

Utilização de Técnicas de Sistemas de Recomendação para Aprimoramento de um E-commerce do Tipo B2b

Anita Maria da Rocha Fernandes
anita.fernandes@univali.br
UNIVALI

Bruno Luiz Linhares
UNIVALI

Resumo: Desde o surgimento das sociedades, sabe-se da existência do comércio, conhecido como o processo de comprar, vender e trocar produtos e serviços. Tendo em vista o grande avanço da internet a cada dia que passa, bem como, seus inúmeros benefícios, diversos produtos e serviços chegam simultaneamente em velocidade rápida em todo o mundo. Não há dúvidas que a internet seja responsável por disseminar informações e mudanças, juntamente com as características do mercado da informática e sua área de programação que são utilizadas hoje para incentivar e ampliar o comércio eletrônico na rede. Para um e-commerce ser considerado eficaz, de maneira geral, deve possuir três características básicas, as quais são: oferecer informações detalhadas sobre os produtos, transmitir confiança ao cliente e ajudar o cliente a tomar decisões. Para esta terceira característica, foco deste trabalho, tem-se pesquisado e utilizado muito técnicas de Sistemas de Recomendação. Essas técnicas auxiliam no aumento da capacidade e eficácia do processo de indicação de produtos e serviços, tendo como principal desafio, realizar o casamento correto entre o que o cliente procura e o que é recomendado a ele. Este artigo apresenta as técnicas de recomendação existentes e a aplicação de uma delas no site de e-commerce da empresa Pauta Distribuição e Logística. Apesar de ainda não se ter concluído todos os testes, os resultados preliminares já são considerados satisfatórios pelo fato de não afetar negativamente o produto como um todo.

Palavras Chave: Comércio Eletrônico - Sist. recomendação - e-commerce - -

1. INTRODUÇÃO

O comércio sempre existiu desde que surgiram as sociedades. Ele é dito como o processo de comprar, vender e trocar produtos e serviços. Inicialmente praticado pelos primeiros povos apenas a troca, foram os Fenícios, Árabes, Assírios e Babilônios que o incrementaram, pois com as expedições e a descoberta de novos mundos, utilização de pedras e metais preciosos, o desenvolvimento científico industrial e os meios de comunicação, eles foram incentivados a comprar e a vender. Os primeiros vendedores viajavam semanas, meses e anos para anunciarem seus lançamentos, por isso seus produtos eram vendidos como novos durante um longo período (LUPPI, 2009).

A partir do Século XX, produtos eletrônicos começaram a chegar a países, como o Brasil. Graças à rede mundial de computadores, a popular Internet que surgiu nos EUA em 1969 durante o projeto da Arpanet na época da guerra fria e que jamais deixou de evoluir, hoje os diferentes tipos de produtos e serviços chegam simultaneamente em velocidade rápida em todo o mundo. Não há dúvidas que ela seja responsável por disseminar informações e mudanças, juntamente com as características do mercado da informática e sua área de programação que são utilizadas hoje para incentivar e ampliar o comércio pela rede (LUPPI, 2009).

Comércio eletrônico (CE) ou e-commerce é a atividade desenvolvida na internet, com a utilização de ferramentas eletrônicas e tecnologias emergentes, que tem por objetivo principal a negociação de compra e venda de bens e serviços (ULBRICH, 2000). Porém, para Luppi (2009), CE não é só pela internet através de conexão com computadores, mas por meio de anúncios na televisão e principalmente o uso de aparelhos celulares e outros equipamentos eletrônicos. Sua tendência é crescer a cada dia, pois a Internet nunca deixará de evoluir, já que surgiu por meio dela a necessidade das empresas e pessoas se comunicarem, interagirem e realizarem transações. Hoje, o comércio em geral está partindo para a internet. Acredita-se que é uma das melhores fases de mudança estrutural na sociedade. Na verdade, é a formação de uma nova sociedade: a Sociedade da Informação, a Era da Digitalização, cujos processos fazem parte de quase todas as atividades do dia a dia de pessoas e profissionais em geral. Pode-se perceber facilmente essa evolução, pois se deixa de escrever cartas, utilizar agendas, papel e caneta para anotações, dinheiro para se negociar mercadorias foi substituído por informações de débito e crédito em contas e cartões (LUPPI, 2009).

Segundo Cameron (1997, *apud* ALMEIDA JR., 2000), CE inclui qualquer negócio transacionado eletronicamente, no qual as transações ocorrem entre um negócio e seus clientes, 2 B2C (Business to Consumer), entre dois parceiros, B2B (Business to Business), entre dois usuários, C2C (Consumer to Consumer), entre usuários e um parceiro, C2B (Consumers to Business), entre outros modelos não tão conhecidos tampouco utilizados.

Conforme Felipini (2006), para um e-commerce ser considerado eficaz o mesmo deve possuir um conjunto de características, sendo que algumas são descritas brevemente a seguir:

- Oferecer informações detalhadas sobre os produtos: diferentemente das lojas tradicionais, as lojas virtuais não possuem vendedores, e as informações relevantes são oferecidas por meio do texto. Por isso é fundamental que o texto forneça todas as informações possíveis sobre o produto, de forma objetiva, para que o cliente ou parceiro possa se convencer da conveniência da compra;
- Ajudar o cliente ou parceiro a tomar decisões: o consumidor típico das lojas virtuais é bem informado, tem boa formação educacional e capacidade de discernimento para saber o que é bom pra si. Mais do que tentar convencê-lo a realizar compras, deve-se fornecer o contexto e as informações que o ajudem a se decidir;

- Transmitir confiança ao cliente ou parceiro: o objetivo da loja virtual deve ser o de criar um ambiente de compras que transmita confiabilidade, e isso pode ser feito por meio de procedimentos como: fornecimento de informações de forma clara e honesta; orientação com relação às medidas de segurança adotadas nas lojas virtuais; garantias de devolução de produtos; depoimentos de clientes; entre outros.

Segundo Forrester (2000, *apud* HORTINHA, 2000), o comércio eletrônico do tipo B2B possui algumas limitações ou problemas, tais como, má navegabilidade e dificuldade de utilização, possuindo erros e omissões. Falta personalização e o número de clicks até chegar ao produto ou serviço pretendido é muito elevado. Além disso, há uma ineficiência nos motores de busca.

Para Luciano (2003), um dos maiores benefícios dos sistemas de computação, aliados ao fato de estarem conectados em rede, é a possibilidade de obtenção de dados, tratamento e disponibilização de conhecimento, extraídos a partir da análise de informações contidas nos dados coletados. Nos negócios on-line, esse processo torna-se ainda mais claro e passível de ser utilizado.

Devido a este benefício definido por Luciano (2003), muitos dos defeitos citados por Forrester (2000, *apud* HORTINHA, 2000), como exemplo a má navegabilidade, a difícil utilização, existência de erros, entre outros, podem ser solucionados com a utilização de sistemas de recomendação.

Os portais de comércio eletrônico são atualmente o maior foco de utilização dos sistemas de recomendação, empregando diferentes técnicas para encontrar os produtos mais adequados para seus clientes ou parceiros e aumentar deste modo sua lucratividade. Introduzido em julho de 1996 o My Yahoo foi o primeiro website a utilizar os sistemas de recomendação em grandes proporções, utilizando a estratégia de customização (MANBER *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005). Hoje em dia, um grande número de portais emprega os sistemas de recomendação para levar aos usuários diferentes tipos de sugestões, como ofertas casadas, itens de sua preferência, itens mais vendidos nas suas categorias favoritas, etc. (REATEGUI & CAZELLA, 2005).

Sistemas de Recomendação podem ser utilizados em diversas áreas, podendo trazer grandes benefícios em determinados casos. Uma das áreas que pode tirar um proveito considerável e satisfatório de sistemas de recomendação é o comércio eletrônico, cujo número de adeptos multiplica-se a cada ano.

Com o estudo de Sistemas de Recomendação e suas diferentes formas de aplicação, observou-se que o mesmo, em certos casos, pode agregar benefícios a diversos sistemas, inclusive a um *e-commerce*. Com isso foi implementada uma técnica de Sistemas de Recomendação no *e-commerce* da empresa Pauta Distribuição e Logística, o qual é o estudo de caso deste trabalho. Esta técnica fez uso das informações armazenadas na base de dados da empresa, tais como o histórico de compras dos seus parceiros. A partir disso, são sugeridos ao cliente, na página inicial, alguns produtos considerados de seu maior interesse, a fim de aumentar a satisfação do mesmo, agilizar o processo de busca por produtos de seu interesse e comprados com mais frequência, reduzindo com isso o tempo em que o cliente necessita navegar pelo site para realizar sua compra, diminuindo conseqüentemente, o fluxo de dados no servidor.

As seções à seguir apresentarão os conceitos relacionados a sistemas de recomendação, o sistema desenvolvido para a empresa Pauta e os resultados obtidos, bem como as conclusões deste trabalho.

2. SISTEMAS DE RECOMENDAÇÃO

Com a quantidade de informações e com a disponibilidade facilitada das mesmas pelo uso da Internet, as pessoas se deparam com uma diversidade muito grande de opções. Muitas vezes um indivíduo possui muito pouca ou quase nenhuma experiência pessoal para realizar escolhas entre as várias alternativas que lhe são apresentadas. Para minimizar as dúvidas e necessidades que se tem frente à escolha entre alternativas, geralmente confia-se nas recomendações que são passadas por outras pessoas, as quais podem chegar de forma direta, cartas de recomendação, opiniões de revisores de filmes e livros, impressos de jornais, entre outros (MAES & SHARDANAND, 1995 *apud* REATEGUI & CAZELLA, 2005).

De acordo com Resnick e Varian (1997, *apud* REATEGUI & CAZELLA, 2005), os sistemas de recomendação auxiliam no aumento da capacidade e eficácia deste processo de indicação já bastante conhecido na relação social entre seres humanos. Em um sistema típico as pessoas fornecem recomendações como entradas e o sistema agrega e direciona para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios deste tipo de sistema é realizar o casamento correto entre os que estão recomendando e aqueles que estão recebendo a recomendação, ou seja, definir e descobrir este relacionamento de interesses (REATEGUI & CAZELLA, 2005).

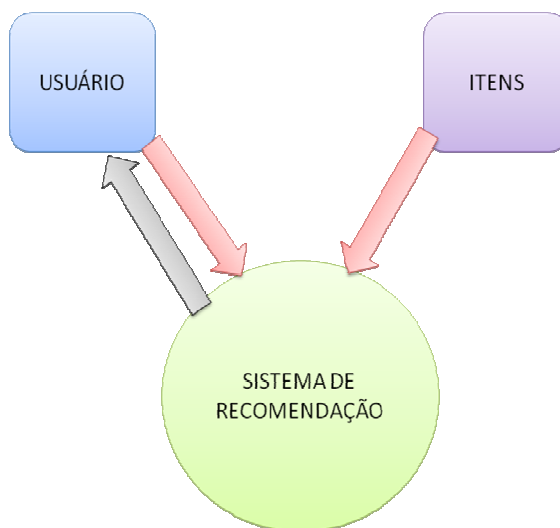


Figura 1: Idéia principal de um Sistema de Recomendação, adaptado de Marchi (2008).

Os portais de comércio eletrônico são atualmente o maior foco de utilização dos sistemas de recomendação, empregando diferentes técnicas para encontrar os produtos mais adequados para seus clientes ou parceiros e aumentar deste modo sua lucratividade. Introduzido em julho de 1996, o My Yahoo foi o primeiro website a utilizar os sistemas de recomendação em grandes proporções, utilizando a estratégia de customização (MANBER *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005). Hoje em dia, um grande número de portais emprega os sistemas de recomendação para levar aos usuários diferentes tipos de sugestões, como ofertas casadas, itens de sua preferência, itens mais vendidos nas suas categorias favoritas, etc. (REATEGUI & CAZELLA, 2005).

2.1. TÉCNICAS UTILIZADAS EM SISTEMAS DE RECOMENDAÇÃO

Segundo Reategui e Cazella (2005), várias tecnologias têm surgido visando à identificação de padrões de comportamento (consumo, pesquisa, etc.) e utilização destes

padrões na personalização do relacionamento com os usuários. Estas técnicas fundamentam o funcionamento dos sistemas de recomendação e são descritas a seguir.

Filtragem de Informações

Filtragem de informação é o nome utilizado para descrever uma variedade de processos que envolvem a entrega de informação para as pessoas que realmente necessitam delas. Com o tempo este termo tornou-se muito popular e artigos técnicos foram escritos descrevendo o emprego de filtragem em diversas aplicações, como e-mail, documentos eletrônicos em escritórios, entre outros. No entanto, a distinção entre filtragem e processos relacionados com recuperação de informações não é sempre clara. As expressões “recuperação de informação” e “filtragem de informação” descrevem soluções que procuram auxiliar na solução de problemas referentes à sobrecarga de informações. A expressão “recuperação de informação” tradicionalmente envolve armazenamento, índices, e tecnologia para recuperação de documentos textuais. Para que a recuperação realmente ocorra o usuário deve descrever sua necessidade de informação em um formulário, com a consulta (*query*) do que necessita. A partir deste momento, o sistema busca casar a necessidade explicitada com os documentos armazenados. Este tipo de abordagem tende a manter uma base com característica mais estática no armazenamento das informações, e o início da interação é provocado pelo usuário (BELVIN & CROFT, 1992 *apud* REATEGUI & CAZELLA, 2005).

Diferentemente da recuperação, segundo Foltz e Dumais (1992, *apud* REATEGUI & CAZELLA, 2005) a filtragem propõe uma abordagem distinta. Esta abordagem geralmente mantém um perfil dos interesses do usuário. A idéia é que esta abordagem tenha como característica um maior tempo de duração no que tange ao casamento de interesses, pois a filtragem não se refere ao momento, mas sim as preferências, enquanto a recuperação baseia-se na percepção de uma necessidade do momento. Esta filtragem deve ser aplicada a cada novo item adicionado procurando verificar se este atende ao usuário. Segundo Bezerra (2002), nem sempre se pode obter o perfil do usuário de forma implícita, como ocorre, por exemplo, na Amazon Book Store, que a cada vez que um usuário compra algum produto o sistema registra essa informação no perfil do usuário de forma positiva. Algumas vezes isso é inviável. Considerando-se, por exemplo, um sítio de recomendação de filmes e assumindo-se que tal sistema não comercializa filmes, não seria possível aplicar o modelo de aquisição implícita de forma adequada. Sendo assim, o sistema necessitaria de uma avaliação do usuário com relação aos filmes apresentados, dando origem a um perfil de usuário constituído dos filmes avaliados e suas respectivas notas atribuídas por esse usuário. A seguir são descritas técnicas de filtragem aplicáveis a sistemas de recomendação.

Filtragem Baseada em Conteúdo (FBC): por muitos anos os cientistas têm direcionado seus esforços para aliviar o problema ocasionado pela sobrecarga de informações através de projetos que integram tecnologias que automaticamente reconhecem e categorizam as informações (HERLOCKER, 2000 *apud* REATEGUI & CAZELLA, 2005). Alguns sistemas têm como objetivo gerar de forma automática descrições dos conteúdos dos itens e comparar estas descrições com os interesses dos usuários visando verificar se o item é ou não relevante para cada um (BALABANOVIC & SHOHAM, 1997 *apud* REATEGUI & CAZELLA, 2005). Esta técnica é chamada de Filtragem Baseada em Conteúdo por realizar uma seleção baseada na análise de conteúdo dos itens e no perfil do usuário (ANSARI *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005).

De acordo com Herlocker (2000, *apud* REATEGUI & CAZELLA, 2005), muitas ferramentas que utilizam esta abordagem, aplicam técnicas como indexação de frequência de termos. Nesta técnica, informações dos documentos e necessidades dos usuários são descritas por vetores com uma dimensão para cada palavra que ocorre na base de dados. Cada



componente do vetor é a frequência que uma respectiva palavra ocorre em um documento ou na consulta do usuário. Claramente, os vetores dos documentos que estão próximos aos vetores da consulta do usuário são considerados os mais relevantes para ele (REATEGUI & CAZELLA, 2005). Outras técnicas aplicadas na Filtragem Baseada em Conteúdo, segundo Marchi (2008), são índices de busca booleana, onde a consulta constitui-se em um conjunto de palavras chaves unidos por operadores booleanos, e sistemas de filtragem probabilística, onde se aplica o raciocínio probabilístico para se determinar a probabilidade que um documento possui para atender as necessidades de informações.

Filtragem Colaborativa (FC): a abordagem da Filtragem Colaborativa, também conhecida como Filtragem Social (BRESSLER, 2004), foi desenvolvida para atender pontos que estavam em aberto na Filtragem Baseada em Conteúdo (ANSARI *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005). A Filtragem Colaborativa se diferencia da Filtragem Baseada em Conteúdo exatamente por não exigir a compreensão ou reconhecimento do conteúdo dos itens (REATEGUI & CAZELLA, 2005). Segundo Kajimoto *et al.* (2007), a Filtragem Colaborativa permite dois tipos de abordagens, a primeira determina recomendações baseadas em conteúdo consumido por usuários com o mesmo padrão de consumo do usuário atual (Figura 2). É utilizada principalmente em sistemas de comércio eletrônico, como Amazon e Submarino. A segunda determina recomendações baseadas nas classificações realizadas por outros usuários dentro de um grupo restrito de conteúdo, ordenadas pela soma da relevância de tais classificações (sistema de votação com pesos) (Figura 3). É utilizada principalmente em sistemas de notícias, como o Digg. A Figura 4 apresenta um exemplo desta abordagem. Em geral, esse tipo de abordagem oferece recomendações menos pessoais e mais dirigidas a um grupo determinado de usuários, restritos a um tema. Em contrapartida, sua implementação enfrenta menos problemas de escalabilidade e em geral é mais viável (KAJIMOTO *et al.*, 2007).



Figura 2: Primeira abordagem, adaptado de Kajimoto *et al* (2007).

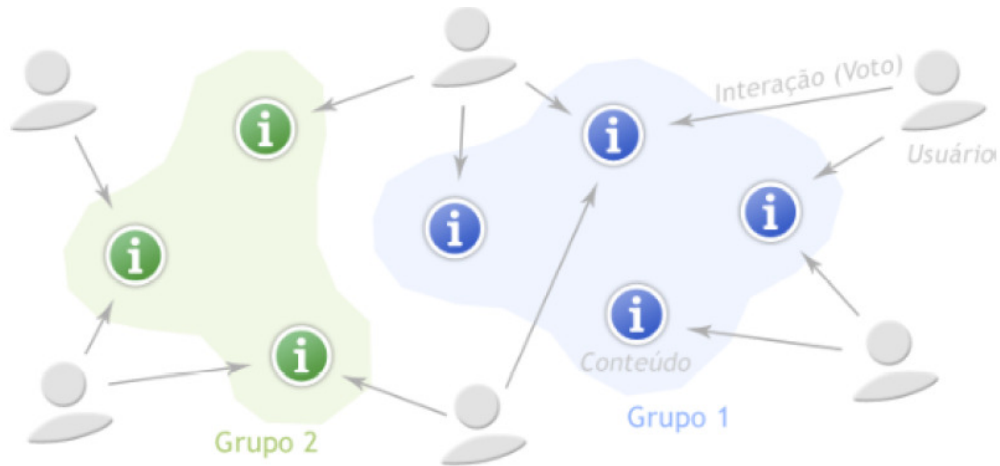


Figura 3: Segunda abordagem, adaptado de Kajimoto *et al* (2007).

Aproveite Também

	+		Harry Potter e as Relíquias da Morte + Chaveiro GRÁTIS - J.K. ROWLING + Harry Potter e o Prisioneiro de Azkaban - vol. 3 - J.K. ROWLING	
			Compre Junto: R\$ 69,10 Economize: R\$ 34,90	
	+		Harry Potter e as Relíquias da Morte + Chaveiro GRÁTIS - J.K. ROWLING + Harry Potter e o Cálice de Fogo - vol. 4 - J.K. ROWLING	
			Compre Junto: R\$ 77,90 Economize: R\$ 40,10	

➤ Harry Potter e a Ordem da Fênix - vol. 5 - J.K. ROWLING : R\$44,60

Figura 4: Exemplo da segunda abordagem, adaptado de Kajimoto *et al* (2007).

Filtragem Híbrida (FH): a abordagem da Filtragem Híbrida procura combinar os pontos fortes da Filtragem Colaborativa e Filtragem Baseada em Conteúdo visando criar um sistema que possa melhor atender as necessidades do usuário (ANSARI *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005). Essa abordagem é constituída de vantagens proporcionadas pela Filtragem Baseada em Conteúdo e pela Filtragem Colaborativa, unindo o melhor das duas técnicas e eliminando as fraquezas de cada uma (Figura 5) (REATEGUI & CAZELLA, 2005).

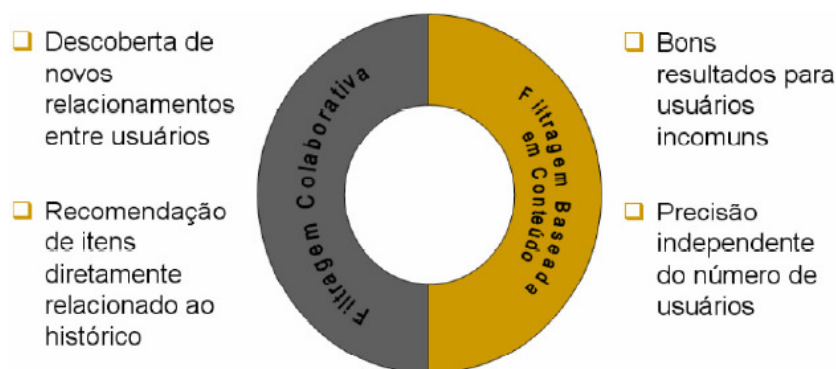


Figura 5: Filtragem híbrida, adaptado de Reategui e Cazella (2005).



Descoberta de Conhecimento em Base de Dados (DCBD)

Com os avanços tecnológicos, os sistemas de armazenamento de dados e monitoramento de fenômenos observados têm acumulado grandes volumes de dados que ocultam conhecimento para o ser humano. Assim, o “KDD – *Knowledge Discovery in Databases*” (Descoberta de Conhecimento em Banco de Dados) é uma área da IA (Inteligência Artificial) que se preocupa em descobrir conhecimento oculto, importante para empresas e órgãos governamentais através do uso de algoritmos eficientes de mineração de dados (SILVA, 2007). A extração do conhecimento é uma área dinâmica e evolutiva, envolvendo integrações com outras áreas de conhecimento como Estatística, Inteligência Artificial e Banco de Dados. Os padrões extraídos devem ser além de confiáveis, compreensíveis e úteis, podendo empregar o conhecimento com utilidade e tirar proveito de alguma vantagem, seja científica ou comercial (LIMA *et al.*, 2008). O KDD pode ser visto como o processo da descoberta de novas correlações, padrões e tendências significativas por meio da análise minuciosa de grandes conjuntos de dados estocados. Este processo se vale de tecnologias de reconhecimento utilizando padrões e técnicas estatísticas e matemáticas. O *Data Mining* é uma das técnicas utilizada para a realização de KDD (NORTON, 1999 *apud* SILVA, 2007).

Segundo Han e Kamber (2000, *apud* TRONCHONI *et al.*, 2008) este processo de descoberta consiste em uma seqüência de seis passos (Figura 6). São eles: limpeza dos dados: para remover ruídos e dados irrelevantes; integração dos dados: onde fontes de dados múltiplos podem ser combinadas; seleção dos dados: onde dados relevantes para a análise são recuperados do banco de dados; transformação dos dados: onde os dados são transformados ou consolidados no formato apropriado para mineração; mineração de dados: é um processo onde métodos inteligentes são utilizados a fim de extrair padrões de dados; e avaliação e representação do conhecimento: onde técnicas de visualização e representação de conhecimento são utilizadas para apresentar o conhecimento extraído para o usuário.

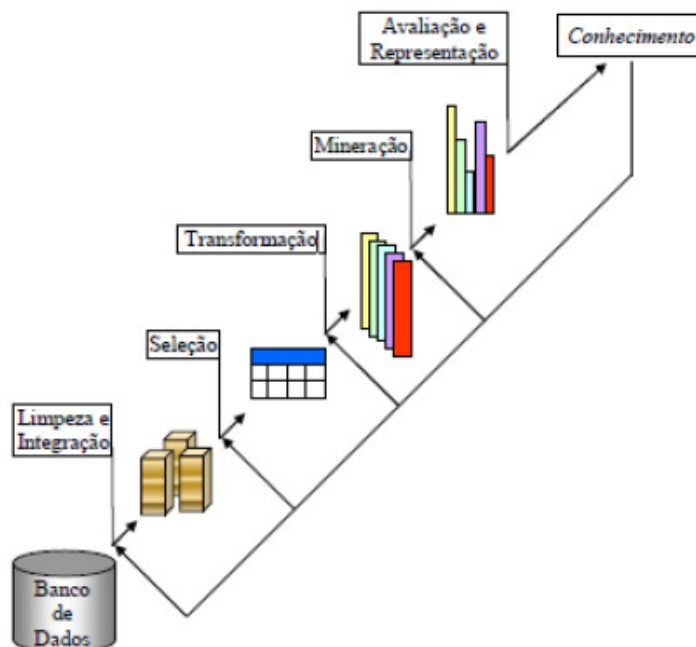


Figura 6: Fluxo do processo de KDD, adaptado de Tronchoni *et al* (2008).

Quando se trabalha com sistemas de recomendação na web, a mineração de dados torna-se um recurso importante para a descoberta de relações entre itens, entre usuários e entre itens e usuários. Através da mineração dos arquivos de log (YANG *et al.*, 2002 *apud*



REATEGUI & CAZELLA, 2005), por exemplo, pode-se obter conhecimentos aprofundados a respeito dos usuários que se conectaram a um website (SPILIOPOULOU, 2000 *apud* REATEGUI & CAZELLA, 2005). Este conhecimento pode ser utilizado para a personalização da oferta de produtos (RUCKER & POLANCO, 1997 *apud* REATEGUI & CAZELLA, 2005), na estruturação de websites de acordo com o perfil de cada internauta e personalização (NULVENNA *et al.*, 2000 *apud* REATEGUI & CAZELLA, 2005) do conteúdo das páginas.

O desenvolvimento de sistemas de KDD está relacionado com diversas áreas de aplicação, como por exemplo: bancária (aprovação de crédito), ciências e medicina (descoberta de hipóteses, diagnóstico, classificação, predição), comerciais (segmentação, localização de consumidores, identificação de hábitos de consumo), engenharia (simulação e análise, reconhecimento de padrões, processamento de sinais e planejamento), financeira (apoio para investimentos, controle de carteira de ações), gerencial (tomada de decisão, gerenciamento de documentos), internet (ferramentas de busca, navegação, extração de dados), manufatura (modelagem e controle de processos, controle de qualidade, alocação de recursos), segurança (detecção de bombas, icebergs e fraudes), etc. (MOTTA, 2005 *apud* SILVA, 2007). Segundo Reategui e Cazella (2005), existem diversos tipos de descoberta de conhecimento aplicáveis nos sistemas de recomendação, que podem ser mais eficientes do que outras, dependendo do domínio de aplicação. Algumas delas são:

Regra de Associação: são técnicas que procuram encontrar padrões entre conjuntos de valores ou em gráficos evolutivos de valores. Uma regra de associação é uma indicação de que existe um certo relacionamento entre fatos diferentes. (BARAZETTI, 2009). Essa regra constitui-se em um relacionamento $X \rightarrow Y$, onde X e Y são conjuntos de itens e a intersecção entre os mesmos constitui-se em um conjunto vazio. Cada regra de associação é associada a um fator de suporte (FSup), e a um fator de confiança (FConf) (REATEGUI & CAZELLA, 2005).

Segundo Reategui e Cazella (2005), o FSup é determinado pela razão do número de tuplas que satisfazem X e Y, sobre o número total de tuplas ($F_{Sup} = |X \vee Y| / N$) (Equação 1). Já o FConf é determinado pela razão do número de tuplas que satisfazem X e Y, sobre o número de tuplas que satisfazem X ($F_{Conf} = |X \vee Y| / |X|$) (Equação 2).

$$\text{Suporte} = \frac{\text{Frequência de X e Y}}{\text{Total da Amostra}} \quad \text{Equação 1. Suporte.}$$

$$\text{Confiança} = \frac{\text{Frequência de X e Y}}{\text{Frequência de X}} \quad \text{Equação 2. Confiança.}$$

A função de extração de regras a partir de um banco de dados consiste em encontrar todas as regras possíveis com FSup e FConf maiores ou iguais a um FSup e FConf especificada pelo usuário.

Classificação: determinam a classe ou categoria de um elemento, a partir das características (ou atributos) deste elemento e com base em classes pré-definidas. Uma nova classe poderia ser acrescida ao conjunto, caso fosse detectado um elemento com características contraditórias (satisfazendo duas classes, mas não podendo ser classificado assim por serem classes disjuntas) ou então quando as características do elemento não se enquadrassem em nenhuma das classes pré-estabelecidas (BARAZETTI, 2009). Um exemplo possível para explicar o uso de classificação é o que segue: supondo que um sistema de recomendação guarde algumas informações sobre os seus usuários (por exemplo, status na Universidade, país e área de interesse), pode ser interessante utilizar esses dados para prever que tipo de usuário está mais interessado em qual área de pesquisa (REATEGUI & CAZELLA, 2005). O sistema pode

então concentrar os esforços de recomendação para classificar os usuários. Para prever se o usuário estará ou não interessado em determinada área, pode ser feito o uso de uma árvore de decisão (HOLTE & YAN, 1996 *apud* REATEGUI & CAZELLA, 2005).

Agrupamento: associa um item a uma ou várias classes categóricas (ou *clusters*), determinando as classes pelos dados, independentemente da classificação pré-definida. Os *clusters* são definidos por meio do agrupamento de dados baseados em medidas de similaridade ou modelos probabilísticos, visando detectar a existência de diferentes grupos dentro de um determinado conjunto de dados e, em caso de sua existência, determinar quais são eles (LIMA *et al.*, 2008). Outra definição de agrupamento, bastante simplificada, segundo Silva (2007), é o processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos. De acordo com Barazetti (2009), este é o processo inverso da classificação. A partir de técnicas de *clustering*, são analisados os atributos de caracterização e os de discriminação para enquadrar os elementos nas suas respectivas classes, que são definidas por essas técnicas. O algoritmo de agrupamento automaticamente deve criar grupos através de partições de um banco de dados em conjunto de tuplas similares. A partição acontece de forma que as tuplas de valores e atributos semelhantes sejam reunidas nos mesmos grupos. Um agrupamento de qualidade surge onde a similaridade intra-classe é alta e a inter-classe é baixa. O agrupamento é um tipo de aprendizado não supervisionado uma vez que fica a cargo do algoritmo a definição de quais atributos e seus valores serão utilizados para montar agrupamentos de dados similares (REATEGUI & CAZELLA, 2005).

2.2. EXEMPLOS DE SISTEMAS DE RECOMENDAÇÃO

A seguir serão apresentados alguns exemplos de sistemas de recomendação, com uma breve descrição dos mesmos.

Referral Web: o projeto Referral Web procura identificar e visualizar redes sociais, que de acordo com o autor seriam grupos de pessoas ligadas por atividades profissionais. Para criar esta rede o sistema utiliza-se de textos já existentes na Web. Uma vez que o usuário entra com seus dados é feita uma busca de documentos que o mencionem. A partir dos documentos recuperados, são extraídos nomes de outros indivíduos citados. No final é construída uma rede global de citações representada na forma de um grafo. Neste sistema a formação da rede está baseada somente na relação entre usuário, documento e citações, a importância e relevâncias dos trabalhos ou indivíduos não é considerada (KAUTZ *et al.*, 1997 *apud* REATEGUI & CAZELLA, 2005).

RINGO: é um sistema desenvolvido para recomendação personalizada de música e foi desenvolvido no Massachusetts Institute of Technology (MAES & SHARDANAND, 1995 *apud* SILVA *et al.*, 2006). O trabalho explora similaridades entre os gostos de diferentes usuários para recomendar itens, baseado no fato de que os gostos das pessoas apresentam tendências gerais e padrões entre gostos, e entre grupos de pessoas. Neste sistema as pessoas descrevem suas preferências musicais através da avaliação de algumas canções. Estas avaliações constituem o perfil dos indivíduos. O sistema usa então estes perfis para gerar recomendações para usuários individuais. Para o seu funcionamento, primeiramente usuários similares são identificados. A partir desta identificação e comparação de perfis o sistema pode prever o quanto o usuário gostaria de um álbum/artista que ainda não foi avaliado pelo mesmo (SILVA *et al.*, 2006). Quando o usuário acessa o RINGO pela primeira vez, é apresentado a ele uma lista de 125 artistas. O usuário avalia estes de acordo com o quanto gosta de ouvi-los. Caso o usuário não esteja familiarizado com o artista ou não possua uma forte opinião sobre o mesmo, é solicitado que não avalie o artista para não gerar distorções. Para a pontuação é aplicada uma escala Likert de 7 pontos, sendo o 1 (não gosta), 4 (indiferente) e 7 (adora) (SILVA *et al.*, 2006). A lista de artistas enviada para os usuários é



dividida em duas partes. A primeira parte da lista é gerada a partir dos artistas mais pontuados (ranking), o que assegura que um novo usuário tenha a oportunidade de pontuar artistas que outros já tenham pontuado. A outra parte da lista é gerada através de uma seleção randômica a partir de um banco de dados sobre os artistas. Uma vez que o perfil inicial da pessoa tenha sido submetido, o usuário pode solicitar ao RINGO algumas predições (SILVA *et al.*, 2006): sugerir novos artistas/álbuns que o usuário gostaria de obter ou ouvir; listar artistas/álbuns que o usuário não gostaria; e realizar uma predição sobre um artista/álbum específico. O retorno dado pelo RINGO aos usuários não inclui nenhuma informação em particular sobre a identidade dos outros usuários que contribuíram com as recomendações. Deve-se enfatizar esta característica em sistemas colaborativos ou de informação social (muitas vezes usados como sinônimo): a identidade de quem avaliou deve ser mantida em segredo (SILVA *et al.*, 2006). O sistema RINGO também permite a inserção de comentários (como *feedback* explícito) sobre o produto recomendado. Os próprios usuários podem incluir novos artistas e álbuns no banco de dados (SILVA *et al.*, 2006).

GroupLens: constituiu-se em um projeto de pesquisa desenvolvido pela Universidade de Minnesota, aplicado à filtragem colaborativa de notícias, auxiliando pessoas a encontrar artigos de interesse entre um grande volume de artigos ofertados (RIEDL *et al.*, 1997 apud SILVA *et al.*, 2006). O Sistema coleta a avaliação dos usuários referente aos artigos lidos (escala de 1-5 pontos) e utiliza esta avaliação para identificar os vizinhos mais próximos com avaliações semelhantes. Desta forma o sistema é capaz de predizer se um usuário pode se interessar por um novo artigo baseando-se nas avaliações dos usuários vizinhos (SILVA *et al.*, 2006).

Fab: é um sistema que procura combinar sistemas de filtragem colaborativa e filtragem baseada em conteúdo (abordagem híbrida), procurando eliminar muitas das fraquezas encontradas em cada uma das abordagens. Foi desenvolvido pela Universidade de Stanford em um projeto para recomendação de páginas web (documentos) para usuários (SILVA *et al.*, 2006). Objetivando a construção de uma arquitetura híbrida, perfis de usuários são gerados a partir da análise de conteúdo e comparados para identificar usuários similares para recomendação colaborativa. O sistema tem uma implementação distribuída e foi construído como parte do projeto da biblioteca digital da Universidade de Stanford (SILVA *et al.*, 2006).

3. O SISTEMA DESENVOLVIDO PARA A EMPRESA PAUTA

O sistema de *e-commerce* baseado em sistema de recomendação desenvolvido para a empresa Pauta, está dividido em três módulos, sendo eles: venda *on-line*, gerenciamento da revenda e portal. O primeiro módulo, “Venda *on-line*”, qual corresponde especificamente ao processo de vendas do *site*, desde a inclusão de produtos no carrinho de compras até o último clique para finalizar de fato o pedido de um determinado parceiro.

O segundo módulo, denominado de “Gerenciamento da Revenda”, corresponde a uma parte específica de cada determinado parceiro, que engloba seus dados cadastrais, a lista dos seus pedidos, cadastro e gerenciamento dos seus usuários, suas notas fiscais, entre outros. O sistema de recomendação atua especificamente no terceiro módulo, na parte do Portal, o qual engloba a página inicial referente ao *e-commerce*.

Esta página possui uma série de funcionalidades e características, dentre elas, uma área específica para exibição de produtos, para qual foi realizado um estudo prévio pela equipe de desenvolvimento da empresa que construiu o portal e por critérios ergonômicos foi definido que seria exibido nesta, seis produtos. Inicialmente, esta área exibia seis produtos randomicamente a partir de uma lista específica que continha apenas produtos em promoção.

Com a implementação do sistema de recomendação, não são mais exibidos somente produtos em promoção, mas sim, seis produtos que o portal considerar de maior interesse para o



determinado parceiro que logar no *site*, exceto para visitantes que continuam visualizando seis produtos em promoção. Estes produtos possivelmente de maior interesse para o determinado parceiro são gerados a partir da utilização de uma técnica de Sistemas de Recomendação denominada Filtragem Baseada em Conteúdo (FBC). O sistema faz uma pesquisa nos últimos pedidos do determinado parceiro que efetuar o *login* e com base no histórico de pedidos, o sistema apresenta os seis produtos considerados de maior relevância, a partir de um cálculo, o qual o número de vezes que o produto foi comprado por este cliente multiplicado por sua quantidade, gerando-se após isto, uma lista em ordem decrescente com os resultados obtidos, exibindo no *site* então os seis primeiros desta lista.

3.1. IMPLEMENTAÇÃO DO SISTEMA

A implementação de um sistema de recomendação inicia com o estabelecimento da janela de tempo de pesquisa para recomendação. A princípio a janela de tempo utilizada para avaliação dos pedidos dos parceiros seria de seis meses, porém após uma melhor avaliação da questão, optou-se em limitar a três meses, tendo em vista que a empresa Pauta é uma distribuidora e conseqüentemente seus clientes são revendas, o que faz com que, teoricamente, os mesmos comprem com mais frequência que usuários finais e também em maior quantidade. Se continuasse sendo utilizada a janela de seis meses para avaliação, isto envolveria um número mais elevado de dados, afetando de maneira negativa no carregamento inicial do *site*, ou ainda corria-se o risco de eleger produtos de não mais interesse por parte da revenda.

O banco de dados relacional utilizado foi o Oracle, pois era o banco que estava em funcionamento na empresa. Sobre a estrutura, relata-se que não foram criadas tabelas e atributos adicionais, ou seja, para a realização do presente trabalho foram utilizadas somente as informações que já estavam sendo armazenadas. Porém com a implantação do sistema de recomendação de produto, essas informações ganharam uma nova importância, podendo trazer um retorno muito positivo para a empresa.

O foco principal do desenvolvimento, em termos de código-fonte, foi a utilização de *Servlets*, que são classes Java utilizadas para ampliar as capacidades e funcionalidades de um servidor *web* através de módulos de aplicação, utilizado basicamente para processar de forma dinâmica requisições e respostas.

Após a criação de todos os controles desenvolvidos em Java, utilizando *servlets*, partiu-se para o layout da tela. Devido o fato da tela já existir, porém com a função antiga do *site*, que exibia apenas seis itens randomicamente da lista de produtos em promoção, foi feita uma cópia desse arquivo JSP, para que se tivesse guardado o modo antigo e foram realizadas basicamente três alterações: o nome do arquivo, o título exibido no site, o qual passou de “DESTAQUES EM PROMOÇÕES” para “PRODUTOS SUGERIDOS” e a referência para o novo *servlet*. A Figura 7 apresenta o *print* da referida área no *site*, após a as referidas alterações:

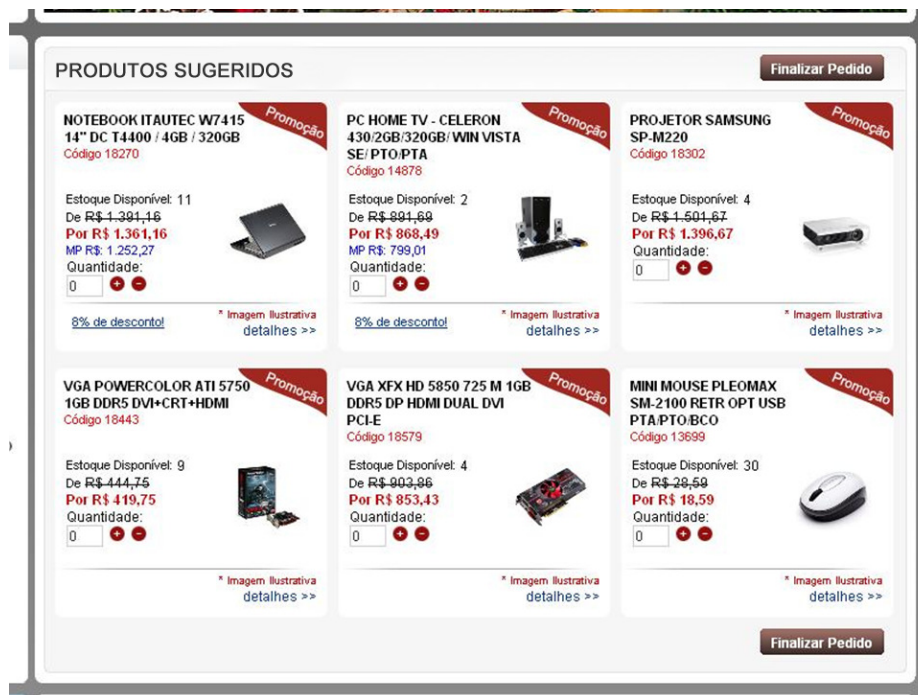


Figura 7: Área de exibição dos produtos sugeridos.

3.2. VERIFICAÇÃO

O processo de verificação do sistema consistiu em certificar-se de que, após todas as alterações realizadas para incluir o módulo de recomendação no sistema já existente na empresa Pauta, não houve impactos negativos no site como um todo, de modo a torná-lo inoperante ou danificado de certa forma. Após toda a etapa de implementação, a qual foi à responsável por todas as alterações de códigos-fonte necessários, a inclusão do sistema de recomendação foi verificado da seguinte forma:

- Foram utilizados para efeitos de testes os três principais navegadores atualmente, os quais são: Internet Explorer, Mozilla Firefox e Google Chrome;
- Quanto as versões, deve-se considerar o Internet Explorer 8, Mozilla Firefox 3.6 e Google Chrome 11, ou qualquer versão superior para ambos os navegadores; e
- Foram testadas todas as funcionalidades, não somente do módulo pertinente a este presente trabalho, mas sim de todo o *site*, utilizando-se os três navegadores citados anteriormente.

Por fim constatou-se que o novo módulo foi acoplado ao projeto do *site* como um todo, não trazendo dano algum a todas as funcionalidades já existentes.

3.3. VALIDAÇÃO

Nesta etapa de validação, foram realizados diferentes testes a fim de garantir que a solução completa produzisse os resultados esperados pelos usuários finais. A validação faz perguntas do tipo: “A aplicação faz o que as pessoas desejam que faça?”, “Está sendo construído o produto correto?”, entre outras.

Segundo a ISO/IEC 9126, para o processo de validação, tem-se seis características gerais a serem consideradas. São elas: Funcionalidade (satisfaz as necessidades?); Confiabilidade (é imune a falhas e estável?); Usabilidade (é fácil de usar?); Eficiência (é rápido e eficiente?); Manutenibilidade (é fácil de alterar e de manter?); e Portabilidade (é fácil



de ser utilizado em outro ambiente?). Com base nessas seis características, foram feitas as validações que seguem descritas à seguir.

Quanto à funcionalidade, pode-se dizer que foi atendida totalmente, pois a necessidade era que fossem exibidos na página principal, os últimos produtos comprados, e isso de fato está ocorrendo. E quando o total de produtos comprados não atingi o número mínimo necessário para exibição, que é seis, o *site* automaticamente completa o restante com uma consulta na lista de produtos em promoção.

Referente à confiabilidade, foram realizados diversos testes a fim de constatar-se que não havia falhas no módulo de recomendação de produtos desenvolvido, bem como, certificar-se que o mesmo estava estável, alcançando uma taxa de 100% de confiabilidade.

No quesito usabilidade, não houveram testes a serem feitos, devido o fato de que o módulo não implantou funcionalidades diferentes das que já existiam no *site*, ou seja, os usuários não tiveram necessidade de adaptação, pois o que mudou foi o tipo de produto exibido na página inicial do portal da empresa.

Sobre a análise da característica eficiência, não se tem muito que falar, pois foram realizados testes e aparentemente o módulo apresentou rapidez e eficiência, mantendo o padrão do *site* antes da introdução do módulo de recomendação, porém ainda se faz necessária uma bateria de testes mais minuciosos, tendo em vista a diversidade de bandas de *link*, versões dos três tipos de navegadores utilizados, servidores de *firewall*, entre outros.

4. CONCLUSÕES

A atividade de comércio eletrônico tem crescido cada vez mais e aumentado significativamente sua fatia de mercado, tendo em vista inúmeras facilidades e comodidades para os indivíduos que utilizam desta prática, seja para vender ou comprar produtos e/ou serviços. Tendo em vista todo esse crescimento exponencial na internet como canal de compra e venda, vários estudos vêm sendo realizados para que este serviço seja melhorado e conquiste cada vez mais adeptos. Com isso diversos conceitos e técnicas têm surgido a fim de aprimorarem os e-commerces, bem como, torná-los mais competitivos.

O sistema de recomendação apresentado neste trabalho apresentou um ganho para a empresa Pauta, no que diz respeito a “cativar” seus clientes. No que diz respeito ao desenvolvimento do sistema, a janela de tempo de análise para a recomendação constitui um fator complicador.

A janela de tempo utilizada para limitar a pesquisa dos pedidos das revendas, que havia sido definida como sendo de seis meses anteriores a data atual, mostrou-se inadequada para o problema, pois o volume de dados a serem processados se tornava extenso, o que estava ocasionando um *delay* muito grande entre o tempo de *login* do cliente e o tempo de carregamento da página inicial do *site* já com a área de produtos sugeridos inclusa. Também, devido ao fato de que a empresa Pauta é uma distribuidora, o volume de itens por pedido é geralmente muito maior do que se fosse para clientes finais, e o tipo de produto varia com facilidade, embalado pelas necessidades dos consumidores, entre outros, então se acreditou que se fosse trabalhado com um espaço de tempo de seis meses para realizar as análises, poderia se estar trabalhando com dados desatualizados, que possivelmente não fossem mais de interesse da revenda, por isso optou-se em alterar esta janela de tempo para três meses.

Como sugestão de trabalhos futuros, é evidenciada a possibilidade de se combinar diferentes técnicas de Sistemas de Recomendação a fim de tornar os resultados ainda mais satisfatórios. Outra possibilidade é utilizar essas técnicas para outras áreas do *site*, transformando-o cada vez mais com o perfil de cada cliente que logar.

Finalmente, nota-se a importância do presente trabalho para as empresas, para os analistas, para os desenvolvedores, para os pesquisadores, para os gerentes de projeto, enfim, à todos que estão envolvidos direta e indiretamente com o comércio eletrônico, pois sua tendência é evoluir a cada dia, e deve-se estar sempre em busca de aprimoramentos, de espaço no mercado e de um produto altamente competitivo.

5. REFERÊNCIAS

ALMEIDA JUNIOR, E. Comércio Eletrônico: e-Commerce. Disponível em: <http://www.consulting.com.br/edsonalmeidajunior/admin/downloads/comercioeletronico.pdf>> Acessado em: 30/09/2010.

BARAZETTI, M. C. Data Mining. Curitiba, 2009. Disponível em: <<http://www.batebyte.pr.gov.br/modules/conteudo/conteudo.php?conteudo=1289>>. Acessado em: 14/10/2010.

BRESSLER, F. Um Protótipo de Aplicação para Recomendação de Produtos Baseado no Interesse e Comportamento de uso do Usuário. São Leopoldo, 2004. Monografia do Curso de Informática. Universidade do Vale do Rio dos Sinos. Disponível em: <http://www.unisinos.br/inf/images/stories/Inf/24tc_fabio_bressler.pdf>. Acessado em: 13/10/2010.

FELIPINI, D. O comércio eletrônico B2B. (2006) Disponível em: <http://www.e-commerce.org.br/artigos/comercioeletronico_B2B.php>. Acessado em: 09/08/2010. Acessado em: 12/08/2010.

HORTINHA, J. E-Marketing: Um Guia para a Nova Economia. Edições Sílabo, 2000. Disponível em: <http://www.e-marketinglab.com/livro_emkt.html>. Acesso em: 16 ago. 2010.

KAJIMOTO, A.P. et al. Sistemas de Recomendação de Notícias na Internet Baseados em Filtragem Colaborativa. São Paulo, 2007. Trabalho de Formatura Supervisionado. Universidade de São Paulo. Disponível em: <http://www.ime.usp.br/~cef/mac499-07/monografias/allan-renatosidney-victor/monografia.pdf>>. Acessado em: 13/10/2010.

LIMA, D. S. et al. O Uso da Descoberta de Conhecimento em Base de Dados para apoiar a tomada de decisões. Simpósio de Excelência em Gestão e Tecnologia (SEGET), 2008. Disponível em: <http://www.economia.aedb.br/seget/artigos08/331_331_Artigo_SEGET_EJDR_67Versao_Final_010808.pdf>. Acessado em: 20/10/2010.

LUCIANO, E. M. Comércio Eletrônico de Produtos Virtuais: Definição de um Modelo de Negócios para a comercialização de software. XV Salão de Iniciação Científica da UFRGS, 2003. Disponível em: <<http://www.buscalegis.ufsc.br/arquivos/27-4.pdf>>. Acessado em: 30/09/2010.

LUPPI, I. Histórico do Comércio Eletrônico. Disponível em: <http://www.oficinadanet.com.br/artigo/1718/historico_do_comercio_eletronico>. Publicado em: 26/05/2009. Acessado em: 26/09/2010.

MARCHI, K. R. C. Sistemas de Recomendação: Uma abordagem geral. Maringá, 2008. Mestrado em Ciência da Computação. Universidade Estadual de Maringá. Disponível em: <<http://kessia.blogs.unipar.br/files/2008/07/sistemas-de-recomendacao.pdf>>. Acessado em: 13/10/2010.



REATEGUI, E. B.; CAZELLA, S. C. Sistemas de Recomendação. XXV Congresso da Sociedade Brasileira de Computação. Universidade do Vale do Rio dos Sinos (UNISINOS). São Leopoldo, 2005.

SILVA, I. A. F. Descoberta de Conhecimento em Base de Dados de Monitoramento Ambiental para Avaliação da Qualidade da Água. Cuiabá, 2007. Programa de Pós-Graduação em Física e Meio Ambiente. Universidade Federal de Mato Grosso.

SILVA, T. M. *et al.* Utilizando Sistemas de Recomendação na Criação de Comunidades Virtuais de Aprendizagem. 2006. Disponível em:
<<http://seer.ufrgs.br/renote/article/view/14178>>. Acessado em: 19/10/2010.

TRONCHONI, A. B. *et al.* Descoberta de Conhecimento em Base de Dados de Eventos de Desligamentos de Empresas de Distribuição. Porto Alegre, 2008. Artigo Científico submetido em: 07/01/2008 e última revisão em: 17/05/2009. Faculdade de Engenharia (PUCRS).

ULBRICH, H. C. *et al.* Guia de Comércio Eletrônico – Ferramentas Linux: e-commerce - Equipe Conectiva. Conectiva: Curitiba, 2000.