forma o *Perfil de Usuário* pode prever necessidades e comportamentos do usuário em um ambiente computacional, enquanto Reputação permite a criação de relação de confiança entre membros de uma sociedade em um ambiente computacional. A Identidade do usuário é muito útil para sua interação social no ambiente computacional.

3.2. Geração e Manutenção de Perfil de Usuário

A personalização de um sistema requer que se possa identificar o usuário no momento em que este acessa o sistema. Na *web*, duas das formas mais habituais de identificação de usuário são:

- Identificação no servidor: normalmente disponibiliza ao usuário uma área de cadastro com informações pessoais, tais como: nome, data de nascimento, sexo, endereço e outros. Além disso, solicita obrigatoriamente um *login* e senha. Estas informações ficam armazenadas em um banco de dados no servidor. Sempre que o usuário acessar o sistema, ele poderá fazer sua identificação/autenticação informando seu *login* anteriormente cadastrado. Este mecanismo permite que o *website* identifique com mais precisão o usuário que nele se conecta.
- Identificação no cliente: utiliza normalmente *cookies*, um mecanismo pelo qual um *website* consegue identificar que determinado computador está se conectando mais uma vez a ele. Este método assume que a máquina conectada é utilizada sempre pela mesma pessoa. Logo, ao identificar a máquina, o *website* está na realidade identificando seu usuário. Trata-se de um mecanismo mais simples do que a identificação através do servidor, porém menos confiável, principalmente se o computador identificado for utilizado por mais de uma pessoa.

Depois de identificado o usuário, é possível coletar dados sobre este de forma implícita ou explícita, permitindo desta maneira a geração e manutenção de seu perfil. Na modalidade de coleta explícita (também conhecida como personalização), o usuário indica espontaneamente o que lhe é importante. No exemplo a seguir (figura 3.3), o usuário do *iGoogle* define como deve ser configurada sua página pessoal, alterando a imagem do cabeçalho da página, inserindo e organizando na tela diferentes ferramentas, tais como: *Gmail*, *Google Docs*, *Agenda* e *Meteorologia*.

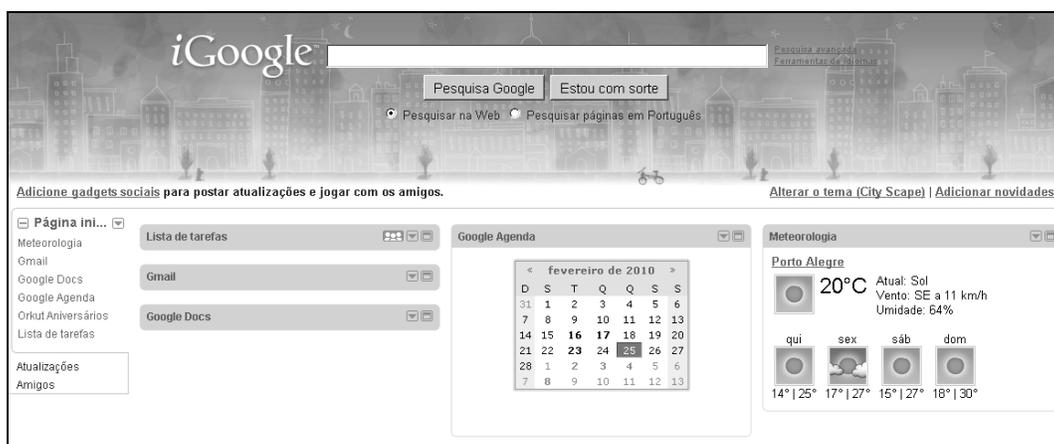


Figura 3.3. Seleção de seções favoritas

Na modalidade implícita, através de ações do usuário inferem-se informações sobre suas necessidades e preferências. Por exemplo, armazenando-se dados de navegação do usuário (páginas consultadas, produtos visualizados e outros) é possível detectar que ele se interessa por determinados assuntos.

Através desta técnica, é possível conhecer melhor a preferência dos usuários sem que eles tenham que fornecer informações explicitamente, e em seguida utilizar estes dados para fazer recomendações. A figura 3.4 mostra uma página da *Amazon.com* personalizada para um usuário que se mostra interessado por diversos assuntos listados na parte inferior da página, com ênfase para aqueles mais consultados. Acima da lista de assuntos, o sistema apresenta alguns livros em destaque nas áreas de interesse do usuário.

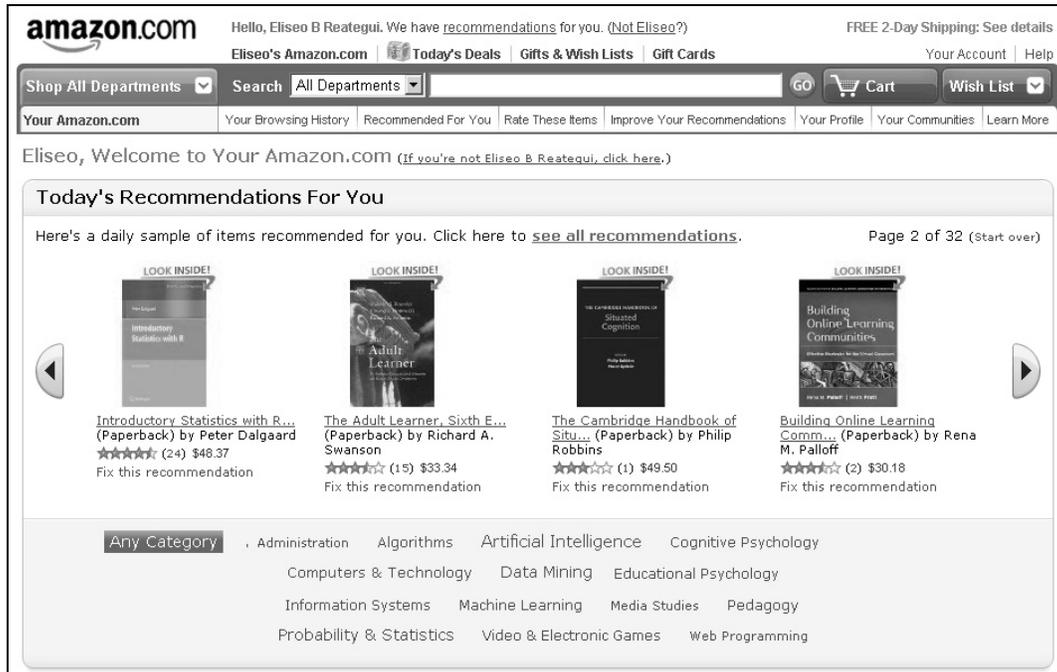


Figura 3.4. Página personalizada a partir de interesses implícitos

3.3. Privacidade em Sistemas de Recomendação

A personalização da pesquisa na *web* é uma das formas mais promissoras de melhorar a qualidade das buscas. Contudo, para que isso possa ser feito, informações sobre os usuários devem ser armazenadas – o que acaba levantando questões sobre a privacidade na *web* [Xu et al. 2007].

Dados comportamentais (navegação, compra, etc.) e pessoais são comumente coletados e armazenados. Esta coleta como visto anteriormente, é feita muitas vezes de forma implícita, i.e. sem que o usuário necessariamente perceba que informações sobre ele estão sendo armazenadas na medida em que utiliza o sistema.

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. A figura 3.5 mostra uma coleta de dados incentivando a participação do usuário para que as recomendações feitas a ele sejam mais eficientes.



Figura 3.5. Coleta de dados para melhoria das recomendações

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. No entanto, 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*. Atitudes como esta, ao invés de estreitar o relacionamento com os clientes, apenas aumentam a distância entre estes e a empresa.

Como atualmente não há leis de privacidade on-line muito claras, é importante que as políticas de privacidade das empresas sejam precisas. A figura 3.6 mostra um trecho da página contendo a política de privacidade do serviço de busca *Buscapé*.

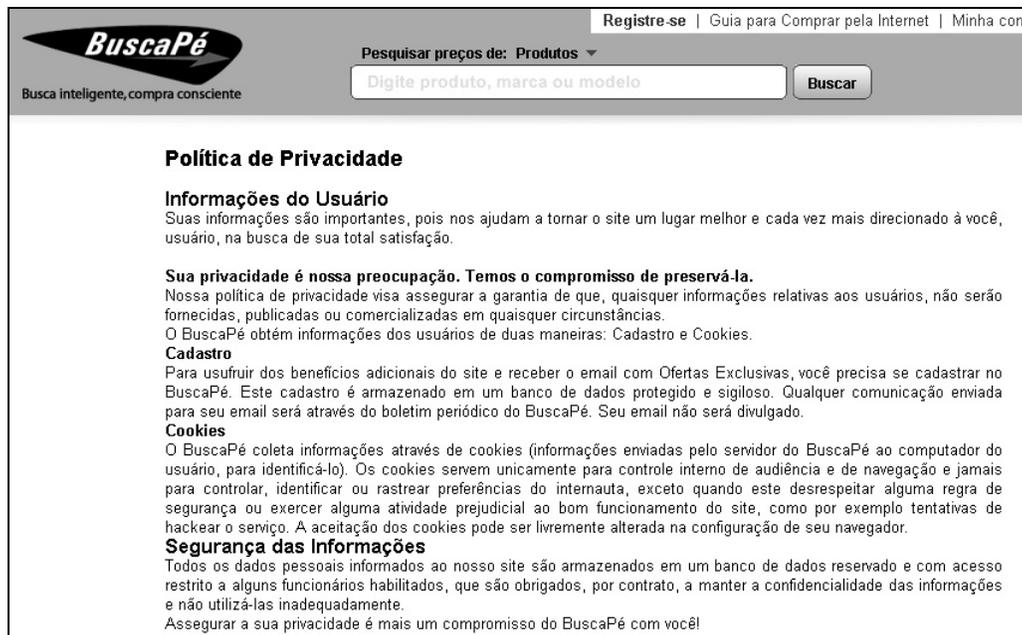


Figura 3.6. Política de privacidade do website Buscapé

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. A

comunidade científica também tem esta preocupação e propõe mecanismos de manutenção da privacidade num mundo onde cada vez mais informações pessoais são compartilhadas.

Algumas organizações propõem selos que garantem que a política de privacidade de um *website* é adequada e cumprida com rigor. A figura 3.7 apresenta dois destes selos, fornecidos por grandes instituições na área (*TRUSTe*, *BBBOnline*).



Figura 3.7. Selos de Privacidade

Apesar das garantias oferecidas por estas instituições, questiona-se o fato de que muitas vezes os padrões exigidos não são rígidos o suficiente, e as estratégias de controle das normas não são rigorosas [Cline 2003]. Mesmo assim, o número de empresas que adota selos de garantia de privacidade é crescente – aumentando o nível de confiança dos usuários na segurança dos *websites* e na manutenção das políticas de privacidade.

4. Estratégias de Recomendação

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

4.1. Reputação do Produto

Uma estratégia bastante utilizada em Sistemas de Recomendação é baseada no uso das avaliações dos usuários para estabelecer a reputação de um item, ou produto. Após conhecer determinado item, consultando-o ou adquirindo-o, o usuário tem a possibilidade de deixar uma avaliação sobre este. É muito comum vermos opinião de usuários na forma de ícones dispostos ao lado do item visualizado, como ilustrado na figura 4.1.



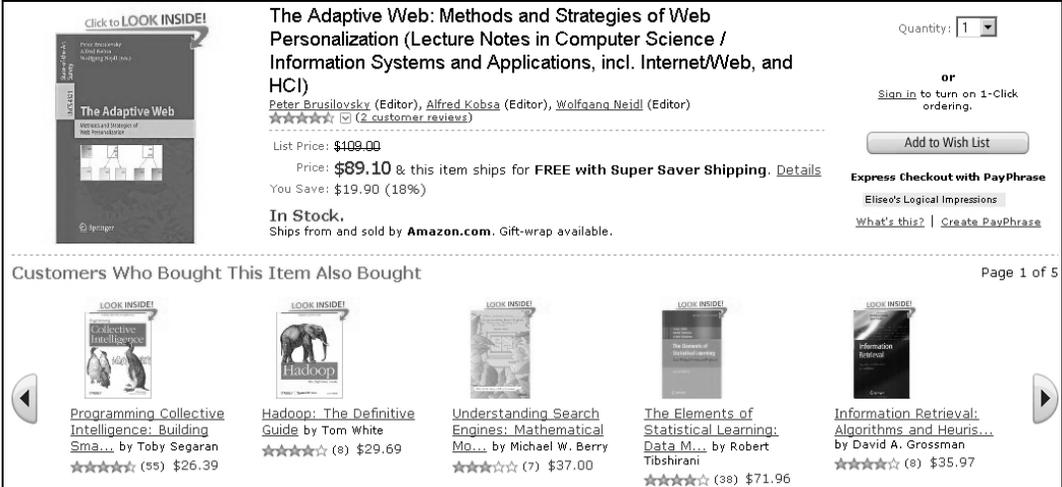
Figura 4.1. Reputação de um livro

Avaliações de 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, entre outros. Do ponto de vista de implementação, este também é um mecanismo fácil de implementar, na medida em que não exige nenhum tipo de dispositivo com maiores capacidades de análise de dados. Basta armazenar e disponibilizar as avaliações de usuários sobre os itens tratados, e apresentar estas opiniões no momento apropriado.

4.2. 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 e outros). É outro tipo de recomendação muito comum em *websites* de comércio eletrônico, como mostra a figura a seguir.



The screenshot displays an Amazon product page for the book "The Adaptive Web: Methods and Strategies of Web Personalization (Lecture Notes in Computer Science / Information Systems and Applications, incl. Internet/Web, and HCI)". The main product is priced at \$89.10, down from a list price of \$109.00. Below the main product, a section titled "Customers Who Bought This Item Also Bought" features five recommended books:

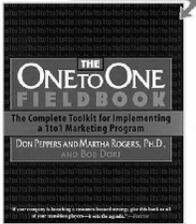
Book Title	Author	Rating	Price
Programming Collective Intelligence: Building Smaller...	Toby Segaran	★★★★☆ (55)	\$26.39
Hadoop: The Definitive Guide	Tom White	★★★★☆ (8)	\$29.69
Understanding Search Engines: Mathematical Models...	Michael W. Berry	★★★★☆ (7)	\$37.00
The Elements of Statistical Learning: Data Mining, Inference, and Prediction	Robert Tibshirani	★★★★☆ (38)	\$71.96
Information Retrieval: Algorithms and Heuristics...	David A. Grossman	★★★★☆ (8)	\$35.97

Figura 4.2. Recomendação por associação

No exemplo ilustrado na figura 4.2, a partir da seleção de um livro de personalização outros livros similares são apresentados. 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.

4.3. Associação por Conteúdo

Também é possível fazer recomendações com base no conteúdo de determinado item, por exemplo, um autor, um compositor, um editor, entre outros. Para possibilitar este tipo de recomendação, é necessário que se encontrem associações num escopo mais restrito. Por exemplo, determinado livro contém assuntos que são relacionados a assuntos contidos em outras obras. A figura 4.3 mostra este tipo de recomendação em uma livraria virtual.



The One to One Fieldbook
 Don Peppers (Author), Martha Rogers (Author)
 ★★★★★ (15 customer reviews)

List Price: ~~\$21.00~~
 Price: **\$17.95** & eligible for **FREE Super Saver Shipping** on orders over \$25.
[Details](#)

You Save: \$3.05 (15%)

In Stock.
 Ships from and sold by Amazon.com. Gift-wrap available.

Books on Related Topics ([learn more](#))

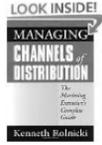
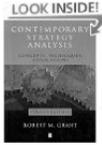
 <p>Business by Editors Of Perseus Publishing</p> <p>Discusses:</p> <ul style="list-style-type: none"> ◆ customer expertise ◆ Harvard Business Review ◆ Harvard Business School Press 	 <p>Managing Channels of Distribution by Kenneth Rolnicki</p> <p>Discusses:</p> <ul style="list-style-type: none"> ◆ customer expertise ◆ channel members ◆ channel partners 	 <p>Angel Customers and Demon Customers by Larry Selden</p> <p>Discusses:</p> <ul style="list-style-type: none"> ◆ most valuable customers ◆ customer manager ◆ Dell Computer 	 <p>Contemporary Strategic Analysis by Robert M. Grant</p> <p>Discusses:</p> <ul style="list-style-type: none"> ◆ divisional integration ◆ Harvard Business Review ◆ Harvard Business School Press
---	--	--	--

Figura 4.3. Recomendação por associação de conteúdo

No exemplo, o sistema recomenda quatro obras de assuntos relacionados aos temas tratados no livro *One to One Fieldbook*. Diferentes técnicas podem ser utilizadas para se obter este tipo de recomendação, apresentadas no próximo capítulo.

4.4. Análise de Sequências de Ações

Sequências são um tipo importante de dados que podem ocorrer com frequência em diferentes domínios, como por exemplo, na medicina, negócios, segurança dentre outros. Na *web*, sequências podem ser utilizadas para capturar o comportamento de usuários através de históricos de atividade temporal, como em *weblogs* e histórico de compras de clientes [Dong and Pei 2007]. As informações encontradas nestas sequências temporais podem ser empregadas nas identificações de padrões de navegação e consumo, que em seguida podem servir aos Sistemas de Recomendação.

Alguns algoritmos buscam agrupar as sequências de ações em *clusters* como forma de identificar e representar os padrões de comportamentos encontrados nos dados. Por exemplo, Demir et al. [2007] representam padrões de sequências em grafos não direcionados e utilizam algoritmos evolucionários com múltiplos objetivos como no processo clusterização. No entanto, um dos algoritmos mais conhecidos na análise de sequência de ações é o *SPADE* (*Sequential PAttern Discovery using Equivalence classes* – Descoberta de Padrões Sequenciais utilizando classes equivalentes) [Zacki 2001]. O algoritmo busca associar a cada sequência encontrada uma lista de objetos nos quais ela ocorre. Através deste mecanismo a busca por objetos e sequências é otimizada. O algoritmo também reduz o número de varreduras na busca por padrões, o que implica em um tempo de execução menor.

Se comparado à recomendação por associação, a análise de sequências permite encontrar padrões que consideram a variável tempo, deste modo levando em consideração a ordem com que determinadas ações ocorrem. Em função do tipo de aplicação, tal característica pode ou não ser importante.

5. Técnicas de Recomendação

Várias técnicas têm surgido visando à identificação de padrões de comportamento (consumo, pesquisa e outros) 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 and Terry 1992]. Peter Denning já escrevia em um volume da *Communications of ACM* do ano de 1982 (Peter Denning apud [Loeb and 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 and 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 and 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.

5.2. 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 and 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 dos itens e no perfil do usuário.

A abordagem baseada em conteúdo tem suas raízes na área de recuperação de informação. Devido aos significativos avanços feitos pelas comunidades de filtragem de informação e filtragem de conteúdo, muitos sistemas baseados em filtragem de conteúdo focam na recomendação de itens com informações textuais, como documentos e *websites*. As melhorias sobre os sistemas tradicionais de recuperação de informação vieram com a utilização do perfil do usuário, que contém suas preferências e necessidades.

As informações sobre o perfil do usuário podem ser obtidas pelo próprio usuário, como uma consulta realizada por ele, ou coletadas através do conteúdo dos itens que o usuário consome. Uma técnica muito comum neste tipo de abordagem é a indexação de frequência de termos (*term frequency indexing*). 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 a respectiva palavra ocorre em um documento ou na consulta do usuário. Sendo assim, os vetores dos documentos que estão próximos aos vetores de consulta são considerados os mais relevantes para ele.

Desta forma, a filtragem baseada em conteúdo parte do princípio de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse no passado, definindo então, a similaridade entre os itens [Herlocker 2000]. Em alguns casos, pode haver maior dificuldade para estabelecer esta similaridade. Para que seja estabelecida a similaridade entre itens como roupas e brinquedos, por exemplo, seria necessária a identificação dos atributos nos itens a serem comparados (peso, preço, marca, etc.). No caso dos itens serem artigos (ou documentos), este processo de comparação pode ser facilitado, pois documentos podem ser considerados similares se compartilharem termos em comum. Sendo assim, a filtragem baseada em conteúdo é mais indicada para a recomendação de itens textuais, onde o conteúdo é geralmente descrito com *keywords* [Salton and McGill 1983]. Um exemplo é o sistema de recomendação *Fab* [Balabanovic and Shoham 1997], o qual recomenda páginas a usuários. Este sistema representa uma página *web* com os seus 100 mais importantes termos.

Outros exemplos de filtragem baseada em conteúdo são sistemas de recuperação booleanos, onde a consulta é um conjunto de palavras chave combinadas com operadores booleanos; sistemas de filtragem probabilística, onde o raciocínio probabilístico é utilizado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário; e linguagem natural [Herlocker 2000].

Conforme mencionado anteriormente, Sistemas de Recomendação baseados em conteúdo podem recomendar itens similares a itens que o usuário gostou no passado. Deste modo, vários itens são comparados com itens que foram avaliados positivamente e os mais similares serão recomendados. Conforme Adomavicius [Adomavicius 2001], podemos formalizar esta definição estabelecendo *ContentBasedProfile(c)* como sendo o

perfil do usuário c . Este perfil é obtido através de uma análise do conteúdo dos itens previamente avaliados pelo usuário utilizando técnicas de recuperação de informação. Por exemplo, $ContentBasedProfile(c)$ pode ser definido como um vetor de pesos $(w_{c_1}, \dots, w_{c_k})$ onde cada peso w_{c_i} denota a importância do termo k_i para o usuário c utilizando-se a medida $TF-IDF$ (*term frequency–inverse document frequency*).

Em Sistemas de Recomendação baseados em conteúdo, a função utilidade $u(c,s)$ é geralmente definida conforme a equação (1):

$$u(c,s) = score(ContentBasedProfile(c), Content(s)) \quad (1)$$

Tanto o $ContentBasedProfile(c)$ do usuário c como o $Content(s)$ podem ser representados como vetores (TF-IDF) de pesos e termos \vec{w}_c e \vec{w}_s . Além disso, a função utilidade $u(c,s)$ normalmente é representada, na literatura de recuperação de informação, por algum tipo de pontuação heurística sobre vetores, como por exemplo, a medida de similaridade do cosseno. O cálculo para a medida de similaridade do cosseno pode ser feito de acordo com a equação (2), onde k é o número total de palavras no sistema [Adomavicius et al. 2005]:

$$u(c,s) = \cos(\vec{w}_c, \vec{w}_s) = \frac{\vec{w}_c \cdot \vec{w}_s}{\|\vec{w}_c\|_2 \times \|\vec{w}_s\|_2} = \frac{\sum_{i=1}^k w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^k w_{i,c}^2} \sqrt{\sum_{i=1}^k w_{i,s}^2}} \quad (2)$$

Desta forma, o cálculo de similaridade é realizado computando o cosseno do ângulo formado pelos dois vetores que representam os documentos (termos e frequências). 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.

Segundo Adomavicius [Adomavicius et al. 2005], a abordagem baseada em conteúdo tem as seguintes limitações:

1. Análise de conteúdo é limitada: o conteúdo de dados pouco estruturados é difícil de ser analisado. A aplicação da filtragem baseada em conteúdo para extração e análise de conteúdo multimídia por exemplo (vídeo, som), é muito mais complexa do que a extração e análise de documentos textuais. Outro problema, relativo a análise de conteúdo textual, é que sistemas baseados em filtragem em conteúdo não conseguem distinguir um artigo bem escrito de um artigo mal escrito se eles utilizam termos muito semelhantes.

2. Super especialização: quando o Sistema de Recomendação pode recomendar somente itens similares a itens avaliados positivamente, pode ocorrer a super especialização. Desta forma, os itens que não fechem com o perfil do usuário não serão apresentados.

5.3. Filtragem Colaborativa

A abordagem da filtragem colaborativa foi desenvolvida para atender pontos que estavam em aberto na filtragem baseada em conteúdo [Herlocker 2000], [Ansari et al. 2000]. A Filtragem Colaborativa se diferencia da filtragem baseada em conteúdo exatamente por não exigir a compreensão ou reconhecimento do 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 nas avaliações feitas pelos usuários. Por exemplo, o primeiro sistema criado com esta abordagem foi o *Tapestry* [Goldberg et al. 1992] que permitia ao usuário especificar um

consulta como: “mostre-me todos os memorandos que uma determinada pessoa considera como importante”. Membros de determinada comunidade podiam, deste modo, ser beneficiados pela experiência de outros.

Segundo Herlocker [Herlocker 2000] os primeiros sistemas de filtragem colaborativa requeriam usuários para especificar o relacionamento de predição entre suas opiniões, ou de modo explícito indicar os itens de interesse. Porém, em seguida estes sistemas automatizaram todo o procedimento através da coleção das pontuações dos itens pelos usuários. Um usuário de um sistema de filtragem colaborativa deve, portanto, pontuar cada item experimentado, indicando o quanto este item casa com sua necessidade de informação. Estas pontuações são coletadas para grupos de pessoas, permitindo que cada usuário se beneficie das pontuações (experiências) apresentadas por outros usuários na comunidade.

Sistemas de filtragem colaborativa simples apresentam para o usuário uma média de pontuações para cada item com potencial de interesse. Esta pontuação permite ao usuário descobrir itens que são considerados de interesse pelo grupo e evitar os itens que são considerados de pouco interesse. Sistemas mais avançados descobrem de maneira automática relações entre usuários (vizinhos mais próximos), baseado na descoberta de padrões comuns de comportamento. Um exemplo de ambiente baseado em filtragem colaborativa é o sistema de recomendação de filmes *MovieLens* [Riedl et al. 1999]. Nele o usuário insere pontuações para filmes que tenha visto e o sistema utiliza estas pontuações para encontrar pessoas com gostos similares. Desta forma o sistema pode recomendar filmes nos quais indivíduos com gostos semelhantes se interessariam, mas não assistiram ainda.

A tabela 5.1 apresenta na prática como a filtragem colaborativa pode funcionar. Por exemplo, se quisermos recomendar um produto ao usuário Mauro procuraremos outros usuários com hábitos de consumo semelhantes. No caso, Paulo e João já compraram produtos que Mauro também comprou (Prod2). Em seguida, recomendamos a Mauro produtos que estes dois outros usuários possuem, mas que Mauro ainda não possui como Prod1 e Prod5. A decisão sobre a recomendação destes produtos baseia-se no histórico de avaliações comuns e o valor de predição calculado.

Tabela 5.1. Recomendação baseada em filtragem colaborativa

Usuário	Prod ₁	Prod ₂	Prod ₃	Prod ₄	Prod ₅	Prod ₆
Paulo		x			X	
João	x	x				
Márcia			x	x	X	
Carlos			x			
Ana	x			x		
Mauro	?	x			?	

A técnica de Filtragem Colaborativa segue basicamente três passos:

1. Calcular o peso de cada usuário em relação à similaridade ao 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*”. No primeiro passo, a definição da similaridade pode ser realizada através da aplicação de diversos coeficientes de similaridade como Coseno, *Pearson*, entre outros. A equação (3) apresenta o coeficiente de *Pearson* (amplamente aplicado na literatura de Sistemas de Recomendação) [Reategui and Cazella 2005]:

$$corr_{ab} = \frac{\sum_i (r_{ai} - \bar{r}_a) (r_{bi} - \bar{r}_b)}{\sqrt{\sum_i (r_{ai} - \bar{r}_a)^2 \sum_i (r_{bi} - \bar{r}_b)^2}} \quad (3)$$

Sendo o $corr_{ab}$ a correlação do usuário alvo a com um determinado usuário b ; r_{ai} : é a avaliação que o usuário ativo a atribuiu para o item i ; r_{bi} : é a avaliação que o usuário ativo b atribuiu para o item i ; \bar{r}_a é a média de todas as avaliações do usuário ativo a , em comum com o usuário b ; \bar{r}_b é a média de todas as avaliações do usuário ativo b , em comum com o usuário a . Note que é preciso mais de uma avaliação em comum para que o índice seja útil, e os resultados variam entre 1 para similaridade total, e -1 para total dissimilaridade.

A predição é feita independentemente do coeficiente utilizado no cálculo da similaridade, pois ela será gerada através de uma média ponderada das avaliações dos vizinhos que obtiveram um coeficiente de similaridade aceitável, ou seja, com limiar igual ou superior, por exemplo, a 0,3. A equação 4 é utilizada para o cálculo da predição [Reategui and Cazella 2005].

$$p_{ai} = \bar{r}_a + \frac{\sum_{b=1}^n (r_{bi} - \bar{r}_b) * corr_{ab}}{\sum_{b=1}^n |corr_{ab}|} \quad (4)$$

Sendo $corr_{ab}$ é a correlação do usuário alvo a com um determinado usuário b ; p_{ai} a predição de um item i para um usuário alvo a ; \bar{r}_a é a média de todas as avaliações do usuário alvo a aos itens que foram pontuados por todos os seus usuários similares; r_{bi} é a avaliação que o usuário alvo b atribuiu para o item i ; \bar{r}_b é a média de todas as avaliações do usuário b , em comum com o usuário a .

A filtragem colaborativa apresenta algumas vantagens, como por exemplo, a possibilidade apresentar aos usuários recomendações inesperadas. O usuário poderia receber recomendações de itens que não estavam sendo pesquisados de forma ativa. Outra contribuição importante dos sistemas de filtragem colaborativa se refere à possibilidade de formação de comunidades de usuários pela identificação de seus gostos e interesses similares.

Uma questão importante em sistemas colaborativos refere-se à coleta de informações dos usuários, que pode apresentar algumas limitações:

1. **Problema do primeiro avaliador:** quando um novo item aparece no banco de dados não existe maneira deste ser recomendado para o usuário até que mais informações sejam obtidas através de outro usuário.

2. Problema de pontuações esparsas: o objetivo dos sistemas de filtragem colaborativa é ajudar pessoas, focando em documentos lidos ou itens adquiridos. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema existe um grande risco das pontuações tornarem-se muito esparsas.
3. Similaridade: caso um usuário tenha gostos que variam do normal este terá dificuldades para encontrar outros usuários com gostos similares, sendo assim suas recomendações podem se tornar pobres.

5.4. Filtragem Híbrida

A abordagem da filtragem híbrida procura, basicamente, 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, conforme apresentado pela figura 5.1.

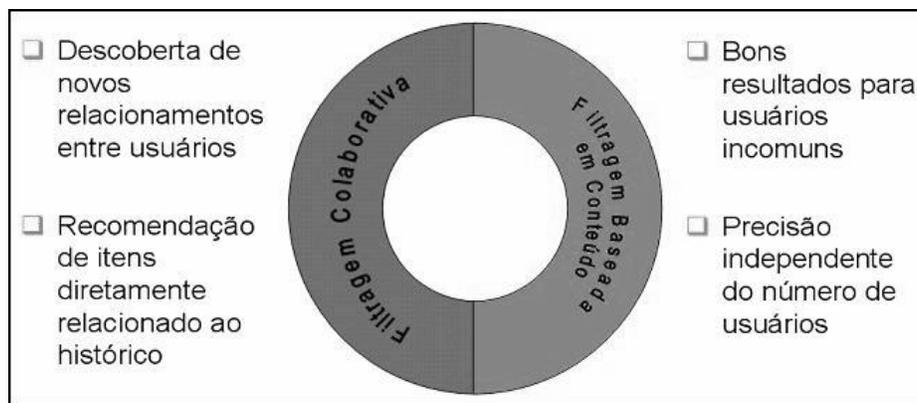


Figura 5.1. Filtragem híbrida

A literatura disponível sobre filtragem híbrida é ampla. Destacam-se os trabalhos produzidos por Adomavicius [Adomavicius et al. 2005], Burke [Burke 2002].

5.5. Filtragem Baseada em outros Contextos

De acordo com a visão de McDonald [McDonald 2003] a mudança mais importante a se desenvolver na nova geração de Sistemas de Recomendação é a devida complexidade na construção do modelo/perfil de usuário e, o uso apropriado desse modelo. Considerando Perugini et al [Perugini et al. 2004] modelos/perfis de usuário propiciam indiretamente conexões entre pessoas possibilitando e direcionando a recomendações mais eficientes. Dessa forma, acredita-se que perfis de usuário devem representar diferentes e ricos aspectos da experiência diária de um usuário, considerando a vida real como modelo.

Considere o percurso de um sistema de computador para atingir a mínima compreensão desta otimizada interação de como os humanos procedem nos seus processos de recomendação na “vida real”. Note que humanos usam em suas recomendações informações mais complexas que informações efetivamente usadas por um computador. Isto é, usualmente, sistemas computacionais usam informações tais

como competências, preferências, informações demográficas dos usuários, entre outras, para a tomada de decisão e posterior recomendação de uma informação, produto ou serviço. Em contraste, humanos, quando tomam decisões, recomendam e personalizam informações, produtos e serviços para outros humanos, além de usar as informações convencionais usadas pelos sistemas computacionais tendem, também, a usar informações adicionais relacionadas a habilidades sociais e psicológicas humanas, tais como, Traços de Personalidade e Emoção [Nunes and Aranha 2009].

Estudos recentes de psicólogos, neurologistas, antropólogos e cientistas computacionais [Damasio 1994], [Simon 1983], [Goleman 1995], [Paiva 2000], [Picard 1997], [Trappl et al. 2003], [Thagard 2006] têm demonstrado quão importante os aspectos psicológicos humanos, tais como Traços de Personalidade e Emoções, são durante o processo de tomada de decisão humana.

Felizmente, algumas pesquisas vêm sendo realizadas por cientistas, principalmente, de Computação Afetiva focando na identificação e na modelagem da Emoção do usuário em sistemas computacionais ([Lisetti 2002], [Picard 1997], [Paiva 2000], [Nunes 2009a]). Pesquisadores pertencentes ao *W3C Emotion Incubator Group* [W3C 2009] têm ampliado esforços para definição de uma padronização de uma *Markup Language* para representar Emoções. Entretanto, na área de Personalidade é ainda incipiente [Nunes 2009].

Mesmo sabendo ser impossível perfeitamente antecipar as necessidades humanas individuais para recomendar o produto certo, sabe-se que quanto mais ricas forem as informações sobre o usuário, mais precisos serão os produtos, serviços e /ou pessoas recomendadas. Basta observar o que ocorre na vida real.

Nessa linha, note que Nunes [Nunes 2009] expande o modelo inicial de técnicas de recomendação proposto por Gonzalez et al [Gonzalez et al. 2007]. Gonzalez por sua vez expande o modelo proposto por Burke [Burke 2002]. O modelo de Burke [Burke 2002] categoriza as técnicas de recomendação em cinco: baseada em conteúdo, filtragem colaborativa, demográfica, baseada em conhecimento e baseada em utilidade. Gonzalez et al [Gonzalez et al. 2007] por sua vez cria uma nova categoria intitulada recomendação baseada em outros contextos, incluindo aspectos Psicológicos, tais como, Inteligência Emocional e Interação Social, usando como fonte um Perfil de usuário baseado em contextos diferenciados dos tradicionais modelos demográficos ou baseados em preferência do usuário.

Considerando esse aspecto Nunes [Nunes 2009] ao expandir o modelo proposto por Gonzalez et al [Gonzalez et al. 2007], apresenta a Personalidade do usuário como um fator relevante na otimização das recomendações ao usuário, como apresentado na figura 5.2.

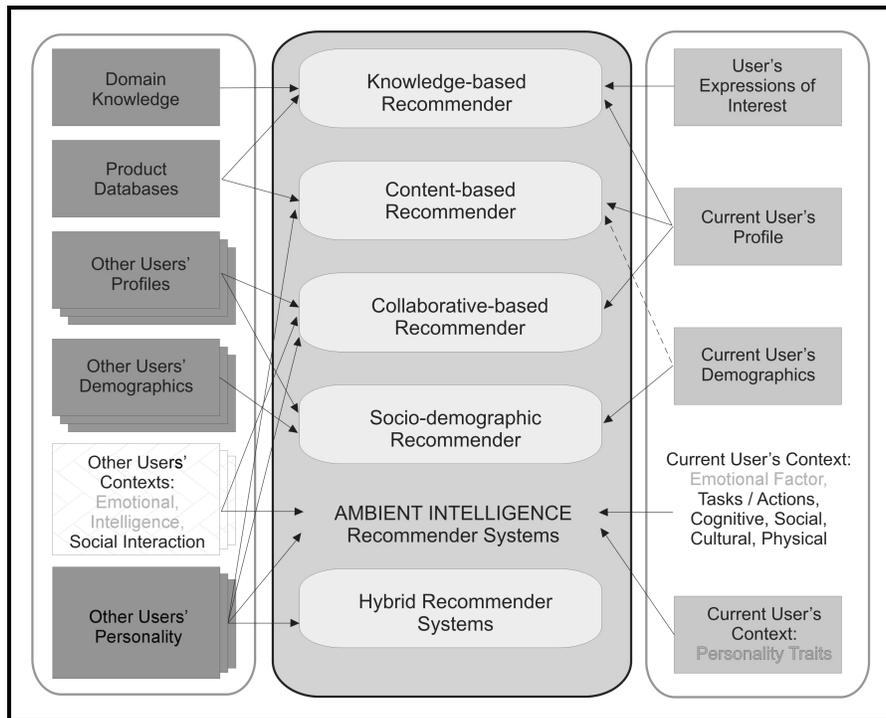


Figura 5.2. Sistemas de Recomendação baseado em Personalidade [Nunes 2009]

A técnica de recomendação baseada em Personalidade geralmente é aplicada juntamente a outra técnica como a filtragem colaborativa e/ou baseada em conteúdo, representando uma técnica de recomendação híbrida.

Note que para que seja viável a recomendação baseada na personalidade do usuário necessita-se a representação da Personalidade, cada projetista de sistema pode usar uma abordagem diferenciada. A seguir apresenta-se uma forma de representação sob a visão de Nunes [Nunes 2009].

- Na Psicologia não existe um consenso para a definição de Personalidade. Burger [Burger 2000] define Personalidade como “um padrão de comportamento consistente e processo intrapessoal que é originado internamente no indivíduo”. A Personalidade é mais que apenas as aparências superficiais e físicas, ela é relativamente estável e previsível em um indivíduo, porém ela não é necessariamente rígida e imutável (geralmente ela permanece estável por um período de 45 anos iniciando na fase adulta).

A Personalidade pode ser definida segundo muitas abordagens. Uma abordagem interessante é a de Traços de Personalidade. Essa abordagem permite diferenciar psicologicamente pessoas usando traços mensuráveis e conceituáveis, chamados Traços de Personalidade. Traços de Personalidade são formados por um conjunto de características humanas factíveis de modelagem e implementação em computadores.

Os Traços de Personalidade foram historicamente definidos por Allport [Allport 1927] que criou 17.953 Traços (Traços “comuns” e “individuais”) para descrever a personalidade de um indivíduo. Considerando que a maioria das diferenças individuais (representadas pelos Traços individuais) eram insignificantes nas interações diárias humanas, objetivando limitar exponencialmente o número de definições de Traços, os pesquisadores assumiram que todos os homens eram identificáveis “como algum outro homem”. Considerando isso, pesquisadores reduziram mais de 99% dos Traços, pois eles consideraram que somente cinco fatores eram replicáveis. Como resultado, o modelo *Big Five* [John and Strivastava 1999] foi criado. Mesmo se o *Big Five* representasse grande eficiência na representação da estrutura de Personalidade, ele não garantia exaustivamente todas as dimensões de personalidade. Dessa forma, *facet*s também foram criadas e usadas pelos psicólogos para dotar o *Big Five* de características mais detalhadas.

Note que exemplos ilustrativos de Sistemas de Recomendação que utilizam Personalidade para no auxílio à tomada de decisão computacional são apresentados na forma de Estudo de Caso em [Nunes 2009], [Nunes 2008a] e [Prada et al. 2009].

5.6. 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].

Quando se trabalha com sistemas de recomendação na *web*, a mineração de dados torna-se um recurso importante para a descoberta de relações entre itens, entre usuários e entre itens e usuários. Através da mineração dos arquivos de *log* [Yang et al. 2002], por exemplo, pode-se obter conhecimentos aprofundados a respeito dos usuários que se conectaram a um *website* [Spiliopoulou 2000]. Este conhecimento pode ser utilizado para a personalização da oferta de produtos [Rucker and Polanco 1997], na estruturação de *sites* de acordo com o perfil de cada internauta e personalização [Nulvenna et al. 2000] do conteúdo das páginas.

As próximas seções apresentam diferentes tipos de descoberta de conhecimento aplicáveis nos Sistemas de Recomendação.

5.6.1. Tipos de Descoberta

O desenvolvimento de sistemas de DCBD está relacionado com diversos domínios de aplicações, como por exemplo: análises corporativas, medicina, biologia, mercado, entre outros [Apte et al. 2002], [Smyth 2002] e [McCarthy 2000]. De acordo com a aplicação algumas técnicas de mineração podem ser mais eficientes do que outras.

5.6.1.1. Regra de Associação

Neste tipo de função cada tupla da base de dados consiste em um conjunto de atributos denominados itens [Cabena 1997], [Plastino 2001], [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 ($FSup = |X \cap Y| / N$). O FConf constitui-se na razão do número de tuplas que satisfazem X e Y sobre o número de tuplas que satisfazem X ($FConf = |X \cap Y| / |X|$).

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 5.2 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 5.2. 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 5.2 são apresentadas a seguir, aplicando-se os valores de FSup e FConf maiores ou iguais aos especificados pelo usuário.

Conjunto de itens freqüentes: Artigo2, Artigo4. FSup = 0.3

Regra: Se (Artigo2) então (Artigo4). FConf = 1.

Conjunto de itens freqüentes: Artigo2, Artigo5. FSup = 0.3

Regra: Se (Artigo2) então (Artigo5). FConf = 1.

Conjunto de itens freqüentes: Artigo4, Artigo5. FSup = 0.4

Regra: Se (Artigo4) então (Artigo5). FConf = 0.8.

Regra: Se (Artigo5) então (Artigo4). FConf = 0.8.

Conjunto de itens freqüentes: Artigo2, Artigo4, Artigo5. FSup = 0.3

Regra: Se (Artigo2 e Artigo4) então (Artigo5). FConf = 1.

Regra: Se (Artigo2 e Artigo5) então (Artigo4). FConf = 1.

Regra: Se (Artigo2) então (Artigo4 e Artigo5). FConf = 1.

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. Entre algoritmos aplicados neste tipo de função destacam-se o *Apriori* e o *Partition* [Plastino 2001].

5.6.1.2. Classificação

Nessa função cada uma das tuplas pertence a uma das classes entre um conjunto pré-definido de classes [Cabena 1997], [Quinlan 93]. 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 and Yan 1996].

Considerando apenas os atributos *status*, país e área de interesse dos usuários como relevantes para a previsão, são apresentados na tabela 5.3 os valores dos atributos selecionados, junto com o valor do atributo objetivo e área de interesse.

Tabela 5.3. Entrada de dados para a tarefa de classificação

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 árvore de decisão gerada com a aplicação de classificação para o exemplo acima poderia apresentar o formato apresentado na figura 5.3.

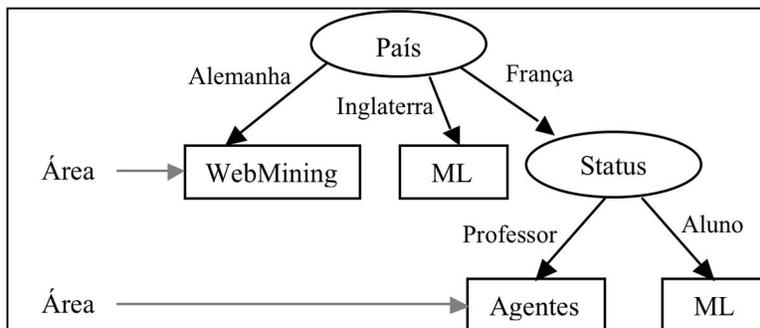


Figura 5.3. Árvore de Decisão

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.

5.6.1.3. 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 5.4 ilustra este processo, onde pode-se observar que cada ponto de treinamento plotado no gráfico representa um objeto qualquer que possui dois atributos que o representa, por exemplo, objeto pessoa com sua altura respectiva (eixo x) e peso (eixo y). 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.

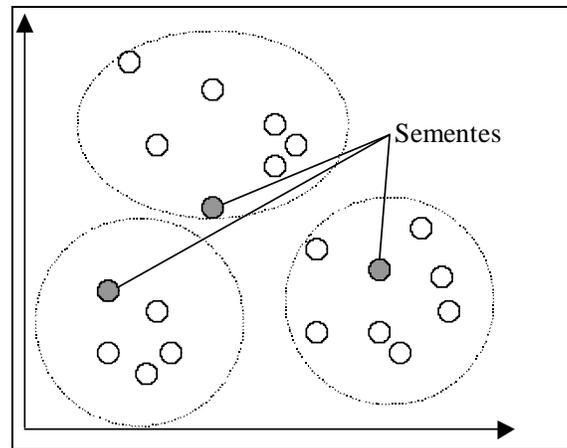


Figura 5.4. Formação inicial de grupos

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 5.5.

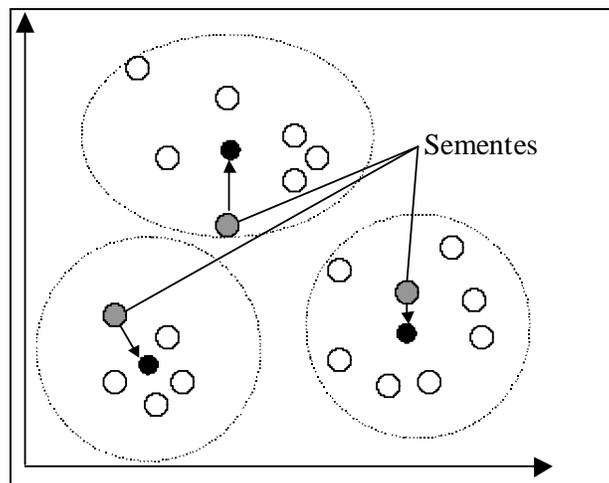


Figura 5.5. Recálculo dos centróides

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.

5.7. Mineração de Textos

A mineração de textos (*text mining*) refere-se ao processo de obtenção de informação de qualidade a partir de texto em linguagem natural. É inspirado na mineração de dados (seção 5.6). A mineração de textos ganhou importância com o crescimento da Internet e dos mecanismos de busca, uma vez que com a mineração de textos é possível extrair informações relevantes de uma grande base de textos não estruturado ou semi-estruturado.

No contexto de Sistemas de Recomendação a mineração de textos é amplamente aplicada na técnica de filtragem baseada em conteúdo (seção 5.2), uma vez que é necessária a análise do conteúdo que descreve o item a ser recomendado. No processo de mineração de textos o pré-processamento é fundamental. Normalmente muitas palavras de um documento não são importantes para representá-lo semanticamente. Podemos afirmar que na maioria das vezes, substantivos ou grupo de substantivos são as palavras mais representativas em um documento. A tarefa de pré-processamento, no contexto de recuperação de informação, possibilita a extração de informações úteis do texto [Cowie and Lehnert 1996]. As operações de pré-processamento referem-se (entre outras) a: remoção de *stopwords*, análise léxica e redução dos termos a um radical comum.

6. Sistemas de Recomendação: exemplos de aplicações

Nesta seção, alguns Sistemas de Recomendação são descritos, enfatizando suas principais características.

6.1. Projetos Acadêmicos

6.1.1. RINGO

Ringo é um sistema desenvolvido para recomendação personalizada de música e foi desenvolvido no *Massachusetts Institute of Technology* [Maes and Shardanand 1995]. O trabalho explora similaridades entre os gostos de diferentes usuários para recomendar itens, baseado no fato de que os gostos das pessoas apresentam tendências gerais e padrões entre gostos, e entre grupos de pessoas. Neste sistema as pessoas descrevem suas preferências musicais através da avaliação de algumas canções. Estas avaliações constituem o perfil dos indivíduos. O sistema usa então estes perfis para gerar recomendações para usuários individuais. Para o seu funcionamento, primeiramente usuários similares são identificados. A partir desta identificação e comparação de perfis o sistema pode prever o quanto o usuário gostaria de um álbum/artista que ainda não foi avaliado pelo mesmo.

Quando o usuário acessa o *Ringo* pela primeira vez, é apresentado a ele uma lista de 125 artistas. O usuário avalia estes de acordo com o quanto gosta de ouvi-los. Caso o usuário não esteja familiarizado com o artista ou não possua uma forte opinião sobre o mesmo, é solicitado que não avalie o artista para não gerar distorções. Para a pontuação é aplicada uma escala *Likert* de 7 pontos, sendo o 1 (não gosta), 4 (indiferente) e 7 (adora).

A lista de artistas enviada para os usuários é dividida em duas partes. A primeira parte da lista é gerada a partir dos artistas mais pontuados (*ranking*), o que assegura que

um novo usuário tenha a oportunidade de pontuar artistas que outros já tenham pontuado. A outra parte da lista é gerada através de uma seleção randômica a partir de um banco de dados sobre os artistas. Uma vez que o perfil inicial da pessoa tenha sido submetido, o usuário pode solicitar ao *Ringo* algumas predições:

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 inclui nenhuma informação em particular sobre a identidade dos outros usuários que contribuíram com as recomendações. Deve-se enfatizar esta característica em sistemas colaborativos ou de informação social (muitas vezes usados como sinônimo): a identidade de quem avaliou deve ser mantida em segredo.

O sistema *Ringo* também permite a inserção de comentários (como *feedback* explícito) sobre o produto recomendado. Os próprios usuários podem incluir novos artistas e álbuns no banco de dados.

6.1.2. GroupLens

O *GroupLens* constituiu-se em um projeto de pesquisa desenvolvido pela Universidade de *Minnesota*, aplicado à filtragem colaborativa de notícias, auxiliando pessoas a encontrar artigos de interesse entre um grande volume de artigos ofertados [Riedl et al. 1997].

O Sistema coleta a avaliação dos usuários referente aos artigos lidos (escala de 1-5 pontos) e utiliza esta avaliação para identificar os vizinhos mais próximos com avaliações semelhantes. Desta forma o sistema é capaz de predizer se um usuário pode se interessar por um novo artigo baseando-se nas avaliações dos usuários vizinhos.

6.1.3. Fab

O *Fab* é 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 [Balabanovic e Shoham 1997]. Foi desenvolvido pela Universidade de *Stanford* em um projeto para recomendação de páginas *web* (documentos) para usuários.

Objetivando a construção de uma arquitetura híbrida, perfis de usuários são gerados a partir da análise de conteúdo e comparados para identificar usuários similares para recomendação colaborativa. O sistema tem uma implementação distribuída e foi construído como parte do projeto da biblioteca digital da Universidade de *Stanford*. O sistema foi implementado através de 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.

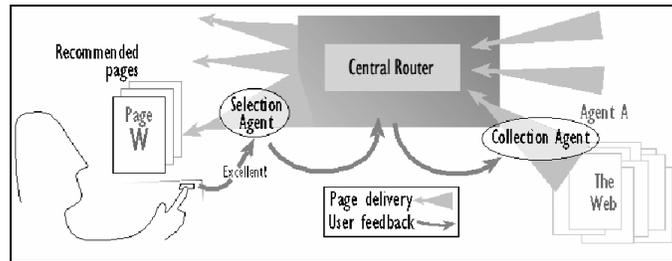


Figura 6.1. Visão geral da arquitetura do Fab

Após a recomendação o usuário tem a possibilidade de avaliar o item aplicando uma escala *Likert* de 1-7 pontos.

6.1.4. Collaborative Recommender Agent CORA

O sistema *Collaborative Recommender Agent CORA* (Christoph apud [Lueg 1998]) é um sistema distribuído assíncrono para filtragem colaborativa ativa de documentos da *web*, desenvolvido na Universidade de Zurique. O *CORA* habilita os usuários a recomendar URLs através de uma simples clicagem de *mouse* [Lueg 1998].

Os usuários eram providos com um agente pessoal de recomendação que monitorava o comportamento no uso do *browser* pelo usuário. O agente reconhecia a página *web* que o usuário estava visualizando e apresentava a mesma em uma janela de recomendação especial. Caso o usuário decidisse recomendar a URL para um grupo de usuários pré-definido, a URL poderia ser enviada pela clicagem de um ícone específico. Assim como o agente monitorava o envio de URLs este também monitorava a chegada de URLs enviadas por outros usuários.

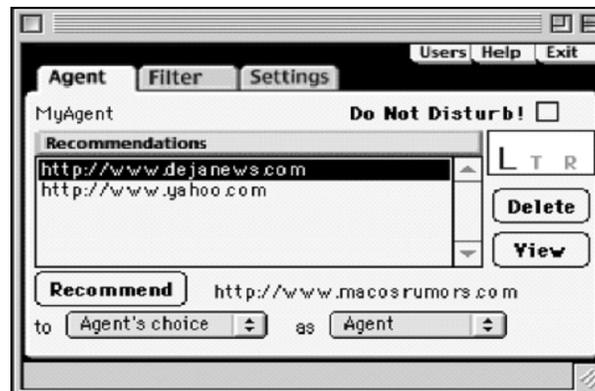


Figura 6.2. Interface da ferramenta CORA

Na interface apresentada na Figura 6.2 pode-se observar três letras L,T e R. Cada uma, de acordo com o seu tamanho, procura simbolizar: o montante de *Links* seguidos tendo as páginas recomendadas como ponto de partida; o Tempo utilizado no acesso da página; e um fator especial de recomendação, o qual é computado com base no uso passado das recomendações. Lueg destaca que sempre que alguém está querendo buscar uma informação de valor, este possui um tempo determinado para aplicar nesta busca, mas este esforço apresenta um limite pequeno [Lueg 1998].

6.1.5. Emagister.com

Emagister.com é uma empresa de *e-commerce* que promove cursos de treinamento *online* para três milhões de usuários [Gonzalez et al. 2007]. Antes da contribuição de

Gonzalez a *Emagister.com* recomendava cursos de treinamento baseado na combinação das preferências explícitas do usuário e também nos *feedbacks* implícitos/explicítos dos mesmos. O *feedback* implícito do usuário era adquirido através do monitoramento da navegação do usuário e seus *clicks* no *site*, enquanto que o *feedback* explícito do usuário era colhido através do ranqueamento que ele fornecia nos itens recomendados a ele. Objetivando melhorar as recomendações do *Emagister.com* os pesquisadores decidiram inovar o processo de recomendação levando em consideração não apenas os interesses e preferências do usuário, mas também decidiram usar a sensibilidade do usuário considerando alguns atributos relevantes na área de treinamento.

A experiência foi realizada considerando 3.162.069 usuários do *Emagister.com*. Setenta e cinco características (objetivas, subjetivas e emocionais) foram extraídas para construir um perfil adequado do usuário. As informações para alimentar o perfil foram extraídas de banco de dados sócio-demográficos (para atributos objetivos) e *WebLogs* baseados em hábitos implícitos de navegação do usuário (para atributos subjetivos e emocionais). As informações extraídas dos *WebLogs* procedem das respostas do usuário considerando um teste gradual de Inteligência Emocional (*EIT test*). A primeira estratégia de marketing para obter os atributos emocionais e seus valores no *EIT test* foi o uso de newsletter. A medida que os usuários respondiam partes do questionário via *newsletter* seu perfil emocional estava sendo completado. Para os usuários que responderam esporadicamente todas as questões do *EIT test* recomendações mais precisas foram geradas. Gonzalez provou que usuários que receberam recomendações dotadas dos aspectos de Inteligência Emocional efetivamente se matricularam em mais cursos no *Emagister.com*.

6.1.6. RECMOBCOLABORATIVA

O Sistema de Recomendação Colaborativa Móvel (RECMOBCOLABORATIVA) [Cazella et al. 2008] constitui-se em um Sistema de Recomendação brasileiro baseado em Filtragem Colaborativa (FC) para ambientes móveis, e teve como objetivo permitir a recomendação de conteúdo interessante para alunos de graduação em Engenharia da Computação. Na figura 6.3 é apresentado o modelo.

Conforme pode ser observado na figura 6.3, o modelo RECMOBCOLABORATIVA compõe-se de:

1. Módulo de Recomendação: responsável pelo cálculo das similaridades e definição dos vizinhos próximos, bem como a predição de itens a serem recomendados;
2. Recomendação Colaborativa: responsável pela interação com o projeto LOCAL e encaminhamento das recomendações produzidas para que o LOCAL encaminhe as mesmas para os dispositivos móveis (maiores informações [Cazella et al. 2008]) ;
3. Aplicativo de Avaliação de recomendações: aplicativo voltado à coleta das avaliações dos usuários aos itens recomendados;
4. Base de dados: com os itens a serem recomendados, os usuários cadastrados e seus perfis, bem como as avaliações fornecidas pelos usuários.

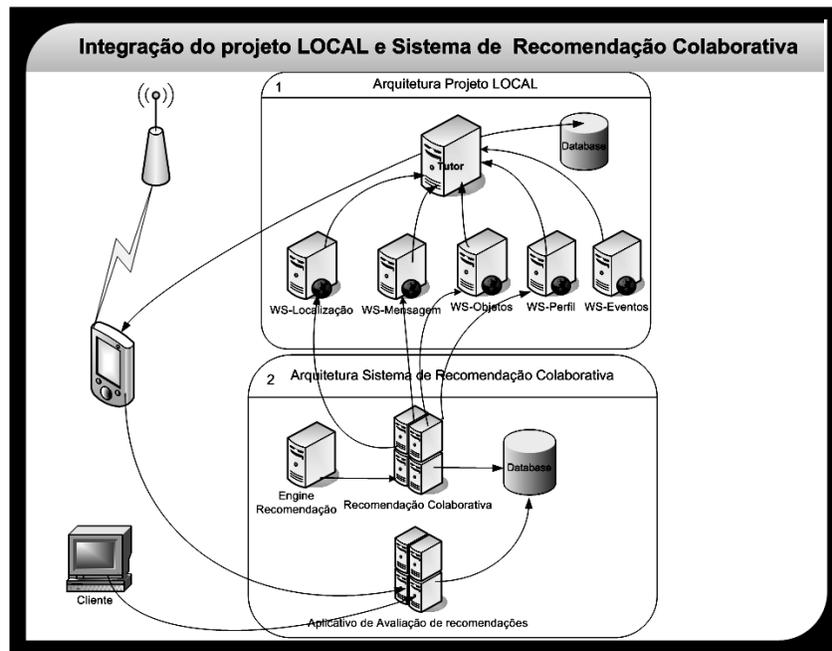


Figura 6.3. Modelo RECMOBCOLABORATIVA

O usuário ao se conectar e se identificar no sistema usando um dispositivo móvel (*palm HP iPAQ*), recebia automaticamente todas as recomendações calculadas pelo protótipo que implementa o modelo RECMOBCOLABORATIVA. O projeto RECMOBCOMPETENCE deu continuidade ao RECMOBCOLABORATIVA, e maiores informações podem ser encontradas em [Cazella et al. 2009].

6.1.7. Sistemas de Combinação Social

Como descrito até então no decorrer desse capítulo, Sistemas de Recomendação são definidos como sistemas que promovem a recomendação de produtos e serviços bem como de pessoas. A Recomendação de pessoas originou um tipo especial de Sistema de Recomendação intitulado de Sistema de Combinação Social.

Terveen e Hill [Terveen and Hill 2001] descrevem recomendação como um recurso que ajuda usuários a fazer escolhas em um universo de alternativas, agindo como uma espécie de filtro. Geralmente, uma recomendação é baseada nas preferências/características do produto, serviço ou pessoa a ser recomendada, considerando também as características do requisitador da recomendação. Terveen e Hill consideram aspectos sociais no processo de recomendação o que evidencia a particularidade do Sistema de Combinação Social.

Antes da criação e definição da abordagem de Sistemas de Combinação Social, surgiram outras abordagens intermediárias como a abordagem intitulada de Sistema de Mineração de Dados Social [Terveen and Hill 2001]. Os Sistemas de Mineração de Dados Social comporta a descoberta de interesses em uma comunidade particular de usuários. Esse sistema se beneficia da experiência e opinião de pessoas especialistas, alguns exemplos são:

1. *PHOAKS* [Terveen et al. 1997] minera mensagens do *Usenet*. A mineração é realizada através de pesquisas realizadas em páginas *web* já postadas em mensagens dos usuários. O *PHOAKS* seleciona as páginas *web* que foram citadas

nas mensagens postadas no *Usenet*. Após a seleção o *PHOAKS* categoriza essas páginas *web* em listas de URLs mais relevantes considerando o domínio específico.

2. *TopicShop* [Amento et al. 2003] minera informações de páginas *web* e, então fornece uma interface para os usuários terem um acesso facilitado à páginas *web* já visitadas por grande parte dos especialistas.

Nos casos descritos acima, além da informação sobre as páginas *web*, a informação sobre usuários que têm selecionado as páginas *web* fica também disponível. Essa informação é importante pois, capacita, implicitamente, a interação entre especialista (que selecionou a página no passado) e a pessoa que esta procurando por um especialista no domínio específico.

No *ReferralWeb* [Kautz et al. 1997], por exemplo, Kautz et al modelaram o sistema para, explicitamente, criar um link entre as pessoas (especialistas). Esses links explícitos são extraídos da Rede Social, ou seja, o sistema analisa os documentos *web* e processa as co-ocorrências de nomes dentro dos documentos e, finalmente, associa esses nomes de pessoas com seus especialistas, criando um relacionamento social, visualizado como uma comunidade de especialistas [Amento et al. 2003].

Em contraste com o *PHOAKS* e o *TopicShop*², os Sistemas de Recomendação como *ReferralWeb*, estão mais interessados em encontrar pessoas do que as informações recomendadas pela pessoa (ou grupo/comunidade).

Como os Sistemas de Recomendação têm mostrado um crescente interesse em elementos sociais, os pesquisadores decidiram re-nomear esse tipo especial de sistema de Recomendação que permite recomendar pessoas a pessoas. Essa nova abordagem permite a construção explícita de link entre pessoas que compartilham as mesmas preferências/características, considerando as pessoas como potencial provedores de serviços. Em 2005, Terveen e McDonald [Terveen and McDonald 2005], definitivamente decidiram nomear essa abordagem de Sistema de Recomendação a intitulando de Sistema de Combinação Social.

6.1.7.1. Definindo Sistemas de Combinação Social

Pessoas são inerentemente criaturas sociais e, por essa razão, pessoas estão constantemente procurando por outras para compartilhar seus interesses, resolver seus problemas, criar encontros, conquistar amigos, conversar informalmente, buscar por um especialista, bem como por diversos outros motivos e/ou interesses. Na vida real, muitas vezes as pessoas procuram por esses serviços de recomendação de pessoas em uma infinidade de contextos, tais como, formação de casais, seleção de melhores pares para composição das mesas em jantares festivos e/ou conferências, encontrar um novo emprego para um alto executivo, encontrar o membro certo para fazer parte de uma equipe de alto nível, entre outros. Dessa forma, com o aumento na demanda de pedidos de combinação social os cientistas da computação resolveram explorar e propor uma forma semi-automática de realizar essa tarefa e então, como extensão aos Sistemas de Recomendação criaram os Sistemas de Combinação Social.

² Ambos sistemas são Sistemas de Recomendação categorizados como Sistema de Mineração de Dados Social.

O serviço mais popular de combinação disponível na *web* é o Sistema de formação de casais. Recentemente as Redes Sociais apareceram como um grande apelo social criando redes de amizade entre pessoas, usado para comunicação entre amigos e também propiciando a criação de novas amizades. Adicionalmente as Redes Sociais, as Redes de Reputação apareceram com intuito de garantir a lealdade e a confiança entre clientes e prestadores de serviço. A seguir descreve-se brevemente cada uma destas perspectivas mais populares:

1. Redes Sociais: contribuem ao permitir a construção de contatos entre “amigos” e “amigos de amigos” baseado em Redes de Amizade (*FOAF Project*). Considerando Redes de Amizade, qualquer pessoa pode encontrar uma outra considerando até no máximo de seis graus de separação [Watts 2003]. A combinação é aplicada com o objetivo de possibilitar a criação de *links* entre amigos da Rede Social, possibilitando a criação de uma Rede de Amizade. Infelizmente, nas Redes Sociais não existe um grande leque para combinações a ser explorado, além da Rede de Amizade. Entretanto, os usuários podem usar técnicas de descobertas diversas, baseada em informação demográfica, por exemplo, procurando randomicamente e livremente por novas pessoas (potenciais futuros amigos). Exemplos de Redes Sociais são: *Orkut*, *FriendFinder*, *Mendeley research networks*, *LinkedIn*.

2. Redes de Encontros/Formação de Casais: são sistemas que combinam casais colocando-os em contato. Uma boa combinação deveria considerar uma variedade complexa de aspectos de usuário, tais como características físicas, preferências, aspectos psicológicos, perspectivas para o futuro e outros. Sistemas de formação de casais deveriam criar e apresentar um perfil de usuário complexo baseado tanto em aspectos convencionais como em aspectos não usuais. Geralmente aspectos não usuais são mais difíceis de extrair do usuário e são extraídos através de testes psico-sociais, como por exemplo, características emocionais [Lopes et al. 2003], identidade [Donath 2000], [Boyd 2002] e Traços de Personalidade [Fiore and Donath 2004], [Fiore 2004], [Allport 1927].

Os aspectos psico-sociais são importantes no momento da combinação dos casais, pois permite gerar recomendações considerando similaridade e/ou complementaridade de algum aspecto determinante para o sucesso afetivo do casal. As técnicas mais eficientes de redes de encontros usam a combinação de casais baseado na complementaridade de aspectos psico-sociais. Alguns exemplos de Redes de Encontros são *Match.com* e *ParPerfeito.com.br* que trabalham com perfil menos rico que o *e-Harmony* e o *e-Chemistry*, por exemplo. Ainda, algumas Redes de Encontros permitem a comunicação entre os usuários apenas a partir da combinação gerada pelo sistema, forçando assim a compatibilidade e a suposta satisfação do cliente. Outras Redes permitem que o usuário, não satisfeito com a combinação gerada pelo sistema, opte por buscar seus parceiros aleatoriamente (geralmente esses sistemas não usam aspectos psico-sociais na combinação social gerada pelo sistema).

3. Redes de Reputação: é uma comunidade que tem como objetivo colecionar, gerenciar e promover usuários como prestadores de serviço e/ou potenciais consumidores (algumas Redes Sociais podem também ser consideradas Redes de Reputação). Geralmente, a Reputação é utilizada para se gerar uma rede de confiança para transações comerciais e/ou profissionais. Isto significa que um

usuário (prestador de serviço) que tem um perfil cadastrado em uma Rede de Reputação, fornece alguns serviços para outro usuário (cliente). O cliente solicita um serviço confiando nas informações declaradas na Rede de Reputação. A Rede de Reputação normalmente deixa claro o índice de satisfação dos clientes sobre a qualidade de serviços já prestados pelo mesmo prestador de serviços em eventos anteriores. Nas Redes de Reputação os clientes podem livremente buscar prestadores de serviço usando *tags*, emails ou ramo de atividades. Alguns exemplos são *IKarma.com*, *Opinity.com*, *Mendeley.com*, *LinkedIn.com* e *eBay.com*. No *IKarma*, por exemplo, o perfil de usuário é simples, é baseado apenas em características básicas do usuário e suas *tags*. O *IKarma* oferece uma máquina de pesquisa convencional capaz de encontrar um usuário considerando informações como seu email ou *tags* (normalmente relacionado aos serviços disponibilizados e/ou competências do usuário). A recomendação baseada em Reputação poderia ser viabilizada se os usuários tivessem sua reputação explicitamente representada diferentemente ao que se apresenta hoje categorizado como “*written reviews*”, que são baseadas em linguagem natural. Atualmente no *IKarma* e *eBay* a recomendação pode ser consultada considerando o ranqueamento feito com base em estrelas, porém o ranqueamento não é relacionado a nenhum outro fator de atributos objetivos como o tipo de serviço a ser fornecido.

Essas três perspectivas apresentadas são os exemplos mais populares de Sistemas de Combinação Social utilizados na *web*. Elas apresentam uma característica comum que é o fato de que todas as três perspectivas permitem a combinação objetivando criar uma rede de possíveis *links* entre as pessoas considerando características/objetivos específicos dos usuários. Entretanto, outras aplicações usando Sistemas de Combinação Social têm aparecido e serão apresentados a seguir.

6.1.7.2. Formalização de Sistemas de Combinação Social

Terveen e McDonald [Terveen and McDonald 2005] criaram o termo Sistemas de Combinação Social objetivando definir os Sistemas de Recomendação capazes de recomendar pessoas às outras ao invés de apenas recomendar informações, produtos ou serviços às mesmas. Os pesquisadores fazem essa denominação diferenciada se calcando no fato de que pessoas são muito mais complexas para serem definidas conceitualizadas e categorizadas que informação, produtos e serviços e, então dessa forma a recomendação também será muito mais complexa e detalhada. Quando se trata de Sistemas de Combinação Social o perfil dos usuários é cuidadosamente descrito e efetivamente mais complexo que o modelo usado tradicionalmente em Sistemas de Recomendação convencionais. Isto significa que o perfil do usuário inclui não somente informações demográficas, preferências e competências, por exemplo, inclui, também, informações mais complexas, tais como, informações sobre aspectos psico-sociais da identidade do usuário, tais como, aspectos de personalidade, emoção, familiaridade, aspectos sociais, entre outros. A combinação deveria incluir regras de atração interpessoal, amizade, compatibilidade conjugal e composição de grupo. Sistemas de Combinação Social foram classificados por Terveen e McDonald [Terveen and McDonald 2005] como:

1. Sistemas de Recomendação Social por Necessidade de Informação: são sistemas capazes de combinar pessoas de acordo com seu relacionamento social e uma informação requerida. Esses sistemas fazem uso de perfis de usuário baseado em

duas características: especialidade do usuário e relações sociais do usuário. A ilustração segue em três exemplos abaixo:

- a. *ReferralWeb* [Kautz et al. 2007] usa uma Rede Social para encontrar um especialista. O sistema minera documentos *web* públicos objetivando encontrar um assunto específico. A co-ocorrência de nomes de especialistas nesses documentos indica uma relação social. O *ReferralWeb* usa a técnica de Sistema de Recomendação Híbrido na sua implementação.
- b. *Expertise Recommender* [McDonald and Ackerman 2000], [McDonald 2001] é um sistema capaz de identificar especialistas para resolver problemas específicos que usuários comuns seriam incapazes de fazê-lo. As informações especialistas são extraídas de um Sistema de Mineração de Dados e, também, de técnicas de observação de Banco de Dados de suporte técnico.
- c. *Pyramid Collaborative Filtering* (PCF) [Rezek et al. 2007] é um sistema que propõe a recomendação de um “ajudante confiável” (professor ou aluno) que poderia, potencialmente, dar suporte a um estudante com algum tipo de deficiência em um conceito específico não suficientemente aprendido durante um curso online em uma comunidade de e-learning³. A recomendação do “ajudante confiável” é combinada considerando características do usuário, tais como, conhecimento do domínio, modelo do aluno (comportamento e estilos de aprendizagem) e credibilidade. A recomendação é gerenciada por um agente guia, inteligente e autônomo, que usa técnica de implementação baseada em filtragem colaborativa.

Esses tipos de sistemas suportam a busca de informação pela identificação da potencial pessoa que é capaz de ajudar através da combinação da sua especialidade.

2. Sistemas de Informação com Combinação Social Implícita: são sistemas onde o foco é na categorização da informação de grandes espaços navegacionais. Ele capacita o reconhecimento de quem postou a mensagem, tornando implicitamente possível encontrar alguém capaz de ajudar o usuário em uma especialidade.
 - a. *PHOAKS* [Terveen et al. 1997] reconhece, categoriza e redistribui recomendações de recursos *web* extraídos de mensagens postadas no *Usenet*. Uma informação importante é que o *PHOAKS* guarda, também, quem postou a mensagem, possibilitando a posterior recomendação implícita dessa pessoa.
 - b. *Answer Garden* [Ackerman and Malone 1990], [Ackerman 1994], [Ackerman and McDonald 1996] é usado nas organizações para organizar e gerenciar sua memória. A memória é hierarquicamente organizada por questões e respostas. Estas questões e respostas são classificadas por domínio, constando também o nome do especialista que postou a informação. Esse tipo de sistema suporta a pesquisa de

³ A comunidade de e-learning é criada e suportada por professores e alunos que usam o *Confidence Intelligent Tutoring System (CITS)* [Rezek et al 2002].

informações trabalhadas na organização. O Sistema capacita usuários a encontrar informações pessoais sobre a pessoa que postou a mensagem, dessa forma promovendo uma combinação social implícita.

3. Sistema de Combinação Social Oportunista: são sistemas que combinam usuários considerando a oportunidade em um dado momento, geralmente considerando interesses em comum, como demonstram os exemplos abaixo:
 - a. *Social Net* [Terry ET al. 2002] combina usuários que estão na mesma localização física. A combinação ocorre considerando os interesses em comum e são determinados pela identificação momentânea da mesma localização física dos usuários. Por exemplo, usuários que são associados ao *Social Net* e estão em um congresso mas não se conhecem. O sistema identifica que os mesmos possuem os mesmos interesses em pesquisas, e, então pode avisar aos usuários que existem potenciais colaboradores no mesmo congresso.
 - b. *I2I* [Budzik et al. 2002] fomenta a comunicação oportunista entre usuários que acessam a mesma página *web*. Interesses compartilhados são descobertos aplicando similaridade nos textos das páginas *web* que os usuários têm visitado.
4. Outras abordagens relacionadas: são pesquisas em áreas correlatas usadas em Sistemas de Combinação Social.
 - a. Recomendação de Grupos [O'Connor et al. 2001], [McCarty and Anagnost 1998]: são Sistemas de Recomendação em que as recomendações são geradas para grupo de usuários e não individualmente para um usuário. Os Sistemas de Combinação Social podem se beneficiar das tecnologias usadas para recomendação de informações de preferência grupal e não individual, como de praxe.
 - b. Comunidades Online: [Preece 1998], [Preece and Maloney-Krichmar 2003]: são espaços virtuais onde usuários se encontram com objetivo de discutir tópicos de interesse mútuos. Os Sistemas de Combinação Social podem contribuir para automatizar a combinação de membros nessas comunidades, objetivando encontrar novas amizades, pares românticos, parceiros profissionais, prestadores de serviço, entre outros.
 - c. Sistemas de Presença [Hudson and Smith 1996]: são sistemas de mensagem instantânea onde usuários podem se comunicar com seus companheiros, familiares, colegas, parceiros, amigos e outros. Os Sistemas de Combinação Social podem ser usados para apresentar pessoas desconhecidas aos usuários através da combinação de interesses e/ou personalidades em comum.
 - d. Visualização Social [Smith and Fiore 2001]: são sistemas que capacitam os usuários a visualizarem as atividades ou comportamento dos outros usuários em uma comunidade online. Esse tipo de sistema apresenta um grafo a fim de auxiliar os usuários a selecionarem a melhor pessoa para interação ou a melhor comunidade para participar. Os Sistemas de Combinação Social podem ser usados com o intuito de automatizar a seleção de pessoas a interagir.

- e. Navegação Social [Wexelblat and Maes 1999]: são sistemas que auxiliam usuários não ficarem perdidos em espaços com grande quantidade de informação. Usualmente esses sistemas direcionam os usuários considerando os caminhos mais popularmente usados por outros usuários. Os Sistemas de Combinação Social poderia colocar em contato os usuários que acessaram caminhos similares.
- f. Modelagem de usuário [Kobsa 2007], [Rich 1979], [Heckmann et al. 2005]: é a maneira para armazenar informações sobre usuários para capacitar sistemas a fornecer eventos personalizados aos mesmos. Os Sistemas de Combinação Social são sistemas construídos para combinarem informações sobre pessoas, então a complexidade das informações das mesmas é de fundamental importância e devem ser otimamente modeladas para garantir o sucesso dos resultados.

6.2. Sites Comerciais

6.2.1. Amazon.comTM

O *site* de comércio eletrônico da *Amazon.comTM*, disponível em www.amazon.com, possui algumas variações de estratégias de recomendação. Em [Riedl et al. 2000] estas variações são apresentadas, destacando as características de cada uma delas:

1. clientes que compraram: são dois tipos de listas de recomendação separadas. A primeira, recomenda livros freqüentemente comprados pelos clientes que adquiriram o livro selecionado; a segunda, recomenda autores cujos livros são freqüentemente comprados por clientes que adquiriram trabalhos de acordo com o autor do livro selecionado;
2. suas recomendações: os clientes qualificam os livros lidos em uma escala de 1 a 5, variando de “eu odeio isto” a “eu adoro isto”. Após a avaliação de um conjunto de livros, os clientes podem requisitar recomendações de livros que lhe agradem;
3. olhos: permite aos clientes receberem e-mails sobre novos itens que foram adicionados no catálogo. Os clientes selecionam as informações, como autor, título, assunto, entre outros, ou utilizam o formato avançado com operadores booleanos (*and / or*) para filtragem de notificações;
4. amazon.com entrega: os clientes selecionam, em caixas de seleção (*checkbox*), os itens de uma lista de categorias e gêneros específica. Periodicamente, os editores da *Amazon.comTM* enviam as suas últimas recomendações, por e-mail, para os inscritos em cada categoria;
5. idéias de presentes da livraria: permite aos clientes receber recomendações dos editores. Os clientes escolhem uma categoria de livros para as quais gostariam de receber algumas sugestões;
6. comentários dos clientes: permite aos clientes receberem recomendações, em formato textual, baseadas em opiniões de outros clientes que, por sua vez, também podem ser qualificadas em uma escala de 1 a 5.

6.2.2. eBayTM

Assim como no *site da Amazon.comTM*, o *site* de leilões on-line *eBay.comTM*, acessível no endereço www.ebay.com, possui mais de uma variação de estratégias de recomendação, também apresentado por [Riedl et al. 2000]. São elas:

1. 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. Quanto mais positiva a pontuação, mais confiável é o negociante, e quanto mais negativa a pontuação, menos confiável é o negociante;
2. 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.

7. Tendências e desafios

A seguir são descritas algumas tendências e desafios na visão dos autores.

7.1. Personalização de Ambientes Virtuais

A maior parte dos exemplos apresentados neste capítulo se referiram à personalização de *websites*. No entanto, outras aplicações dos Sistemas de Recomendação são possíveis e encontram cada vez mais mercado. Uma destas aplicações é a de *call-centers*. A maior parte das empresas que lida diretamente com os clientes possuem 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.



Figura 7.1. Recomendações encaminhadas por um assistente virtual

A necessidade de personalizar o relacionamento com os clientes é cada vez mais crescente nos dias de hoje. Ela é na verdade uma forma de volta ao passado, onde o

vendedor de um pequeno mercado conhecia cada um de seus clientes e conseguia, de forma bastante eficiente, lembrar-se das preferências de cada um para fazer recomendações que aumentavam a lucratividade do negócio e promoviam a fidelização.

Hoje em dia, para o vendedor de uma grande corporação, não é possível lembrar dos principais interesses dos clientes bem como dos itens por eles adquiridos nas últimas compras. Por isso, os bancos de dados, as ferramentas de mineração e recomendação servem como memória da empresa, provendo informações valiosas relacionadas aos hábitos de consumo e preferências dos consumidores. A utilização destas informações na comercialização de produtos não implica na substituição do homem por computadores com maior capacidade de armazenamento de informações, mas sim na adaptação de um tipo de serviço que era prestado por comerciantes e que, hoje em dia, não é mais possível dado o crescente tamanho da população e, conseqüentemente, o número de consumidores.

7.2. Web Semântica

Lojas virtuais são amplamente difundidas na Internet, muitas são populares como, por exemplo, o sucesso das lojas virtuais do “Magazine Luiza” no Brasil. A loja virtual é uma das principais responsáveis pelo relacionamento entre a empresa e o consumidor.

A tarefa do novo consumidor agora é encontrar o produto que procura, na quantidade e qualidade desejada, em determinado tempo, e pelo menor preço. Com pouca experiência para escolhas, o indivíduo se depara diante de muitas alternativas. Os Sistemas de Recomendação auxiliam no aumento da capacidade e eficácia do processo de iniciação conhecido na relação social entre seres humanos, onde recomendações são realizadas entre as pessoas. As recomendações podem levar à personalização no atendimento ao consumidor.

É um desafio para a máquina, saber o contexto em que uma palavra está sendo utilizada. Hoje, a maioria dos Sistemas de Recomendação utiliza algoritmos para estruturar informações diretamente em uma base dados. Porém, a representação dos dados geralmente não permite a identificação do contexto em que uma palavra está sendo usada. Contudo, ontologias podem ser utilizadas para melhorar a filtragem baseada em conteúdo (Guarino, 1999 apud [Middleton et al. 2004]). Ao trabalhar com ontologias, pode-se fazer uso da *Web Semântica*, uma vez que a *Web Semântica* desenvolve linguagens para expressar informações de forma processável pela máquina. Deste modo é possível descrever produtos da loja virtual de uma forma padronizada e de fácil compreensão para a máquina.

7.3. Sistemas de Recomendação ou Combinação Social baseado em Perfil Psicológico

Sistemas de Recomendação são considerados como uma rica área de pesquisa. Eles têm várias aplicações práticas também definidas como sistemas que promovem recomendação de pessoas (em Redes Sociais) tão bem como promove a recomendação de produtos e serviços (em *e-training* e *e-commerce*). Em 2005, Terveen and McDonald [Terveen and McDonald 2005] re-definiu esses Sistemas de Recomendação específicos, chamado-os de Sistemas de Combinação Social.

Como comentado anteriormente, Sistemas de Recomendação e de Combinação Social não usam aspectos psicológicos em suas recomendações. Entretanto, mesmo que

os aspectos psicológicos humanos sejam difíceis de extrair intencionalmente do usuário, sua relevância é altamente significativa em processos de tomada de decisão para serem ignorados por Sistemas de Recomendação. Nesse contexto, em 2005, Gonzalez [Gonzalez et al. 2007] propõe um primeiro modelo baseado em aspectos psicológicos, ele usa Inteligência Emocional para melhorar a recomendação de cursos em um *site* de *e-training*. Em 2008, Nunes [Nunes 2009] [Nunes ET al. 2008] publica o primeiro trabalho que usa Traços de Personalidade para recomendar produtos, serviços ou pessoas em Sistemas de Recomendação. Este trabalho revelou indícios de que o uso de Traços de Personalidade efetivamente proporciona uma melhora na recomendação. Um dos exemplo descritos no trabalho de Nunes [Nunes 2009] identifica a Reputação baseada em Traços de Personalidade como fator determinante na eficiência da recomendação.

A seguir apresentam-se vários possíveis cenários onde o conhecimento aqui apresentado, se aplicado, pode ter impacto imediato.

7.3.1 Cenários de Aplicação

7.3.1.1. Recomendação de Produtos Baseadas em suas Características Subjetivas

Este cenário é original e, futuramente, será usado largamente em Sistemas de Recomendação comerciais (*e-commerce*). Atualmente os Sistemas de Recomendação convencionais usados em *e-commerce* oferecem produtos e serviços para os seus clientes, geralmente, baseados em informações convencionais do usuário, tais como, informação demográfica, preferências, entre outras. Como exemplos: livros, na *Amazon.com*, Submarino.com.br, musicas no *MyStrands* e filmes no *MovieLens*.

Na intenção de fornecer uma melhor otimização e personalização nas recomendações prestadas pelos Sistemas de Recomendação os *sites* comerciais deveriam mudar drasticamente a maneira que tem representado as informações de usuários e de produtos, que deveriam ser representados como segue:

1. Dados do usuário: devem ser enriquecidos com aspectos psicológicos, tais como, Traços de Personalidade, Inteligência Emocional e Habilidades Sociais [Nunes 2009];
2. Informações de produtos: uma descrição mais subjetiva deve ser adicionada a descrição tradicional de dados do produto. Características subjetivas em dados convencionais podem ser descritas como meta-dados dos dados, baseados na perspectiva dos aspectos psicológicos já medidos em humanos.

Na tabela 7.1, apresenta-se uma hipótese de como os dados convencionais podem ser representados subjetivamente usando características psicológicas.

Tabela 7.1. Representação das informações convencionais e subjetivas de produtos e usuários [Nunes 2009].

Nowadays		Future	
<i>Books</i>	-number of pages -language -category -textual description	<i>Books</i>	-the author writing style -desired Emotions after reading -desired psychological aspects as pre-condition to read a book the designing of the book Personality of the book and the characters
<i>Reader/ User</i>	-subject interests -favorite artists, writers - demographic information	<i>Reader/ User</i>	-Personality Traits of each user emotional Intelligence of user Soft Skills of user

7.3.1.2. Recomendação de Membros como parte de Grupos Sociais

O cenário da busca do membro certo para fazer parte de um grupo social é um problema antigo e freqüente na vida em comunidade. A formação de grupos começa já no início da vida social humana, quando somos convidados a participar e compartilhar atividade social com outras pessoas que tem comportamentos similares ou complementares aos nossos. A diversidade humana é de extrema valia em grupos sociais. Perfis humanos similares/complementares são fundamentais para atingir sucesso em atividades sociais compartilhadas em ambientes como comunidades virtuais, comunidades reais, escolas, universidades e empresas.

7.3.1.3. Recomendação de Novos Amigos em Redes Sociais

A teoria de atração interpessoal pode predizer a maneira como personalidades humanas interagem entre si. Nass [Nass et al 1995] tem demonstrado um extenso estudo na literatura psicológica que indica forte relacionamento entre similaridade e atração, isto significa que pessoas preferem interagir com outros (estranhos ou não) que tenham personalidade similar a si, a interagir com outros que tenham personalidade diferente. Normalmente, as pessoas preferem outras que tenham a personalidade similar a sua [Reeves and Nass 1996].

Sistemas de Recomendação aplicado a Redes Sociais podem ser considerados com uma alternativa em potencial para encontrar amigos com Traços de Personalidade, bem como, outros interesses, similares. Para alcançar recomendações promissoras de potenciais futuros amigos, os usuários de Redes Sociais (como o *Orkut*, por exemplo) deveriam melhorar seus perfis de usuário incluindo suas características psicológicas, tais como Traços de Personalidade. A efetividade dessa suposta recomendação poderá ser medida pelo grau de satisfação do usuário após a criação de novas e potenciais amizades.

7.3.1.4. Recomendação em Ambiente Educacional

Nos cenários universitários, estudantes tendem a construir grupos para interação e realização de trabalhos extraclasse. Esses grupos são criados, principalmente, baseados em empatia, similaridade e atração [Lisetti 2002]. Os estudantes nem mesmo consideram compatibilidade como um atributo sério. Estudantes estão muito mais interessados em parceria e amizade do que na eficiência de um grupo de trabalho. Dessa forma, estudantes universitários tendem muito mais a procurar parceiros com personalidades similares (como fariam em comunidades virtuais e Redes Sociais, como apresentado

anteriormente). Um sistema nessa linha está sendo proposto por Nunes [Nunes et al. 2009] na Universidade Federal de Sergipe.

7.3.1.5. Recomendação em Ambiente Empresarial

Em escolas e universidades a constituição de um grupo/equipe eficiente de trabalho, se desejado, é realizado mais facilmente que em empresas. Isto acontece porque em uma escola/universidade cada aluno na sala de aula possui, a priori, o mesmo *background* e o mesmo nível de conhecimento que seus colegas. Por isso que, nesse caso, o principal diferencial usado para a criação dos grupos são as características psicológicas dos alunos. Nesse contexto, traços psicológicos são mais importantes que informações demográficas e competências. Em empresas, competência é extremamente relevante bem como traços psicológicos. Para construir Sistemas de Recomendação capazes de gerar grupo/times de trabalho eficientes em empresas deve-se combinar os membros considerando competências e Traços de Personalidade complementares.

A similaridade e/ou complementaridade na composição de grupos/equipes pode prever o desempenho no trabalho. Psicólogos têm estudado como Traços de Personalidade influenciam no desempenho da equipe. Uma equipe que é construída considerando valores de complementaridade aumenta o nível de colaboração, pois, normalmente, as pessoas têm diferentes *backgrounds*, pontos de vista, personalidade complementar e conseqüentemente contribuições diferentes em um mesmo contexto. Geralmente, complementaridade em equipes de trabalho é saudável e são usadas para atingir metas organizacionais específicas. Prada [Prada et al. 2009] está realizando um estudo onde Agentes Virtuais personificam características de Agentes Reais para tomada de decisão de uma equipe, questões como personalidade, motivação entre outros aspectos psicológicos estão envolvidos nos aspectos de dinâmica de grupo incorporados no modelo proposto. Esse modelo servirá como padrão inicial às pesquisas onde a dinâmica de grupo será fundamental para realização de trabalho otimizado em equipes empresariais com metas pré-estabelecidas.

7.3.1.6. Recomendação de casais em Sites de Relacionamento

Sistemas de Recomendação baseadas em Traços de Personalidade aplicada em *sites* de Relacionamentos podem ser uma alternativa para pessoas que procuram por parceiros românticos em cujo relacionamento exista compatibilidade. Existem pesquisas que comprovam que *sites* de relacionamentos que consideram aspectos psicológicos na busca de pares compatíveis tem gerado um maior número de pessoas (casais) satisfeitas que os *sites* tradicionais de relacionamento [Fiori and Donath 2004], [Fiori 2004].

Note que exemplos ilustrativos de Sistemas de Recomendação que utilizam Personalidade para no auxílio à tomada de decisão computacional são apresentados na forma de Estudo de Caso em [Nunes et al. 2008], [Nunes 2009] e [Prada et al. 2009]. Esses trabalhos também descrevem como ocorreu a modelagem da Personalidade no Perfil do Usuário.

8. Conclusões e Comentários Finais

Vimos neste capítulo como os Sistemas de Recomendação podem ser utilizados para que se possa conhecer melhor os hábitos de consumo e interesses dos usuários, e como este tipo de conhecimento pode ser empregado para personalizar o relacionamento.

Ao longo deste capítulo foram discutidos mecanismos para coleta de informação do usuário visando à formação de seu perfil, bem como técnicas específicas para recomendação e as mais variadas estratégias de recomendação. Vários exemplos de Sistemas de Recomendação foram abordados, desde os sistemas de caráter acadêmico, caráter comercial, até sistemas de combinação social.

Do ponto de vista mercadológico, a identificação do comportamento social assim como a interação entre as pessoas sempre foi algo valioso a ser explorado pela área de marketing das empresas. Saber identificar o público alvo e a disseminação viral de uma informação pode significar altas margens de lucro. Com a *web*, essas informações começaram a ser mais facilmente mapeadas digitalmente. Não demorou para aparecer ferramentas como *IRC*, *ICQ*, *MSN* que centralizam a comunicação das pessoas em um único mapa. Logo depois *ParPerfeito* começa a relacionar as pessoas segundo afinidades conjugais. O *MySpace* encontrou a música como um grande fio condutor de relacionamentos afetivos e popularizou o conceito. O *Orkut* e o *Facebook* deram continuidade e aproximaram pessoas com afinidades perdidas no tempo. O último chegou a ser avaliado em 15 bilhões de reais. O fenômeno chamou tanta atenção para as redes sociais que iniciativas simples como o sistema *Geni* que faz a disseminação viral traçando um mapa genealógico global foi logo em seguida avaliado em 100 milhões de reais.

Encontrar afinidades psicológicas continua entre os itens de maior valor. Os novos nichos são *LinkedIn* -afinidade profissional, *CineDica*-afinidades por filmes e *GoodReads*-afinidades por livros. Como as iniciativas começaram a se proliferar, as novas tendências são novamente de centralização. O forte crescimento do *Twitter* que abrange qualquer informação e o lançamento do *Google Wave/Buzz* como unificador de *Gmail*, *MSN* e *Twitter*. A grande vantagem do *Twitter* é que ele é partidário de serviços abertos, sendo assim, todos terão acesso às informações para construir sua própria aplicação de afinidades psicológicas sem se preocupar na divulgação do canal [Nunes and Aranha 2009].

Dessa forma, conclui-se que existe uma tendência e necessidade de que os programas *web* se utilizem na essência de seus processos internos de tomada de decisão de tecnologias que permitam incorporar a extração, modelagem e uso de características psicológicas do usuário, produtos ou serviços em questão nos processos recomendatórios. Sem dúvida, essas novas representações criarão uma nova perspectiva de como essas informações serão tratadas, representadas e manipuladas na *web* em um futuro muito próximo. Nesse contexto serão criadas novas técnicas de mineração que possam contextualizar conhecimento afetivo, seja através de *web-mining* na análise de *clickstreams* associados a símbolos de conotação afetiva, ou seja, através de mineração de textos, analisando o sentimento do autor expresso em um texto.

Essas tecnologias inteligentes, juntamente com novas *markup languages* criadas para representar características psicológicas, seguindo-se o exemplo da *W3C* [W3C 2009] que padroniza Emoções, podem ajudar a enriquecer os textos na *web* com informações relevantes para a tomada de decisão. Finalmente, tudo isso que foi abordado, focando na personalização, aparecerá como algo que poderá potencializar e, muito, os lucros empresariais.

Referências Bibliográficas

- Ackerman, M. S. and Malone, T. W. (1990). Answer garden: a tool for growing organizational memory. In Proceedings of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems, pages 31–39, New York, NY, USA. ACM.
- Ackerman, Mark S. (1994). Augmenting the organizational memory: a field study of answer garden. In CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work, pages 243–252, New York, NY, USA. ACM.
- Ackerman, Mark S. and McDonald, David W. (1996). Answer garden 2: merging organizational memory with collaborative help. In CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work, pages 97–105, New York, NY, USA. ACM.
- 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. 6p. 734-749.
- Agrawal, R. (1993) “Mining Association Rules between sets of items in large databases”, In: International Conference Management of Data (SIGMOD-93). Proceedings, p.207-216.
- Allport, G. W.(1927). Concepts of trait and personality. Psychological Bulletin, (24):284–293. (Available at <http://psychclassics.yorku.ca/Allport/concepts.htm>).
- Amento, B.; Terveen, L.; Hill, W.; Hix, D.; Schulman, R. (2003). Experiments in social data mining: The TopicShop system. ACM Transactions Computer-Human Interaction, 10(1):54–85.
- Ansari, A. et al. (2000). Internet Recommendation Systems. Journal of Marketing Research, v.37, n.3, p. 363-375, Aug.
- Apte, C. et al. (2002). Business Application of Data Mining. Communications of the ACM, New York, v.45, n.8, p. 49-53, Aug.
- Balabanovic, M.; Shoham, Y. (1997). Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3):66-72.
- 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.
- Boyd, D. (2002). Faceted id/entity: Managing representation in a digital world. Master's thesis, Cambridge, MA.
- Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual websearch engine. Computer Network ISDN Systems, 30(1-7):107-117.
- Budzik, Jay; Bradshaw, Shannon; Fu, Xiaobin and Hammond, Kristian J. (2002). Clustering for opportunistic communication. In WWW '02: Proceedings of the 11th international conference on World Wide Web, pages 726–735, New York, NY, USA. ACM.
- Burger, J.M. (2000). Personality. Wadsworth, fifth edition.

- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331-370.
- Cabena, P. et al. (1997). *Discovering data mining: from concept to implementation*, Prentice-Hall.
- Carreira, R., Crato, J.M., Gonçalves, D. and Jorge, J. A. (2004). Evaluating adaptive user profiles for news classification. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*, pages 206–212, New York, NY, USA. ACM Press.
- Cazella, S. C.; Correa, I. ; Barbosa, J. ; Reategui, E. (2009) “Um modelo para recomendação de artigos acadêmicos baseado em filtragem colaborativa aplicado à ambientes móveis.” *Revista Novas Tecnologias na Educação*, v. 7, p. 12-22, 2008.
- Cazella, S. C.; Reategui, E.; Machado, M.; Barbosa, J. (2009) *Recomendação de Objetos de Aprendizagem Empregando Filtragem Colaborativa e Competências*. In: *Simpósio Brasileiro de Informática na Educação (SBIE)*.
- Cline, J. (2003) *Web Site Privacy Seals: Are they worth it?* *Computer World*. May, 8.
- Cowie, J.; Lehnert, W. (1996). “Information Extraction”. *Communications of the ACM*, vol. 39, New York, USA, p. 80-91.
- Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Quill, New York.
- Demir, G. N., Uyar, A. S., Oguducu, S. G. (2007) Graph-based sequence clustering through multiobjective evolutionary algorithms for web recommender systems. In *ACM Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*. London, England.
- Donath, J. S. (2000). Being Real: Questions of Tele-Identity. In: Ken Goldberg, editor, *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*, chapter 16, pages 296–311. The MIT Press, first edition.
- Donath, J.S. (1999) *Identity and Deception in the Virtual Community*. In M. A. Smith and P. Kollock, editors, *Communities in Cyberspace*, chapter 2, pages 29–59. Routledge, London, first edition.
- Dong, G., Pei, J. (2007) *Sequence Data mining*. Berlin: Springer.
- Erikson, Erik H.(1980). *Identity and the Life Cycle*. Norton.
- Fiori, A. R. T. (2004). *Romantic regressions: An analysis of behavior in online dating systems*. Master's thesis, MIT Media Lab.
- Fiori, A. R. T.; Donath, J. S (2004). Online personals: an overview. In *CHI'04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1395–1398, New York, NY, USA. ACM Press.
- 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.
- Giddens, A. (1991) *Modernity and Self-Identity. Self and Society in the Late Modern Age*. Stanford university Press, Stanford, California.

- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor Book.
- Goldberg, D., Nichols, D., Oki, B. M., 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.
- Goleman, D. (1995). *Emotional Intelligence - Why it can matter more than IQ?* Bloomsbury, London, first edition.
- Gonzalez, G., de la Rosa, J.L., and Montaner, M. (2007). Embedding Emotional Context in Recommender Systems. In *The 20th International Florida Artificial Intelligence Research Society Conference-FLAIRS*, Key West, Florida.
- Guttman, R. H., Moukas, A. G. and Maes, Pattie. (1998) Agent-mediated electronic commerce: a survey. *Knowl. Eng. Rev.*, 13(2):147-159.
- Gyara, F., Sachdev, T. (2008) Win in the flat world. White paper - Infosys Technologies. Disponível em: <http://www.infosys.com/offerings/it-services/information-management/white-papers/documents/personalizing-portals.pdf>
- Heckmann, D. (2005). *Ubiquitous User Modeling*. Phd thesis, Technischen Fakultäten der Universität des Saarlandes, Saarbrücken-Germany.
- Heckmann, D. and Kruger, A. (2003). A user modeling markup language (UserML) for ubiquitous computing. In *8th International Conference on User Modeling*, LNAI 2702, page 393-397, Johnstown, PA, USA. Springer, Berlin Heidelberg.
- Heckmann, D., Brandherm, B., Schmitz, M., Schwartz, T. and von Wilamowitz-Moellendorf, B.M. (2005). GUMO – the general user model ontology. In *10th International Conference on User Modeling*, LNAI 3538, page 428-432, Edinburgh, Scotland. Springer Verlag.
- 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., Konstan, J. A., Terveen, L. G. and Riedl, J. T. (2004). Evaluating collaborative Filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5-53.
- Holte, R. C. ; Yan, J. Y. (1996). “Inferring What a User is Not Interested”, In: *Advances in Artificial Intelligence, Proceedings*, p. 159-171.
- Hudson, Scott E. and Smith, Ian. (1996). Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 248–257, New York, NY, USA. ACM.
- Jensen, C.; Davis, J. and Farnham, S. (2002). Finding others online: reputation systems for social online spaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 447-454, New York, NY, USA. ACM.
- John, O.P. and Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. The Guilford Press, New York, second edition.

- Josang, A., Ismail, R. and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644.
- 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.
- Kobsa, A. (2007). Generic user modeling systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web*, volume 4321 of *Lecture Notes in Computer Science*, chapter 4, pages 136–154. Springer Verlag.
- Konstan, J. A.; Miller, B. N., Maltz, D.; Herlocker, J. L.; Gordon, L. R. and Riedl, J. (1997). GroupLens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77-87.
- Lisetti, Christine L. (2002). Personality, affect and emotion taxonomy for socially intelligent agents. In *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, pages 397–401. AAAI Press.
- Loeb, S.; Terry, D. (1992). Information Filtering. *Comunications of ACM*, New York, v.35, n.12, p.26, Dec.
- Lopes, P. N.; Salovey, P. and Straus, R.(2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences-Elsevier*, 35:641–658.
- Lueg, C. (1998) “Considering Collaborative Filtering as Groupware: Experiences and Lessons Learned”, In: 2nd. International Conference on Practical Aspects of Knowledge Management, *Proceedings*, Basel, Switzerland, p. 29-30.
- Maes, P.; Shardanand, U. (1995). “Social information filtering: Algorithms for automating "word of mouth”, In: *Human Factors in Computing Systems. Proceedings....*, 1995, p. 210-217.
- Manber, U.; Patel, A.; Robison, J. (2000). Experience with Personalization on Yahoo! *Communication of the ACM*, New York.
- Masthoff, J. (2005) *The pursuit of satisfaction: Affective state in group recommender systems*. In: *User Modeling 05*. Springer Verlag.
- Mccarthy, J. (2000). Phenomenal Data Mining. *Communications of the ACM*, v.43, n.8, pp. 75-79.
- McDonald, David W. (2001). Evaluating expertise recommendations. In *GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, pages 214–223, New York, NY, USA. ACM.
- McDonald, David W. (2003). Recommending collaboration with social networks: a comparative evaluation. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 593–600, New York, NY, USA. ACM.
- McDonald, David W. and Ackerman, Mark S.(2000). Expertise recommender: a flexible recommendation system and architecture. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 231–240, New York, NY, USA.ACM.
- Mead, G. H. (1934). *Mind, Self, and Society*, volume 1. Univeristy of Chicago,Chicago, charles w. morris edition.

- Middleton, S.; Shadbolt, N.; De Roure, D. (2004) Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems*. [S.l.]: [s.n.]. v.22, n.1, pp.54-88.
- Montaner, M., López, B., de La Rosa, J.L. (2003). A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*. Netherlands : Kluwer Academic Publishers, pp. 285-330, Aug.
- Nass, C et al. (1995). Can computer personalities be human personalities? *International Journal Human-Computer Studies*, 43(2):223–239.
- Nulvenna, M. et al. (2000). Personalization on the Net Using Web Mining. *Communications of the ACM*, v.43, n.8, pp. 122-125, Aug.
- Nunes, M. A. S. N. (2009). Recommender Systems based on Personality Traits: Could human psychological aspects influence the computer decision-making process?. 1. ed. Berlin: VDM Verlag Dr. Müller. v.1. 140 p.
- Nunes, M. A. S. N. ; Aranha, C. N. (2009) .Tendências à Tomada de Decisão computacional . In: W3C, 2009, São Paulo. W3C.
- Nunes, M. A. S. N. ; Cerri, Stefano A. ; Blanc, N. (2008a). Towards User Psychological Profile. In: Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais, 2008, Porto Alegre. VIII Simpósio sobre Fatores Humanos em Sistemas Computacionais IHC 2008. Porto Alegre : Sociedade Brasileira da Computação. v. 1. p. 196-203.
- Nunes, M. A. S. N. ; Cerri, Stefano A. ; Blanc, N. (2008) . Improving recommendations by using Personality Traits in. In: International Conference on Knowledge Management-In: International Conference on Knowledge Management-I.KNOW08, 2008, Graz-Austria. International Conference on Knowledge Management-I.KNOW08. v. 1. p. 92-100.
- Nunes, M. A. S. N. ; Machado, G. J. C. ; Schneider, H. N. (2009). Repensando os ambientes virtuais de aprendizagem: o caso da . In: XX Simpósio Brasileiro de Informática na Educação - II Workshop sobre Modelos Pedagógicos em Educação a Distância, 2009, Florianópolis. XX SBIE-SBC.
- O'Connor, Mark; Cosley, Dan; Konstan, Joseph A. and Riedl, John (2001) Polylens: a recommender system for groups of users. In *ECSCW'01: Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work*, pages 199–218, Norwell, MA, USA. Kluwer Academic Publishers.
- Paiva, A. (2000). *Affective interactions: towards a new generation of computer interfaces*. Springer-Verlag New York, Inc., New York, NY, USA.
- Paiva, A. and Self, J.A. (1995). Tagus - a user and learner modelling workbench. *User Model. User-Adapt. Interact.*, 4(3):197–226.
- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13(5-6):393-408.
- Perugini, Saverio; Gonçalves, Marcos André and Fox, Edward A. (2004). Recommender systems research: A connection-centric survey. *Journal of Intelligent Information Systems*, 23(2):107–143.

- Picard, R.W. (1997). *Affective computing*. MIT Press, Cambridge, MA, USA.
- Plastino, A. (2001) “Regras de Associação e Algoritmos de Mineração de Dados”, Mini-Curso 2, SBBD, 2001. Instituto Militar de Engenharia. Rio de Janeiro: Rio de Janeiro.
- Poo, D., Chng, B. and Goh, J.M. (2003). A hybrid approach for user profiling. In HICSS '03: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) Washington, DC, USA. IEEE Computer Society.
- Prada, R., MA, S., Nunes, M. A. S. N. (2009) .Personality in Social Group Dynamics In: International Conference on Computational Science and Engineering- CSE '09, 2009, Vancouver. International Conference on Computational Science and Engineering- CSE '09. v.4. p.607 – 612.
- Preece, Jenny and Maloney-Krichmar, Diane. (2003). Online communities: focusing on sociability and usability. pages 596–620.
- Preece, Jenny. (1998). Empathic communities: reaching out across the web. *Interactions*, 5(2):32–43.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*, San Mateo: Morgan Kaufmann.
- Reategui, E.; Cazella, S. (2005) *Sistemas de Recomendação*. Minicurso do IV Encontro Nacional de Inteligência Artificial (ENIA). São Leopoldo, p. 306-348.
- Reeves, B. and Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, New York, NY, USA.
- Rein, G. L. (2005). Reputation information systems: A reference model. In HICSS'05: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 1, page 26, Washington,DC, USA. IEEE Computer Society.
- Resnick, P. e Varian, H. R. (1997). Recommender Systems. *Communications of the ACM*, New York, v.40, n.3, pp. 55-58, Mar.
- Resnick, P.; Kuwabara, K.; Zeckhauser, R. and Friedman, E. (2000). Reputation systems. *Commun. ACM*, 43(12):45-48.
- Resnick, P.; Zeckhauser, R.; Swanson, J. and Lockwood, K.. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*,9(2):79-101.
- Rich, Elaine. (1979). User modeling via stereotypes. *Cognitive Science*, 3:329–354.
- Riedl, J. et al. (1997). GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, New York, v.40, n.3, pp. 77-87, Mar.
- Riedl, J. et al. (1999). “Combining Collaborative Filtering with Personal Agents for Better Recommendations”, In: *Proceedings of AAI, Proceedings*, vol. 35, pp.439-446. Press.
- Riedl, J. et al. (2000). Electronic Commerce Recommender Applications. *Journal of*

Data Mining and Knowledge Discovery, vol. 5 nos. 1/2, pp. 115-152, Aug.

- Rousseau, B., Browne, P., Malone, P., Foster, P. and Mendis, V. (2004). Personalised resource discovery searching over multiple repository types: Using user and information provider profiling. In ICEIS (5), pages 35–43.
- Rucker, J.; Polanco, M. J. (1997). SiteSeer: Personalized Navigation for the Web. Communications of the ACM, New York, v.40, n.3, pp. 73-75, Mar.
- Saari, T., Ravaja, N., Laarni, J., Turpeinen, M., Kallinen, K. (2004). Psychologically targeted persuasive advertising and product information in e-commerce. In ICEC '04: Proceedings of the 6th international conference on Electronic commerce, pages 245–254, New York, NY, USA. ACM Press.
- Salton, G., McGILL, J. M. (1983) Introduction to Modern Information Retrieval. McGraw Hill, New York, USA.
- Schafer, J. B.; Konstan, J. and Riedl, J. (1999). Recommender systems in e-commerce. In EC '99: Proceedings of the 1st ACM conference on Electronic commerce, pages 158-166, New York, NY, USA. ACM.
- Schafer, J. B.; Konstan, J. and Riedl, J. (2001). E-commerce recommendation applications. Data Mining Knowledge Discovering, 5(1-2):115-153.
- Simon, H.A. (1983). Reason in Human Affairs. Stanford University Press, California.
- Smith, Marc A. and Fiore, Andrew T. (2001). Visualization components for persistent conversations. In CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 136–143, New York, NY, USA. ACM.
- Smyth, P. (2002). Business Applications of Data Mining. Communications of the ACM, v.45, n.45, p. 49-53, Aug.
- Spiliopoulou, M. (2000). Web Usage Mining for *website* Evaluation. Communications of the ACM, New York, v.43, n.8, pp. 127-134, Aug.
- Terry, Michael; Mynatt, Elizabeth D.; Ryall, Kathy and Leigh, Darren. (2002). Social net: using patterns of physical proximity over time to infer shared interests. In CHI'02: CHI '02 extended abstracts on Human factors in computing systems, pages 816–817, New York, NY, USA. ACM.
- Terveen, L.; Hill, W. (2001). Human-Computer Collaboration in Recommender Systems. In J. Carroll (Ed.), HCI on the new Millennium: Addison Wesley.
- Terveen, L.; Hill, W.; Amento, B.; McDonald, D. and Creter, J. (1997). Phoaks: a system for sharing recommendations. Communications of the ACM, 40(3):59–62.
- Terveen, L. and McDonald, D.W. (2005) Social matching: A framework and research agenda. ACM Trans. Comput.-Hum. Interact., 12(3):401–434.
- Thagard, Paul. (2006). Hot Thought: Mechanisms and Applications of Emotional Cognition. A Bradford Book- MIT Press, Cambridge, MA, USA.
- Torres, R. Personalização na Internet. Editora Novatec, 2004.
- Trappl, Robert; Payr, Sabine and Petta, Paolo (2003). Emotions in Humans and Artifacts. MIT Press, Cambridge, MA, USA.

- W3C (2009). Emotion incubator Group. (Available at <http://WWW.w3.org/2005/Incubator/emotion>).
- Watts, D. J. (2003). *Six Degrees - The science of a connected age*. Norton.
- Wexelblat, Alan and Maes, Pattie. (1999). Footprints: history-rich tools for information foraging. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 270–277, New York, NY, USA. ACM.
- Xu, Y., Zhang, B., Chen, Z., Wang, K. (2007) Privacy Enhancing Personalized Web Search. *ACM WWW 2007*, May 8–12, Banff, Alberta, Canada.
- Yang, H. et al. (2002) “On the use of constrained association for Web log mining”, In: *Web Mining for Usage Patterns & User Profiles, WebKDD, 4.*, 2002, Edmonton. Proceedings Alberta, Canadá: ACM SIGKDD, pp. 100-114.
- 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.
- Zacki, M. J. (2001) Spade: An efficient algorithm for mining frequent sequences. In: *Machine Learning*. Vol. 42. pp. 31–60.