



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS



# **INVESTIGAÇÃO DE PADRÕES DE COMPORTAMENTO ARQUITETURAL DE AMBIENTES DE TESTE DE PROJETOS DE ENGENHARIA DE SOFTWARE**

**RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA  
(PIBIC/CNPq/INPE)**

Carla Cristina Doescher Fernandes (UNIFESP, Bolsista PIBIC/CNPq)  
E-mail: c.fernandes11@unifesp.br

Nandamudi Lankalapalli Vijaykumar (LAC/CTE/INPE, Orientador)  
E-mail: vijay@lac.inpe.br

## **COLABORADORES**

Me. Érica Ferreira de Souza (LAC/INPE)  
Me. Luciana Rebelo (LAC/INPE)

Julho de 2011

## **AGRADECIMENTOS**

Agradeço ao meu orientador Dr. Nandamudi Lankalapalli Vijaykumar, a Me. Érica Ferreira de Souza e a Me. Luciana Rebelo pela paciência em me ajudar e me ensinar neste trabalho, ao CNPq pelo apoio financeiro e ao INPE pela disponibilização de recursos físicos e ferramentais para a realização desse projeto.

## RESUMO

Atualmente, nas grandes organizações, há uma enorme quantidade de informações propaladas em diversos setores e armazenadas em bancos de dados operacionais. Tais informações são de extrema importância para auxiliar no processo de tomada de decisões estratégicas. Neste contexto, surge o *Data Warehouse* (DW) , um banco de dados informacional que consiste em organizar os dados de maneira integrada, isto é, todas as informações dispersas são tomadas e passam por uma transformação, sendo otimizadas para um processo de consulta. Depois de armazenadas, tais informações precisam ser recuperadas. Uma das formas mais comuns para realizar o processo de extração de informações é através de técnicas de *Data Mining* (Mineração de Dados) capazes de reconhecer padrões descobrindo relações em um banco de dados e possibilitando assim a obtenção do conhecimento. Um ambiente que associa o apoio à tomada de decisão com as técnicas de *Data Mining*, possibilita aplicações em diversas áreas, tais como telecomunicação, varejo, saúde e transporte. Este trabalho tem o objetivo de fazer um estudo detalhado sobre DW e técnicas de mineração dos dados identificando, dessa forma, padrões de comportamento entre os dados. Tal trabalho será utilizado em pesquisas desenvolvidas no curso de Computação Aplicada do INPE.

# **INVESTIGATION OF ARCHITECTURAL BEHAVIOR PATTERNS OF TEST ENVIRONMENTS OF SOFTWARE ENGINEERING PROJECTS**

## **ABSTRACT**

Nowadays, large corporations store a huge amount of information from several sectors in operational databases. Such information are extremely important to support the process of strategic decision making. In this context, Data Warehouse (DW) concept becomes important, which is an informational database that consists in organizing the data in an integrated form. In other words, all dispersed information are taken and undergo a transformation, being optimized for a query process. Once stored, this information needs to be recovered. One of the most common forms to do the process of information extraction is employing Data Mining techniques that are able to recognize patterns, discovering relations in a database and enabling the learning. An environment that associates the support of decision making with Data Mining techniques enables applications in a wide range of areas, for example telecommunication, retails, health, transport and others. This project aims to conduct a detailed study about DW and Data Mining techniques, identifying behavior patterns among the data. Such project will be used in qualifying embedded software of Space Applications.

## LISTA DE FIGURAS

**Pág.**

|                                                         |    |
|---------------------------------------------------------|----|
| Figura 2.1 - Exemplo de dados orientados por temas..... | 4  |
| Figura 2.2 - Exemplo de variação no tempo.....          | 5  |
| Figura 2.3 - Exemplo de <i>Data Mart</i> .....          | 6  |
| Figura 2.4 - Arquitetura de um DW.....                  | 7  |
| Figura 3.1 - Pirâmide da informação .....               | 11 |

## LISTA DE SIGLAS E ABREVIATURAS

|      |                                           |
|------|-------------------------------------------|
| DM   | <i>Data Mining</i>                        |
| INPE | Instituto Nacional de Pesquisas Espaciais |
| BD   | Banco de Dados                            |
| DW   | <i>Data Warehouse</i>                     |

## SUMÁRIO

|                                                                            | <u>Pág.</u> |
|----------------------------------------------------------------------------|-------------|
| <b>1 INTRODUÇÃO .....</b>                                                  | <b>1</b>    |
| 1.1. Objetivos do trabalho.....                                            | 2           |
| 1.2. Organização do texto .....                                            | 2           |
| <b>2 O CONCEITO DE DATA WAREHOUSE .....</b>                                | <b>3</b>    |
| 2.1. Características.....                                                  | 3           |
| 2.2. <i>Data Marts</i> .....                                               | 5           |
| 2.3. Arquitetura de um <i>Data Warehouse</i> .....                         | 6           |
| <b>3 COMBINANDO TÉCNICAS DE MINERAÇÃO DE DADOS E DW .....</b>              | <b>8</b>    |
| 3.1. Conceito de Mineração de dados .....                                  | 8           |
| 3.2. Técnicas e algoritmos.....                                            | 8           |
| 3.3. Aplicações .....                                                      | 9           |
| 3.4. DW e <i>Data Mining</i> .....                                         | 10          |
| <b>4 DW E DATA MINING APLICADOS EM AMBIENTES DE TESTE DE SOFTWARE.....</b> | <b>12</b>   |
| <b>5 CONSIDERAÇÕES FINAIS .....</b>                                        | <b>14</b>   |
| <b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>                                    | <b>15</b>   |

## 1 INTRODUÇÃO

A informação é uma das mais importantes ferramentas que uma empresa pode dispor para tomar decisões estratégicas, é possível obtê-la analisando dados históricos sobre vendas, produção, clientes, dentre outros. Mas, em qualquer grande organização há dificuldades para analisar tais dados, pois eles crescem com a expansão dos negócios e com o passar dos anos, além disso, dados conflitantes vindos de fontes diferentes podem gerar informações desencontradas.

Para facilitar esse processo criou-se o *Data Warehouse* (DW), um banco de dados utilizado para manter os dados históricos, sendo assim uma fonte de consulta, por isso o DW é classificado como um banco de dados informacional.

Tais dados passam por um processo de transformação, na qual são padronizados e armazenados e ficarão disponíveis apenas para consulta, sendo impossível a alteração dos mesmos.

A extração das informações contidas nos dados armazenados em um DW pode ser feita de diversas maneiras, uma bastante conhecida e que será tratada neste trabalho é através do processo de *Data Mining* (mineração de dados) que realiza técnicas capazes de reconhecer padrões descobrindo relacionamentos em bancos de dados, e assim, possibilitando a obtenção do conhecimento.

Essas tecnologias possuem uma grande quantidade de aplicações, tais como *marketing*, vendas, finanças, manufatura, saúde, energia. O sistema de apoio à tomada de decisão unido com técnicas de mineração de dados também tem sido utilizado em pesquisas para melhoria de ambientes de desenvolvimento de *softwares* e também ambientes de teste de *software*.

Este trabalho tem o objetivo de fazer um estudo detalhado sobre o DW e técnicas de mineração dos dados, identificando, dessa forma, padrões de comportamento entre os dados.

### **1.1. Objetivos do trabalho**

O objetivo deste projeto é estudar detalhadamente o DW, que armazena dados históricos de organizações, e então fazer uma investigação de técnicas de *Data Mining* que possibilitam a análise das informações contidas no DW e encontram padrões de comportamento entre esses dados. Um ambiente que associa o apoio à tomada de decisão com as técnicas de *Data Mining* possibilita uma vasta área de aplicação; este trabalho será utilizado em pesquisas desenvolvidas no curso de Computação Aplicada do INPE.

### **1.2. Organização do texto**

A divisão dos capítulos está descrita a seguir:

- CAPÍTULO 2: Este capítulo explica de forma clara o conceito de *Data Warehouse*, bem como suas principais características e sua forma arquitetural.
- CAPÍTULO 3: É apresentada neste capítulo a definição e aplicações de *Data Mining* e a relação entre do banco de dados DW e as técnicas de mineração de dados.
- CAPÍTULO 4: Descreve de uma forma breve como tais tecnologias podem ser utilizadas em ambientes de engenharia de *software*, em especial, em ambientes de teste de *software*.
- CAPÍTULO 5: Neste capítulo, encontram-se as considerações finais e trabalhos futuros, subsequente a este capítulo encontram-se as referências bibliográficas utilizadas na elaboração desse documento.

## **2 O CONCEITO DE DATA WAREHOUSE**

Em qualquer grande organização, é de extrema importância o armazenamento de dados que estão dispersos em diversos setores dessa organização, armazenados em banco de dados operacionais. Tais dados podem ser informações de clientes e funcionários, quantidade de consumidores dos produtos oferecidos, relações entre produto e cliente, dentre outros.

Com o intuito de trabalhar nesses diferentes dados para uma visão ampla de toda informação guardada nesses setores, surge o DW.

Um DW é um banco de dados informacional alimentado com os dados dos bancos de dados operacionais da empresa, cujo objetivo é agrupar todas as informações de forma consolidada, voltada a tomada de decisões e embasada no histórico dos dados (KIM, 98).

### **2.1. Características de um Data Warehouse**

Um DW possui as seguintes características:

- **Orientado por temas**

O DW é orientado aos principais assuntos, temas ou área de negócios do empreendimento. Sistemas comerciais clássicos são organizados em torno das aplicações da empresa.

A Figura 2.1 exemplifica essa diferença em uma organização bancária, onde os temas podem ser contas, clientes e atividades, diferente de operações funcionais como empréstimos, investimentos e seguros.

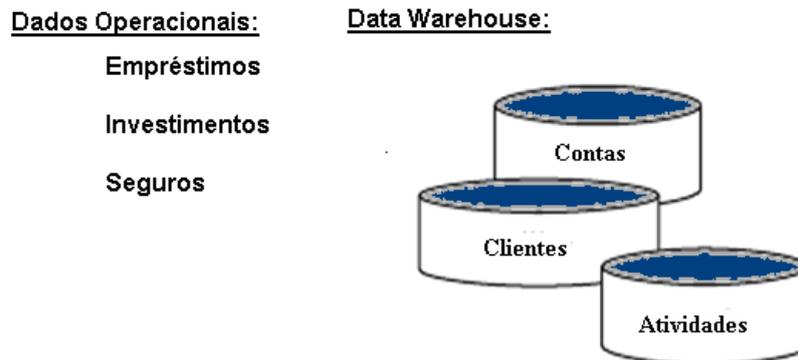


Figura 2.1 – Exemplo de dados orientados por temas  
 Fonte: adaptada de (Ferreira, 2002)

- **Integração**

Os dados provêm de diversas fontes, podendo divergir na forma de representação dos nomes das variáveis . Tais dados devem ser uniformizados de forma consistente. Por exemplo, uma fonte pode codificar o sexo como M/F, outra como 1/0, ou alguma outra como H/M. Na transição desses dados para um DW eles são convertidos em um estado uniforme, sendo assim, no DW todos eles terão apenas uma forma de codificação.

- **Não volatilidade**

Todas as informações guardadas em um DW são mantidas nele, ou seja, não é possível atualizar tais informações, o DW permite ao usuário somente a consulta a esses dados. Caso haja necessidade de alteração nos dados, é criada uma nova entrada para retratar essa mudança.

- **Variação no tempo**

Os dados referem-se a um dado instante de tempo . Na Figura 2.2 nota-se a diferença entre um BD operacional e o DW quanto à variação no tempo, enquanto o BD operacional guarda apenas o preço atual, o DW guarda os preços antigos também.

| Produto | Preço   |
|---------|---------|
| Camisa  | R\$ 90  |
| Toalha  | R\$ 39  |
| Calça   | R\$ 110 |

BD Operacional

| Produto | Jan/07 | Jan/08 | Jan/09  | ... |
|---------|--------|--------|---------|-----|
| Camisa  | R\$ 70 | R\$ 87 | R\$ 90  |     |
| Toalha  | R\$ 23 | R\$ 34 | R\$ 39  |     |
| Calça   | R\$ 89 | R\$ 98 | R\$ 110 |     |

Data Warehouse

Figura 2.2 – Exemplo de variação no tempo

- **Granularidade**

Representa o nível de detalhes em um DW. Quanto mais detalhes, menor o nível de granularidade e maior quantidade de dados armazenados. É essencial que a granularidade seja definida para que um DW atenda seus objetivos pois é através dela que se sabe o tipo de consulta que pode ser atendida.

- **Metadados**

São os dados que definem como os dados em um DW foram alterados e como devem ser interpretados, pode ser definido como um dicionário de informações e devem compreender informações como: origem dos dados, fluxo dos dados, formato dos dados, definições de negócio, regras de transformação, dentre outros.

## 2.2. *Data Marts*

Quando armazenados em um DW, os dados são separados por assunto em subconjuntos de acordo com a estrutura da organização, esses subconjuntos dos dados são chamados de *Data Mart*. Em tese, um *Data Mart* é um DW específico de um determinado departamento (DMA, 2011). A Figura 2.3 exemplifica três diferentes tipos de *Data Mart*.

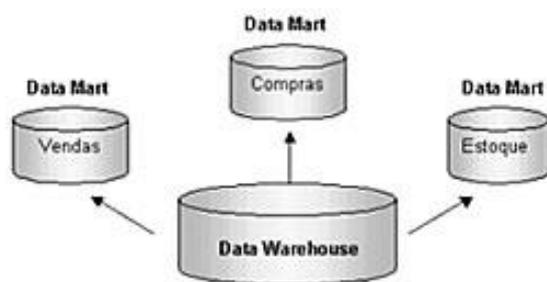


Figura 2.3- Exemplo de data mart.

Fonte:(DMA, 2011)

### 2.3. Arquitetura de um *Data Warehouse*

Um DW deve ter possuir mecanismos para analisar e transformar dados, ferramentas para armazenar e analisar dados, a arquitetura de um DW pode ser dividida em duas partes : estrutura interna e estrutura externa.

Na estrutura interna é feita a aquisição dos dados extraídos de fontes externas, os dados são formatados para permanecer no repositório de dados do DW, e armazenado na área de *Staging*. Então, ocorre a transformação, limpeza e agregação dos dados para então serem carregados ao DW.

Depois disso, as informações precisam ser extraídas e analisadas, tais atividades são feitas na estrutura externa onde ocorre a interface do usuário com o sistema. Fazem parte dessa área as ferramentas de acesso aos dados e geradores de relatórios, uma dessas ferramentas é chamada *Data Mining* e será apresentada no capítulo posterior. A Figura 2.4 exemplifica esses processos.

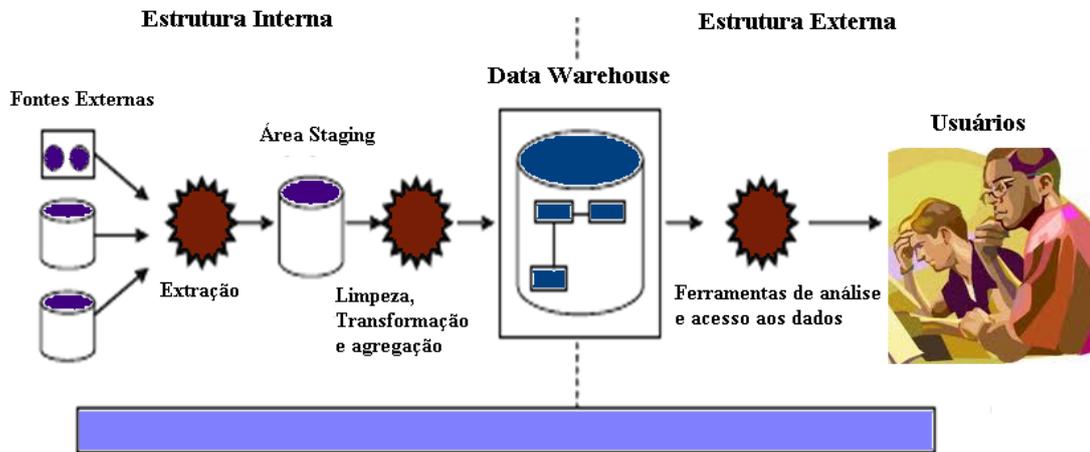


Figura 2.4 – Arquitetura de um DW  
Fonte: adaptada de (Ferreira, 2002)

### **3 COMBINANDO TÉCNICAS DE MINERAÇÃO DE DADOS E DW**

#### **3.1. Conceito de mineração de dados**

*Data Mining* (mineração de dados) é um processo no qual os dados são extraídos e analisados afim de obter padrões consistentes de comportamento como regras de associações ou sequências temporais, permitindo assim ao usuário, a exploração da informação previamente desconhecida (DMI, 2011).

Um ambiente que associa o apoio à tomada de decisão (DW) com as técnicas de *Data Mining*, possibilita aplicações em diversas áreas, tais como telecomunicação, varejo, saúde, transporte.

#### **3.2. Técnicas e algoritmos**

A obtenção de novas informações a partir de dados já existentes é feita pelo *Data Mining* através do uso de algoritmos capazes de reconhecer padrões de comportamento. Por meio de sistemas de redes neurais, esses dados são avaliados e observados, obtendo uma nova informação que pode associar-se a outros dados. Além disso, a formação de estatística é uma de suas funções; são gerados números estatísticos que trazem resultados comparativos possibilitando uma melhor tomada de decisão.

Os algoritmos mais utilizados são os de regressão e classificação.

##### **Algoritmos de Regressão**

Regressão é uma técnica bastante antiga que se baseia em analisar um conjunto de dados numéricos e então gerar uma fórmula matemática que se ajusta a eles. Para prever o comportamento futuro, aplicam-se os novos dados à fórmula desenvolvida. Um dos principais defeitos dessa técnica é que ela é eficiente apenas para dados quantitativos (DMI, 2011).

## **Algoritmos de Classificação**

Um algoritmo de classificação examina um conjunto de dados de treinamento (classificação é conhecida) e , embasado nos atributos dos dados constrói um modelo para cada classe. Esse processo gera uma árvore de decisão ou um conjunto de regras de classificação, e com isso é possível analisar melhor cada classe no banco de dados para classificação de dados subsequentes. Por exemplo, uma classe doença ajuda a prever tipo de doenças baseados nos sintomas dos pacientes (DMI, 2011).

### **3.3. Aplicações**

*Data Mining* é uma ferramenta que possui uma vasta área de aplicações. A área financeira, médica e vendas e *marketing* são alguns exemplos onde o DM pode ser muito útil.

- **Vendas e Marketing**

Nessa área, um DM pode ser utilizado para identificar os consumidores “leais”, relacionar as características demográficas com comportamentos dos clientes, bem como hábitos e preferências deles.

- **Área médica**

Embora pouco utilizado na área médica, o DM é útil para identificar fraudes em planos de saúde, terapias de sucesso para diferentes tratamentos, comportamento de pacientes e de usuários de planos de saúde.

- **Área financeira**

Na área financeira o DM é capaz de detectar padrões de fraude, analisar características de correntistas e carteiras de ações.

- **Área de Engenharia de Software**

A mineração de dados tem se posicionado para exercer um papel fundamental para as organizações, tornando-se um dos fatores mais importantes no desenvolvimento de soluções e promover o surgimento de conhecimento novo. Na área de engenharia de software não é diferente. Atualmente, diversas organizações de desenvolvimento de software buscam utilizar essa tecnologia para desenvolverem competências específicas e capacidade inovadora que se traduzem em serviços com menor custo e tempo, e que ajudem na busca de conhecimento para o processo de tomada de decisão.

### **3.4. *Data Warehouse e Data Mining***

A informação é o bem mais valioso para uma organização. Em geral, tais organizações dispõem de uma enorme gama de dados espalhados em diversos setores; o intuito de um DW é agregar esses dados de forma padronizada e armazená-los, e para que os dados guardados gerem informações, são utilizadas técnicas de *Data Mining*. Tal processo é descrito a seguir:

- Todos os dados são coletados e passam por uma transformação (limpeza, padronização).
- Cria-se os repositórios organizados (*Data Marts* e DW).
- Utilizam-se as técnicas e algoritmos de *Data Mining*.
- Interpretam-se os resultados.
- Informação é gerada.

A Figura 3.1 representa a pirâmide informação, é notório o aumento da abstração à medida que se caminha da base para o topo.



Figura 3.1 – Pirâmide da informação

DW e *Data Mining* são tecnologias modernas que, em conjunto, exploram além do armazenamento de dados, as técnicas para facilitar a compreensão e reconhecimento de padrões de vários tipos de bases de dados.

#### **4 DATA WAREHOUSE E DATA MINING APLICADOS EM AMBIENTES DE TESTE DE SOFTWARE**

Uma das características de projetos de engenharia de software é o elevado número de informações que são geradas e manipuladas. Os envolvidos no projeto enfrentam problemas, tais como: dificuldade de sistematizar as informações geradas ao longo dos processos de software; dificuldade para reutilizar o conhecimento gerado de um projeto em outro, por falta de vocabulário comum; perda de capital intelectual da organização, devido à rotatividade das equipes; e a não representação do conhecimento (Andrade, 2010). A expectativa é que o emprego de tecnologias, como, por exemplo, DW e *Data Mining*, amenize ou resolva, pelo menos parcialmente, estes problemas. Em um ambiente de engenharia de software uma quantidade enorme de informação é produzida ao longo do processo de desenvolvimento de um software, estas informações precisam ser compartilhadas para auxiliarem em processos de aprendizagem, para tornar mais ágil a resolução de problemas e para novos produtos. Tais informações podem ser armazenadas em uma base de conhecimento, um DW, por exemplo. E algoritmos de mineração de dados podem ser utilizados para buscar conhecimento novo nessas bases criadas.

Atualmente, é possível encontrar alguns trabalhos voltados para área de teste de software, utilizando as tecnologias de DW e mineração de dados. No trabalho de (Ferreira, 2010), por exemplo, foram utilizadas técnicas de Mineração de Dados, afim de levantar informações relevantes para melhoria do processo de teste empregando o algoritmo de associação Apriori (Agrawal,1993). O objetivo era identificação de regras ou correlação entre os dados de teste, tais como, resultados dos casos de teste exercitados, gravidade do defeito, prioridade de correção, dentre outros. O trabalho mostrou como resultados preliminares a coerência existente entre os padrões encontrados e o processo de teste, e também mostrou ser um bom ponto de partida para realizar estudos semelhantes. Sendo a área de teste de software tão crítica na garantia da qualidade de um software, principalmente aqueles

considerados críticos e complexos, é importante o investimento de pesquisas nessa área. Dessa forma, um grupo de alunos do curso de Pós-Graduação do curso de Computação Aplicada do INPE, tem iniciado estudos relacionados com melhoria de processos de teste de software, a partir do uso de tecnologias de DW, Mineração de dados, Gestão de Conhecimento (Natali, 2002) e Ontologia (Rios,2005). E este trabalho de Iniciação Científica irá contribuir com pesquisas futuras realizadas por esses alunos.

## 5 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as principais conclusões relacionadas a esse estudo. Ao final, apresenta também sugestões para trabalhos futuros.

Este trabalho teve como principal objetivo fazer um estudo detalhado sobre DW e técnicas de mineração dos dados, identificando, dessa forma, padrões de comportamento entre os dados.

Durante o projeto foram estudados tipos de testes de *software*, o banco de dados DW, suas características, formas de implementação e técnicas de mineração adaptadas aos dados armazenados em um DW.

A mineração de dados é uma área de pesquisa atual, e ainda existem alguns problemas a serem resolvidos, no entanto, representa um importante papel nos processos de tomada de decisão relacionados ao DW.

As pesquisas envolvendo mineração de dados e DW têm crescido muito pela sua importância, principalmente na área de engenharia de *software*, em particular, a área de teste de *software*. Hoje, é possível encontrar um aumento de trabalhos publicados nessa área.

Este trabalho será utilizado em pesquisas desenvolvidas no curso de Computação Aplicada do INPE.

Para trabalhos futuros, é interessante envolver mais pesquisas voltadas para aplicações dessa tecnologia na área de engenharia de *software*. E também pesquisas para melhoria do processo de teste de *software* que envolva essas tecnologias.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, M. T. T.; FERREIRA, C. V.; PEREIRA, H. B. B. Uma ontologia para a gestão conhecimento no processo de desenvolvimento de produto. Gest.Prod., Universidade Federal de São Carlos, São Carlos, v. 17, No 3, p. 537-551,2010.

AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Mining association rules between sets of itens in large database. Conference Management of Data (SIGMOD-93), Washington DC, USA, p. 207-216, 1993.

DMA – Data MArt, 2011. Acesso em julho de 2011.

Disponível em:

<[http://www.blink.com.br/index2.php?pag=ver\\_artigo&codigo=13](http://www.blink.com.br/index2.php?pag=ver_artigo&codigo=13)>

DMI - Data Mining, 2011. Acesso em julho de 2011.

Disponível em:

<<http://www.oficinadanet.com.br/artigo/tecnologia/data-mining-uma-introducao>>

FERREIRA, E. Uso de Mineração de Dados para Análise do Processo de Teste. São José dos Campos - SP, 2010. Acesso em maio 2011.

Disponível em: <<http://urlib.net/8JMKD3MGP7W/38CDFNB>>.

FERREIRA,R.G.C.. Data Warehouse na prática: Fundamentos e Implementação. Universidade Federal do Rio Grande do Sul, 2002.

KIMBALL, R. Data warehouse toolkit São Paulo: Makron Books, 1998.

NATALI, A. C. C.; FALBO, R. Gerência de conhecimento de qualidade de software. Proceedings of the 2nd Ibero-American Symposium on Software Engineering and Knowledge Engineering, JIISIC'2002, Salvador BA, 2002.

RIOS, J. A. Ontologias: alternativa para a representação do conhecimento explícito organizacional. In Proceedings CINFORM – Encontro Nacional da Ciência da Informação VI, Salvador-Bahia, 2005.