

# Revista Eletrônica de Sistemas de Informação

## ISSN 1677-3071

No 2 (7)

2008

---

### Sumário

#### Editorial

Primeira edição no SEER

*Alexandre Reis Graeml*

Sobre o conteúdo desta edição

*Alexandre Reis Graeml*

---

#### Artigos

Comunicação Organizacional e Processo Decisório em Cooperativa

*André Luis Marra do Amorim, Gustavo Abib, Sergio Bulgacov*

Decorrências da Utilização de Tecnologia de Informação Móvel e Sem Fio em uma Concessionária de Energia

*Oswaldo Garcia, Maria Alexandra Viegas Cortez da Cunha*

Os Processos Produtivos e as Exportações da Indústria Brasileira de Software

*Marcia May Gomel, Roberto Sbragia*

A importância dos sistemas de informação para o planejamento de ações e políticas de saúde nos PSFS do município de Porto Alegre

*Licélia de Lima Mendonça, Marie Anne Macadar*

Data Mining – Aplicação em uma Base de Dados Real com Dados de Usuários dos Faróis do Saber

*Samuel Diego Sandri*

Indicadores para o uso de sistemas interorganizacionais e relacionamento em rede entre empresas.

*Marco Antônio Pinheiro da Silveira, Elisabete Camilo Rigolon Lança, André Luiz Pestana, Cláudio Pereira dos Santos*

Sistemas de Informação de Apoio à Decisão e suas Relações com Processos Decisórios: Estudo de Caso em um Município Paranaense

*Leandro Wzorek, Denis Alcides Rezende, Jefferson Marcel Gross Mendes*

---

#### Ensaios

Padrões Tecnológicos e a Produção Social da Tecnologia

*Milton Adrião*

Afinal, o que é Business Process Management (BPM)? Um novo conceito para um novo contexto

*Manoel Veras Sousa Neto, Josué Vitor Medeiros Junior*



# Data Mining – Aplicação em uma Base de Dados Real com Dados dos Usuários dos Faróis do Saber

(artigo submetido em 26/10/2008)

**Samuel Diego Sandri**

Universidade Tecnológica Federal do Paraná

sdsandri@yahoo.com.br

## **ABSTRACT**

*This paper presents the results of the experimentation with Apriori, a data mining software, in an attempt to demonstrate its effective use as a decision support tool. The data that was analyzed was related to a survey that was carried out with users of Internet access facilities freely provided by the local government to the citizens of Curitiba, in the so called "Faróis do Saber". Some interesting patterns were detected, such as the relationship of the users' age, income and schooling with the reasons for which they use the Internet.*

*Key-words: telecenters; data mining.*

## **RESUMO**

Este trabalho realiza um estudo sobre data mining, demonstrando a sua eficácia na análise de dados, além de verificar o seu efetivo resultado no suporte à decisão. Foram analisados os dados contidos em uma base de dados, contendo registros relacionados aos resultados de uma pesquisa sobre o uso da internet nos Faróis do Saber da cidade de Curitiba. Foi possível identificar padrões interessantes escondidos nos dados brutos como a relação entre faixa etária, renda e escolaridade dos usuários com a finalidade para a qual a internet é utilizada.

Palavras-chave: telecentros; mineração de dados.

## **1 Introdução**

Antigamente, as organizações possuíam preocupações mais concentradas em seus concorrentes diretos, que disputavam, na grande maioria dos casos, a liderança pelo mercado regional. Contudo, após o fenômeno da globalização, a fronteira de atuação das corporações passou a se estender a todos os continentes. Em função disso, a tarefa de administrar tornou-se algo muito mais complexo e valioso. Uma boa administração e a capacidade de tomar decisões coerentes são ingredientes básicos para quem quer sobreviver, crescer e se perpetuar no mercado.

Até certo período da história, os administradores eram capazes de processar as informações necessárias à administração das empresas de forma satisfatória, sem a necessidade de recursos tecnológicos sofisticados. Tanto

a coleta de dados, quanto a sua análise e interpretação eram executadas sem maiores dificuldades. Os problemas começaram a aparecer quando o número de informações coletadas se expandiu drasticamente e a velocidade com a qual deveriam ser acessados e analisados, em contrapartida, diminuiu (AURÉLIO; VELLASCO e LOPES, 1999).

Inicialmente, a criação dos bancos de dados propiciou uma enorme facilidade, já que dinamizou o armazenamento de registros e sua busca. Os profissionais que necessitavam acessar as informações podiam fazê-lo mais rapidamente, sem precisar vasculhar papéis e arquivos físicos. Mais adiante, percebeu-se que os dados contidos nestas bases poderiam conter padrões de informações que, após identificados e interpretados, seriam capazes de gerar conhecimento.

Freitas (*apud* GONÇALVES, 2001) diz que o ser humano possui limitações para analisar grandes quantidades de dados, portanto a alternativa para se tratar a informação contida nos repositórios eletrônicos que estavam sendo criados pela própria informatização das empresas e dos seus negócios veio, novamente, da tecnologia da informação, com a criação da ferramenta de mineração de dados ou *data mining*. Este novo tipo de *software* tem por função vasculhar bases de dados e descobrir informações que, por si sós, não se tornariam evidentes de forma clara.

Informações são algo útil gerado a partir de dados (REZENDE, 2005). Com base nelas, decisões podem ser tomadas. Quanto mais dados disponíveis houver, mais se justificam os investimentos em ferramentas de tecnologia que possam auxiliar as organizações a tratá-los, transformando-os em informação, primeiramente, e em decisões que possam auxiliar as empresas a obter bom desempenho no mercado, na seqüência.

## 2 Data Mining

Em um mercado que necessita de decisões rápidas e confiáveis é quase impossível uma organização se destacar, e até mesmo sobreviver, sem a utilização das ferramentas disponibilizadas pela tecnologia da informação, que podem proporcionar vantagem competitiva, segundo Oliveira (*apud* GONÇALVES, 2001). Os bancos de dados são uma das inúmeras vantagens que a TI fornece a seus usuários, já que o armazenamento convencional de documentação é oneroso e nada prático para consulta.

Contudo, os dados armazenados nestas bases não contêm apenas valores, características ou nomes isolados para eventuais pesquisas sobre o his-

tórico de um cliente, por exemplo. Eles também possuem informações que, quando cruzadas, podem possibilitar a identificação de padrões que as auxiliem a decidir que direção seguir ou que produto priorizar. Também é possível, a partir da análise das informações de mercado, antecipar as ações dos concorrentes, estabelecer formas de se aproximar dos clientes e até mesmo definir se é necessário ou não mudar o foco do negócio.

De acordo com Harrison (1998), o *data warehouse* (depósito de dados) fornece memória, mas não inteligência. A mineração de dados é uma ferramenta que permite adicionar inteligência à memória que a organização possui, permitindo realizar descobertas importantes dentro de grandes volumes de dados.

Processo de descoberta de conhecimento em base de dados

Hoje, a tecnologia existente para a coleta e armazenamento de dados é muito mais avançada do que a tecnologia de processamento desses dados. Partindo-se do princípio de que as coleções de dados já existentes são fontes de conhecimento útil que pode ser explorado, as pesquisas em descoberta do conhecimento em base de dados (KDD – *Knowledge Database Discovery*) crescem e atraem muitos esforços (FELDENS, 1997).

A KDD é um processo de natureza interativa e iterativa, pois é composto de etapas seqüenciais, podendo existir retorno às etapas anteriores, ou seja, descobertas podem levar a novas hipóteses de descoberta. Além disso, ainda segundo Feldens (1997), o usuário pode retomar à mineração dos dados ou realizar uma nova seleção de atributos para, por exemplo, validar uma nova hipótese surgida durante o processo.

Os desenvolvedores projetam as bases de dados, descrevem problemas e definem objetivos, enquanto os computadores processam os dados e procuram padrões que coincidam com os objetivos estabelecidos. Isto faz da KDD

um processo cooperativo, segundo Brusso (1998). A Figura 1, a seguir, mostra as etapas do processo de KDD, segundo Fayyad (*apud* GONÇALVES, 2001).

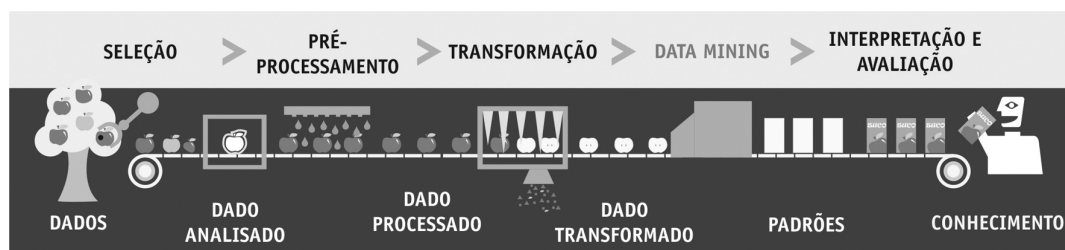


Figura 1 Processo de KDD (FAYYAD, *apud* GONÇALVES, 2001)

De acordo com Figueira (1998), podem-se dividir as etapas de KDD da seguinte forma:

- Seleção: é a etapa de agrupamento (organizado) dos dados;
- Pré-processamento: neste momento os dados passam por uma adequação. Ao final do processo, devem possuir o formato correto e não apresentar duplicidade, entre outras características;
- Transformação: é a etapa de armazenamento dos dados de forma a facilitar o uso das técnicas de *data mining*;
- *Data mining*: é a principal atividade da descoberta do conhecimento, aplicando, para este fim, algoritmos de descoberta de padrões;
- Interpretação e avaliação: como o nome diz, esta fase consiste em interpretar os dados gerados e verificar se possuem alguma validade para o problema proposto.

### 3 Metodologia

Após a análise teórica das possibilidades da ferramenta *data mining*, o próximo passo se constituiu na aplicação de um programa que utiliza esta tecnologia, para identificar a efetiva

capacidade de descobrimento de padrões interessantes. Para tanto, foi escolhido o programa *Apriori*, que faz parte de uma iniciativa de pesquisadores da Universidade de Waikato, na Nova Zelândia. Também foi definida a base de dados a ser utilizada no exercício, a qual possui registros coletados por meio de um questionário respondido por mais de 500 usuários dos Faróis do Saber, em Curitiba.

Os Faróis do Saber proporcionam aos seus usuários oportunidade de empréstimo de livros (originalmente, esses equipamentos urbanos foram concebidos como bibliotecas), local para realização de trabalhos escolares e disponibilizam acesso à internet para os cidadãos. Geralmente, são localizados próximos a escolas municipais, possuindo um vínculo com a instituição de ensino. Outras vezes, situam-se em praças ou outras áreas públicas, com vistas a proporcionar atendimento à população em geral.

O processo de montagem, validação e aplicação foi realizado pelo criador do questionário, já que faz parte da sua pesquisa de doutoramento. Para facilitar o entendimento das etapas de extração do conhecimento, a partir da base de dados obtida com a aplicação



do questionário, elas são descritas nos tópicos a seguir.

### 3.1 Pré-processamento

Após verificar a base de dados existente, houve a necessidade de realizar a limpeza dos registros, bem como a eliminação de alguns campos. Com a limpeza, foi possível padronizar a forma como os dados estavam assinalados, corrigindo respostas em branco ou itens fora do formato escolhido. Este procedimento permitiu que os registros fossem manipulados sem duplicidade e dentro de padrões estabelecidos, evitando inconsistências e ruídos, conforme afirma Gonçalves (2001).

Os atributos retirados foram os campos *nome do respondente* e *período em que respondeu o questionário*, uma vez que não eram úteis para a

análise. Além disso, o campo *bairro do respondente* foi alterado para *região do respondente*, pois, desta forma, pôde-se agrupar os respondentes em apenas nove regiões, que compõem os distritos da cidade de Curitiba.

### 3.2 Transformação

Antes da aplicação do algoritmo de mineração de dados, foi executada a transformação dos registros. Isso significa que a forma como eles são apresentados foi modificada para que o algoritmo os reconhecesse e identificasse os padrões neles existentes.

A função utilizada para efetuar tal modificação foi o concatenamento de células, presente no programa *Excel*. Para exemplificar esta etapa, apresenta-se a tabela seguinte.

Tabela 1 Dados normais e modificados

<b><i>Campo Normal</i></b>	<b><i>Campo Modificado</i></b>
Profissao	Profissao
Estudante	Profissao_Estudante
Estagiario	Profissao_Estagiario

Após estas alterações, o arquivo, que estava em formato XLS, foi salvo no formato CSV, que apresenta configuração de texto sem formatações, finalizando a etapa de transformação.

### 3.3 Ferramenta escolhida

Como já foi mencionado, o *software* utilizado para a mineração da base em questão foi o *Apriori*. Este programa é gratuito e está disponível para *download* na internet, junto com seu manual do usuário. Algumas características importantes do *Apriori* são:

- Trabalha com apenas 1 (um) conseqüente e N antecedentes. Isto se deve ao fato de estudos mostrarem que as informações mais relevantes e oportunas são aquelas em que o conseqüente é único. Os antecedentes são re-

gistros da base de dados que implicam em outro registro distinto da mesma base (conseqüente). Por exemplo, se o antecedente é composto por pão e manteiga e o conseqüente por leite, pode-se dizer que os consumidores que compram pão e manteiga também compram leite;

- Permite a configuração do *suporte* e da *confiança*, que, no seu padrão, são 10% e 80% respectivamente. Os resultados são mostrados pelo programa apenas se atenderem aos requisitos pré-definidos, ou seja, se o valor destas duas variáveis for igual ou superior aos limites mínimos configurados. Considerando-se a expressão (A→B), o algoritmo entende *suporte* e *confiança* como se apresenta no Quadro 1.

Quadro 1 - Suporte e confiança no Apriori

Suporte = Número de A / Número Total (Facilita a compreensão)

Confiança = Número de A e B / Número de A

- É uma ferramenta que pode apresentar padrões corretos, úteis e compreensíveis, facilitando sua aceitação pelos gestores.

### 3.4 Mineração dos dados

A etapa seguinte, no processo de KDD, é a mineração da base em si. Contudo, alguns problemas surgiram e acabaram dificultando o *data mining*. As dificuldades encontradas e as soluções adotadas para resolvê-las estão separadas e explicadas na seqüência.

#### 3.4.1 Número de campos da base

Como a base de dados apresentava, originalmente, um número de campos elevado (72 colunas), o arquivo gerado pelo processo de *data mining* ultrapassou, em uma primeira abordagem, o tamanho de 4GB e o procedimento não pôde ser completado. A solução para este primeiro empecilho foi aumentar o *suporte* utilizado, que, originalmente, havia sido mantido no valor *default*, ou seja, 10%.

Para gerar um arquivo passível de análise o valor do *suporte* foi configurado para 70%. Por este valor ser alto, dados que poderiam ser interessantes foram desconsiderados, gerando então o segundo problema.

#### 3.4.2 Descoberta de padrões com suporte baixo

No caso específico deste estudo, a descoberta de padrões com suporte baixo também era interessante. Um exemplo que ilustra esta necessidade pode ser a descoberta do perfil de pessoas que compram bebidas importadas e de alto valor em supermercados, o que exige um valor de suporte baixo

visto que são poucos clientes que adquirem tais produtos.

A solução, para tanto, foi recorrer à segmentação da base de dados. Verificaram-se quais eram os registros mais relevantes para o estudo, permitindo que minerações fossem executadas em conjuntos de dados específicos e com maior interesse para os objetivos da pesquisa que determinou a coleta e análise dos dados.

Este procedimento gerou mais duas bases, portanto, ao final da mineração de dados, o resultado foi a criação de três bases para análise, sendo que as duas derivadas da principal puderam ser trabalhadas com suporte igual a 10%, que já contemplava as necessidades relacionadas aos objetivos originais da pesquisa.

## 4 Resultados

O próximo passo consistiu na interpretação e avaliação dos resultados obtidos pela análise da base principal, criada com suporte 70% e confiança 80%, e das duas sub-bases criadas posteriormente, para as quais foi utilizado suporte 10% e confiança 80%. Antes de partir para a apresentação dos resultados e da análise, é necessário mostrar a forma como o *Apriori* apresenta os seus resultados:

Conseqüente  $\leq$  Antecedente(s) (suporte, confiança)

Com isto esclarecido, passa-se, a seguir, à análise dos resultados obtidos a partir da aplicação do algoritmo aos dados da base. Os padrões significativos extraídos são apresentados nos quadros, a seguir, com as regras descobertas e conclusões sobre elas.

Quadro 1 - Conjunto de padrões 01

- 1 - Mot\_Enc\_Amigos\_Nao <- RendaAte1Sal (17.5, 91.9)
- 2 - Mot\_AcessoInternet\_Sim <- Renda1a3Sal (42.2, 89.4)
- 3 - Mot\_AcessoInternet\_Sim <- Renda3a9Sal (21.8, 81.3)
- 4 - Mot\_AcessoInternet\_Sim <- EnsinoMedioIncompleto (26.3, 85.3)
- 5 - Mot\_AcessoInternet\_Sim <- EnsinoMedioCompleto (16.3, 91.3)

O nível de renda dos frequentadores dos Faróis do Saber é, em geral, baixo. 17,5% deles são sustentados com até 1 salário mínimo (regra 1) e quase 60% dos utilizadores tem renda inferior a 3 salários mínimos (regra 1 + regra 2).

Verifica-se que a escolaridade dos utilizadores está entre o ensino médio incompleto e o completo, em 42,6% dos casos (regra 4 e regra 5). Chama a atenção o fato de 16,3% dos usuários

terem completado o ensino médio, mas não estarem cursando o superior (regra 5). Nos casos acima (regras 2 a 5) se percebe que o uso da internet é um importante motivo para a visita ao Farol do Saber. Contudo, mais uma vez é importante lembrar que a pesquisa foi realizada com pessoas que estavam utilizando os computadores nos Faróis do Saber e não com todos os usuários desse espaço público.

Quadro 2 - Conjunto de Padrões 02

- 1 - Tipo\_Farol\_escola <- FaixaEtaria10\_19 (71.1, 81.9)
- 2 - Tipo\_Farol\_escola <- SabiaUsarInt\_Sim (74.9, 80.4)
- 3 - Tipo\_Farol\_escola <- Out\_Loc\_Casa\_Nao (76.8, 82.0)
- 4 - Tipo\_Farol\_escola <- Red\_Uso\_Prefeitura\_Nao (93.9, 80.9)
- 5 - Tipo\_Farol\_escola <- Red\_Uso\_Banco\_Nao (93.3, 80.3)
- 6 - Out\_Loc\_Não\_SoFarol\_Nao <- Red\_Uso\_Cinema\_Nao (94.5, 80.4)
- 7 - Out\_Loc\_Escola\_Nao <- Red\_Uso\_Supermercados\_Nao (95.3, 82.5)
- 8 - Out\_Loc\_Escola\_Nao <- Red\_Uso\_CasaAmigo\_Nao (81.3, 82.5)
- 9 - Out\_Loc\_Escola\_Nao <- Red\_Uso\_Biblioteca\_Nao (83.9, 84.0)

Percebe-se que os usuários estão, em 71,1% dos casos, na faixa etária entre 10 e 19 anos. Destes, 81,9% utilizam os faróis do tipo escola, ligados diretamente a uma instituição de ensino (regra 1). O curioso, neste caso, é que todos os respondentes do questionário participaram da pesquisa porque estavam utilizando um computador no farol, quando abordados pela pesquisadora, ou seja, há gente que utiliza a infraestrutura computacional disponibilizada para outros fins, como digitar um texto, atualizar e imprimir um currículo, etc. 74,9% dos participantes declararam já saber usar a internet, antes de começar a utilizá-la nos Faróis do Saber (regra 2), embora 76,8% deles não possuam acesso à internet a partir de casa (regra 3).

Nota-se, também, que, com a utilização da internet a partir dos Faróis do Saber, não há uma redução significativa no número de vezes que as pessoas vão à prefeitura, ao banco, ao cinema e ao supermercado (regras 4 a 7). Contudo, quando se fala em ir à casa de amigos (regra 8) ou à biblioteca (regra 9), a redução chega a 18,7% e 16,1%, respectivamente.

A maior parte destas conclusões está relacionada apenas à análise da frequência com que o antecedente aparece na base de dados, ou seja, ao *suporte* obtido pela variável em estudo, que é o primeiro dos números apresentados dentro dos parênteses.



Quadro 3 Conjunto de padrões 03

1 - Mot\_TrabalhoEscolar\_Nao <- EnsinoMedioCompleto Mot\_Estudar\_Nao (10.8, 81.1)  
 2 - Mot\_Outro <- EnsinoMedioCompleto Mot\_AcessoInternet\_Sim (14.9, 93.2)  
 3 - Mot\_Enc\_Amigos\_Nao <- EnsinoFundamentalIncompleto Mot\_AcessoInternet\_Sim (30.3, 91.3)  
 4 - Freq\_Usa\_Curriculo\_Nunca <- EnsinoFundamentalIncompleto (35.2, 87.9)  
 5 - Mot\_Ler\_Nao <- EnsinoMedioIncompleto Mot\_Estudar\_Nao (19.1, 84.0)  
 6 - Mot\_Enc\_Amigos\_Nao <- EnsinoMedioIncompleto Mot\_AcessoInternet\_Sim (22.4, 85.5)  
 7 - Mot\_AcessoInternet\_Sim <- EnsinoMedioIncompleto (26.3, 85.3)  
 8 - Mot\_AcessoInternet\_Sim <- EnsinoFundamentalInc Mot\_Est\_Nao Mot\_Ler\_Nao (18.1, 85.4)  
 9 - Mot\_Outro <- EnsinoFundamentalInc Mot\_TrabEsc\_Sim Mot\_Est\_Nao Mot\_Ler\_Nao (11.6, 100.0)  
 10 - Mot\_Outro <- EnsinoFundamentalInc Mot\_TrabalhoEscolar\_Sim Mot\_AcessoInternet\_Sim (20.2, 100.0)  
 11 - Mot\_Outro <- EnsinoMedioInc Mot\_Estudar\_Nao Mot\_Ler\_Nao Mot\_AcessoInternet\_Sim (14.1, 97.1)  
 12 - Freq\_Usa\_ComprasInternet\_Nunca <- Freq\_Usa\_OportEmprego\_Nunca Freq\_Usa\_PesqPreco\_Nunca Freq\_Usa\_AcoesGov\_Nunca Freq\_Usa\_PagContas\_Nunca (31.8, 98.1)  
 13 - Freq\_Usa\_CursosOnline\_Nunca <- Freq\_Usa\_InscConcurso\_Nunca Freq\_Usa\_Batepapo\_Nunca Freq\_Usa\_Reclamacao\_Nunca Freq\_Usa\_ServGov\_Nunca (30.3, 95.3)

A regra 1 mostra que 10,8% dos participantes possui ensino médio completo e não vão ao Farol do Saber para estudar. Deste grupo, 81,1% declaram que não usam os computadores para realizar trabalhos escolares. Um número relevante é a porcentagem de 30,3% de pessoas com ensino fundamental incompleto que vão ao Farol para acessar a Internet (regra 3), seguido de 22,4% de usuários com ensino médio incompleto que fazem a mesma coisa (regra 6), pois estes representam, respectivamente, 35,2% e 26,3% da amostra (regras 4 e 7).

Existe um valor elevado, para cada faixa de escolaridade, de usuários que não usam o Farol do Saber para estudos ou leitura (regras 8 a 11). Ou seja, a Internet, que poderia ser um valioso meio de aprendizado, não é manuseada com esta finalidade. Um comportamento, quase padrão, demonstra que pessoas que nunca utilizam o computador para executar atividades como serviços, compras, informações do governo, entre outras, também não o usam para outras funções correlacionadas (regras 12 e 13).

## 5 Conclusão

Devido à grande quantidade de dados armazenados, o uso de ferramentas de tecnologia da informação torna-se prático e necessário na busca

pela informação e pelo conhecimento. O *data mining* é realmente útil, quando se pretende extrair padrões em bases de dados. Entretanto, sua aplicação exige um trabalho minucioso e bem estudado sob a perspectiva do problema a ser tratado.

As fases de pré-processamento e transformação são essenciais para que o algoritmo utilizado possa reconhecer os dados, analisá-los corretamente e fornecer resultados passíveis de compreensão. Além disso, permitem que registros desnecessários sejam eliminados e outros relevantes sejam preenchidos corretamente.

Mesmo após a conclusão destas duas etapas o processo de *data mining* ainda está sujeito a problemas, como constatado na primeira mineração da base utilizada neste estudo. A solução envolveu a divisão da base principal em outras menores, com informações consideradas essenciais. Este procedimento comprovou que a participação de quem coletou e tem mais experiência com os dados é muito importante, por possuir maior domínio do assunto estudado e ter condições, portanto, juntamente com o responsável pela mineração, de decidir como resolver os problemas que porventura ocorram, com o menor impacto sobre a qualidade da análise proporcionada.

Finalizada a extração dos padrões, para que eles realmente se transformem em conhecimento, devem atender a três quesitos: ser tão corretos quanto possível, compreensíveis e úteis. A forma como eles serão apresentados, ao final do processo, depende muito das duas primeiras etapas da KDD. A real utilidade dos dados precisa ser avaliada pelo decisor, reafirmando, portanto, a importância do trabalho conjunto de quem conhece o problema com quem tratará os dados.

Os resultados alcançados atenderam à expectativa de descoberta de informações úteis e demonstram que esta técnica, de fato, auxilia no processo de análise de dados, podendo ser utilizada como subsídio para decisões em empresas públicas e privadas, ou na solução de outros problemas específicos.

## 6 Referências

APRIORI. Disponível em: [http://universitario.educacional.com.br/include/download\\_page.asp?FILE=/dados/materialapoio/47180001/985526/apriori.html](http://universitario.educacional.com.br/include/download_page.asp?FILE=/dados/materialapoio/47180001/985526/apriori.html). Acesso em 4 ago. 2006.

AURÉLIO, Marco; VELLASCO, Marley; LOPES, Carlos H. **Descoberta de Conhecimento e Mineração de Dados**. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 1999.

BRUSSO, Marcos José. **O paralelismo na mineração de regras de associação**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1998.

FELDENS, Miguel Artur. Knowledge discovery in databases. Disponível em: <http://www.ufrgs.br/~feldens>. Acesso em 1 ago. 2006.

FIGUEIRA, Rafael. **Mineração de dados e bancos de dados orientados a objetos**. Dissertação de Mestrado – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.

GONÇALVES, Loren P. F. **Avaliação de ferramentas de mineração de dados como fonte de dados relevantes para a tomada de decisão**: aplicação na rede união de supermercados São Leopoldo. Dissertação de Mestrado – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2001.

HARRISON, Thomas H. **Intranet data warehouse**. São Paulo: Bekerley, 1998.

REZENDE, Denis Alcides. **Sistemas de Informações Organizacionais**: guia prático para projetos em cursos de administração, contabilidade e informática. São Paulo: Atlas, 2005.