

# **MAC0499 – Trabalho de Formatura Supervisionado**

## **MONOGRAFIA**

**Título:** As aplicações de B.I. em vida pessoal

**Supervisor:** Prof. Dr. Flávio Soares Corrêa da Silva

**Aluno:** Xieli Zhaofu

## **Sumário**

### **I Introdução**

Foco do Conceito de Business Intelligence (B.I.)

Um estudo sobre B.I.

Aplicações de B.I. em vida pessoal

### **II Projeto**

Objetivos

Ferramentas usadas para implementação

Estrutura de dados

Organização de dados

Conceitos de B.I. no projeto

Andamento do projeto

### **III Impressões**

### **IV Frustrações**

### **V Referências**

# I Introdução

## *Foco do Conceito de Business Intelligence (B.I.)*

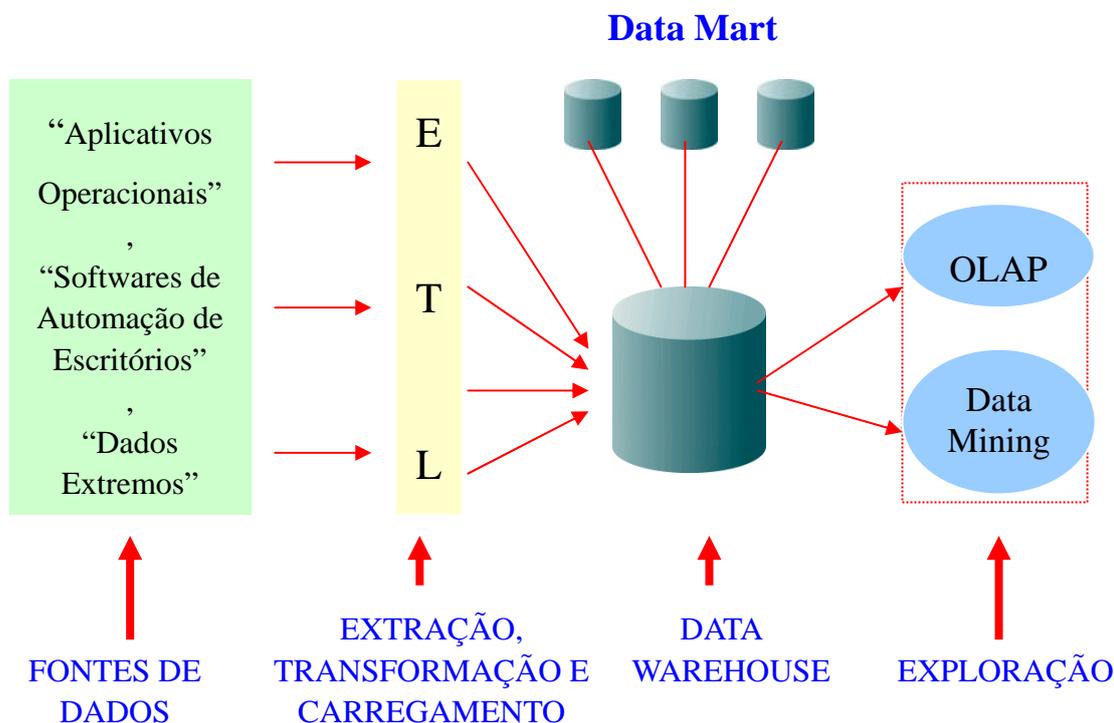
O termo *Business Intelligence* pode ser traduzido como *Inteligência de Negócios*, refere-se ao processo de coleta, organização, análise, compartilhamento e monitoração de informações que oferecem suporte a gestão de negócios.

O universo empresarial hoje padece de um mal clássico. Possui uma montanha de dados, mas enfrenta grande dificuldade na extração de informações a partir dela. Essa crescente inundação de informações dificulta o processo de tomada de decisão, na medida em que a alta e média gerência se sentem impotentes no processo de sua busca e recuperação. As informações vitais para tomadas de decisões estratégicas estão escondidas em milhares de tabelas e arquivos inacessíveis aos mortais, ligadas por relacionamentos e correlações transacionais, numa anatomia inadequada para os tomadores de decisão. Dessa forma, o conhecimento corporativo e as informações externas não estão prontamente disponíveis. O jogo de palavras que melhor define essa situação é: “Não se sabe o que se sabe e não se sabe o que não se sabe”. O objetivo maior das técnicas de B.I. está exatamente na definição de regras e técnicas para a formatação adequada destes volumes de dados, visando transformá-los em depósitos estruturados de informações, independentemente de sua origem. Os dados poderão vir das técnicas

emergentes de garimpo de informações via CI (*Competitive Intelligence*), ou de amplas fontes conceituais como KMS (*Knowledge Management System*). Em qualquer situação, a definição de estruturas modeladas dimensionalmente, armazenadas em *Data Warehouse* ou *Marts* e interpretadas pela ótica analítica das ferramentas de OLAP (*On Line Analytical Processing*) ou pelo prisma inferencial das ferramentas de *Data Mining*, atinge o objetivo proposto pelas premissas de BI.

### **Um estudo sobre B.I.**

Podemos dividir as técnicas de B.I. em quatro partes: “Fontes de Dados”, “E.T.L”, “Data Warehouse” e “Exploração”.



#### **I - Fontes de Dados**

Fontes de Dados representam as fontes que fornecem entradas para o processo de BI. Eles podem ser aplicativos operacionais, software

automação de escritórios (ex: Word, Outlook-Express etc.), ou até outros bancos de dados.

## II - E.T.L.

E.T.L. é acrônimo em inglês para “*Extract, Transform and Load*”, ou em português “Extração, Transformação e Carregamento”.

*Extração* - É o processo que responsabiliza para extrair os dados de fontes de dados e fornece dados para o próximo processo *Transformação*.

*Transformação* - É o processo que faz “limpeza” aos dados externos. Essa “limpeza” pode se referir uma tradução de dados, alterações de acordo com regras de negócios etc. Por exemplo, suponhamos o processo extraiu dados de um outro banco de dados onde usa “M” para representar “Male” e “F” para “Female”, mas no nosso banco de dados definimos “1” para “Male” e “2” para “Female”. Então, essa tradução é a tarefa do processo *Transformação*.

*Carregamento* é o processo que carrega os dados fornecidos pelo processo *Transformação* para nosso banco de dados - *Data Warehouse/Data Marts*.

## III - Data Warehouse

Um *Data Warehouse* é um sistema de computação utilizado para armazenar informações relativas às atividades de uma organização em banco de dados, de forma consolidada. O desenho da base de dados favorece os relatórios, a análise de grandes volumes de dados e a

obtenção de informações estratégicas que podem facilitar a tomada de decisão.

Um *Data Warehouse* pode armazenar grandes quantidades de informação, às vezes divididas em unidades lógicas menores que são chamadas de *Data Marts*. O esquema de dados mais utilizado é a Modelagem Multidimensional. Apesar de bastante utilizado, não existe um padrão na indústria de software para o armazenamento de dados. Existem, na verdade, algumas controvérsias sobre qual a melhor maneira para estruturar os dados em um *Data Warehouse*.

Geralmente, o *Data Warehouse* não armazena informações sobre os processos correntes de uma única atividade de negócio, mas sim cruzamentos e consolidações de várias unidades de negócios de uma empresa.

#### IV - Exploração

O ambiente de exploração objetiva disponibilizar, através de ferramentas apropriadas, o acesso a informação.

Duas ferramentas usadas no ambiente de exploração são OLAP e *Data Mining*.

OLAP é um acrônimo em inglês para *Online Analytical Processing*, ou processamento analítico on-line. É uma abordagem tecnológica para gerar respostas rápidas a consultas analíticas de natureza tipicamente dimensional. Aplicações típicas de OLAP são relatórios de negócios,

marketing, relatórios gerenciais, *Business Performance Management* (BPM), orçamento e previsão, relatórios financeiros e áreas similares.

Bases de dados mais adequadas para OLAP empregam um modelo de base de dados dimensional, que permite consultas analíticas complexas ou ad-hoc, com um tempo de execução pequeno.

*Data Mining* é o processo de varrer grandes bases de dados a procura de padrões como regras de associação, seqüências temporais, para classificação de itens ou agrupamento (*clustering*).

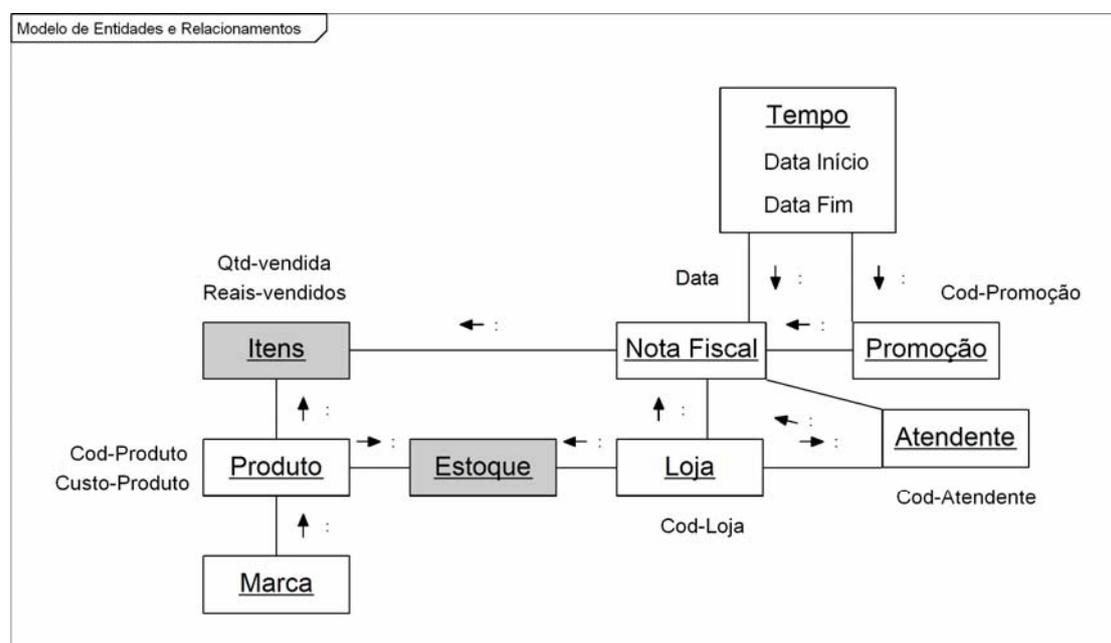
Na etapa de *Data Mining*, várias técnicas são utilizadas como da Estatística, Recuperação de informação, Inteligência Artificial e reconhecimento de padrões. São exemplos de técnicas de mineração de dados: *Redes Neurais, Regras de Associação, Clustering, Algoritmo genético, Árvores de decisão, Indução de Regras* e outras.

#### V – Sumário do estudo

O conceito de B.I. está na sua essência relacionado com formas alternativas de tratamento de informações – sua visão dimensional. A estrutura dimensional modifica a ordem de distribuição de campos por entre as tabelas, permitindo uma formatação estrutural mais voltada para os muitos pontos de entradas específicos (as dimensões) e menos para os dados granulares em si (os fatos). A sua vantagem é que oferece clara e diretamente os elementos que se precisa para buscar as informações sobre fatos via dimensões de referências.

Um exemplo do livro “B.I. – Business Intelligence” pode mostrar com mais clareza sobre as diferenças (as vantagens) entre o modelo tradicional de modelagem de dados (o modelo de Entidades e Relacionamentos) e o modelo dimensional na hora de usar BI.

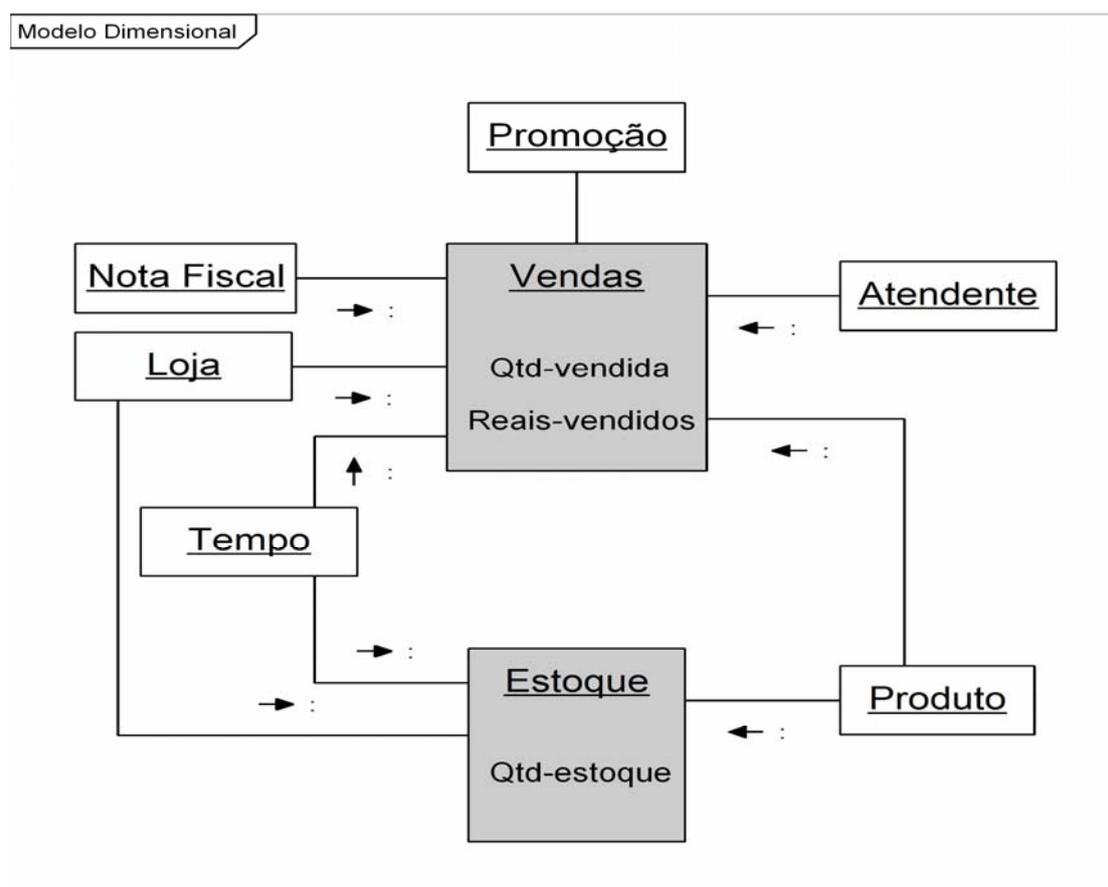
Observe o modelo da *Figura 1*. O modelo representado é um sistema de vendas a varejo. As principais entidades são: “Loja” que emite várias “Notas Fiscais”, cada qual com vários “Itens”, cada qual está associando a um “Produto”. Existe uma “Promoção” associada a um período de tempo. A entidade “Estoque” representa a resolução do relacionamento mxn entre “Loja” e “Produto” e a figura do “Atendente” daquela venda específica.



*Figura 1 – Modelo de Entidades e Relacionamentos*

Observe agora o mesmo modelo, na forma dimensional, na Figura 2. Neste modelo temos duas tabelas de Fato (Vendas e Estoque) e seis

Tabelas dimensões (Tempo, Produto, Loja, Promoção, Atendente e Nota-Fiscal). A tabela Fato representa as vendas de um “Produto”, num tempo feito numa “Promoção”, numa “Loja”, realizada por um atendente



*Figura 2 Modelo Dimensional equivalente ao modelo E/R da Figura 1*

e registrada em uma nota Fiscal. Observe que a tabela Fato relativa a Vendas tem como métricas desejadas somente os campos Quantidade-Vendida e o Valor em Reais, que na realidade eram os dados do item da “Nota Fiscal”. Outra tabela Fato, surgiu também para registrar os aspectos de Estoque. Essa segunda tabela Fato também se relaciona

com algumas das dimensões já identificadas para a primeira (Tempo, Loja e Produto) e registra a quantidade em estoque, como métrica principal.

A sua vantagem, embora mais leve e “clean” do que o modelo E/R, é que agora é possível acessar as métricas desejadas, entrando e combinando atributos das diversas dimensões, de uma forma muito mais objetiva e direta, sem grandes percursos de navegação.

A tabela a seguir, mostra as diferenças entre modelo E/R e Dimensional.

<b>Modelo Dimensional</b>	<b>Modelo E/R</b>
<i>Padrão de estrutura mais fácil e intuitiva</i>	<i>Modelo mais complexo</i>
<i>Anterior ao modelo ER, anos 60</i>	<i>Ênfase nos Bancos de Dados Relacionais, anos 70</i>
<i>Tabelas Fato e tabelas Dimensão</i>	<i>Tabelas que representam Dados e Relacionamentos</i>
<i>Tabelas Fato são o núcleo – normalizadas</i>	<i>Todas as tabelas são comumente normalizadas</i>
<i>Tabelas Dimensão são os pontos de entrada</i>	<i>As tabelas são indistintamente acessadas e de filtro inicial</i>
<i>Tabelas Dimensão opcionalmente</i>	<i>Todas as tabelas são comumente</i>

<i>normalizada</i>	<i>normalizadas</i>
<i>Modelo mais facilmente “joined”</i>	<i>Maior dificuldade de “join” pelo número maior de tabelas</i>
<i>Leitura mais fácil do modelo por usuários não especializados</i>	<i>Maior dificuldade de leitura pelo usuário não especializado</i>

### ***Aplicações de BI em vida pessoal***

Na era da informação, com avanços tecnológicos em diversas áreas, uma pessoa física recebe um volume cada vez maior de informações. Isto conduz a necessidade de buscar formas mais eficiente de coletar e processar apenas as informações necessárias no nosso cotidiano.

Uma das aplicações é controle financeiro pessoal. Nos dias de hoje, por falta de informações e análises adequadas, é difícil de responder algumas perguntas simples mas essenciais, como: Quanto você gasta por mês? Quanto você tem de dívida? Onde você aplica seu dinheiro? E o resultado de falta de informações é que o dinheiro que faz tanto esforço para obter parece desaparecer muito rapidamente.

Com as aplicações das técnicas de B.I., é possível coletar e analisar as informações financeiras, e fornecer uma visão geral sobre receitas e gastos.

## **II Projeto**

## ***Objetivos***

O objetivo desse projeto foi aprender as principais técnicas de B.I. e construir um sistema de controle financeiro pessoal.

## ***Ferramentas usadas para implementação***

Para atingir os objetivos do projeto, várias ferramentas estão sendo utilizadas, elas são:

Java IDE – Eclipse é o ambiente de programação do projeto.

UML Tool – “Poseidon For UML CE” é usado para desenhar os diagramas do projeto. Esses diagramas são: Main Classes diagram, Use Case diagrams e Sequence diagrams.

Chart Tool Package – ILOG JViews Charts é usado para gerar digramas de estatística que representam os conjuntos de dados.

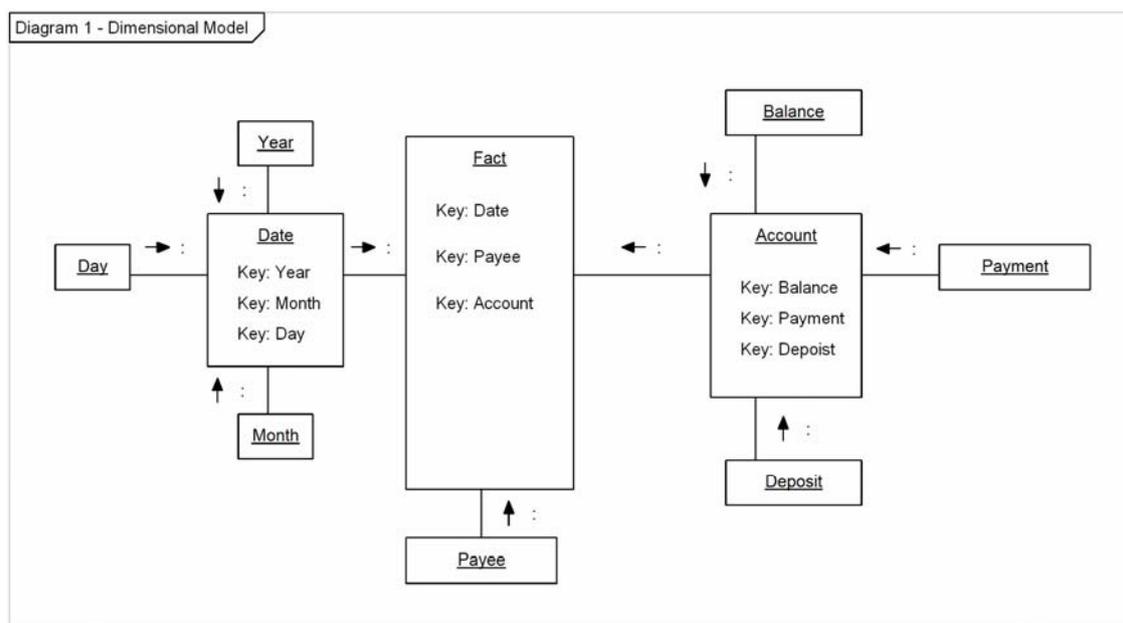
## ***Estrutura de dados***

O conceito de B.I. está na sua essência relacionado com formas alternativas de tratamento de informações.

A abordagem tradicional de dados é imperfeita para os processamentos demandados pela ótica dimensional. Então, precisa de uma estrutura dimensional que modifica a ordem de distribuição de campos por entre as tabelas, permitindo uma formatação estrutural mais voltada para os muitos pontos de entradas específicos, as chamadas

dimensões, e menos para os dados granulares em si, os chamados fatos.

O modelo no *Diagrama 1* objetiva uma estrutura dimensional que permita o sistema, gerenciando a financeira, por conta, por período e por tipo de gasto/receita (Payee).



*Diagrama 1 – Modelo Dimensional*

Podemos observar que o modelo tem seguintes dimensões: a dimensão Date, que contém três sub-dimensões Year, Month e Day; a dimensão Payee, que caracteriza destino de gastos e origem de receitas; e a dimensão Account, que também possui três sub-dimensões: Balance, Payment e Deposit.

Esse modelo, no fundo, representa uma visão agregada de um modelo granular e permite o acompanhamento do desempenho financeiro de uma pessoa física.

## **Organização de dados**

A organização de dados segue a padrão de ebXML (*e-business XML*). Sua principal vantagem é que os dados podem ser compartilhados com outros principais sistemas usados no mundo de e-business e vice versa.

Atualmente, o sistema guarda as informações de uma conta financeira num arquivo XML. Utilizando “tag”, os dados e as informações são divididos por categorias/dimensões, e assim, facilita que o sistema usa as técnicas de BI mencionadas acima para carregar, processar e explorar as informações.

## ***Conceitos de B.I. no projeto***

As seguintes técnicas de B.I. são usadas no programa:

- Processo de extração dados de fonte: o sistema é projetado para extrair e processar dois tipos de fonte de dados. O primeiro tipo de fonte é “fonte interna”. Esse tipo de fonte é o arquivo XML que pode conter informações de conta bancaria de uma pessoa física, e é gerado pelo programa usando formato pré-definido. Outro tipo de fonte é “fonte externa”. Uma fonte externa pode ser um “web feed” (RSS, por exemplo) que fornece informações de bolsas do mundo, valor de cambio e outras informações financeiras, ou um arquivo XML que contem informações em formato de ebXML. Para as informações desse tipo de fonte, o sistema precisa extrair e

processar as informações e transformá-las para um arquivo do tipo fonte interna.

- Data Warehouse em dimensões: usando “tag”s de XML, o sistema guarda e analisa as informações por categorias.
- Uso das técnicas de OLAP para fazer resumo estatístico: com as implementações das duas primeiras técnicas, agora o sistema pode utilizar a técnica de OLAP para explorar, por categorias/dimensões, as informações de fonte interna. Por exemplo, para explorar o peso de gasto de transporte no orçamento do mês específico X, o sistema simplesmente procura tag “month” com valor “X” e divide o valor de sub-tag “transport” por todos os gastos dos sub-tags do tag “month” X. Com o resultado, o sistema pode usar classes e funções do pacote JViews Charts para gerar gráficos estatísticos.

### *Andamento do projeto*

O projeto ainda está na fase de desenvolvimento, muitas detalhas ainda precisam ser implementadas. Mas com a toda minha vontade, eu pretendo continuar desenvolvendo esse projeto e torná-lo útil.

## **III Impressões**

Ficou muito claro para mim que a maior parte das técnicas de programação que me aprendi no BCC é utilizada no dia-a-dia do

desenvolvimento.

No meu projeto, foram mais úteis os conhecimentos obtidos nas disciplinas tais como:

MAC110 - Introdução à Computação

MAC211, MAC242 - Laboratório de Programação I - II.

MAC323 - Estruturas de Dados

MAC426 - Sistemas de Bancos de Dados

MAC332 - Engenharia de Software

MAC441 - Programação Orientada a Objetos

Por fim, devo dizer que considero os conhecimentos adquiridos no BCC como estritamente positivos.

#### **IV Frustração**

A maior frustração é que apesar de muitas dedicações ao TCC, muitas coisas do projeto ainda estão em desenvolvimento no final do semestre.

#### **V Referências**

Booch, Grady. Rumbaugh, James & Jacobson, Ivar. The Unified Modeling Language User Guide. Addison Wesley, 1998.

Palestino, Carlos B. BI-Business Intelligence – Modelagem & Tecnologia. Axcel Books, 2001.

Muller, Robert J. Database Design for Smarties: Using UML for Data

Modeling. Morgan Kaufmann Publishers, 1999.

Reese, George. Database Programming with JDBC and Java, 2<sup>nd</sup>  
Edition. O'Reilly, 2000.