

Predição de Páginas Web Baseadas em Regras de Associação

Tatiane Crescencio¹, Cláudio Ratke²

¹ Pós-Graduanda em Tecnologia para Desenvolvimento de Aplicações Web - VII
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brasil

²Departamento de Sistemas e Computação
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brasil

{tatiane@inf.furb.br, ratke@furb.br}

Abstract. *The present paper aims at explaining the concepts of Data Mining and consequently, their application in the World Wide Web through Association Rules. For this reason the Apriori algorithm is used, which is implemented to identify and make the prediction of products groups to be included in the product catalog in the Cremer S.A. website.*

Resumo. *O presente artigo visa explicar os conceitos de Mineração de dados e conseqüentemente, sua aplicação na Web através das Regras de Associação. Para isto utiliza-se do Algoritmo Apriori, que é implementado a fim de identificar e fazer a predição de agrupamentos de produtos a serem disponibilizados no catálogo de produtos publicado na Web pela Cremer S.A..*

1. Introdução

Com o passar dos anos, tornou-se perceptível o crescimento exponencial da Internet, pela expansão de sua massa de informações juntamente com o número de usuários que dela participam. Um fator que ajudou neste crescimento foi a diminuição do custo de armazenamento e processamento de dados nos últimos anos (TURBAN et al, 2009).

Os avanços na tecnologia de coleta e armazenamento de dados permitiram que as organizações acumulassem uma vasta quantidade de dados. A extração de informações úteis, no entanto, tem provado ser extremamente desafiadora. Para Davenport e Prusak (1998) a informação é vista como dados que fazem diferença. Enquanto os dados são um conjunto de fatos, a “informação é uma mensagem, geralmente na forma de um documento ou uma comunicação audível ou visível. (DAVENPORT; PRUSAK, 1998).

A informação tem por finalidade mudar o modo como o destinatário vê algo, exercer algum impacto sobre seu julgamento e comportamento. Ela deve informar. O significado original da palavra “informar” é “dar forma a”, sendo que a informação visa a modelar a pessoa que a recebe no sentido de fazer alguma diferença em sua perspectiva (DAVENPORT; PRUSAK, 1998).

O grande problema das organizações é como transformar esses dados em informações valiosas. Por isto gerenciar a informação é primordial e torna-se uma vantagem competitiva. Silva e Tomaél (2007, p. 1) afirmam que “para ser utilizada estrategicamente, é fundamental que a informação seja gerida em favor da sobrevivência e competitividade organizacional.” A gestão da informação que possibilita as organizações gerir tanto os recursos internos quanto os externos à organização.

A gestão da informação engloba a prospecção, seleção e obtenção da informação; o mapeamento e reconhecimento dos fluxos formais de informação; o tratamento, análise e armazenamento da informação utilizando tecnologias de informação; a disseminação e mediação da informação ao público interessado, e a criação e disponibilização de produtos e serviços de informação (VALENTIM, 2002).

Para Mena (1999), a mineração de dados e a identificação de perfis de visitantes pode fornecer uma percepção sobre quais tipos de mensagens, banners, anúncios, promoções, incentivos, produtos e serviços que possam ou queiram ser colocados na linha de frente. Através da análise dos dados de um site, a empresa pode obter várias informações relevantes e começar a compilar um perfil de seus potenciais clientes ou usuários futuros. A mineração de dados da Web também é uma necessidade estratégica para a criação de uma relação duradoura com os clientes ou visitantes atuais e estabelecer um site rentável.

Segaran (2008) comenta que organizações, como Amazon, já tem como prática armazenar os dados de seus clientes em bases de dados, a fim de identificar possíveis informações relevantes e quando os mesmos retornam ao site, informando seus dados, tentam fazer alguma recomendação relacionada com compras passadas, utilizando algoritmos para classificação dos dados.

2. Mineração de Dados

A “mineração de dados, como indica o próprio nome, se refere à mineração ou descoberta de novas informações em função de padrões ou regras em grandes quantidades de dados” (ELMASRI; NAVATHE, 2005, p. 624).

Nos últimos anos, a mineração de dados tem atraído uma grande atenção da indústria de informação e da sociedade como um todo, devido à grande disponibilidade de grandes quantidades de dados e a necessidade iminente de transformar esses dados em informações úteis para tomada de decisão. As informações e conhecimentos adquiridos podem ser usados para aplicações que variam de mercado: exemplo análise de detecção de fraude e retenção de clientes (HAN; KAMBER, 2006).

Para auxiliar na descoberta de novas informações, existem várias técnicas de mineração de dados ou descoberta do conhecimento. Han e Kamber (2006) afirmam que a mineração de dados pode ser utilizada para controle de qualidade, marketing, comércio, análise de dados financeiros, indústria das telecomunicações, medicina, aplicações e pesquisas científicas diversas, detecção de fraudes e intrusão. Com essas técnicas a mineração de dados consegue focar nas variáveis mais importantes para a identificação dos padrões. O Quadro 1, a seguir, apresenta um resumo das técnicas de mineração de dados existentes.

Funções da Mineração de dados	Algoritmo	Exemplo de Aplicação
Associação	Estatística, teoria dos conjuntos.	Analisar conteúdo das cestas de compras
Classificação	Árvores de decisão, redes neurais, controle, avaliação de riscos e regras.	Marketing direcionado, qualidade.
Agrupamento	Redes neurais, estatística, otimização, análise discriminante.	Segmentação de mercado.

Descoberta de sequência	Estatística, teoria dos conjuntos.	Análise de cesta de supermercado ao longo do tempo, análise do ciclo de vida do cliente.
Modelagem	Regressão linear e não linear, ajuste da curva e redes neurais.	Previsão de vendas, nível de interesse, predição e controle de estoque.
Drill-down e visão agregada dos dados	Visualização, usando muitas abordagens diferentes.	Pode ser aplicada para quase todos os exemplos anteriores.
Lógica Fuzzy	Lógica nebulosa, permite modelagem do modo aproximado	Extensão lógica clássica, extensão dos conceitos imprecisos.

Quadro 1 Funções da mineração dos dados.

3. Regras de Associação

Uma das principais tecnologias em mineração de dados é a descoberta das regras de associação. Um banco de dados é considerado uma coleção de transações, cada uma envolvendo um conjunto de itens. O objetivo das regras de associação é encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados. O exemplo mais comum é o do carrinho de supermercado, onde o carrinho corresponde aos itens que um consumidor compra em um supermercado (ELMASRI; NAVATHE, 2005).

O algoritmo percussor das regras de associação é o Apriori, que busca encontrar os *itemsets* (conjunto de itens) frequentes e utilizar estes *itemsets* para gerar as regras de associação. Outros algoritmos como o FUP (Fast Update Algoritmo), Partition, FP-Growth e Zig-Zag são modificações da forma de execução do Apriori.

As regras de associação, através do Apriori,, neste caso, serão aplicadas para influenciar na estrutura das ligações entre as páginas web, facilitando a navegação e recomendando caminhos de navegação específicos ou ainda apresentar propagandas ou outras mensagens promocionais (VERCELLIS, 2009, p. 278).

3.1. Algoritmo Apriori

Este algoritmo foi proposto por R. Agrawal e R. Srikant em 1994, para mineração de *itemsets* para regras de associação. O algoritmo baseia-se no fato de que um *itemset* pode ser freqüente somente quando cada um dos seus subconjuntos é freqüente, caso contrário, o *itemset* é freqüente.

Na primeira passagem, o algoritmo Apriori constrói e conta todos os 1-*itemsets*. (Ak-*itemset* é um *itemset* que inclui itens k). Depois de ter encontrado todos os *itemsets* freqüentes 1, o algoritmo junta-se ao 1-*itemsets* freqüentes com os outros candidatos para formar 2-*itemsets*. Apriori varre o banco de dados de operação e contagem do candidato 2-*itemsets* para determinar qual dos 2-*itemsets* são freqüentes. Os outros passes são feitos em conformidade. Freqüente (k - 1)-*itemsets* são unidas para formar k-*itemsets* cujo primeiro k-1 itens são idênticos. Se k 3, Apriori ameixa alguns dos k-*itemsets*, destas, k (-1)-*itemsets* ter pelo menos um subconjunto freqüentes. Todas as restantes k-*itemsets* candidatos constituem k-*itemsets*. O processo é reiterado até que não exista mais nenhum candidato para ser gerado (CHANG; LI; LEE, 2005, p. 2)

A seguir, é apresentado um exemplo para melhor entendimento do conceito de regras de associação com base no Apriori. Este exemplo é uma adaptação de Tan et al. (2009) e Elmasri & Navathe (2005).

O quadro abaixo apresenta um exemplo prático de regras de associação. Cada linha representa uma transação e seu identificador e o conjunto de itens comprados por um determinado consumidor.

TID	Itens
1	Pão, Leite
2	Pão, Fraldas, Cerveja, Ovos
3	Leite, Fraldas, Cerveja, Cola
4	Pão, Leite, Fraldas, Cerveja
5	Pão, Leite, Fraldas, Cola

Quadro 2 - Exemplo de transações de cestas de compras.

No exemplo citado, observa-se que existe uma forte relação entre fraldas e cerveja (fralda => cerveja), sendo que, do total de cinco transações, quatro delas havia fraldas, e dessas quatro três havia cerveja. Com base nessas informações, pode-se afirmar que, dos consumidores que compram fraldas, 75% deles também tendem a comprar cerveja.

Seja $I = \{i_1, i_2, \dots, i_m\}$ um conjunto de itens e D um conjunto de transações, onde cada transação T é um conjunto de itens (*itemset*) tal que $T \subseteq I$. Associado a cada transação está um único identificador, chamado identificador de transação (TID). Assim, uma regra de associação é uma implicação na forma expressa por $X \Rightarrow Y$, onde $X \subseteq I, Y \subseteq I$ e $X \cap Y = \emptyset$.

A representação binária, zero ou um também pode ser usada para apresentar os dados, onde cada linha é equivalente a uma transação e cada coluna a um item, e os itens por sua vez assumem valores como verdadeiro (um) se estiver presente na transação ou falso (zero) se não estiver. (TAN; STEINBACH; KUMAR, 2009, p. 391). Assim o Quadro 2 poderia ser representado da seguinte forma no Quadro 3.

TID	Pão	Leite	Fraldas	Cerveja	Ovos	Cola
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

Quadro 3 - Exemplo de representação binária de transações realizadas.

“O contador de suporte é uma propriedade importante de um conjunto de itens, que se refere ao número de transações que contém um determinado conjunto de itens (k itens)” (TAN; STEINBACH; KUMAR, 2009, p. 392) como, no exemplo citado no Quadro 2, {fraldas} tem contador do suporte igual a quatro, pois está presente em quatro das cinco transações.

Em resumo, as regras de associação possuem duas medidas utilizadas para análise, o suporte e a confiança. O suporte determina a frequência na qual uma regra é aplicável a um determinado conjunto de dados, e a confiança determina a frequência na qual os itens em Y aparecem em transações que contenham X. (TAN; STEINBACH; KUMAR, 2009, p. 392).

O suporte da regra $X \Rightarrow Y$ é definido na expressão como $\text{sup}(X \Rightarrow Y) = \sigma(X \cup Y) / D$, ou seja, o contador de suporte em questão, dividido pelo total de transações, obtém-se o percentual do suporte da regra. E a confiança conf da regra $X \Rightarrow Y$ é definida na expressão como $\text{conf}(X \Rightarrow Y) = \sigma(X \cup Y) / \sigma(X)$, dividindo o contador do suporte em questão (dentre as transações utilizadas anteriormente), pelo contador do suporte utilizado para encontrar o suporte. (TAN; STEINBACH; KUMAR, 2009, p. 392).

A seguir, no Quadro 4 é representado em números o exemplo para a regra $\{\text{fraldas}\} \Rightarrow \{\text{cerveja}\}$, para encontrar o suporte e a confiança desta regra.

Regra: $\{\text{fraldas}\} \Rightarrow \{\text{cerveja}\}$ Suporte da regra: $\{\text{fraldas}\} = 4$ Total de transações = 5 $4 / 5 = 0,8$ (80% de suporte em relação ao banco de dados) Confiança da regra: $\{\text{cerveja}\} = 3$ $\{\text{fraldas}\} = 4$ $3/4 = 0,75$ (75% de grau de confiança para esta regra)

Quadro 4 - Cálculo para encontrar suporte e confiança de uma regra.

Uma regra de baixo suporte pode indicar uma simples coincidência e também não ser interessante dependendo do problema em questão, por isso o suporte é muitas vezes utilizado para limitar regras sem interesse. Por outro lado, a confiança, mede a confiabilidade da inferência feita por uma regra, seja uma regra $X \Rightarrow Y$. Neste caso, quanto maior a confiança, maior a probabilidade de que Y esteja presente em transações que possuam X. A confiança pode fornecer também uma avaliação da probabilidade condicional de Y dado X. (TAN; STEINBACH; KUMAR, 2009, p. 393).

4. Estudo de Caso

A Web contém informações interessantes sobre como inúmeras pessoas têm relacionamentos com uma página de Web. Cada visitante de um site, cada pesquisa em um mecanismo de busca, cada clique em um link e cada operação em um site de comércio eletrônico leva à criação de dados adicionais. A análise dessa informação pode identificar, por exemplo, a melhor forma de uso de um site e ajudar a fornecer um maior valor para os visitantes do mesmo (TURBAN, et al ,2009).

Pensando dessa forma, utilizou-se o algoritmo Apriori sobre uma amostra coletada de uma base de dados de um catálogo de produtos comercializados pela Cremer S.A.. Conforme o seu site para investidores, a Cremer S.A. é uma das empresas fornecedoras líderes de produtos para cuidados com a saúde nas áreas de primeiros socorros, cirurgia, tratamento e higiene. Esta amostra é constituída pelo histórico de acesso aos agrupamentos de produtos pelos visitantes do site nos últimos dois meses.

Este catálogo de produtos é composto por categoria ou canais de agrupamentos de produtos, fabricantes e agrupamentos de produtos. Conforme apresentado na Figura 1. Ao clicar em um agrupamento de produtos, é exibida uma lista com os respectivos códigos dos produtos associados.



Figura 1: Tela principal do catálogo de produtos.

Para que fosse possível identificar a relação entre os agrupamentos de produtos, foi necessário monitorar as páginas que os visitantes acessaram no site. Para isto, precisamos identificar e armazenar em base de dados quais os agrupamentos de produtos que cada visitante acessou no site em um determinado período.

Segundo MENA, 1999, existem vários métodos para capturar informações sobre os usuários em nível de servidor. Estes métodos incluem dados provenientes de arquivos de histórico, *cookies*, formulários, entre outros. Podem também ser usados scripts desenvolvidos em linguagens para Web como PHP, JavaScript, ASP, JSP, entre outras.

Para coleta e armazenamento dos dados dos visitantes do catálogo, ou seja, da sequência de cliques originada das consultas aos agrupamentos, foi utilizado um script desenvolvido na linguagem PHP. Este script tem a finalidade de identificar dados da sessão do usuário, bem como o endereço ip de origem, data e hora das consultas e através de comando SQL, armazenar os dados em uma tabela no banco de dados DB2.

Com base nos dados armazenados, foi aplicado um script SQL, este desenvolvido de acordo com o conceito do algoritmo Apriori, cuja sua finalidade é a de identificar de maneira rápida e precisa quais as ocorrências de combinação de um determinado agrupamento em relação a outro qualquer e contar quantas transações em que ambos apareceram juntos.

No Quadro 5, temos o agrupamento com código 2012, descrito como Lençol e fabricado pela POÇOSPEL. Para este agrupamento calculamos o suporte e a confiança com base em um conjunto de 2855 transações e apresentamos as quatro principais regras ou predições para este agrupamento, sendo os agrupamentos a seguir

apresentados por código, descrição e fabricante: 340 - Lençol Saúde – POÇOSPEL, 355 - Luva de Procedimento – LEMGRUBER, 786 - Toalha Interfolha Saúde – POÇOSPEL e 25 - Agulha Descartável – INJEX.

<p>Regra: {2012} => {340} Suporte da regra: {2012} = 2855 Total de transações = 20906 $2855 / 20906 = 13,66\%$ de suporte em relação ao banco de dados Confiança da regra: {340} = 214 {2012} = 2855 $214/2855 = 7,49\%$ de grau de confiança para esta regra</p> <p>Regra: {2012} => {355} Suporte da regra: {2012} = 2855 Total de transações = 20906 $2855 / 20906 = 13,66\%$ de suporte em relação ao banco de dados Confiança da regra: {355} = 52 {2012} = 2855 $52/2855 = 1,82\%$ de grau de confiança para esta regra</p>	<p>Regra: {2012} => {786} Suporte da regra: {2012} = 2855 Total de transações = 20906 $2855 / 20906 = 13,66\%$ de suporte em relação ao banco de dados Confiança da regra: {786} = 38 {2012} = 2855 $38/2855 = 1,33\%$ de grau de confiança para esta regra</p> <p>Regra: {2012} => {25} Suporte da regra: {2012} = 2855 Total de transações = 20906 $2855 / 20906 = 13,66\%$ de suporte em relação ao banco de dados Confiança da regra: {25} = 35 {2012} = 2855 $35/2855 = 1,23\%$ de grau de confiança para esta regra</p>
---	--

Quadro 5 - Cálculo para encontrar suporte e confiança das regras propostas para 2012 - Lençol – POÇOSPEL.

Como podemos perceber, o suporte foi baixo para esta situação. O volume transações é alto, porém, a empresa disponibiliza através do site mais de seis mil produtos ativos, o que pode representar uma gama imensa de combinações. Entretanto, o Quadro 6 a seguir, apresenta o agrupamento mais visitado pelos visitantes contido na amostra: 514 - Fralda Cremer com Bainha – CREMER.

<p>Regra: {514} => {515}</p> <p>Suporte da regra: {514} = 3500 Total de transações = 20906 $3500 / 20906 = 16,74\%$ de suporte em relação ao banco de dados</p> <p>Confiança da regra: {515} = 642 {514} = 3500 $642/3500 = 18,34\%$ de grau de confiança para esta regra</p> <p>Regra: {514} => {296}</p> <p>Suporte da regra: {514} = 3500 Total de transações = 20906 $3500 / 20906 = 16,74\%$ de suporte em relação ao banco de dados</p> <p>Confiança da regra: {296} = 325 {514} = 3500 $325/3500 = 9,29\%$ de grau de confiança para esta regra</p>	<p>Regra: {514} => {298}</p> <p>Suporte da regra: {514} = 3500 Total de transações = 20906 $3500 / 20906 = 16,74\%$ de suporte em relação ao banco de dados</p> <p>Confiança da regra: {298} = 277 {514} = 3500 $277/3500 = 7,91\%$ de grau de confiança para esta regra</p> <p>Regra: {514} => {293}</p> <p>Suporte da regra: {514} = 3500 Total de transações = 20906 $3500 / 20906 = 16,74\%$ de suporte em relação ao banco de dados</p> <p>Confiança da regra: {293} = 271 {514} = 3500 $271/3500 = 7,74\%$ de grau de confiança para esta regra</p>
--	---

Quadro 6 - Cálculo para encontrar suporte e confiança das regras propostas para 514 - Fralda Cremer com Bainha – CREMER.

Complementando o Quadro 5, podemos concluir que para o agrupamento 514 - Fralda Cremer com Bainha – CREMER foram feitas quatro predições, sendo elas os agrupamentos: 515 - Fralda Luxo – CREMER, 296 - Fralda Especial Cremer – CREMER, 298 - Fralda Pinte e Borde – CREMER e 293 - Fralda Colorida Cremer – CREMER.

O resultado prático e visual destas análises, já implementadas no site do catálogo, podem ser vistos a seguir respectivamente nas Figuras 2 e 3:



Figura 2: Predições para o agrupamento 2012 - Lençol - POÇOSPEL

The screenshot shows the 'Catálogo de Produtos' page for Cremer. On the left, there is a search bar and a list of channels: Destaques, Hospitalar, Varejo, Dental, Saúde Animal, Fitas Industriais, and Exportação. The main content area features a product titled 'Fralda Cremer com Bainha - CREMER'. Below the title is a description in Portuguese: 'A Fralda Cremer com Bainha é confeccionada com tecido duplo de fibras 100% algodão no formato 70cm x 68cm e possui costura em duas das laterais para evitar o desfiamento. Extremamente macia e absorvente, é ideal para uso noturno. Devido a sua alta capacidade de reter líquido, permite uma troca mais prolongada, o que proporciona maior conforto e segurança ao bebê, deixando-o sequinho por mais tempo.' Below the description is the product code '332803 - FRALDA LUXO BRANCA 70X68-BAINHA C/ 5UN' and a 'Veja também' section with four product thumbnails: 'Fralda Luxo CREMER', 'Fralda Especial CREMER', 'Fralda Pinte e Borde CREMER', and 'Fralda Colorida CREMER'. The footer contains the copyright information: 'Copyright - Cremer S.A. Área de Informática'.

Figura 3: Predições para o agrupamento 514 - Fralda Cremer com Bainha - CREMER

Uma situação interessante é a apresentada na Figura 2: para o agrupamento 2012 - Lençol – POÇOSPEL foram feitas quatro predições, onde duas delas poderia se dizer que não fariam sentido, uma vez que se pensasse o que tem haver lençol com luva e agulha? Pois bem, para a empresa em questão isso representa provavelmente visitantes do ramo hospitalar, como diretores ou compradores responsáveis pelo estoque de produtos em um hospital.

5. Conclusão

Dadas as características próprias e o cenário da pesquisa, pode-se dizer que a mesma não se aplica a sites que tem seu conteúdo gerado de forma dinamicamente, mas nada impede que seja desenvolvida alguma outra forma aplicando os mesmos conceitos.

Um ganho significativo para a empresa com a aplicação desta estratégia foi apresentado nas estatísticas de acesso do catálogo. Através da ferramenta da Google, conhecida como Google Analytics, é possível monitorar o tráfego de dados do catálogo. Para esta análise, utilizou-se dados obtidos no mês de Agosto (mês este em que o site não possuía a implementação do Algoritmo) e de Setembro, exatamente um mês após a implementação do algoritmo e publicação das predições. Outro fator importante é que nesta análise não estão considerados os endereços ips utilizados pela companhia para acesso ao catálogo.

Embora um mês não ser um tempo muito longo, observou-se um crescimento nos acessos de 3,11% um mês após a implantação. Junto com este crescimento, observou-se que houve uma queda de 6,78% das *Page views* do site, bem como o tempo médio de permanência no site caiu aproximadamente 26,92% juntamente com o número de páginas por visitante que obteve queda de 9,56%. Isto significa que os visitantes passaram a encontrar as informações importantes para eles com mais agilidade, diminuindo assim a utilização dos mecanismos de busca do catálogo.

Apesar do crescimento dos acessos, até a conclusão deste artigo, não foram registradas alterações no comportamento das regras apresentadas nos exemplos. Isto indica que o suporte e a confiança para estes agrupamentos de produtos apresentados estão crescendo ainda mais.

6. Referências

- AGRAWAL, Rakesh et al. Fast discovery of association rules. In: FAYYAD, Usama M. et al. *Advances in knowledge discovery and data mining*. Menlo Park: AAAI: MIT, 1996. p. 307-328.
- CARVALHO, Elizabeth Leão de. Importância da gestão da informação para o processo decisório nas organizações. In: VALENTIM, Marta Lígia Pomim (org.). *Informação, conhecimento e inteligência organizacional*. Marília: FUNDEPE ED, 2006. p. 81-98.
- CHANG, Chin-Chen; LI, Yu-Chiang; LEE, Jung-San. An Efficient Algorithm for Incremental Mining of Association Rules. *Proceedings of the 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications*. IEEE Computer Society v. 1, p. 3-10, 2005.
- CREMER.S.A. Relações com investidores. Disponível em: <<http://www.cremer.com.br/ri>>. Acesso em: 09 out 2010.
- DAVENPORT, Thomas H; PRUSAK, Laurence. *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. Rio de Janeiro: Campus, 1998. 237p.
- ELMASRI, Ramez; NAVATHE, Shamkant B. *Sistemas de banco de dados*. 4. ed. São Paulo: Pearson Education do Brasil: Addison Wesley, 2005. xviii, p. 724.
- HAN, Jiawei; KAMBER, Micheline. *Data Mining: Concepts and Techniques*, Second Edition. 2006.
- MENA, Jesus. *Data Mining your website*. Boston: Digital Press, 1999. 368p.
- SEGARAM, Toby; *Programando a Inteligência Coletiva: Desenvolvendo Aplicativos WEB 2.0 Inteligentes*. Rio de Janeiro: Alta Books, 2008. 282p.
- SILVA, Terezinha Elisabeth da; TOMAÉL, Maria Inês. A gestão da informação nas organizações. *Informação e Informação*, Londrina, v. 12, n. 2, p. 1-2, jul./dez. 2007.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. *Introdução ao data mining: mineração de dados*. Rio de Janeiro: Ciência Moderna, 2009.
- TURBAN, Efraim et.al. *Business intelligence: um enfoque gerencial para a inteligência do negócio*. Porto Alegre : Bookman, 2009. xi, 253 p, il. (Gestão do conhecimento/sistemas de informação).
- VERCELLIS, Carlo. *Business Intelligence: Data Mining and optimization for decision making*. Local: Wiley, 2009. 436p.