

e-Recommender: Sistema Inteligente de Recomendação para Comércio Eletrônico

Trabalho de Conclusão de Curso
Engenharia da Computação

Valmir Macário Filho
Orientador: Prof. Fernando Buarque de Lima Neto

Recife, novembro de 2006



e-Recommender: Sistema Inteligente de Recomendação para Comércio Eletrônico

Trabalho de Conclusão de Curso

Engenharia da Computação

Este Projeto é apresentado como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Valmir Macário Filho
Orientador: Prof. Fernando Buarque de Lima Neto

Recife, novembro de 2006



Valmir Macário Filho

**e-Recommender: Sistema Inteligente
de Recomendação para Comércio
Eletrônico**

Resumo

No comércio eletrônico, ofertas de infindáveis produtos estão a um simples clique do mouse. A partir da personalização, o usuário pode encontrar facilmente o item que necessita ou deseja, sendo um diferencial competitivo que possibilita a fidelização de clientes. Sistemas de recomendação implementam a personalização.

Este trabalho apresenta um estudo da área de sistemas de recomendação para comércio eletrônico e teve como propósito desenvolver uma ferramenta que utilizou uma estratégia de recomendação inteligente. Para isso foram utilizados dois algoritmos de inteligência artificial: KNN e simbólico.

A abordagem híbrida utilizada procura combinar as vantagens e também atenuar as desvantagens da abordagem colaborativa e baseada em conteúdo. O algoritmo KNN implementa o componente colaborativo, e o algoritmo de análise simbólica implementa o componente baseado em conteúdo da ferramenta.

Como resultado deste estudo, um sítio de comércio eletrônico foi preparado para atuar com recomendação e a ferramenta foi conectada com o mesmo. Deste modo foi possível que usuários realizassem compras no sítio e recebessem recomendações, as quais se mostraram interessantes para os usuários que avaliaram a qualidade da recomendação.

Abstract

In e-commerce, a massive number of products are separated only by a simple mouse click. With personalization, the user can easily find items that he/she needs or desires. This is a competitive differential that helps on shorten customer-to-business relations. Recommendations systems implement the personalization.

This work presents a recommendation system to be used in e-commerce. This software construction was carried out using intelligent recommendation strategy. For this, two algorithms of artificial intelligence were utilized: Symbolic and KNN.

The hybrid solution combines the advantages and also attenuates the disadvantages of collaborative and content-based approaches. The KNN algorithm implements the collaborative component and the symbolic algorithm implements the content based component.

As a result, an e-commerce shop was set up to deal with recommendations. The conceived tool was plugged in the e-commerce site where users can experiment its functionality. The recommendations provided by the tool put forward here were judged to be relevant to users who have evaluated them well in a usability test also carried out in this work.

Sumário

Índice de Figuras	v
Índice de Tabelas	vi
1 Introdução	8
2 Sistema de Recomendação aplicado ao comércio eletrônico	10
2.1 Personalização	10
2.2 Sistemas de Recomendação	11
2.3 Taxonomia de Sistemas de Recomendação	11
2.3.1 Entrada/Saída	12
2.3.2 Métodos de Recomendação	15
2.3.3 Outros aspectos de projeto	17
3 Técnicas de Sistemas de Recomendação	19
3.1 Filtragem Colaborativa	19
3.1.1 Funcionamento da Filtragem Colaborativa	20
3.1.2 Vantagens da Filtragem Colaborativa	22
3.1.3 Desvantagens da Filtragem Colaborativa	22
3.2 Filtragem Baseada em Conteúdo	23
3.2.1 Funcionamento da Filtragem Baseada em Conteúdo	24
3.2.2 Vantagens da Filtragem Baseada em Conteúdo	25
3.2.3 Desvantagens da Filtragem Baseada em Conteúdo	25
3.3 Filtragem Híbrida	25
3.3.1 Classificação de Sistemas Híbridos	26
3.3.2 Exemplo de Sistemas Híbridos	26
3.4 Dados Simbólicos	27
3.4.1 Análise de Dados Simbólicos	27
3.4.2 Entrada do SDA	27
3.4.3 Objetos Simbólicos	28
4 e-Rec recommender: Uma Ferramenta Híbrida para Recomendação Inteligente de Produtos	30
4.1 Características	30
4.2 Arquitetura	32
4.3 Sítio de Comércio Eletrônico	33
4.4 Preparação da Base de Dados	35
4.5 Perfil do Usuário	36
4.5.1 Pré-processamento	36
4.5.2 Generalização	37
4.6 Recomendando Produtos	38

5 e-Recommender: Experimentos e Avaliações	41
5.1 Experimentos e Validação	41
5.1.1 Metodologia dos Experimentos	42
5.1.2 Realização dos Experimentos	43
6 Conclusão e Trabalhos Futuros	46
6.1 Conclusão	46
6.2 Discussão	47
6.3 Trabalhos Futuros	47

Índice de Figuras

Figura 1. Taxonomia dos sistemas de recomendação	12
Figura 2 . Predição de filmes para usuário	14
Figura 3. Diagrama de estados do e-Recommenderr	31
Figura 4. Arquitetura <i>blackboard</i>	32
Figura 5. Arquitetura da Aplicação e-Recommenderr	33
Figura 6. Mostra <i>link</i> para usuário requisitar recomendação	34
Figura 7. Recomendações personalizadas	35
Figura 8. Velocidade do algoritmo	44
Figura 9. Quantidade de livros que o usuário compraria	45

Índice de Tabelas

Tabela 1. Grau de personalização x método de recomendação	17
Tabela 2. Tabela clássica de dados	28
Tabela 3. Dados simbólicos modais	29
Tabela 4. Descrição simbólica dos tributos do domínio livro	36
Tabela 5. Livros no perfil do usuário	37
Tabela 6. Perfil modal simbólico de um usuário	37
Tabela 7. Pesos de cada atributo do produto livro	42
Tabela 8. Recomendações geradas a partir de três configurações dos parâmetros para o usuário 1	43
Tabela 9. Recomendações geradas a partir de três configurações dos parâmetros para o usuário 2	44
Tabela 10. Média e desvio padrão dos resultados da pergunta 3	45

Agradecimentos

Gostaria de agradecer ao meu pai, Valmir Macário, que tornou possível eu estudar numa cidade como Recife, ofereceu todo o suporte para que eu tivesse uma boa educação e nunca deixou faltar nada na caminhada para o término desse curso.

Agradeço a minha Mãe que apesar de longe me forneceu apoio e amor para que conseguisse passar por todas as dificuldades.

Agradeço a Patrícia Maia pelo incentivo, ombro amigo, compreensão e ajuda incondicional durante todo o período de realização deste trabalho.

Agradeço a Byron Leite, que desde o primeiro momento se prontificou a me ajudar, fornecendo valiosas dicas e materiais para tornar essa monografia possível, mesmo sem receber nada em troca.

Agradeço a Adélia Caroline, que me ajudou lendo e corrigindo cada capítulo escrito, com um senso de humor incrível nas críticas e correções, as quais foram valiosas.

Agradeço a todos os meus amigos da Poli, que fizeram parte da minha história nessa universidade. Com eles eu me diverti, trabalhei, aprendi e ensinei. Pode ter certeza que foram de grande importância para meu crescimento. Foram tempos inesquecíveis.

Agradeço a meu orientador, Fernando Buarque, que aceitou me ajudar passado o período estabelecido para isto e não mediu esforços para isso durante todo esse ano. Sou muito grato por suas aulas, seus grandes ensinamentos, que me ajudaram e continuarão a me ajudar daqui para frente.

Por fim agradeço a todos que me ajudaram num menor ou maior grau, pois foram anos muito importantes para mim e é impossível agradecer um por um a todos que fizeram parte dessa etapa da minha vida.

Muito obrigado a Deus, que guia minha vida!

Capítulo 1

Introdução

O advento e o desenvolvimento da internet proporcionaram de uma forma nunca vista na história da humanidade a evolução da comunicação eletrônica, e como consequência, mais acesso à informação. Na medida em que a tecnologia evolui e que mais pessoas têm acesso a mesma, altera-se continuamente a forma de interação, busca e disponibilização da informação.

Assim, a internet forneceu suporte ao surgimento e desenvolvimento do comércio eletrônico. Turban e King [1] definem comércio eletrônico como o processo de compra, venda e troca de produtos, serviços e informações por redes de computadores ou pela própria internet.

Através do comércio eletrônico, ofertas de infindáveis produtos estão a um simples clique do mouse. Nas televisões, outdoors, e outros meios de comunicação, é utilizado o marketing em massa [2], para atrair atenção de grande quantidade de pessoas. Entretanto, o marketing em ascensão no comércio eletrônico é um marketing direto [3], um marketing que utiliza a personalização para atrair clientes e ao mesmo tempo tornar esses clientes fiéis. De acordo com pesquisa da revista Info [4], a personalização é utilizada em 60 % dos sítios de comércio eletrônico pesquisados, ficando em primeiro lugar entre as tecnologias utilizadas por esses sítios.

Sistemas de recomendação [5][6][7][8][9][10] podem fazer a diferença em sítios de comércio eletrônico, proporcionando uma personalização para que o usuário encontre e tenha acesso facilmente ao que necessita ou deseja. Os sistemas de recomendação surgiram na tentativa de ajudar o usuário a escolher a melhor opção de compra e ajudar empresas de comércio eletrônico a sugerir produtos e serviços para seus clientes de forma personalizada. Há diversas estratégias na implementação de sistemas de recomendação, dentre elas as mais conhecidas são a filtragem baseada em conteúdo [1][11], a filtragem colaborativa [5][12][13] e a filtragem híbrida [5][6][14][15].

Dada a importância dos sistemas de recomendação, este projeto descreve o protótipo de uma ferramenta que utiliza a filtragem híbrida para recomendar produtos de um sítio de comércio eletrônico. O perfil do usuário será modelado baseando-se no histórico de compras do usuário. O aprendizado da ferramenta será implementado pelos

algoritmos KNN-FC [16] e simbólico [14], utilizados para formar grupos de usuários com preferências semelhantes e assim recomendar itens que sejam de interesse e ou necessidade dos usuários.

Este trabalho está organizado em 6 capítulos. Neste capítulo são apresentadas a motivação, objetivos e contexto no qual esse trabalho está inserido juntamente com uma breve descrição dos trabalhos relacionados.

No Capítulo 2 são revisados alguns conceitos inerentes a sistemas de recomendação aplicados ao comércio eletrônico e é apresentada uma taxonomia para estes sistemas.

No Capítulo 3 são descritos o funcionamento, vantagens e desvantagens das abordagens colaborativa, baseada em conteúdo e híbrida. Essas abordagens são as mais utilizadas em sistemas de recomendação. Ainda é descrita a técnica de abordagem simbólica, utilizada para construir a ferramenta fruto desse trabalho.

No Capítulo 4 a ferramenta fruto deste trabalho é apresentada em detalhes. Sua arquitetura e seu funcionamento são explicados detalhadamente. O sítio de comércio eletrônico utilizado para testar a ferramenta e sua estrutura também é apresentada nesse capítulo, cuja funcionalidade de recomendação foi desenvolvida especialmente para o trabalho.

O capítulo 5 traz o resultado dos experimentos que foram realizados com a finalidade de testar a eficácia e validar a ferramenta.

No Capítulo 6 são discutidas as conclusões, principais contribuições e propostas para trabalhos futuros.

Capítulo 2

Sistema de Recomendação aplicado ao comércio eletrônico

O advento e o desenvolvimento da internet proporcionaram de uma forma nunca vista na história da humanidade a evolução da comunicação eletrônica, e como consequência mais acesso à informação. Na medida em que a tecnologia evolui e que mais pessoas têm acesso a mesma, altera-se continuamente a forma de interação, busca e disponibilização da informação.

A cada segundo novas notícias são colocadas na internet em sítios como terra (www.terra.com.br), novos produtos são colocados à venda em sítios de comércio eletrônico como a Amazon (www.amazon.com) e Submarino (www.submarino.com), novos filmes são lançados em locadoras, cinemas, etc. Portanto, atualmente existem muitas opções de produtos e serviços e apenas poucos segundos para escolher os produtos ou serviços que os usuários desejam. Assim, é necessário filtrar essas possibilidades para o interesse do usuário, oferecendo opções que o interessem.

Nesse capítulo são descritos os conceitos de sistemas de recomendação aplicados no domínio de comércio eletrônico e ainda é apresentada uma taxonomia para estes sistemas.

2.1 Personalização

A maior parte das propagandas existentes hoje em dia é voltada para atingir o público em geral, voltado para ter sucesso na televisão, *outdoors*, rádios, revistas, entre outros. Esse *marketing* é chamado de *marketing* de massa [3]. Alguns problemas são apontados nesse tipo de *marketing* [17]: (i) elevado custo; (ii) muitas vezes, esses anúncios não são criativos o suficiente para atrair atenção dos consumidores; (iii) curta duração, nem sempre consegue passar a mensagem para o consumidor; (iv) o cliente é bombardeado com propagandas todos os dias. Com os problemas citados, as empresas cada vez mais caminham para um tipo de *marketing* que tenha como objetivo atingir diretamente o cliente com menor custo e maior nível de atração da atenção do usuário, o *marketing* direto [2].

O *marketing* direto é voltado para o indivíduo. Ele lida diretamente com o usuário, tentando personalizar a loja ou o sítio com informações específicas daquele usuário. A personalização é conseguida graças ao maior conhecimento a respeito do cliente. O armazenamento de histórico de compras, por exemplo, é importante para a personalização, pois a partir daí, se pode inferir o que o cliente prefere comprar [18].

A personalização é uma maneira de conseguir vantagem competitiva [1], pois ajuda o cliente a encontrar produtos de seu interesse facilmente, diminuindo com isso o tempo de escolha ao oferecer para o cliente um bom produto, tornando a compra conveniente. Por conseguinte, aumentando a sua satisfação.

O principal objetivo da personalização é a fidelização dos clientes. Para empresas que querem se manter no mercado, a conquista e a fidelização de clientes deve ser o maior objetivo. Segundo Stan Rapp e Tom Collins [19], os clientes querem sentir que as empresas se importam com eles, e então as empresas devem estar dispostas a manter um bom relacionamento com os mesmos. As empresas podem aumentar essa fidelização através da personalização e interação com necessidades e peculiaridades de seus clientes [1]. Para empresas de comércio eletrônico essa ação é chamada de *e-loyalty* [20].

2.2 Sistemas de Recomendação

O usuário, frequentemente, procura se informar que item seria interessante para ele antes de comprar, geralmente com indicações de pessoas conhecidas, sobre a melhor opção para sua compra (i.e. o chamado *Word of Mouth* [21]) ou ainda olhando críticas, jornais, revistas, entre outros recursos para obter informações.

Sistemas de recomendação podem automatizar parcial ou totalmente a busca por informações relevantes ao usuário. Esses sistemas ajudam ao usuário a escolher produtos e serviços dentre um universo de possibilidades. Sistemas de recomendação aplicam a personalização, recomendando produtos de interesse do usuário baseado, por exemplo, no histórico de interação do mesmo com o sistema. Com a facilidade de escolha de um produto proporcionada pelos sistemas de recomendação e o bom relacionamento oferecido ao cliente, as lojas que utilizam esses sistemas podem aumentar suas vendas e também fidelizar ainda mais seus clientes.

Sistemas de recomendação se tornaram uma das áreas de estudo mais importantes da inteligência artificial [22]. A Internet impulsionou as aplicações de sistemas de recomendação, pois se trata de um ambiente ideal para a personalização, principalmente no comércio eletrônico com sua altíssima diversidade de produtos e serviços e a necessidade cada vez maior da fidelização de clientes.

2.3 Taxonomia de Sistemas de Recomendação

Uma taxonomia foi produzida por Shafer [23] para classificar a arquitetura de um sistema de recomendação. Roberto Torres [18] produziu também uma classificação de sistemas de recomendação baseado na taxonomia de Shaffer. Nessa seção é descrita a taxonomia desenvolvida por Shaffer e acrescentando alguns pontos propostos por Roberto Torres.

A arquitetura da taxonomia envolve 3 blocos distintos e complexos, pois cada um deles pode ter implementações e classificações distintas. A divisão se torna necessária para um melhor entendimento desses sistemas, pois cada bloco é um componente significativo do sistema de recomendação. A Figura 1 apresenta a taxonomia.

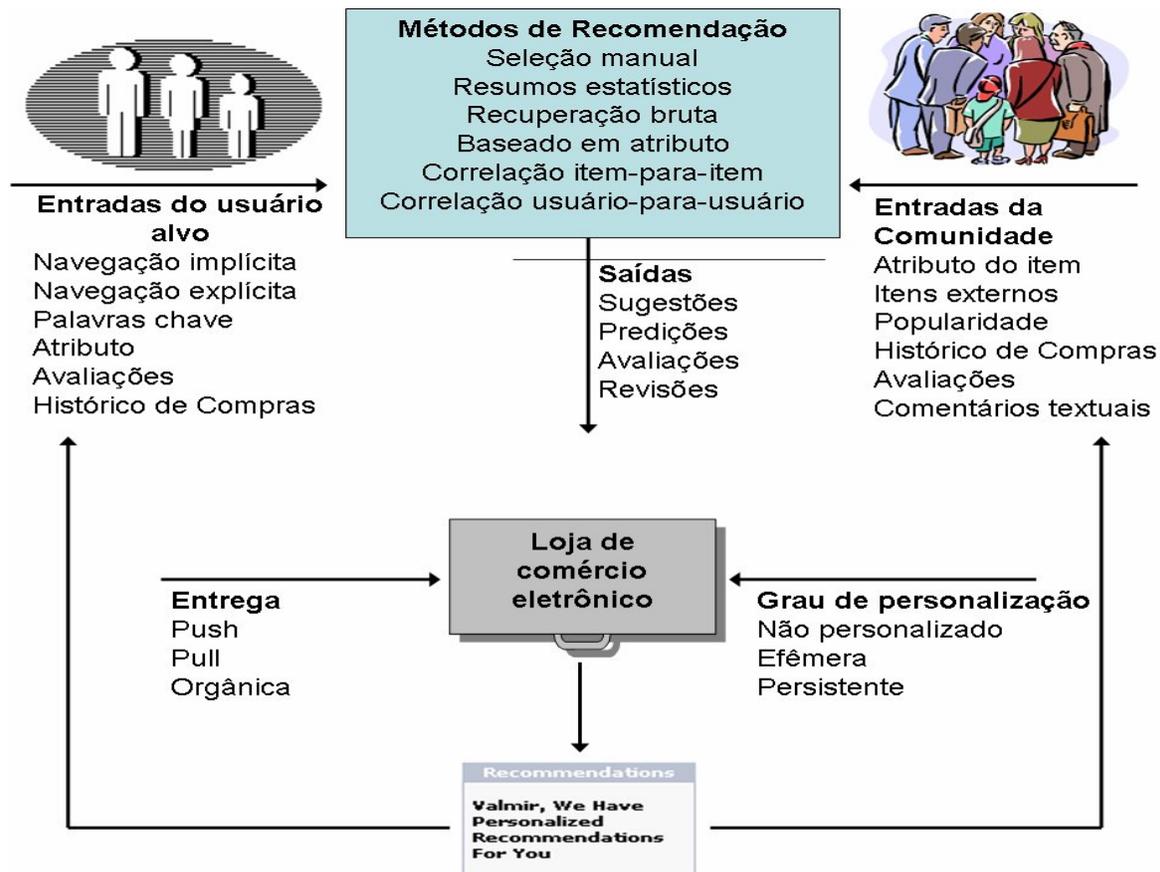


Figura 1. Taxonomia dos sistemas de recomendação

2.3.1 Entrada/Saída

Aqui é mostrado como se processa o fluxo das informações de entrada e saída do sistema. Esta classificação leva em conta 3 tipos de informações: (i) informações do usuário-alvo (ativo), (ii) informações da comunidade de usuários e (iii) informações de saída.

- **Informações do usuário-alvo**

Um sistema de recomendação precisa das informações do usuário alvo por se tratar de uma personalização para esse usuário. Então, quanto mais informações obtivermos sobre esse usuário, mais o sistema poderá recomendar produtos ou serviços para ele de acordo com suas preferências.

Os usuários-alvo devem ser identificados pelo sistema. Atualmente é comum o cadastro de usuários em sítios de comércio eletrônico, fóruns, entre outros. Pode-se utilizar *cookies* [24], que armazenam informações sobre o usuário ligado aquele computador. O problema dos *cookies* é que nem sempre o computador é utilizado por uma única pessoa, podendo então, confundir o sistema com informações de outras pessoas.

Há duas formas de coletar informações sobre o usuário-alvo: implícita e explícita. Se uma aplicação não contém informações sobre o usuário-alvo, apenas poderá ter resultados impessoais, ou seja, não-personalizados.

As informações implícitas são coletadas de maneira que o usuário não perceba que o sistema está coletando informações sobre o mesmo. Essas informações são coletadas durante a navegação do usuário, tais como: tempo de visita a uma página de um certo produto, movimento do *mouse*, entre outros [25][26][27]. Outra forma implícita é o histórico de compras do usuário.

Informações implícitas são ideais para usuários que não têm tempo suficiente para interagir com o sistema, pois essas informações são coletadas sem que o usuário tenha conhecimento que o sistema está armazenando informações sobre o mesmo e são coletadas sem um esforço do mesmo. A desvantagem é que o usuário não fornece informações diretamente, fica a cargo de o sistema presumir o que aquela avaliação implícita significa, se é boa ou ruim.

Na segunda forma de coletar informação, a explícita, as informações são fornecidas intencionalmente pelos usuários, nela o usuário expressa de alguma forma se gosta ou não do produto quantitativamente, por exemplo, fornecendo notas aos produtos. Esse tipo de informação é considerada mais confiável, já que o usuário é quem fornece a informação, mas o custo desse tipo de procedimento é o esforço do usuário que nem sempre está disposto a colaborar com mais informações para o sistema. Geralmente são coletadas informações como notas atribuídas para um produto, essas notas são obtidas por interfaces simples, por exemplo, com possibilidades de notas variando de 1 a 5 [11], ou indicações binárias simples informando se gosta ou não gosta de uma lista de produtos indicados.

Existem sistemas de recomendações que mesclam essas duas formas de coleta de informações [28], pois a forma implícita nem sempre é confiável enquanto a explícita tem a vantagem mencionada acima.

- **Informações da comunidade.**

Informações de todos os usuários que interagem de alguma forma com o sítio também podem ser utilizadas para calcular recomendações de produtos. Existem três tipos de informação da comunidade, são elas: comentário sobre produtos, avaliações e popularidade do produto:

- Os comentários sobre produtos são colocados por usuários do sítio, os mesmos devem conhecer o produto e expressar num comentário sua opinião sobre o mesmo, seja ela positiva ou negativa. Apesar de algum esforço ser necessário do usuário, esse tipo de comentário garante a outros usuários uma opinião neutra que pode ajudá-lo a escolher entre a compra ou não desse produto.

- Avaliações são mais simples que os comentários, pois o usuário geralmente avalia com uma nota numérica esse produto, expressando o quanto gosta ou não deste. Dessa forma, elimina-se um pouco do esforço que é o de escrever um comentário para expressar sua opinião a respeito de um produto.
- A popularidade é uma maneira de oferecer um produto quando não se tem informações suficientes no perfil de um usuário. Podem ser utilizados, por exemplo, os produtos mais vendidos para oferecer ao usuário.

• Saída

A saída representa a resposta do sistema, variando em tipo, quantidade e aparência. A saída mais comum é a sugestão, que aparece na forma: “Temos uma recomendação para você!” ou ainda “Experimente este produto!”.

Com uma sugestão, geralmente o usuário é atraído para ela e pode realmente considerar essa sugestão, obviamente existe o risco dele simplesmente descartar essa sugestão por já possuir esse item ou simplesmente não gostar. Outra estratégia é recomendar uma lista de produtos e oferecer descontos para os mesmos, assim aumentar as chances do usuário poder se interessar por algum deles.

Predictions for you ↕	Your Ratings	Movie Information	Wish List
★★★★★	Not seen ▾	Herod's Law (La Ley de Herodes) (2000) DVD VHS info imdb add tag Comedy, Crime, Mystery - Spanish	<input type="checkbox"/>
★★★★★	Not seen ▾	Star Wars (a.k.a.: Star Wars: Episode IV - A New Hope) (1977) DVD VHS info imdb Action, Adventure, Sci-Fi [add tag] Popular tags: sci-fi , space , George Lucas	<input type="checkbox"/>
★★★★★	Not seen ▾	Star Wars: Episode V - The Empire Strikes Back (1980) DVD VHS info imdb Action, Adventure, Sci-Fi [add tag] Popular tags: sci-fi , space , George Lucas	<input type="checkbox"/>
★★★★★	Not seen ▾	Lord of the Rings: The Two Towers, The (2002) DVD VHS info imdb Action, Adventure, Fantasy [add tag] Popular tags: fantasy , high fantasy , Peter Jackson	<input type="checkbox"/>
★★★★★	Not seen ▾	Raiders of the Lost Ark (1981) DVD VHS info imdb Action, Adventure [add tag] Popular tags: indiana jones , archaeology , ark of the covenant	<input type="checkbox"/>
★★★★★	Not seen ▾	Shawshank Redemption, The (1994) DVD VHS info imdb Drama [add tag] Popular tags: prison , classic , Stephen King	<input type="checkbox"/>
★★★★★	Not seen ▾	Silence of the Lambs, The (1991) DVD VHS info imdb Crime, Horror, Thriller [add tag] Popular tags: serial killer , cannibalism , crime	<input type="checkbox"/>
★★★★★	Not seen ▾	American History X (1998) DVD VHS info imdb Drama	<input type="checkbox"/>

Figura 2. Predição de filmes para usuário

A predição é outro tipo de estratégia que pode ser empregada por sítios de comércio eletrônico. Na predição o sistema oferece uma avaliação do produto, que seria uma suposta avaliação que o usuário produziria caso avaliasse o produto. Para a predição pode ser considerada a nota de um usuário com preferência semelhante ou, ainda, pode-se calcular a similaridade desse produto com o perfil do usuário. A Figura 2 acima, demonstra a predição de um item do sítio *movielens* [29], que é um sítio especializado em recomendações de filmes.

Outras formas de saída são resenhas e avaliações que são baseadas nas informações obtidas pela comunidade de usuários. As resenhas são opiniões que os próprios usuários escrevem acerca de algum produto conhecido. Elas são isentas e conferem senso de comunidade na loja virtual. Alternadamente, resenhas podem ser apresentadas em forma de estrelas que indicam a média de avaliações dadas por usuários da comunidade para aquele produto.

Todas essas formas podem ser utilizadas pelo sítio, em maior ou menor nível, dependendo da estratégia de marketing do sítio para melhor fidelizar seus clientes. Lojas como a Amazon (www.amazon.com) utilizam diversas formas de apresentação das recomendações.

2.3.2 Métodos de Recomendação

Aqui são descritas categorias que representam métodos utilizados atualmente para gerar recomendações. Esses métodos representam as diferentes maneiras que as informações coletadas no perfil do usuário são utilizadas para gerar recomendações.

- **Seleção Manual**

Método de recomendação mais simples, onde editores e especialistas que conhecem o público-alvo propõem as recomendações.

- **Resumos Estatísticos**

Método de recomendação que visa complementar os métodos de seleção manual, pois nem sempre editores e especialistas acertam as necessidades e preferências de todos os usuários. Exemplos de resumos estatísticos são índices de satisfação do usuário e índice produtos mais vendidos.

- **Recuperação Bruta**

Sistemas de recuperação bruta implementam uma interface simples para busca por palavras-chave numa base de dados. O usuário procura um produto ou serviço por uma palavra-chave e o sítio exhibe uma lista dessa busca em sua base de dados de acordo com essa palavra-chave. Atualmente esse método é bastante utilizado em sítios de comércio eletrônico e não representa propriamente uma recomendação, já que esse sistema apenas recupera informações de uma base de dados e não realiza nenhuma filtragem a não ser à busca das palavras chave [18]. Este método é muito utilizado também em sítios de busca na

Internet, como o Google (www.google.com), por exemplo. Os 3 primeiros métodos de recomendação citados são considerados não-personalizados, pois todos os usuários que receberem recomendações de um desses 3 métodos receberão a mesma recomendação.

- **Baseado em Atributos**

Este tipo de sistema recomenda produtos baseado em suas características. Para um usuário que escuta muitos CDs da banda U2, por exemplo, um lançamento de CD desta banda seria uma possível recomendação do sistema para ele. Outro exemplo é a oferta de desconto no valor de um CD de rock, enquanto o usuário navega numa seção do sítio que vende esse gênero musical.

- **Correlação item-para-item**

Esse método tenta identificar produtos que de alguma forma tem relação com produtos que o usuário demonstrou interesse no seu histórico de interação com a loja virtual. Um exemplo desse tipo de recomendação é a venda casada, por exemplo, a oferta de um par meias quando o usuário compra um tênis. Aplicações podem inspecionar itens dentro do carrinho de compras para identificar essas associações. Esse tipo de recomendação tem como objetivo principal aumentar as vendas e oferecer um serviço diferencial aos clientes. No momento em que o usuário está realizando alguma compra, o sítio tenta convencer-lhe a comprar mais um produto.

- **Correlação usuário-para-usuário**

Esse método leva em conta a colaboração de usuários que têm preferências parecidas, visto que, estes provavelmente manterão a coerência em seus produtos comprados. Esta técnica recomenda produtos de acordo com a similaridade do usuário-alvo com outros clientes do sítio de comércio eletrônico.

Um fator importante a ser considerado nas recomendações é se a geração da recomendação pode ser realizada *on-line*, ou se é necessário um pré-processamento *off-line*. Recomendações *on-line* utilizam informações mais atualizadas, pois utilizam as informações do usuário até aquele momento da geração da recomendação. Resumos estatísticos e recomendação baseada em atributos são simples de processar e podem ser geradas *on-line*. As correlações item-para-item e usuário-para-usuário requerem cálculos mais complexos e demorados. Usualmente elas são pré-processadas para que a geração da recomendação não seja tão demorada. Deve-se ter cuidado com a atualização desses sistemas para que o mesmo utilize informações as quais correspondam ao comportamento atual do usuário.

2.3.3 Outros aspectos de projeto

Alguns outros aspectos do sistema de recomendação que não tem relação com a implementação, mas diretamente com sua qualidade. Esses aspectos são considerados importantes num projeto de sistema de recomendação, são eles: grau de personalização e entrega das recomendações.

- **Grau de Personalização**

O grau de personalização descreve o quanto o sistema é personalizado e pode ser: (i) não-personalizado, (ii) efêmero e (iii) persistente.

O sistema que gera recomendações iguais para todos os usuários é considerado não-personalizado. Sistemas que utilizam recomendações escolhidas por editores, seleção manual, e sistemas que recomendam livros mais vendidos, resumos estatísticos, também são sistemas considerados não-personalizados.

Um sistema efêmero leva em consideração necessidades e desejos momentâneos do usuário. Por exemplo, utilizam-se informações da navegação atual do usuário, assim, dependendo da seção que o usuário esteja navegando gera-se uma recomendação baseada em atributo. Ainda, outro exemplo é a recuperação de informação do carrinho de compras do usuário para gerar uma recomendação, utiliza-se a correlação item-para-item.

Um sistema personalizado leva em conta informações mais consistentes do histórico de interação do usuário com o sítio de comércio eletrônico para calcular as recomendações. Cada usuário recebe uma recomendação diferente, de acordo com seu perfil armazenado nesse sítio. O sistema deve ser capaz de alterar as preferências do usuário com o passar do tempo para continuar atendendo aos desejos e necessidades do seu cliente. A Tabela 1 destaca o grau de personalização versus o método de recomendação.

Tabela 1. Grau de personalização x método de recomendação

Grau de Personalização	Métodos de Recomendação
Não-personalizada	Seleção Manual, resumos estatísticos
Efêmera	Item-para-item e baseada em atributo
Persistente	Usuário-para-usuário

- **Entrega das Recomendações**

A maneira que o sistema apresenta as recomendações ao usuário é uma decisão crítica da estratégia de personalização [2]. Pode ser classificado como *push*, *pull* e orgânica.

A tecnologia *Push* recomenda produtos ao usuário sem que o mesmo peça as recomendações. A vantagem é que o usuário não precisa interagir com o sistema para receber uma recomendação. O envio de emails com recomendações é a principal forma desse tipo de entrega de recomendação

A tecnologia *Pull* proporciona ao usuário o controle de quando a recomendação é apresentada. Geralmente o usuário clica num *link* que apresenta a recomendação que o sítio de comércio eletrônico o oferece.

A forma orgânica é apresentada ao usuário de forma leve, geralmente com janelas ao lado do sítio que o usuário está navegando, sem avisar ao usuário e sem que permita controle do usuário sobre a requisição da recomendação. Esse tipo de recomendação pode não deixar claro para o usuário que ele está tendo um tratamento diferenciado e por isso é pouco utilizado.

Capítulo 3

Técnicas de Sistemas de Recomendação

Nesse capítulo é explicado o funcionamento das técnicas de sistemas de recomendação. São abordadas as três técnicas mais utilizadas que são a filtragem colaborativa, a filtragem baseada em conteúdo e a filtragem híbrida.

3.1 Filtragem Colaborativa

A filtragem colaborativa parte do princípio de que pessoas pedem recomendações sobre produtos que irão consumir a outras pessoas que tenham gostos parecidos com elas. Dá-se o nome de filtragem colaborativa à automação desse processo realizado por computadores.

Um dos primeiros sistemas de recomendação criaram a primeira, e a mais utilizada atualmente [6], abordagem dos sistemas de recomendação, ela foi denominada de “filtragem colaborativa”. Esse sistema era o Tapestry [8][13], que recomendava emails que vinham de diversas listas e utilizavam uma filtragem colaborativa para organizar os emails de interesse ao usuário-alvo. Após esse trabalho foi desenvolvido o sistema GroupLens [30] que impulsionou a pesquisa dessa área sendo citado por diversos outros trabalhos. Esse sistema recomendava notícias avaliadas quantitativamente por notas de 1 a 5 pelos usuários relacionando o quanto o usuário gostou daquela notícia e utilizando a técnica de filtragem colaborativa para apresentar notícias personalizadas

Esse tipo de abordagem tenta identificar indivíduos que possuem preferências comuns. Após identificar esses indivíduos, forma-se um grupo de indivíduos com gosto comum, depois os indivíduos desse grupo colaboram entre si através dos produtos que os mesmos gostam para encontrar um produto que um usuário-alvo ainda não possua. O produto, ou os produtos são então recomendados a este usuário, que provavelmente irá julgar interessante.

3.1.1 Funcionamento da Filtragem Colaborativa

- **Perfil do Usuário**

A primeira etapa dessa técnica é o armazenamento dos perfis dos usuários que geralmente é composto por avaliações destes para alguns produtos da loja de comércio eletrônico. Após o armazenamento dessas informações, é realizado um processamento para identificar pessoas com gostos semelhantes utilizando técnicas de inteligência artificial. O passo final é a apresentação dos produtos para o usuário-alvo, são escolhidos os produtos melhor avaliados no perfil dos vizinhos mais próximos.

Existem diversas técnicas de modelagem do perfil do usuário para a filtragem colaborativa. Nos primeiros sistemas desse tipo, os perfis eram vetores de avaliações dos produtos com notas ou simplesmente com informações binárias informando se gosta ou não do produto [13][30]. Com o passar do tempo novas técnicas de modelagem surgiram, utilizando também o acréscimo de informações conseguidas implicitamente de acordo com a navegação do usuário no sítio de comércio eletrônico [6][25][26]. Outra técnica que vem alcançando bons resultados é a formação do perfil do usuário utilizando técnicas de análise de dados simbólica, esse perfil também conta com as informações de avaliações do usuário, mas ele é formado de acordo com pesos calculados para cada produto conjuntamente com as avaliações dos usuários [12].

- **Gerando Recomendações**

Dependendo da técnica utilizada, a etapa seguinte é a geração de grupos de usuários com preferências semelhantes. Uma das técnicas mais utilizadas para essa finalidade é a KNN (K vizinhos mais próximos) [31].

Para identificar a vizinhança, é medida a similaridade de um usuário-alvo com os outros usuários da loja virtual. A função mais conhecida para medir a similaridade é a distância euclidiana [31], que calcula a distância entre dois vetores que representam os perfis do usuário. Quanto menor a distância calculada de um vetor a outro vetor, mais semelhantes são os usuários. O valor da distância varia de 0, maior similaridade, a 1, menor similaridade. A equação (1) define a distância euclidiana:

$$w_{a,u} = \sqrt{\sum_{i=1}^m (r_{a,i} - r_{u,i})^2} \quad \text{Equação (1)}$$

Na equação (1), $w_{a,u}$ a distância entre o usuário ativo a com um determinado usuário u , $r_{a,i}$ é a avaliação que o usuário ativo deu para o item i , $r_{u,i}$ é a avaliação de um outro usuário para o mesmo item.

Existem diversas funções que podem ser utilizadas para medir essa similaridade, mais uma função bastante utilizada é a cossenoidal, que calcula o cosseno de dois vetores

n-dimensionais que representam o perfil do usuário. Os valores do cosseno indicam a similaridade dos usuários. Quando o valor é próximo ou igual a 1, são usuários muito próximos, e quanto mais perto ou igual a 0, são usuários que possuem nenhuma ou pouca correlação. O cálculo do cosseno é definido pela equação (2)

$$Cos_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} * r_{u,i})}{\sqrt{\sum_{i=1}^m (r_{a,i})^2} \sqrt{\sum_{i=1}^m (r_{u,i})^2}} \quad \text{Equação (2)}$$

Na equação (2), o $Cos_{a,u}$ é o cálculo do cosseno entre um usuário ativo a e determinado usuário u , $r_{a,i}$ é a avaliação que o usuário ativo deu para o item i , $r_{u,i}$ é a avaliação de um outro usuário para o mesmo item.

Ainda sobre funções de similaridade, uma das mais conhecidas é o coeficiente de Pearson, que mede a “força” de duas variáveis. Para se calcular a similaridade de dois usuários, calcula-se esse coeficiente para itens que ambos os usuários avaliaram e depois é calculado a média das avaliações de todos os coeficientes calculados, fornecendo o quanto um usuário é similar a outro. O valor do coeficiente de Pearson varia de -1, indicando ausência de correlação, a 1, indicando total correlação. Ele é ideal quando há produtos avaliados positivamente e negativamente em perfis dos usuários. O coeficiente do Pearson é definido pela equação (3):

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}} \quad \text{Equação (3)}$$

Na equação (3), o $w_{a,u}$ é a correlação entre um usuário ativo a e determinado usuário u , $r_{a,i}$ é a avaliação que o usuário ativo deu para o item i , e \bar{r}_a é a média de todas as avaliações do usuário ativo, $r_{u,i}$ e \bar{r}_u possuem a mesma definição apresentada anteriormente sendo para o usuário u .

No caso da análise simbólica é utilizada uma fórmula adaptada da distância euclidiana que analisa a similaridade entre dois perfis simbólicos dos usuários [12]. O valor dessa função de similaridade, assim como a distância euclidiana varia de 0 a 1, indicando ausência e forte relação, respectivamente. A equação de similaridade simbólica é apresentada em detalhes no capítulo 4.

A última etapa é a apresentação das recomendações ao usuário. A recomendação é apresentada de acordo com a estratégia utilizada pelo sítio de comércio eletrônico explicados no capítulo 2, seção 2.3.1.. Os produtos recomendados são escolhidos entre os produtos melhor avaliados contidos nos perfis dos usuários semelhantes.

3.1.2 Vantagens da Filtragem Colaborativa

A filtragem colaborativa possui 3 principais vantagens: (i) independência de conteúdo, (ii) geração de recomendações baseadas em preferências dos usuários, (iii) possibilidade de produzir recomendações inesperadas e de alta qualidade.

A técnica de filtragem colaborativa se baseia em avaliações de usuários para gerar a recomendação. Essa característica fornece a esses sistemas uma independência de conteúdo, podendo gerar recomendações de diversos tipos de produtos. Filmes, CDs e livros, por exemplo, podem ser recomendados pelo mesmo sistema que utiliza a filtragem colaborativa.

A capacidade de produzir recomendações de qualidade depende das avaliações positivas e negativas fornecidas pelo usuário. Para um sistema automatizado é muito difícil perceber a qualidade de um livro, por exemplo, baseando-se na análise do texto do livro, quando o usuário avalia esse livro fica mais simples para o sistema produzir recomendações de qualidade.

A terceira vantagem é a capacidade de o sistema recomendar produtos inesperados, mas com qualidade, ou seja, uma recomendação que surpreenda positivamente o usuário, essa ação é chamada de *serendipity*.

3.1.3 Desvantagens da Filtragem Colaborativa

Há 3 desvantagens inerentes à filtragem colaborativa: (i) o problema do avaliador, (ii) a dispersão da base de dados, (iii) ovelha negra e (iv) custo de processamento.

A base da filtragem colaborativa é a avaliação de produtos por usuários, então, um produto que não tenha sido avaliado nunca será recomendado, pois esse produto nunca irá aparecer como opção de recomendação para o sistema de recomendação.

Sítios de comércio eletrônico possuem uma base de dados, comumente, muito grande. Para um usuário formar um perfil com 0,1% dos produtos de uma loja que tenha um milhão de produtos em seu catálogo é preciso que o usuário compre ou avalie 1000 produtos. Esse problema é chamado de dispersão dos dados e devido a isso é difícil encontrar bons vizinhos pelo baixo número de produtos em comum consumidos ou avaliados (baixo *overlapping*).

O problema do falso vizinho ou ovelha negra acontece quando usuários são considerados semelhantes para o sistema, mas na verdade esses usuários não possuem preferências parecidas para outros itens. Pode ocorrer que usuários, coincidentemente, avaliem alguns itens iguais e nos perfis dos mesmos possuam poucos itens. Usuários com mais produtos avaliados igualmente têm muito mais chance de realmente serem semelhantes. O que pode-se fazer é multiplicar por uma constante $n/50$ quando o número de itens avaliados for inferior a 50, onde n é número de produtos avaliados. Usuários com mais de 50 itens avaliados não são influenciados na sua similaridade [32].

Outro problema inerente a filtragem colaborativa é que o cálculo para formar o grupo é computacionalmente caro, pois para formar um grupo de pessoas com gostos semelhantes é necessário calcular a similaridade de um usuário-alvo com todos os outros usuários da loja virtual, que pode ser um processo demorado e caro computacionalmente.

3.2 Filtragem Baseada em Conteúdo

A filtragem baseada em conteúdo é uma técnica que, ao contrário da filtragem colaborativa, faz as comparações de similaridade entre o usuário e um item da loja virtual, sendo recomendado o item que for mais similar com o perfil do usuário. A filtragem colaborativa parte do princípio de que se o usuário gosta de um produto, provavelmente irá gostar de um produto similar a esse produto. Para produtos que contenham alguma informação textual que o categorize ou forneça informação suficiente para uma classificação essa técnica é bastante utilizada.

Sistemas de recuperação de informação (RI) [33] foram os primeiros sistemas a apresentar para o usuário respostas baseadas em seu conteúdo. Através de uma consulta por palavras-chave, uma busca na base de dados é realizada e assim, esses sistemas apresentam como resposta para o usuário os itens da base de dados que contenham essa palavra-chave em seu conteúdo.

Sistemas de filtragem de informação (FI) acrescentaram mais complexidade aos sistemas de recuperação de informação [34]. Os sistemas de filtragem de informação são utilizados constantemente pelos usuários e as informações buscadas neles são informações de longo prazo, ao contrário dos sistemas de recuperação de informação onde as informações buscadas são de curto prazo. As consultas do sistema FI também diferem do sistema RI. No primeiro, o sistema geralmente oferece uma resposta baseada num perfil que foi construído para o usuário. No segundo, a consulta por respostas é realizada através de *queries*, que são palavras-chaves criadas pelos usuários.

Para exemplificar a diferença entre os sistemas RI e FI, vamos utilizar um cenário onde consultas são realizadas em uma biblioteca virtual. Uma consulta é feita a cada um desses sistemas, da seguinte forma: para o sistema RI, o usuário busca no sistema um livro que ele está precisando no momento para uma prova que fará no dia seguinte. Desse modo, ele fornece ao sistema o assunto da prova através de uma *query*. A partir da *query* fornecida, o sistema recupera na biblioteca virtual livros que contenham essa *query* (palavra-chave) em seu conteúdo. Enquanto para o sistema FI, o usuário consulta um livro para ler sem precisar fornecer informação para o sistema. O sistema realiza uma filtragem baseada no perfil de livros já lidos por esse usuário na biblioteca virtual e, assim, fornece como resposta, livros que mais se assemelham com o perfil desse usuário.

A diferença entre FI e RI é apresentada para entender as características da filtragem baseada em conteúdo. O sistemas baseados em conteúdo contém exatamente todas as características citadas para os sistemas FI, por isso a filtragem baseada em conteúdo é considerada um sistema FI utilizado em sistemas de recomendação [18].

3.2.1 Funcionamento da Filtragem Baseada em Conteúdo

- **Perfil do Usuário**

A construção do perfil do usuário é um dos pontos-chave desse método. O perfil do usuário contém características descritivas (palavras-chave) dos produtos que o usuário deseja ou não deseja. As descrições dos produtos são textos que expressam o conteúdo do produto.

A técnica TF-IDF (*Term-frequency Inverse-Document-Frequency*) [35] é uma das técnicas mais utilizadas hoje em dia para a filtragem baseada em conteúdo. Essa técnica calcula similaridade de textos baseada na frequência que palavras chave aparecem nos textos. Para textos de informática, por exemplo, as palavras software e JAVA, provavelmente serão frequentes, distinguindo-se dos demais.

Outra técnica em ascensão é a técnica que forma o perfil do usuário baseada em análise simbólica das variáveis que descrevem o produto. O perfil do usuário é formado de variáveis multivaloradas, onde uma palavra chave suporte, é associada a um peso que representa o peso dessa variável no perfil do usuário. Nessa técnica é levada em conta a nota que o usuário forneceu para o produto, sendo construído o perfil positivo, para produtos que o usuário gostou, e o negativo, para produtos que o usuário não gostou [11].

- **Geração de Recomendações**

Diferente da técnica colaborativa, na filtragem baseada em conteúdo não há formação de vizinhança entre usuários semelhantes. As comparações são realizadas comparando-se o perfil do usuário a todos os produtos, identificando os produtos mais similares ao perfil desse usuário.

Para a filtragem que utiliza a técnica TF-IDF, o perfil do usuário irá conter os textos que descrevem os produtos que o mesmo gosta. Assim, cada texto da base de dados é comparado com o perfil do usuário que leva em conta a frequência das palavras chave que o texto contém. Além da frequência, essa técnica também discrimina as palavras que podem ou não descrever certo tipo de texto, a palavra tecnologia nada diz do seu contexto, mas se temos a palavra JAVA nós sabemos se tratar de um texto que fala da tecnologia da informática. Além disso, normalização é realizada para que textos grandes ou pequenos tenham o mesmo peso.

A técnica que utiliza análise simbólica compara os perfis do usuário com os perfis formados para cada um dos produtos da base. As comparações são realizadas para cada tipo de atributo que descreve o perfil do usuário e do produto. Compara-se cada tipo de atributo com o mesmo tipo de atributo no outro perfil. Se um atributo descreve a categoria de um livro, *ficção*, por exemplo, esse atributo vai ser comparado com o atributo que também descreve a categoria do livro, sendo utilizada fórmulas que levam em conta se há interseção entre os mesmos e os pesos que são associados a cada um desses atributos. Utilizando o mesmo exemplo, para elevar a similaridade entre um perfil do usuário que contenha *ficção* no seu atributo categoria, no perfil do produto comparado, *ficção* também deve estar presente.

Identificados os produtos que têm mais similaridade com o perfil do usuário, o último passo é a apresentação das recomendações ao usuário que utiliza as técnicas explicadas no capítulo 2 na seção 2.3.1.3.

3.2.2 Vantagens da Filtragem Baseada em Conteúdo

São duas as principais vantagens da filtragem baseada em conteúdo: (i) não possui o problema do primeiro avaliador e (ii) possui capacidade de recomendar todos os itens [18].

Por não precisar que algum usuário avalie um produto para gerar a recomendação, a filtragem baseada em conteúdo não possui o problema do primeiro avaliador mencionado em 3.1.3. Ademais, produtos disponíveis na base são todos comparados com o perfil do usuário, assim, todos os produtos possuem chances de serem recomendados.

3.2.3 Desvantagens da Filtragem Baseada em Conteúdo

Existem duas principais desvantagens na filtragem baseada em conteúdo: (i) não-consideração de aspectos como qualidade do texto e renome do autor e (ii) superespecialização [18].

Apesar de a técnica analisar o conteúdo de um texto, sua qualidade, clareza ou elegância não é levado em conta em sistemas que utilizam essa técnica, e como o usuário não fornece notas nessa filtragem, fica mais difícil ainda para um sistema saber se o usuário achou esse um texto bem escrito ou não, além disso não é levado em conta o renome de quem produziu o produto. Por exemplo, diretores de filmes ou autores de renome em determinada área são tratados igualmente com os outros nesse tipo de filtragem.

A superespecialização acontece porque os sistemas de recomendação de filtragem baseados em conteúdo fazem comparação de palavras-chave, então as recomendações geradas são semelhantes com as palavras-chave encontradas no perfil do usuário. Se um usuário possui no seu perfil livros de JAVA, mas também goste de C#, talvez nunca receba recomendação de um livro de C# apesar das duas opções serem linguagens de programação. Palavras diferentes que possuem significados iguais também podem ser consideradas de assuntos diferentes para um sistema baseado em conteúdo.

3.3 Filtragem Híbrida

As filtragens mais utilizadas, citadas nas seções anteriores, possuem desvantagens conhecidas. A filtragem híbrida faz uso de umas ou mais técnicas de recomendação, geralmente combinando uma de abordagem colaborativa com outra técnica, para minimizar suas limitações individuais [6].

Na filtragem colaborativa há a vantagem da independência de conteúdo, mas ao mesmo tempo existe a dependência da avaliação do usuário para um item ser recomendado. Enquanto na filtragem baseada em conteúdo, existe uma superespecialização, podendo recomendar qualquer produto da base de dados, eliminando o problema da dependência que algum usuário o avalie. As desvantagens de sistemas colaborativos e baseados em conteúdo

são complementares. Desvantagens de um são vantagens de outro. Então uma das técnicas híbridas mais utilizada é a junção dessas duas abordagens.

A única desvantagem que a junção com sistemas colaborativos não resolve é o *start-up*, inerente a praticamente todos os tipos de sistemas de recomendação [18]. O problema do *start-up*, refere-se ao tempo que o sistema leva para conseguir informações relevantes do usuário para gerar boas recomendações. Até que o sistema consiga essas informações alguns autores sugerem recomendar uma lista de itens mais vendidos ou utilizar dados demográficos como idade, sexo, cidade onde mora, para gerar uma recomendação[36].

3.3.1 Classificação de Sistemas Híbridos

Em seu artigo [6], Burke faz um estudo sobre sistemas híbridos. Nele é proposta uma classificação desses sistemas com sete classes possíveis:

1. **Ponderado:** A similaridade de um item é calculada combinando-se várias técnicas de recomendação, com pesos diferentes para cada uma delas.
2. **Alternado:** O sistema alterna a técnica que gera a recomendação dependendo de um critério escolhido.
3. **Cascata:** Uma técnica de recomendação refina as recomendações fornecidas por outra técnica.
4. **Combinação de características:** Características de diferentes fontes de recomendação são acopladas num único algoritmo
5. **Aumento de características:** As recomendações geradas por uma técnica são utilizadas como entrada para outra técnica.
6. **Misto:** Recomendações de várias técnicas são apresentadas na mesma lista.
7. **Meta-Level:** O modelo aprendido por uma técnica de recomendação é utilizado como informação de entrada para outra técnica.

3.3.2 Exemplo de Sistemas Híbridos

- **Fab:** Sistema desenvolvido na Universidade de Stanford, recomenda páginas da Internet a usuários [37]. Esse sistema utiliza análise de conteúdo para criar perfis de usuários e compara esses perfis para determinar usuários semelhantes. Por utilizar um modelo aprendido por uma técnica baseada em conteúdo como entrada de uma técnica colaborativa, esse sistema é considerado um sistema *Meta-Level*.
- **P-tango:** Sistema desenvolvido no Instituto Politécnico de Worcester, recomenda notícias em um jornal on-line [15]. As recomendações geradas por esse sistema utilizam uma média entre sugestões de filtragem colaborativa e baseada em conteúdo. Os usuários avaliam as notícias recomendadas, então o peso da técnica depende do usuário, esse tipo de sistema é classificado como ponderado.
- **TechLens:** Sistema desenvolvido na Universidade Federal do Rio Grande do Sul, recomenda artigos científicos para pesquisadores e estudantes [38]. O sistema combina sistemas baseados em conteúdo e colaborativos de diversas maneiras. O usuário seleciona um artigo que já tenha lido e recebe a recomendação. TechLens utiliza algoritmos de aumento de características e misto.

- **FEERS:** Sistema desenvolvido na Universidade Federal de Pernambuco, recomenda filmes para usuários baseados em avaliações de usuários [14]. Esse sistema utiliza um modelo aprendido por um método baseado em conteúdo como entrada de um método baseado em conteúdo, portanto é um sistema *Meta-Level*.
- **e-Recommender:** Novo sistema desenvolvido na Universidade de Pernambuco, fruto trabalho de conclusão de curso, recomenda produtos de uma loja de comércio eletrônico baseando-se nos produtos comprados pelo usuário. Esse sistema é baseado no FEERS, utilizando os mesmos algoritmos. A diferença está na formação do perfil do usuário que não leva em conta avaliações dos usuários, leva em conta itens comprados pelo usuário numa loja virtual. Também é considerado um sistema do tipo *Meta-Level*, pois utiliza um modelo aprendido pelo algoritmo da filtragem baseada em conteúdo como entrada para o algoritmo da filtragem colaborativa. Esse sistema é apresentado em detalhes no capítulo 4.

3.4 Dados Simbólicos

Dados simbólicos são uma nova área de interesse no que diz respeito à descoberta de conhecimento e, portanto têm sido objeto de estudo pelo meio acadêmico. É particularmente interessante pelo fato de poder ser aplicado em diversas áreas da ciência. Análise multivariada, reconhecimento de padrões, banco de dados e inteligência artificial são algumas das áreas em que a análise simbólica é utilizada com sucesso.

Nessa seção é explanado como é realizada a análise e modelagem de dados simbólicos. Modelagem esta que dá suporte ao algoritmo utilizado na ferramenta desenvolvida para realizar filtragem de informação do usuário.

3.4.1 Análise de Dados Simbólicos

A análise simbólica de dados, SDA (*Symbolic Data Analysis*), surgiu da influência de vários campos de atividade: (i) explanação clássica de análise de dados [39][40], (ii) inteligência artificial [22] e (iii) taxonomia numérica na biologia [41].

Dados simbólicos são utilizados para descrever estruturalmente dados complexos. Uma tabela simbólica contém células, onde cada uma delas pode conter distribuições ou intervalos categóricos ou ainda pesos ligados a regras lógicas ou taxonomias [42].

São duas etapas para construção de tabelas simbólicas: (i) extração de características das bases de dados, (ii) aplicação de ferramentas para estender a mineração de dados para extração de conhecimento dessa base de dados utilizando objetos simbólicos (SOs) [42].

3.4.2 Entrada do SDA

As tabelas do SDA contêm variáveis simbólicas que descrevem uma unidade chamada de indivíduo. As linhas são chamadas de descrições simbólicas dessas unidades e as colunas são as variáveis. Os tipos de valores que podem ser encontrados em cada célula são exemplificados a seguir:

- a) Valor quantitativo simples: se a altura é uma variável e h é um indivíduo: $Altura(h) = 6.8$.
- b) Variável categórica simples: $Cidade(h) = Recife$.
- c) Conjunto de valores ou categorias (Variáveis Multivaloradas): $Altura(h) = \{1.5, 1.65, 1.75\}$ onde a altura de h pode ser 1.5, 1.65 ou 1.75.
- d) Intervalo: $Altura(h) = [1.5, 1.75]$ significa que a altura de h varia no intervalo de 1.5 a 1.75.
- e) Conjunto de valores associados com pesos: $altura(u) = \{1.5, 1.75\}$ que representa a altura que os usuários podem possuir é associado ao conjunto $peso(u) = \{0.2, 0.8\}$ onde a altura h igual a 1.5 tem o peso associado de 0.2 e para a altura h igual a 1.75 tem o peso associado 0.8.

Exemplo de tabela simbólica:

Uma célula da tabela 2 é um valor quantitativo simples ou uma categoria:

Tabela 2. Tabela clássica de dados

Pessoa	Idade	Altura(m)	Peso(kg)	Sexo
W_1	20	1.68	50	F
W_2	25	1.75	70	M
W_3	12	1.50	45	M
W_4	60	1.65	70	F
W_5	15	1.63	55	F

3.4.3 Objetos Simbólicos

Objetos simbólicos resolvem o problema de descrever consistentemente a generalização de uma descrição de uma determinada Classe. Diday [43] define um objeto simbólico como uma tripla $s=(a, R, d)$, onde d é uma descrição, R é uma relação entre descrições, e a é um mapeamento definido por Ω em L dependendo de R e d . Ω é um conjunto de indivíduos (população ou universo, amostra) e $L = \{\text{verdadeiro, falso}\}$, onde é verdadeiro quando há uma conexão entre duas descrições e falso quando não há uma conexão. A relação R pode ser uma das relações do conjunto $\{=, \equiv, \leq, \subseteq, \Rightarrow\}$, onde \Rightarrow é uma implicação ou pode significar também, que combina com. Objetos Simbólicos se distinguem em dois tipos básicos: (i) objetos simbólicos booleanos e (ii) objetos Simbólicos modais:

i) Objeto Simbólico Booleano:

se $[y(w) R d] \in L = \{\text{verdadeiro, falso}\}$. $y(w)$ são dos tipos (a) até (d) definidos na seção 3.4.2. Um exemplo de objeto simbólico booleano é descrito no parágrafo a seguir:

seja $a(u) = [y(i) R d]$ com $R: [d' R d] = \bigvee_{i=1,2}$ onde \vee é um símbolo lógico. $R_i = \subseteq$. Se $y(i) = \{cor(i), peso(i)\}$, $cor(u) = \{azul, vermelho\}$, $peso(u) = \{21\}$. Então $a(u) = [cor(u) \subseteq \{vermelho, azul, amarelo\}] \vee [peso(u) \subseteq [10,15]] = verdadeiro \vee falso = verdadeiro$.

ii) Objeto Simbólico Modal: se $[y(w) R d] \in L = [0,1]$.

Por exemplo: utilizando a Tabela 3, vamos construir uma tabela simbólica modal onde vamos categorizar os dados. A idade é categorizada por pré-adolescente (10 a 14 anos), adolescente (15 a 18 anos), jovem (19 a 24 anos), adulto (25 a 60 anos) e terceira idade (a partir dos 60 anos). O peso é categorizado em {magro, normal, gordo, obeso} e altura {baixo, médio, alto}. Dessa forma a tabela de dados modais a qual o peso representa frequências é ilustrada na Tabela 3 abaixo:

Tabela 3. Dados simbólicos modais

G	Idade	Altura (m)	Peso (Kg)	Sexo
G_1	{(Pré-adolescente,1/5), (adolescente,3/5), (adulto,1/5)}	{(baixo,2/5), (médio,2/5), (alto,1/5)}	{(magro,2/5), (normal,2/5), (gordo,1/5)}	{(M,2/5)}, (F,3/5)}

A descrição modal descreve a probabilidade de encontrarmos respostas para consultas como a apresentada no exemplo acima num dado conjunto de dados. Geralmente os pesos são associados às frequências que uma variável aparece na base de dados, mas também podem ser utilizados: “capacidade”, “necessidade”, “possibilidades”, etc [43].

Capítulo 4

e-Recommender: Uma Ferramenta Híbrida para Recomendação Inteligente de Produtos

Sítios de comércio eletrônico procuram a personalização com a finalidade de conquistar e tornar o novo usuário um cliente fiel. Com a quantidade de produtos e serviços oferecidos por esses sítios, ferramentas de personalização inteligentes auxiliam a árdua tarefa que seria personalizar uma loja virtual para cada um de seus clientes.

Uma maneira de adquirir personalização é a utilização de sistemas de recomendação inteligentes. Esses sistemas atuam sugerindo produtos de uma loja eletrônica a seus respectivos usuários. Sistemas inteligentes de recomendação como o proposto nesse trabalho utilizam técnicas de inteligência artificial para fornecer recomendações relevantes ao cliente, baseado em um histórico de interações desse usuário com o sítio referido.

Esse trabalho sugere o uso de uma técnica relativamente nova para calcular essas recomendações. A abordagem simbólica representa usuários e produtos como um conjunto multivariado de características e pesos associados a essas características, permitindo que informações relevantes sejam modeladas e utilizadas para calcular uma recomendação que atenda às expectativas de um usuário, com a finalidade principal de que esse usuário se torne um cliente assíduo da loja.

4.1 Características

Como mostrado no Capítulo 3, há várias abordagens possíveis para um sistema de recomendação. Há uma gama de áreas de sistemas híbridos, que ainda não foram explorados, em particular há diversos exemplos de ferramentas que utilizam a abordagem colaborativa em conjunto com a abordagem baseada em conteúdo [6]. O sistema de

recomendação inteligente para produtos, e-Rec recommender, é um exemplo de um sistema que utiliza essa combinação de abordagem colaborativa e baseada em conteúdo.

e-Rec recommender é uma ferramenta de recomendação proposta e implementada nesse trabalho que utiliza uma abordagem colaborativa para encontrar usuários com gostos semelhantes. Ela também utiliza uma abordagem de conteúdo para comparar características descritivas dos produtos, para encontrar produtos interessantes ao usuário. Na abordagem colaborativa é utilizada a técnica KNN-FC [16], já na abordagem de conteúdo, é utilizada a técnica baseada em análise simbólica. De acordo com a classificação de sistemas híbridos, esse sistema é *Meta-Level*, pois o modelo aprendido pela filtragem de conteúdo é utilizado como entrada para a filtragem colaborativa.

O usuário interage com o sítio de comércio eletrônico e assim, o sítio armazena informações dessas interações. As interações como já explicado no capítulo 2 são explícitas ou implícitas. A informação utilizada na ferramenta é implícita, pois são utilizados os produtos que o usuário comprou para construir o perfil desse usuário. Por exemplo, a quantidade de livros que o usuário compra influi no resultado da recomendação. Como toda ferramenta de inteligência artificial, quanto mais informações a ferramenta dispuser para realizar seu treinamento, melhor a qualidade dos resultados.

A ferramenta foi desenvolvida utilizando a linguagem JAVA [44] e o banco de dados mysql [45] para armazenar a base de dados e as recomendações geradas. Um sítio de comércio eletrônico foi utilizado para aplicar a ferramenta e-Rec recommender. Uma adaptação nesse sítio de comércio eletrônico foi realizada para acomodar a funcionalidade de recomendação

Para um melhor entendimento do funcionamento do sistema, um diagrama de estados é apresentado na Figura 3.

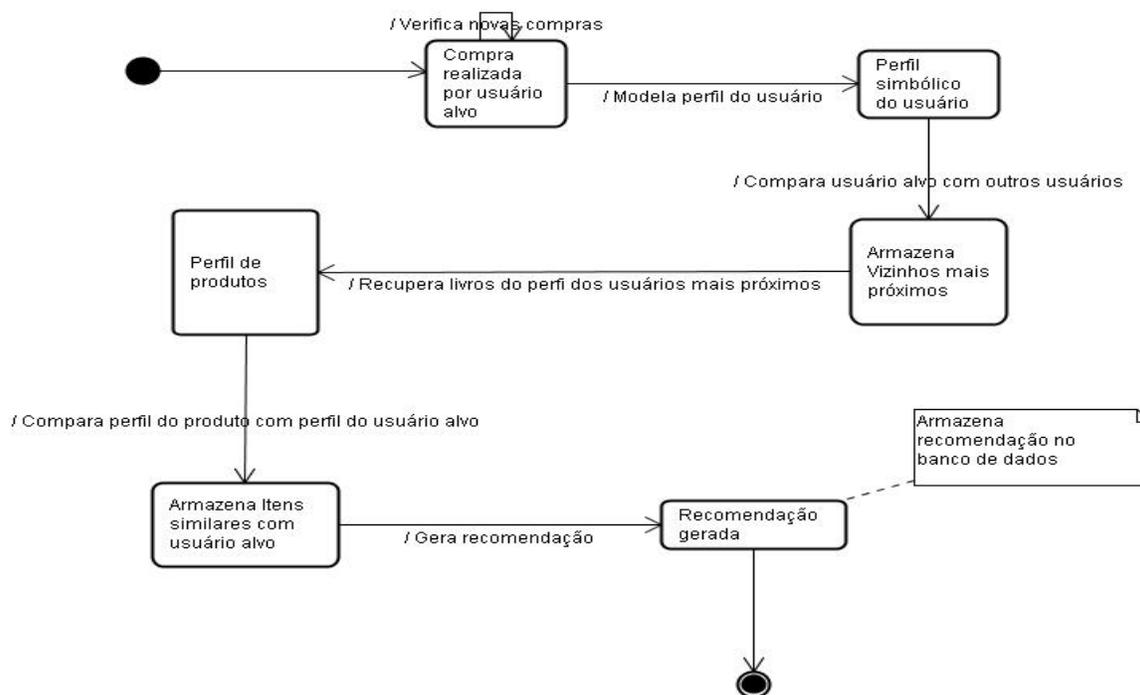


Figura 3. Diagrama de estados do e-Rec recommender

Primeiro, a ferramenta verifica se há novas compras realizadas no sítio de comércio eletrônico. Uma *thread* [44] realiza essa tarefa, verificando modificações na tabela do banco de dados que armazena as compras realizadas. O cliente que realizou a compra passa a ser chamado de usuário alvo, pois para ele a recomendação será gerada. Depois desta etapa, o perfil simbólico do usuário alvo é modelado.

A ferramenta utiliza uma função de similaridade para encontrar usuários com preferências semelhantes, armazenando k usuários mais próximos. A variável k se refere ao número de usuários semelhantes que o algoritmo utiliza para propor itens contidos nos perfis desses usuários. Na medida em que mais compras são realizadas por um usuário, os itens comprados são adicionados em seu perfil, proporcionando novas informações à ferramenta que procura por novos usuários mais próximos.

Após encontrar os vizinhos mais próximos do usuário alvo, a ferramenta modela o perfil dos produtos contidos nos perfis desses vizinhos. Estes produtos são comparados com o usuário alvo num próximo passo. Assim, a ferramenta armazena os n produtos que obtiverem maior similaridade. A variável n é o tamanho da lista de recomendações, essa lista de recomendações é armazenada no banco de dados.

4.2 Arquitetura

Para a integração do e-Rec recommender com os sítios de comércio eletrônico, a arquitetura *blackBoard* [46] foi estudada e desenhada. *BlackBoard* é uma arquitetura muito utilizada em sistemas distribuídos. Sua característica principal é que um recurso compartilhado, chamado *blackboard*, é utilizado por dois ou mais componentes. Os componentes que utilizam o recurso compartilhado não se comunicam, ou seja, não possuem nenhum tipo de ligação que permita a sincronização da utilização do componente compartilhado. Para manter um estado válido, um outro componente, chamado de controlador, sincroniza o acesso dos diversos componentes ao *blackboard*. A Figura 4 ilustra esse conceito:

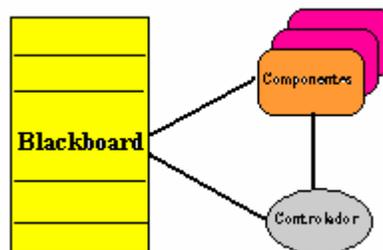


Figura 4. Arquitetura *blackboard*

O e-Rec recommender é escrito em JAVA, enquanto o sítio de comércio eletrônico pode ser escrito em qualquer linguagem. O recurso compartilhado que facilita a comunicação entre as duas aplicações é o banco de dados. No sítio de comércio eletrônico, as compras realizadas pelo usuário são armazenadas no banco de dados. Essas compras são as entradas que o e-Rec recommender utiliza para gerar as recomendações.

No e-Rec recommender, uma *thread* verifica alterações na tabela onde o sítio virtual armazena as compras do usuário. Deste modo, no mesmo momento que uma nova compra é realizada, a *thread* chama o método para gerar a recomendação para aquele usuário.

A recomendação utiliza o histórico de compras do usuário, recuperando todas as compras realizadas pelo usuário na loja virtual. Após as recomendações serem geradas, o e-Rec recommender se encarrega de armazená-las numa tabela no banco de dados, onde ficam armazenadas todas as recomendações geradas pela ferramenta. Desse modo, a recomendação destinada para algum usuário que efetuou uma compra no sítio de comércio eletrônico estará disponível após poucos segundos.

Após essas etapas, o usuário já pode requisitar uma recomendação personalizada. O sítio de comércio eletrônico seleciona as recomendações e apresenta ao usuário da maneira que for a estratégia de entrega do mesmo. A Figura 5 ilustra a arquitetura descrita acima.

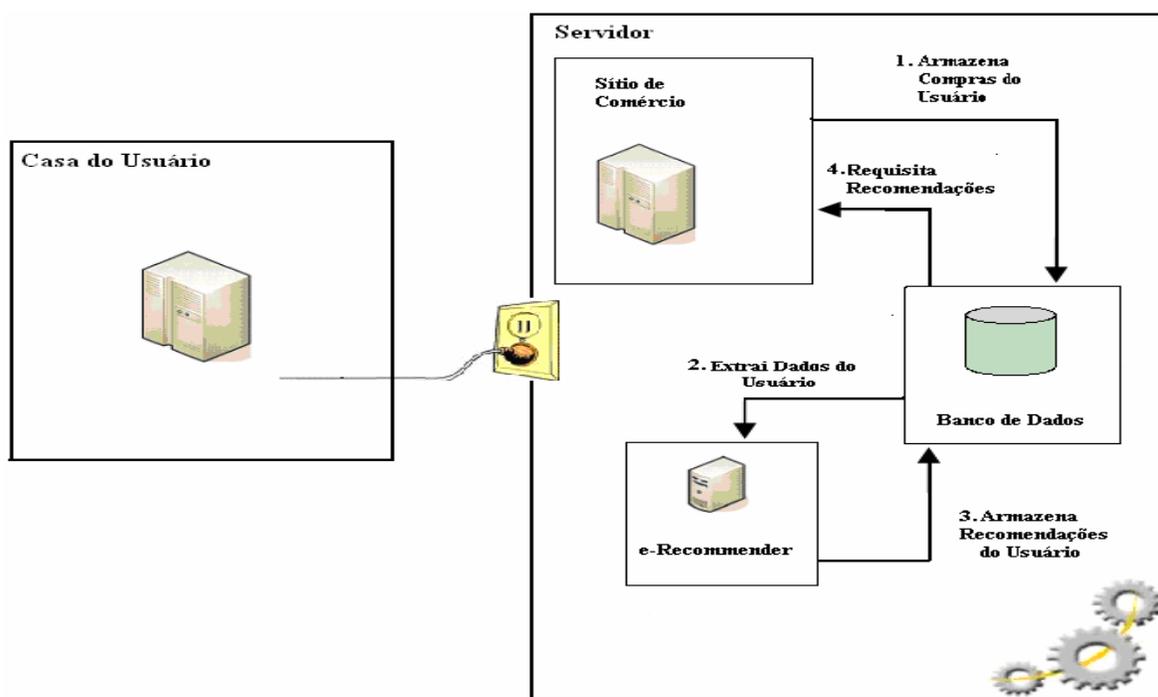


Figura 5. Arquitetura da aplicação e-Rec recommender

4.3 Sítio de Comércio Eletrônico

Para ilustrar a utilização da ferramenta, um sítio de comércio eletrônico foi integrado ao e-Rec recommender. O osCommerce [47] é um software de código aberto¹ [48] que implementa todas as funcionalidades de uma loja virtual.

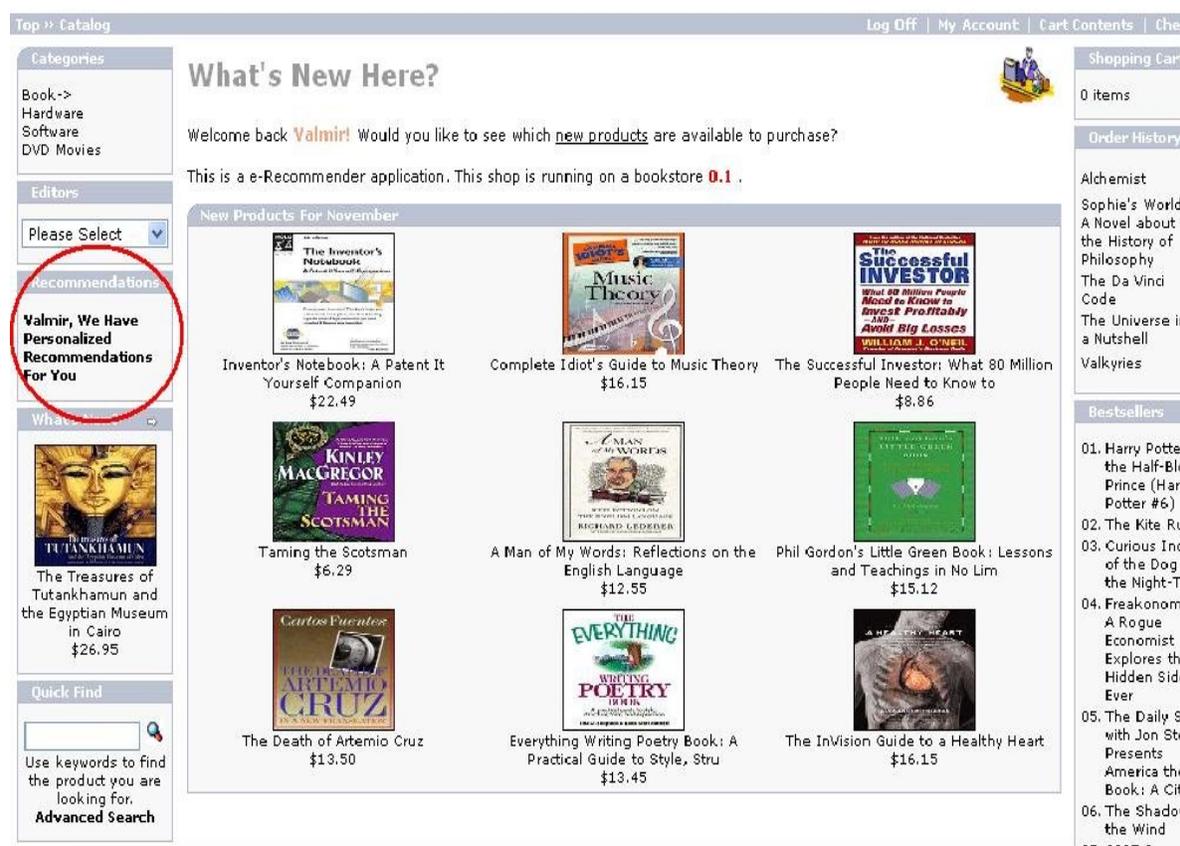
O servidor apache 2 (Apache Server) [49] foi utilizado como servidor do sítio de comércio eletrônico osCommerce.

¹ Software que possui licença gratuita e código disponível para alterações

PHP (*Hypertext Preprocessor*) [50] é uma linguagem de programação muito utilizada para criar sítios dinâmicos na internet. As principais características do PHP são: linguagem interpretada, tipagem fraca e portabilidade. O sítio de comércio eletrônico foi escrito utilizando PHP versão 4. O arquivo de instalação do PHP versão 4 é facilmente encontrado no sítio (www.php.net).

O osCommerce apresenta todas as funcionalidades de um sítio de comércio eletrônico e por se tratar de um projeto código aberto, pode ser facilmente adaptado para qualquer necessidade. A funcionalidade de recomendação é inerente a grandes sítios de comércio eletrônico e a adaptação da recomendação inteligente de produtos para o osCommerce é uma contribuição respeitável para essa ferramenta de código livre.

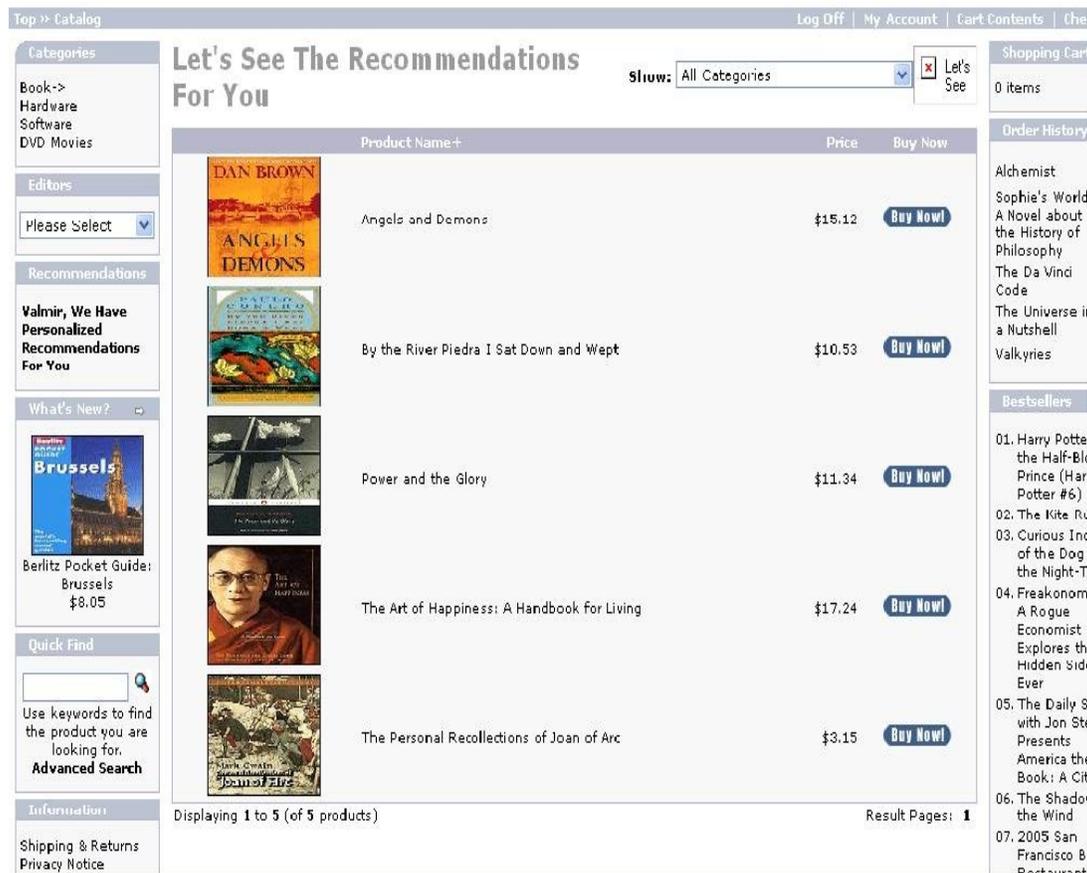
A recomendação é ativada através de um botão na interface do sítio de comércio eletrônico por um usuário cadastrado e ativo no sítio. A Figura 6 exibe a interface do sítio.



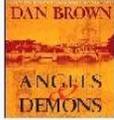
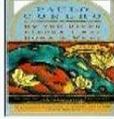
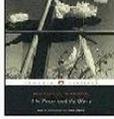
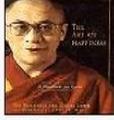
The screenshot shows a web page with a top navigation bar containing 'Top >> Catalog', 'Log Off', 'My Account', 'Cart Contents', and 'Che...'. On the left, there are sections for 'Categories' (Book, Hardware, Software, DVD Movies), 'Editors' (Please Select), and 'Recommendations' (highlighted with a red circle and containing the text 'Valmir, We Have Personalized Recommendations For You'). Below that is a 'Quick Find' search box. The main content area is titled 'What's New Here?' and includes a welcome message: 'Welcome back Valmir! Would you like to see which new products are available to purchase?'. Below this, it states 'This is a e-Recommender application. This shop is running on a bookstore 0.1'. The 'New Products For November' section displays a grid of book covers with titles and prices: 'The Investor's Notebook: A Patent It Yourself Companion' (\$22.49), 'Complete Idiot's Guide to Music Theory' (\$16.15), 'The Successful Investor: What 80 Million People Need to Know to Invest Profitably and Avoid Big Losses' (\$8.86), 'Taming the Scotsman' (\$6.29), 'A Man of My Words: Reflections on the English Language' (\$12.55), 'Phil Gordon's Little Green Book: Lessons and Teachings in No Lim' (\$15.12), 'The Death of Artemio Cruz' (\$13.50), 'Everything Writing Poetry Book: A Practical Guide to Style, Stru' (\$13.45), and 'The InVision Guide to a Healthy Heart' (\$16.15). On the right, there are sections for 'Shopping Cart' (0 items), 'Order History', a list of authors (Alchemist, Sophie's World, etc.), and 'Bestsellers' (Harry Potter, etc.).

Figura 6. Mostra *link* para usuário requisitar recomendação

A apresentação das recomendações geradas pelo e-Recommender são apresentadas de forma ordenada no sítio. O usuário facilmente obtém informações acerca do produto que é recomendado pela ferramenta, podendo ainda escolher comprar algum item facilmente. A Figura 7 exibe a recomendação gerada para o usuário com perfil do exemplo utilizado na seção 4.5.2, os produtos contidos em seu perfil são expostos na Tabela 5.



The screenshot shows a web page titled "Let's See The Recommendations For You". The main content is a table of recommended books:

Product Name+	Price	Buy Now
 Angels and Demons	\$15.12	Buy Now!
 By the River Piedra I Sat Down and Wept	\$10.53	Buy Now!
 Power and the Glory	\$11.34	Buy Now!
 The Art of Happiness: A Handbook for Living	\$17.24	Buy Now!
 The Personal Recollections of Joan of Arc	\$3.15	Buy Now!

Additional elements in the interface include a left sidebar with navigation menus (Categories, Editors, Recommendations, What's New?, Quick Find, Information), a top navigation bar with links (Log Off, My Account, Cart Contents, Che), and a right sidebar with sections like Shopping Cart (0 items), Order History, and Bestsellers (listing books like Harry Potter, The Kite Runner, etc.).

Figura 7. Recomendações personalizadas

4.4 Preparação da Base de Dados

A fim de ilustrar a preparação da base de dados em nossa ferramenta uma loja de livros foi selecionada para esse fim. A base original só possuía o ISBN (*Internacional Standard Book Number*), um código do usuário e a data que a compra foi realizada. Como a ferramenta possui um caráter híbrido, colaborativo e baseado em conteúdo, e além de informações de quais livros os usuários compraram, ela também utiliza como entrada informações das características dos livros. Para complementar a base, um *crawler* [51] teve que ser desenvolvido.

O *crawler* produzido nesse trabalho é um programa que visita sítios de informações públicas de livros na internet, por exemplo, o www.barnesandnoble.com, procurando por livros a partir do ISBN. Quando informações sobre o livro são encontradas, elas são armazenadas numa nova base, adicionando assim novas características a base de dados inicial para que a ferramenta seja capaz de gerar boas recomendações utilizando essas informações. As informações acrescentadas são autores, categorias, resumo, avaliações, ano, preço, páginas, editor e livros que foram comprados em conjunto com o livro encontrado.

O *crawler* faz parte do pacote da ferramenta final ao usuário, pois pode ser estendido para acrescentar informações de qualquer produto que se queira recomendar e não possua informações suficientes sobre o produto.

4.5 Perfil do Usuário

O perfil do usuário é a informação mais importante que a ferramenta utiliza como base para fazer o cálculo da recomendação. A representação do perfil do usuário é um conjunto de descrições simbólicas modais dos produtos que o mesmo comprou na loja eletrônica. A descrição modal dos produtos envolve os atributos e os pesos desses atributos para o usuário. A construção do perfil do usuário envolve duas etapas: (i) pré-processamento e (ii) generalização.

4.5.1 Pré-processamento

O pré-processamento envolve a construção de conjuntos de descrições simbólicas modais para cada um dos produtos que fazem parte do histórico de compras para cada usuário. Os atributos são descritos através do seu nome e peso que esse atributo representa no produto. A descrição simbólica y_i de um item i é representada por um subconjunto que está contido no domínio de uma variável categórica D_j , onde $x_i = (X_i^p, \dots, X_i^p)$ com $X_i^j = X_i$ ($i \subseteq D_j$, com i variando no intervalo de 1 até n , enquanto X_j ($j=1, \dots, p$) representa uma variável categórica simples ou multivariada. A Equação (4) retirada de BEZERRA [14] representa o peso $w(m)$ de cada categoria $m \in X_j^i$. Na equação (4) $|X_j^i|$ representa a cardinalidade de X_j^i :

$$w(m) = \frac{1}{|X_j^i|}, \text{ se } y_j \text{ é qualitativo simples ou multivalorado.} \quad \text{Equação (4)}$$

Tabela 4. Descrição simbólica dos tributos do domínio livro

Atributo	Livro
Título	Da Vinci Code
Ano	{2001=1}
Páginas	{464=1}
Preço	{15.71=1}
Editor	{Doubleday Publishing=1}
Autor(s)	{Dan Brown=1}
Categoria(s)	{Religion Beliefs - Fiction = 0.3333333, Arts Entertainment – Fiction=0.333333, Settings Atmosphere - Fiction=0.3333}
Livros comprados	0

Na Tabela 4 um livro é representado simbolicamente na etapa de pré-processamento, o atributo título não recebe nenhum peso porque não é considerado no cálculo da recomendação. Para os demais atributos são associados números que variam de 0 a 1 que representam os pesos inerentes a variável naquele atributo. Nos atributos simples, esses pesos serão sempre 1; nos atributos que não possuem valores, o valor associado será 0 e nos atributos multivalorados, os pesos serão divididos pelo número de variáveis presente no atributo, representado por X_i^j na equação (4).

4.5.2 Generalização

A generalização transforma o perfil do usuário numa descrição simbólica, formando pares de características de produtos e seus pesos na importância da formação do perfil do usuário, englobando todos os produtos do perfil do usuário, ou seja, aqueles que o usuário comprou.

Para melhor entendermos como a generalização funciona, vamos introduzir um exemplo resumido de um perfil de usuário. Um usuário que possua os livros apresentados na Tabela 5 em conjunto com seus respectivos autores, após a generalização para esse atributo o perfil simbólico desse usuário é apresentado na Tabela 6.

Tabela 5. Livros no perfil do usuário

Título	Autor
Código Da Vinci	Dan Brown
Anjos e Demônios	Dan Brown
O Alquimista	Paulo Coelho
As Valquírias	Paulo Coelho
O Mundo de Sofia	Jostein Gaarder e Paulette Moller
Universo Numa Casca de Noz	Stephen Hawking

Note que a generalização utiliza a frequência dos itens em cada variável do perfil. O nome dos autores Dan Brown e Paulo Coelho, por aparecerem em dois livros, possui um peso duas vezes maior que os demais autores, que aparecem apenas uma vez. A generalização é realizada para todos os atributos do livro, exceto título. Para simplificar, apenas o atributo autor é apresentado na Tabela 6.

Tabela 6. Perfil modal simbólico de um usuário

Atributo	Perfil Simbólico
Autor(s)	{Dan Brown = 0.286, Jostein Gaarder = 0.143, Paulette Moller = 0.143, Paulo Coelho = 0.286, Stephen Hawking=0.143}

As equações utilizadas para fazer o cálculo da generalização foram retiradas de BEZERRA [14]. A descrição simbólica modal do perfil do usuário u é representada por $(S_j(u), q_j(u))$ com $S_j(u)$ representando o suporte, ou seja, a descrição dos atributos do produto e $q_j(u)$ representando uma distribuição de pesos do suporte com j variando no

intervalo de 1 até n . Seja $m \in S_j(u)$ que é uma categoria pertencente ao domínio D_j , o peso $W(m) \in q_j(u)$ da categoria é dado pelas equações (5) e (6):

$$W(m) = \frac{1}{|u|} \sum_{i \in u} \delta(i, m) \quad \text{Equação (5)}$$

$$\delta(i, m) = \begin{cases} w(m) \in q_j(i), & \text{se } m \in S_j(i) \\ 0, & \text{cc} \end{cases}$$

Equação (6)

4.6 Recomendando Produtos

Na recomendação de produtos é gerada uma lista de produtos que é apresentada ao usuário. Essa lista pode ser de tamanho variável. A lista é ordenada de acordo com a similaridade que a ferramenta estabelece ser mais relevante para o usuário. Para cada usuário é realizado um cálculo de similaridade entre o produto e o perfil do usuário sendo gerado um número que indica a similaridade do produto com o usuário, é com esse número que quanto maior, mais similar o produto é do perfil do usuário que a ferramenta ordena a lista de produtos.

O primeiro passo para recomendação de produtos é calcular a similaridade entre usuários para formar um grupo onde os usuários mais próximos possuam maior similaridade entre si. Para realizar esse cálculo é utilizada uma função de similaridade para variáveis simbólicas modais. O valor dessa função aumenta de acordo com interseções entre as descrições simbólicas modais dos perfis dos usuários. As interseções são itens iguais que estão contidos no perfil do usuário ativo e também do usuário que é comparado. O peso associado a cada variável é levado em conta, aumentando proporcionalmente o valor da similaridade de acordo com este peso. A similaridade de um usuário é calculada com todos os outros clientes na loja virtual. A ferramenta encontra os K usuários mais próximos, onde K pode variar de 1 ao número de usuários da loja virtual.

Sendo $y_{uk} = (Y_{uk}^1, \dots, Y_{uk}^p)$ e $y_{vk} = (Y_{vk}^1, \dots, Y_{vk}^p)$ as descrições simbólicas modais de dois usuários u_k e v_k respectivamente, o usuário v_k é um aspirante a vizinho do usuário ativo u_k . As equações de (7) a (10) retiradas de BEZZERRA [14], descrevem formalmente a função que computa a similaridade entre dois usuários:

$$\omega(u, v) = \sum_{k=1}^k \delta(k) \quad \text{Equação (7)}$$

onde,

$$\delta(k) = \begin{cases} 0, & \text{Se } Y_{uk} \text{ (ou } Y_{vk}) \notin \text{ ao perfil do usuário } u \text{ (ou } v) \\ w_k ((1 - \phi(Y_{uk}, Y_{vk})), cc & \end{cases} \quad \text{Equação (8)}$$

A função de similaridade é composta de dois componentes, o primeiro compara os conjuntos de suporte e o segundo compara as distribuições de peso:

$$\phi(y_{uk}, y_{vk}) = \frac{1}{p} \sum_{j=1}^p \frac{1}{2} [\phi_{cf}(S_j(u_k), S_j(v_k)) + \phi_{cd}(q_j(u_k), q_j(v_k))] \quad \text{Equação (9)}$$

onde ϕ_{cf} mede as diferenças em posições no caso em que conjuntos $S_j(u_k)$ e $S_j(v_k)$ são ordenados e ϕ_{cd} mede as diferenças entre o conteúdo entre y_{uk} e y_{vk} . A variável p representa o número de itens comparados nos perfis dos usuários. A função de similaridade varia entre 0 e 1.

O componente ϕ_{cd} é definido como:

$$\phi_{cd}(q_j(u_k), q_j(v_k)) = \frac{1}{2} \left(\frac{\gamma + \delta}{\alpha + \gamma + \delta} \right) + \left(\frac{\gamma + \delta}{\beta + \gamma + \delta} \right) \quad \text{Equação (10)}$$

Os componentes na equação (10) são definidos com base da interseção ou não de suporte dos atributos do perfil do usuário alvo com o usuário comparado. Os tópicos a até d apresentam formas de medir acordos (α e β) e desacordos (δ e γ) entre as distribuições de pesos $q_j(u_k)$ e $q_j(v_k)$.

a) Quando o suporte existe nos perfil dos dois usuários comparados:

$$\alpha = \sum_{m \in S_j(u_k) \cap S_j(v_k)} W(m) \in q_j(u_k)$$

$$\beta = \sum_{m \in S_j(u_k) \cap S_j(v_k)} W(m) \in q_j(v_k)$$

b) Quando o suporte existe no usuário ativo e não existe no usuário alvo:

$$\delta = \sum_{m \in S_j(u_k) \cap \overline{S_j(v_k)}} W(m) \in q_j(u_k)$$

c) Quando o suporte não existe no usuário ativo e existe no usuário alvo:

$$\gamma = \sum_{m \in \overline{S_j(u_k)} \cap S_j(v_k)} W(m) \in q_j(v_k)$$

Para variáveis categóricas ordenadas é utilizada a fórmula definida por Ichino e Yaguchi [52], onde $m_l = \min(S_j(v))$, $m_u = \max(S_j(v))$, $c_l = \min(S_j(u))$ e $c_u = \max(S_j(u))$. A operação de junção é dada por:

$$S_j(v) \oplus S_j(u) = \begin{cases} S_j(v) \cup S_j(u), & \text{se } D_j \text{ é não ordenado} \\ \{\min(m_l, c_l), \max(m_u, c_u)\}, & \text{cc} \end{cases} \quad \text{Equação (11)}$$

Então, o componente livre de contexto ϕ_{cf} é definido como:

$$\phi_{cf}(S_j(u_k), S_j(v_k)) = \begin{cases} 0, & \text{se } (S_j(u_k) \cap S_j(v_k)) \neq \Phi \\ \frac{|S_j(u_k) \oplus S_j(v_k)| - |S_j(u_k)| - |S_j(v_k)|}{|S_j(u_k) \oplus S_j(v_k)|}, & \text{cc} \end{cases} \quad \text{Equação (12)}$$

Após calcular os usuários mais próximos do usuário ativo, é realizada uma seleção dos produtos a serem recomendados. A fase de pré-processamento explicada na seção 4.1.2 é aplicada para todos os produtos presentes nos perfis desses usuários vizinhos, esse produto é chamado de g_t . A equação (13) realiza o cálculo da relevância de cada produto g_t para o usuário ativo. Os produtos que alcançarem uma relevância mais alta são ordenados em uma lista e logo em seguida, armazenados num banco de dados para serem apresentados ao usuário.

$$\rho(u, g) = \sum_1^t \omega(u, g) \quad \text{Equação (13)}$$

Cada vez que um usuário realizar uma nova compra no sítio de comércio eletrônico, uma nova recomendação pode ser gerada. Desse modo, o sistema irá produzir recomendações com base em informações recentes acerca desse usuário.

No próximo capítulo serão apresentados experimentos realizados com o e-Recommendier, para descobrir os parâmetros padrões da ferramenta. Os resultados serão apresentados logo em seguida.

Capítulo 5

e-Recommender: Experimentos e Avaliações

Neste capítulo, apresentamos os experimentos realizados, que consistiram na realização de testes variando os parâmetros utilizados no algoritmo de análise simbólica.

Para validar a ferramenta, apresentamos resultados de questionários respondidos por alguns usuários que avaliaram as diferentes listas de recomendações geradas pela variação desses parâmetros.

5.1 Experimentos e Validação

Os testes foram realizados com a finalidade de descobrir a eficácia da ferramenta e também validar a ferramenta. Pelo fato de o algoritmo utilizado ser não supervisionado [31], não cabe um teste que compare o resultado obtido por outro resultado esperado, já que não existe resultado esperado. Os testes desse tipo de algoritmos são empíricos, principalmente se tratando de sugestões pessoais. A maneira escolhida para validar a ferramenta foi a entrevista de pessoas para saber a opinião delas sobre as listas de recomendações personalizadas.

Os testes realizados variaram os pesos de cada atributo do produto na ferramenta, que no caso, foram livros. Assim, foram testadas variações de pesos sobre esses atributos que geraram resultados diferentes. Utilizando estes resultados foram realizadas entrevistas com 25 usuários que colaboraram com um *feedback* da experiência deles com a ferramenta respondendo a um formulário. De acordo com esse *feedback*, os pesos dos atributos foram convergindo. Esses valores se tornaram o padrão da ferramenta para esse produto.

5.1.1 Metodologia dos Experimentos

A primeira etapa foi identificar parâmetros que poderiam influenciar na resposta da ferramenta. Os parâmetros identificados são os pesos associados a cada atributo do livro, utilizados para o cálculo da recomendação na equação (6) do capítulo 4. As variações desses valores acarretam em diferentes respostas do sistema. Esses pesos foram escolhidos de forma aleatória e são apresentadas na Tabela 7.

Tabela 7. Pesos de cada atributo do produto livro

Atributos	Teste 1	Teste 2	Teste 3
Ano	0.05	0	0.05
Páginas	0.025	0	0.05
Preço	0.075	0.05	0.1
Editor	0.15	0.1	0.2
Autor	0.3	0.35	0.2
Categoria	0.225	0.3	0.2
Livros Comprados Juntos	0.175	0.2	0.2

Uma lista de recomendação foi gerada para 25 usuários para cada um dos três testes. As respostas foram apresentadas aos usuários que responderam ao formulário do anexo A tendo como objetivo avaliar questões de qualidade, robustez, usabilidade e velocidade do e-Recommend. A partir da análise do resultado do formulário, os valores que obtiveram melhores resultados foram escolhidos como o padrão para a ferramenta.

O Formulário é composto por sete questões. O sítio de comércio eletrônico deve possuir uma interface amigável para o usuário realizar compras. A primeira pergunta do formulário analisa esse ponto, questionando o usuário qual a avaliação para a interface do sítio de comércio eletrônico.

A pergunta 2 analisa a velocidade de processamento da ferramenta, questionando que nota o usuário atribui a velocidade da recomendação. Vale lembrar, que na arquitetura do e-Recommend uma *thread* monitora a realização de novas compras do sítio de comércio eletrônico. Então, em poucos segundos, o usuário possui uma lista recomendações a seu dispor.

A pergunta chave para a avaliação da qualidade da lista de recomendação é a terceira, a qual pergunta quantos livros o usuário compraria da lista de itens apresentada para ele. Quanto mais produtos apontados pelo usuário, melhores os resultados, pois o objetivo da recomendação é convencer o usuário a consumir algum item da lista.

A pergunta 4 questiona quantos livros o usuário não compraria, deste modo avalia a quantidade de resultados negativos, que poderia melhorar.

Levando em consideração que alguns livros, o usuário fica em dúvida na compra, a pergunta 5 serve o usuário atribuir uma nota, avaliando todos os itens da lista. Na pergunta 5 o usuário atribui uma nota para avaliar quantitativamente a lista de recomendação.

Pelo fato da base de dados ser americana e a grande maioria dos itens nela ser estrangeiro, inclusive no idioma inglês, achou-se importante avaliar se o usuário conhecia o livro indicado, por isso foram introduzidas as perguntas 6 e 7. A pergunta 6 questiona

quantos livros o usuário já leu e a pergunta 7 questiona quantos livros o usuário conhecia de alguma forma.

5.1.2 Realização dos Experimentos

As variações dos parâmetros, apresentados na Tabela 7 da seção anterior, acarretaram mudanças sensíveis no resultado das recomendações. Os resultados foram avaliados por 25 usuários que responderam ao formulário do anexo A.

Para ilustrar a variação dos resultados, são apresentadas listas geradas para cada um dos testes realizados para dois usuários. O usuário 1, possui em seu perfil os livros: “O Código Da Vinci”, “O Universo numa Casca de Noz”, “O Mundo de Sofia”, “O Alquimista” e “Só o Amor é Real”. As recomendações para esse perfil são exibidas na Tabela 8.

Tabela 8. Recomendações geradas a partir de três configurações dos parâmetros para o Usuário 1

Teste 1	Teste 2	Teste 3
Alquimista: Uma Fábula Sobre Seguir Seus sonhos	Alquimista: Uma fábula sobre seguir seus sonhos	Alquimista: Uma fábula sobre seguir seus sonhos
Muitas Vidas, Muitos Mestres	Realidade	Sobre o Rio Piedra Sentei e Chorei
Anjos e Demônios	Muitas Vidas, Muitos Mestres	A Arte da Felicidade
Harry Potter e a Ordem do Fênix	O Templo Interior da Bruxaria	Poder e Glória
The Monks of New Skete: How to Be Your Dog's Best Friend.	A Arte da Felicidade	Anjos e Demônios

Os títulos dos livros apresentados ao usuário nas 3 listas contêm valores em comum, mas ainda sim, livros diferentes aparecem em apenas uma das listas. Esse fato se repetiu em outras listas geradas com preferências de usuários diferentes. Os livros estão sendo apresentados na ordem de decrescente de similaridade com o usuário, muitas vezes essa ordem também varia de uma lista de recomendações para outra.

Outro exemplo apresentado na Tabela 9 é para um usuário que possui em seu perfil os livros: “Ensaio sobre a cegueira”, “O guia dos Mochileiros da Galáxia”, “Crime e Castigo”, “A Metamorfose” e “Memórias das Minhas Putas Tristes”. Note que do mesmo modo da lista da Tabela 8, livros em comum e diferentes são apresentados nas listas, ainda há a mudança na ordem de similaridade do livro com o usuário.

Tabela 9. Recomendações geradas a partir de três configurações dos parâmetros para o Usuário 2

Teste 1	Teste 2	Teste 3
A queda de Neskaya	O Idiota	O Idiota
Dragon's Kin	Anjos e Demônios	O Restaurante no Fim do Universo
O Restaurante no Fim do Universo	As Valquírias	O Mundo de Sofia
Quicksilver: Volume um do Ciclo Barroco	A Arte da Felicidade	Don Quixote
Discworld: Volume 1	Histórias com Matisse	Amor em Tempos e Ódio

A partir daqui, são apresentados gráficos que ilustram as respostas a perguntas do formulário do anexo A. O objetivo da realização dos testes é a avaliação do algoritmo, então aqui nós apenas apresentamos os gráficos das perguntas que nos auxiliam a tirar as conclusões necessárias para escolher os valores padrão da ferramenta, essas perguntas são a 2 e a 3:

- Pergunta 2: O que você achou da velocidade que foi gerada a recomendação?

A Avaliação da velocidade do algoritmo foi avaliada utilizando as graduações a seguir: “rápido” que na figura é representado por 4; “normal” que é representado pelo valor 3; “pouco lento” que é representado pelo valor 2; “lento” que é representado pelo valor 1. A Figura 8 apresenta os resultados.

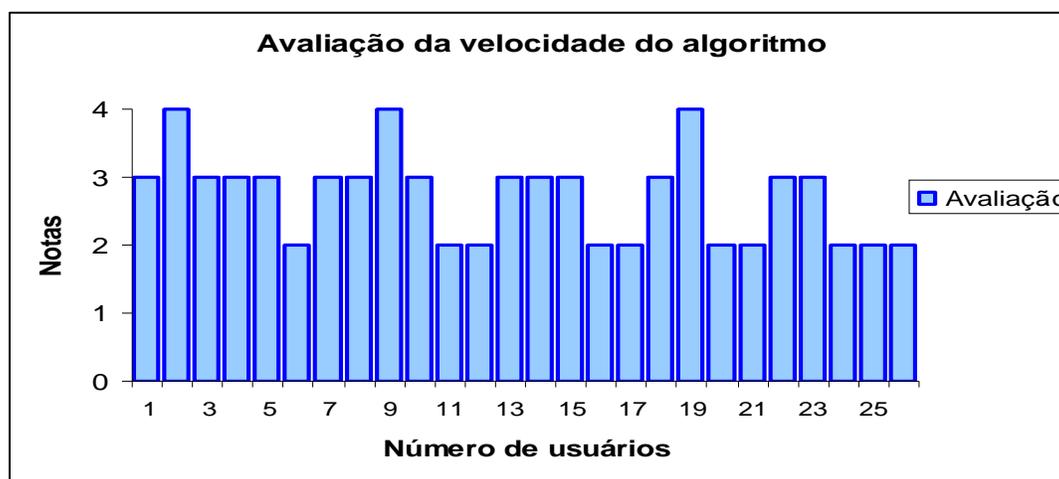


Figura 8. Velocidade do Algoritmo

Alguns usuários acharam um pouco lenta a geração da lista de recomendação, isso acontece porque o número de comparações realizadas é muito grande, demorando em média 10 segundos para uma recomendação ser gerada. Vale ressaltar que esses segundos

são apenas percebidos para aquele usuário que, logo após realizar a compra, requerer uma recomendação.

- Pergunta 3: Quantos livros dessa lista você compraria?

A pergunta de quantos livros o usuário compraria é a mais importante do formulário para avaliar as recomendações, pois o objetivo do algoritmo é antever compras que o usuário desejaria realizar. Na Figura 9 é apresentada a quantidade de livros que o usuário compraria, variando de 0 ao número de livros presentes na lista, 5.

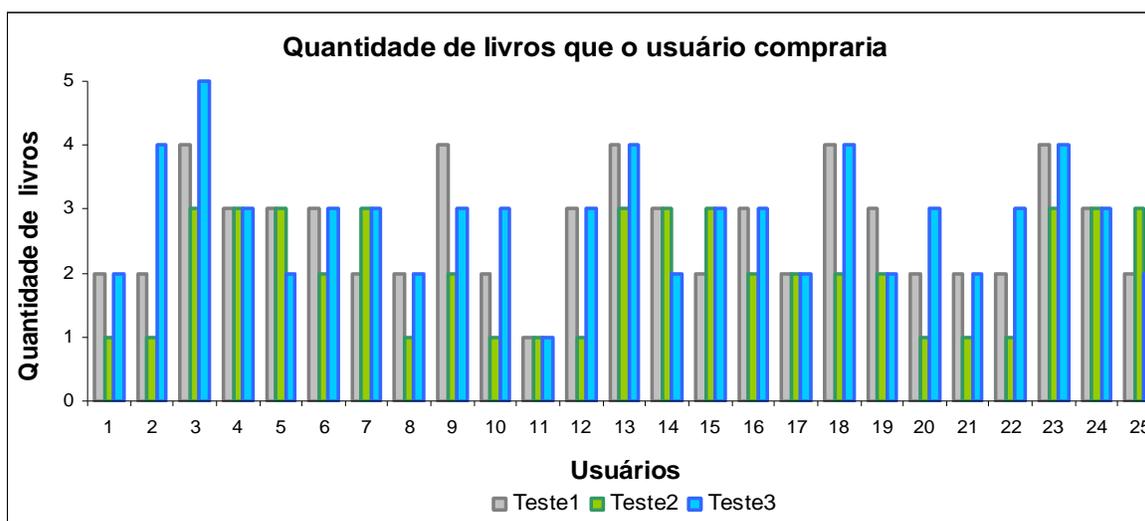


Figura 9. Quantidade de Livros que usuário Compraria

Na Figura 9, os valores para o Teste 3 e para o Teste 1 demonstram um equilíbrio. Para identificar o melhor dentre os testes, a média e o desvio padrão foram calculados e apresentados na Tabela 10.

Tabela 10. Média e desvio padrão dos resultados da pergunta 3

Teste	Média	Desvio Padrão
Teste 1	2.68	0.852
Teste 2	2.04	0.889
Teste 3	2.84	0.898
Média Geral	2.52	0.879

Percebe-se nas tabela 8 e 9 que as listas geradas variaram sensivelmente. Essa sensível diferença foi percebida nos resultados dos formulários que apresentassem um ligeiro equilíbrio, principalmente entre as listas apresentadas no Teste 1 e no Teste 3. Contudo, a partir dos resultados apresentados na Tabela 10, o Teste 3 foi o que apresentou melhores resultados.

Capítulo 6

Conclusão e Trabalhos Futuros

6.1 Conclusão

Com o crescimento do comércio eletrônico e a necessidade de clientes fiéis a esses sítios, a personalização vêm se tornando uma das características mais importantes do comércio eletrônico. Esse trabalho contribui com uma ferramenta inteligente que é capaz de recomendar produtos baseados num histórico de compras do usuário, sendo uma opção para sítios de comércio eletrônicos escritos em qualquer linguagem de programação.

A ferramenta implementa uma nova técnica proposta por BEZERRA [14], se trata da abordagem simbólica, ideal para lidar com problemas de análise de dados complexos. A técnica utilizada mostrou-se bastante pertinente, já que a avaliação de 25 pessoas aponta que em 100% dos casos há pelo menos um item desejado na lista de recomendações.

Para elucidar o emprego da ferramenta, um sítio real de comércio eletrônico foi preparado para trabalhar com recomendações. A ferramenta foi conectada a este sítio com muito sucesso. A arquitetura *blackboard* permitiu a comunicação do sítio e da ferramenta através do banco de dados. Através dessa arquitetura, a ferramenta pode ser acoplada a qualquer sítio de comércio eletrônico que possua acesso a banco de dados.

Este trabalho contribui com uma abordagem de alta qualidade para sistemas de recomendação, abordando conceitos atuais e importantes de sistemas de recomendação. A ferramenta, fruto deste trabalho, adiciona personalização para sítios de comércio eletrônico, uma grande vantagem competitiva que possibilita tornar seu cliente fiel. Além disso, proporciona recomendações de alta qualidade para que o usuário encontre itens de seu interesse e ou necessidade.

6.2 Discussão

A utilização desta ferramenta por um sítio de comércio eletrônico conseguirá alcançar os resultados esperados pela personalização, que é tornar seu cliente fiel. O usuário será beneficiado, pois itens de seu interesse serão recomendados pela ferramenta.

Foram encontradas algumas dificuldades no desenvolvimento deste trabalho. A principal delas foi gerada pela base de dados de comércio eletrônico. A base de dados, por conter muitos itens, cerca de 41000, é muito esparsa. A quantidade de usuários é da ordem de 11000. Deste modo, os livros presentes na base foram consumidos geralmente por poucos usuários. Esse problema é chamado de esparsidade dos dados. Encontrar um algoritmo que pudesse gerar boas recomendações foi uma tarefa difícil.

Com relação à base de dados, por não ser uma base de livros nacionais, muitas vezes o consumidor brasileiro, que foi alvo dos testes, nunca ouviu falar naquele livro ou autor. Muitos dos livros da base sequer chegam ao mercado brasileiro e isto foi uma das dificuldades encontradas durante a avaliação pelos usuários,

O algoritmo escolhido foi o simbólico, que possui uma função de similaridade muito complexa, difícil de entender e implementar numa linguagem computacional. A confecção da ferramenta em si foi complicada, pois foi pensada para ser uma ferramenta portátil para qualquer base de dados e, além disso, integrável com qualquer sítio de comércio eletrônico.

Outra dificuldade encontrada foi preparação do sítio osCommerce para operar com recomendações. O sítio é escrito em PHP, linguagem que o autor não dominava.

Pra finalizar, uma dificuldade foi encontrada pelos usuários durante o teste da ferramenta. Pela linguagem da base ser inglês, nem todos entendiam o título e a sinopse dos livros. Muitas vezes, recomendou-se títulos que traduzidos não coincidiam com seus títulos em português, além de ser uma base americana, de cultura diferente, contendo alguns livros desconhecidos para o público brasileiro.

6.3 Trabalhos Futuros

A ferramenta, pelo modo que foi construída, permite a adição de novas funcionalidades. Podem ser adicionadas novas técnicas de inteligência artificial com a finalidade de produzir melhores resultados, de acordo com o domínio do produto.

Outra funcionalidade a ser acrescentada é uma tela de configuração para o acréscimo de novas bases, de formatos e tipos diferentes. Esta funcionalidade aumentaria os tipos de produtos que a ferramenta seria capaz de fornecer recomendações. O protótipo construído neste trabalho não concluiu esta funcionalidade por não estar no escopo da mesma.

Avaliações explícitas não são consideradas nessa ferramenta, pois as avaliações implícitas são mais importantes no domínio de comércio eletrônico. Facilmente esse tipo de avaliação pode ser acrescentado na ferramenta.

Pra complementar o poder da ferramenta, mais opções de avaliações implícitas podem ser levadas em conta no cálculo da recomendação. Como exemplo, temos o tempo

que o usuário passa visitando certo produto, número de vezes que usuário visita certo produto, entre outros.

Outra proposta é a criação de técnicas que modelem o perfil do usuário com base em seu perfil psicológico, utilizando a psicologia para ajudar a construção de perfis dessa natureza.

Bibliografia

- [1] TURBAN, E.; KING, D. Comércio Eletrônico: estratégia e gestão. São Paulo, Prentice Hall, 2004.
- [2] KENDALL, J.E.; KENDALL, K.E. Information Delivery Systems: An Exploration of Web Pull and Push Technologies. Communications of AIS, Vol. 1, Article 14, 1999.
- [3] RIZZO, E. Uma análise comparativa entre o marketing de massa e o “One to One” Marketing, no cenário de empresas competitivas. Disponível em: <<http://www.bocc.ubi.pt/pag/rizzo-esmeralda-analise-comparativa.pdf>> Acesso em: 03/10/2006.
- [4] INFO EXAME: revista mensal de informática. São Paulo, Ed. Abril, n. 245, Agosto. 2006. 114 p.
- [5] ADOMAVICIUS, G.; TUZHILIN, A. 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, Vol. 17, n. 6, 2005.
- [6] BURKE, R. Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction, Vol. 12, N. 4, p. 331-370, 2002.
- [7] REATEGUI, E. B.; CAZELLA, S. B. Sistemas de Recomendação. Mini-Curso, Enia, 2005.
- [8] RESNICK, P.; VARIAN, H. R. Recommender Systems. Communications of the ACM, New York, v.40, n.3, pp. 55-58, 1997.
- [9] TERVEEN, L.; HILL, W. Beyond Recommender Systems: Helping people help each other. HCI in the millennium, Addison Wesley, 2001.
- [10] KONSTAN, J. A. Introduction To Recommender Systems: Algorithms and Evaluation. ACM Transactions in Information Systems, Vol. 22, N. 1, 2004.
- [11] BEZERRA, B.L.D.; De CARVALHO, F.A.T. A Symbolic Approach for Content-Based Information Filtering. Information Processing Letters, Amsterdam, V.25, n. 8, p. 911- 921, 2004.
- [12] BEZERRA, B.L.D.; De CARVALHO, F.A.T.; MACÁRIO, V. F. C²:: A Collaborative Recommendation System Based on Modal Symbolic User Profile. IEEE/ACM International Conference on Web Intelligence, 2006.
- [13] GOLDBERG, D., et al. Using collaborative filtering to weave an information Tapestry. Communications of the ACM, New York, v.35, n.12, p. 61-70, 1992.

- [14] BEZERRA, B. L. D.; De CARVALHO, F. de A. T.. A Symbolic Hybrid Approach to Face the New User Problem in Recommender Systems. In: Australian Joint Conference on Artificial Intelligence - AI2004, Lectures Notes on Artificial Intelligence. Springer-Verlag, 2004.
- [15] CLAYPOOL, M., et al. Combining content-based and collaborative filters in a online newspaper. ACM SIGIR Workshop on Recommender Systems, 1999.
- [16] BEZERRA, B.L.D ; CARVALHO, F.A.T. Maximizando o Conhecimento Sobre o Usuário com o Mínimo de Informação Disponibilizada. Enia, 2005.
- [17] NETO, B. O Marketing Direto para o próximo século. Disponível em: <http://www.mktdireto.com.br/MAT%C9RIAS/materia19.html>> Acesso em: 03/10/2006;
- [18] TORRES, R. Personalização na Internet. São Paulo, Novatec, 2004.
- [19] RAPP, S; COLLINS, T. 5º geração do marketing: maximarketing II. São Paulo: Makron Books, 1991. Maximarketing. São Paulo: Mc Graw-Hill, 1988.
- [20] E-Loyalty Resource, Disponível em: <<http://www.e-loyaltyresource.com/>> Acesso em: 12/12/2005.
- [21] ENDO, H.; NOTO, M. A Word-Of-Mouth Information Recommender System Considering Information Reliability and User Preferences. Proceedings of CHI 95, 1995.
- [22] RUSSEL, S.; PETER, N. Inteligência Artificial. São Paulo, Campus, 2004.
- [23] SCHAFER, J.B., et al. E-Commerce recommendation Applications. Data Mining and Knowledge Discovery, Vol. 5, Nº 1-2. p. 115-153, 2001.
- [24] ARAÚJO, J. Dominando os cookies. Rio de Janeiro: Ciência Moderna, 2003.
- [25] CLAYPOOL, M.; LE, P.; WASEDA, M. Implicit interest indicators. Internacional Conference on intelligent User Interfaces, 2001.
- [26] OARD, D.W.; KIM, J. Implicit Feedback for Recommender Systems. Madison: Proceedings of the AAAI Workshop on Recommender Systems, p. 81-83, 1998.
- [27] STUART E.M., et al. Capturing knowledge of user preferences: ontologies in recommender systems. International Conference On Knowledge Capture, p. 100-107, 2001.
- [28] SHAHABI, C.; SHEN, Y. An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback. Distributed and Parallel Databases, Vol. 14, p.173–192, 2003.
- [29] MOVIELENS. Disponível em: <<http://www.movielen.umn.edu>> Acesso em: 19/11/2006.
- [30] RESNICK, P., et al. GroupLens: Na open achitecture for collaborative filtering od netnews. ACM Workshop on recommender systems, 2001.
- [31] DUDA, R.O.; HART, P.E.; STORK, D.G. Pattern Classification. John Wiley & Sons, 2001.
- [32] HERLOCKER, J., et al. An algorithmic framework for performing collaborative filtering. ACM SIGIR Conference on Research and Development in Information Retrieval. Berkeley, 1999.
- [33] LOPES, I.L. Estratégia de busca na recuperação da informação: revisão da literatura. Ci. Inf., vol.31, no.2, p.60-71, 2002.

- [34] BELKIN, J. N.; CROFT, B. W. Information Retrieval and Information Filtering: Two sides of the same Coin?. Communications of the ACM, Vol. 35, N. 12, 1992.
- [35] SALTON, G.; BUCKLEY, C. Term weighting approaches in automatic text retrieval. Information Processing and Management, Vol. 5, p. 513-523, 1988.
- [36] PAZZANI, M.J. A framework for collaborative, content-based and demographic filtering. Artificial Intelligencet Review, Vol. 13, p. 393-408, 1999.
- [37] BALABANOVIC, M.; SHOHAM, Y. Fab: Content-based, collaborative recommendation. Communication of the ACM, Vol. 40, p. 66-72, 1997.
- [38] TORRES, R., et al. Enhancing Digital Libraries With TechLens. Internacional Conference on Digital Libraries, 2004.
- [39] DIDAY, E. Une représentation visuelle des classes empiétantes: les pyramides. INRIA, Rapport de Recherche N.291, 1984.
- [40] LEBART,L.;MORINEAU,A.;D'AMBRA, L. Statistique explanatoire multidimensionnelle. Dunod, Paris, 1995.
- [41] SNEATH, P.H.A.; SOKAL, R.R. numerical taxonomy. Freeman, Sam Francisco, 1973.
- [42] BOCK, H.H.; DIDAY, E. Analysis of Symbolic Data, Springer, Berlim, 2000.
- [43] DIDAY, E. A introduction to Symbolic Data Analysis and the Sodas Software. Journal of Data Analysis, 2002.
- [44] DEITEL, H.M.; DEITEL, P.J. Java: Como Programar. Porto Alegre: Bookman, 2002.
- [45] MYSQL. Disponível em: <<http://www.mysql.com>> Acesso em: 09/12/2005.
- [46] ENGELMORE, R.; MORGAN, T. Black Board Systems. Addison Wesley, 1988.
- [47] OSCOMMERCE. Disponível em: <www.oscommerce.org> Acesso em: 12/11/2006.
- [48] GUESSER, A.H. Software Livre & Controvérsias Científicas. Curitiba, Juruá, 2006.
- [49] APACHE HTTP SERVER, Disponível em: <<http://httpd.apache.org>> Acesso em: 12/11/2006.
- [50] CHOI, et al. Beginning PHP 4 Programando. Makron Books, 2001.
- [51] HEATON, J. Programming Spiders, Bots, and Aggregators in Java. Alameda: Sybex, 2002.
- [52] ICHINO, M.; YAGUSHI, H. Generalized Mikowsky Metrics for Mixed Feature Type Data Analysis. IEEE Transactions System, Man and Cybernetics, Vol. 24, p. 698-708, 1994.

Anexo A

Formulário



Universidade de Pernambuco
Escola Politécnica de Pernambuco
Departamento de Sistemas Computacionais
Trabalho de Conclusão de Curso

Aluno: Valmir Macário Filho
Orientador: Fernando Buarque, Phd

Questionário de avaliação da ferramenta e-Recommend

Nome: _____

Número do Teste: _

Que tipo de consumidor virtual você se considera?

- Casual. Compra de vez em quando.
- Adepto. Compra regularmente.
- Não compra em sítios virtuais.

Descrição da ferramenta:

e-Recommend é uma ferramenta de recomendação inteligente para lojas de comércio eletrônico. As recomendações geradas pelo e-Recommend são personalizadas, a ferramenta constrói um perfil para cada usuário, baseando-se em no histórico de compras de cada usuário. A resposta da ferramenta é uma lista de recomendações, que espera-se que o usuário goste.

Após ter analisado as recomendações, responda as seguintes questões:**1. Como você avalia a interface do sítio?**

- Não gostei.
- Gostei pouco.
- Gostei.
- Gostei Muito

2. O que você achou da velocidade que foi gerada a recomendação?

- Lenta
- Pouco Lenta
- Normal.
- Rápida.

3. Quantos livros dessa lista você compraria?

- Nenhum.
- De 1 a 2
- De 3 a 4
- Todos.

4. Quantos livros dessa lista você não compraria?

- Não Compraria nenhum
- Apenas 1
- De 2 a 3
- 4

5. Que nota de 1 a 5 você avaliaria a lista de recomendações?

- 0
- 1-2
- 3-4
- 5

6. Você já leu os livros recomendados?

- Não li Nenhum.
- Li poucos.
- Li a maioria.
- Já li todos.

7. Já conhecia de alguma forma os livros recomendados?

- Não conhecia nenhum.
- Conhecia poucos.
- Conhecia a maioria.
- Conhecia todos.

Observações:
