

**UNIVERSIDADE REGIONAL DE BLUMENAU**  
**CENTRO DE CIÊNCIAS EXATAS E NATURAIS**  
**CURSO DE CIÊNCIAS DA COMPUTAÇÃO**  
(Bacharelado)

**PROTÓTIPO DE UM SISTEMA DE INFORMAÇÃO  
GERENCIAL APLICADO A CENTRAL DE INFORMAÇÃO  
AOS ALUNOS DA FURB UTILIZANDO  
DATA MINING**

TRABALHO DE CONCLUSÃO DE CURSO SUBMETIDO À UNIVERSIDADE  
REGIONAL DE BLUMENAU, PARA OBTENÇÃO DOS CRÉDITOS NA DISCIPLINA  
COM NOME EQUIVALENTE NO CURSO DE CIÊNCIAS DA COMPUTAÇÃO —  
BACHARELADO

**BIANCCA NARDELLI**

BLUMENAU, DEZEMBRO/2000.

2000/2-11

**PROTÓTIPO DE UM SISTEMA DE INFORMAÇÃO  
GERENCIAL APLICADO A CENTRAL DE INFORMAÇÃO  
AOS ALUNOS DA FURB UTILIZANDO  
DATA MINING**

**BIANCCA NARDELLI**

ESTE TRABALHO DE CONCLUSÃO DE CURSO FOI JULGADO ADEQUADO  
PARA OBTENÇÃO DOS CRÉDITOS NA DISCIPLINA DE TRABALHO DE  
CONCLUSÃO DE CURSO OBRIGATÓRIA PARA OBTENÇÃO DO TÍTULO DE:

**BACHAREL EM CIÊNCIAS DA COMPUTAÇÃO**

---

Prof. Oscar Dalfovo — Orientador na FURB

---

Prof. José Roque Voltolini da Silva — Coordenador do TCC

**BANCA EXAMINADORA**

---

Prof. Oscar Dalfovo – Orientador na FURB

---

Prof. Everaldo Artur Grahl

---

Prof. Roberto Heinzle

**A meus pais Adolar e Maria Aparecida, irmãs Candice e Greyce, meu noivo Ozéias e a todos que contribuíram, direta ou indiretamente, para a realização deste trabalho.**

## AGRADECIMENTOS

Ao orientador e amigo Oscar Dalfovo, pelo exemplo acadêmico e por me incentivar a não desistir de meus sonhos.

A meu pai Adolar, pelo exemplo de força e determinação passado durante todos os anos de minha vida.

À minha mãe Maria Aparecida, por seu amor verdadeiro, carinho, compreensão e por estar sempre ao meu lado, me apoiando e me fazendo acreditar que nada é impossível quando temos vontade de vencer.

Às minhas irmãs Candice e Greyce, pela alegria que transmitem em todos os momentos.

A meu noivo Ozéias, o companheiro de toda uma vida, por todos os bons momentos que passamos juntos e todos que ainda virão.

À equipe da Pró-Reitoria de Ensino da FURB, de quem espero estar junto por muito tempo, trabalhando em prol de uma Universidade melhor e que, apesar das dificuldades, não esquecem do lado humano, do companheirismo que faz a força para alcançarmos nossos objetivos.

A todas as pessoas que, de alguma forma, participaram da realização deste trabalho, entre eles: amigos, professores, colaboradores, monitores, colegas de trabalho e outros, ainda que não estejam aqui relacionados.

E sobretudo a Deus, que me conduziu em todas as horas, durante todo o tempo, dando-me boas oportunidades e ensinando-me a aproveitá-las.

# SUMÁRIO

AGRADECIMENTOS .....	IV
SUMÁRIO .....	V
LISTA DE FIGURAS .....	VII
LISTA DE TABELAS .....	VIII
LISTA DE ABREVIATURAS.....	IX
RESUMO .....	XI
ABSTRACT .....	XII
1 INTRODUÇÃO.....	1
1.1 OBJETIVO.....	2
1.2 ORGANIZAÇÃO DO TEXTO .....	3
2 SISTEMAS DE INFORMAÇÃO .....	4
2.1 CONCEITOS .....	4
2.2 DADOS.....	4
2.3 INFORMAÇÃO.....	5
2.4 SISTEMA .....	5
2.5 ELEMENTOS DO SISTEMA DE INFORMAÇÃO.....	7
2.6 TIPOS DE SISTEMAS DE INFORMAÇÃO.....	9
2.7 SISTEMAS DE INFORMAÇÃO GERENCIAL .....	10
2.7.1 COMPONENTES DO SISTEMA DE INFORMAÇÃO GERENCIAL.....	11
2.7.2 FASES DO DESENVOLVIMENTO E IMPLANTAÇÃO DO SIG .....	13
2.7.2.1 FASE I – CONCEITUAÇÃO DO SIG.....	14
2.7.2.2 FASE II – LEVANTAMENTO E ANÁLISE DO SIG .....	14
2.7.2.3 FASE III – ESTRUTURAÇÃO DO SIG.....	15
2.7.2.4 FASE IV – IMPLANTAÇÃO E AVALIAÇÃO DO SIG .....	16
2.7.3 PROTOTIPAÇÃO.....	16
3 DATA MINING .....	19
3.1 PROSPECÇÃO DE CONHECIMENTO .....	20
3.2 AS ETAPAS DO PROCESSO DE KDD .....	22
3.3 FUNÇÕES DO DATA MINING.....	23
3.3.1 CLASSIFICAÇÃO.....	24
3.3.2 ESTIMATIVA.....	24
3.3.3 AGRUPAMENTO POR AFINIDADE.....	24
3.3.4 PREVISÃO.....	25
3.3.5 SEGMENTAÇÃO.....	25
3.4 TÉCNICAS DE DATA MINING.....	26
3.4.1 REDES NEURAIS ARTIFICIAIS.....	26
3.4.2 ALGORITMOS GENÉTICOS.....	27
3.4.3 INDUÇÃO DE REGRAS.....	27
3.4.4 ANÁLISE ESTATÍSTICA DE SÉRIES TEMPORAIS .....	28
3.4.5 VISUALIZAÇÃO .....	28
3.4.6 ÁRVORES DE DECISÃO.....	29
4 PRÓ-REITORIA DE ENSINO – PROEN .....	31
4.1 CENTRAL DE INFORMAÇÃO AOS ALUNOS – CIA & CIA.....	32
5 TECNOLOGIAS E FERRAMENTAS UTILIZADAS .....	35

5.1	ANÁLISE ESTRUTURADA .....	35
5.1.1	CONCEITOS.....	35
5.1.2	DIAGRAMA DE FLUXO DE DADOS .....	36
5.1.3	MODELO ENTIDADE-RELACIONAMENTO .....	37
5.1.4	DICIONÁRIO DE DADOS .....	38
5.2	FERRAMENTAS CASE.....	38
5.2.1	POWER DESIGNER .....	39
5.3	FERRAMENTAS DE PROGRAMAÇÃO – AMBIENTE VISUAL DELPHI .....	39
5.4	TRABALHOS CORRELATOS .....	40
6	DESENVOLVIMENTO DO PROTÓTIPO.....	42
6.1	METODOLOGIA SIG.....	42
6.2	ESPECIFICAÇÃO.....	43
6.2.1	DIAGRAMA DE CONTEXTO .....	43
6.2.2	DIAGRAMA DE FLUXO DE DADOS .....	44
6.2.3	MODELO ENTIDADE-RELACIONAMENTO .....	46
6.2.4	DICIONÁRIO DE DADOS .....	47
6.3	APRESENTAÇÃO DAS TELAS.....	50
7	CONCLUSÕES E SUGESTÕES.....	59
7.1	CONCLUSÃO .....	59
7.2	LIMITAÇÕES .....	60
7.3	SUGESTÕES .....	60
	ANEXO 1 – QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA DE DEMANDAS POR CURSOS.....	61
	ANEXO 2 – TABELAS DE PRIORIDADES .....	64
	REFERÊNCIAS BIBLIOGRÁFICAS .....	68

## LISTA DE FIGURAS

Figura 1 – Componentes de um Sistema .....	6
Figura 2 - Elementos de um Sistema de Informação.....	8
Figura 3 – Componentes do SIG .....	12
Figura 4 - Processo KDD .....	21
Figura 5 - As etapas do processo de KDD.....	22
Figura 6 – Estrutura de Rede Neural Artificial.....	27
Figura 7 - Fórmulas para calcular entropia e <i>gain</i> .....	30
Figura 8 - Diagrama de Fluxo de Dados .....	36
Figura 9 - Área de trabalho do Delphi.....	40
Figura 10 - Diagrama de Contexto do Sistema de Informação Gerencial.....	44
Figura 11 - Diagrama de Fluxo de Dados (Nível 0).....	44
Figura 12 - Diagrama de Fluxo de Dados (Nível 1).....	45
Figura 13 - Diagrama de Fluxo de Dados (Nível 2).....	45
Figura 14 - Diagrama de Fluxo de Dados (Nível 3).....	46
Figura 15 - Modelo Entidade-Relacionamento .....	47
Figura 16 - Tela de Abertura do Protótipo .....	50
Figura 17 - Tela Principal.....	50
Figura 18 - Questão 9 dos Módulos Questionários .....	51
Figura 19 - Resultado da Questão .....	51
Figura 20 - Questão 16 do Módulo Comparativos .....	52
Figura 21 - Algoritmo da Árvore de Decisão .....	53
Figura 22 - Preferência dos Alunos pelas Instituições de Ensino Superior (Questão 13).....	54
Figura 23 - Preferência dos Alunos na Área de Ciências Biológicas, da Saúde e da Terra (Questão 17a).....	55
Figura 24 - Preferência dos Alunos na Área de Ciências Humanas (Questão 17b).....	55
Figura 25 - Preferência dos Alunos na Área de Ciências Sociais Aplicadas (Questão 17c)....	56
Figura 26 - Preferência dos Alunos na Área de Ciências Exatas (Questão 17d).....	56
Figura 27 - Preferência dos Alunos na Área de Ciências Tecnológicas (Questão 17e).....	57
Figura 28 - Preferência dos Alunos na Infra-Estrutura de um Curso (Questão 20) .....	58

## LISTA DE TABELAS

Tabela 1 - Níveis dos Sistemas de Informação .....	8
Tabela 2 - Dicionário de Dados .....	48
Tabela 3 - Atributos da Questão 13 .....	64
Tabela 4 - Atributos da Questão 17a .....	64
Tabela 5 - Atributos da Questão 17b .....	65
Tabela 6 - Atributos da Questão 17c .....	66
Tabela 7 - Atributos da Questão 17d .....	66
Tabela 8 - Atributos da Questão 17e .....	66
Tabela 9 - Atributos da Questão 20 .....	67

## LISTA DE ABREVIATURAS

CIA & CIA	- Central de Informação aos Alunos
CASE	- <i>Computer Aided Software Engineering</i>
COMAVI	- Comissão de Avaliação Institucional
DFD	- Diagrama de Fluxo de Dados
DIAEG	- Divisão de Administração do Ensino de Graduação
DRA	- Divisão de Registros Acadêmicos
FIC – FURB	- Fórum de Informação dos Cursos de Graduação
IR	- Indução de Regras
KDD	- <i>Knowledge Discovery in Databases</i>
MER	- Modelo Entidade-Relacionamento
PROAD	- Pró-Reitoria de Administração
PROEN	- Pró-Reitoria de Ensino
PROERC	- Pró-Reitoria de Extensão e Relações Comunitárias
PROPEP	- Pró-Reitoria de Pesquisa e Pós-Graduação
SAD	- Sistema de Apoio à Decisão
SAE	- Sistema de Automação de Escritórios
SAP	- Seção de Apoio Pedagógico
SATE	- Seção de Apoio Técnico ao Ensino
SE	- Sistema Especialista
SI	- Sistema de Informação
SIE	- Sistema de Informações Executivas
SIG	- Sistema de Informações Gerenciais
SPT	- Sistema de Processamento de Transações

SRD - Seção de Registros de Docentes  
SUPRA - Sistema Universitário Prova por Área

## RESUMO

O trabalho tem como objetivo principal gerar um modelo de classificação de dados utilizando técnicas de *Data Mining*, mais especificamente árvores de decisão. Para auxiliar esta tarefa, foi implementado um protótipo que permite ao usuário definir um valor-prioridade para cada atributo que fará parte do modelo de classificação. Para a elaboração do protótipo, foram analisadas as características de Sistemas de Informação e técnicas de *Data Mining*, e montada uma base de dados fornecida pela CIA & CIA, que será aplicada à classificação. Foram realizados testes e foi possível desenvolver modelos de classificação, nos quais colocou-se em prática o uso de árvores de decisão.

## **ABSTRACT**

The work has as main objective to generate a model of data classification using techniques of *Data Mining*, more specifically, trees of decision. To aid this task a prototype was implemented that allows the user to define a priority value for each attribute that will be part of the classification model. For the elaboration of the prototype, the characteristics of Systems of Information and techniques of *Data Mining* were analyzed and a base of data supplied by CIA & CIA was mounted which will be the applied base to the classification. Tests were accomplished and it was possible to develop models of classification where the use of trees of decision was put into practice.

# 1 INTRODUÇÃO

A CIA & CIA (Central de Informação aos Alunos) da Pró-Reitoria de Ensino da FURB foi criada pela necessidade de divulgação e esclarecimento das principais dúvidas que os vestibulandos possuem sobre os cursos oferecidos pela Universidade Regional de Blumenau. Com o evento do FIC – FURB (Fórum de Informação dos Cursos de Graduação), realizado todos os anos e reunindo mais de 3 mil alunos, entre os que estão estudando ou já concluíram o 3º ano do ensino médio, vários dados são coletados por meio da aplicação de questionários. Existe, hoje, uma necessidade grande de gerenciamento desses dados, de forma a serem transformados em informações úteis, visto ser um processo que abrange investimentos altos, tanto em recursos financeiros, quanto em recursos humanos.

De acordo com [DAL2000], atualmente organizações como a FURB vêm passando por processos de reestruturação e mudanças na área de informática e é natural que elas busquem alternativas de armazenamento de dados e o seu acesso de maneira rápida e confiável. Para isso, o Sistema de Informação veio com o objetivo de auxiliar as organizações, mais especificamente as pessoas responsáveis por divulgar informações importantes. Sistemas de Informação são sistemas de coleta, armazenamento, processamento, recuperação e propagação das informações. As pessoas que se utilizam desse sistema são tanto do nível operacional quanto tático e ainda do nível estratégico. É possível integrar as pessoas envolvidas com relações públicas por esse sistema informacional, fornecendo informações úteis e objetivas para suas necessidades estratégicas e operacionais.

De acordo com [OLI1992] toda organização tem informações que proporcionam a sustentação para as suas decisões. Entretanto, apenas algumas têm um sistema estruturado de informações gerenciais que possibilita otimizar o seu processo decisório. As que estão neste estágio evolutivo seguramente possuem vantagem empresarial interessante. Para o processo decisório as empresas precisam de informações históricas e fazer uma garimpagem sobre os dados (que pode ser traduzido como *Data Mining*).

Segundo [GRO1997], a tecnologia em torno de um *Data Mining* envolve recursos de inteligência artificial e análises estatísticas, busca correlações de dados dentro de um ambiente de Banco de Dados, fornecendo informações consideradas relevantes para o negócio. De acordo com [OLI1996], a criação de *Data Mining* vem de encontro às necessidades atuais das grandes organizações em obter informações que podem gerar um grande diferencial, numa velocidade maior do que as formas tradicionais, com o objetivo de encontrar facilidades na hora de utilizar estes dados na tomada de decisões. Para auxiliar na garimpagem dos dados pode utilizar-se de técnicas tais como árvores de decisão, agentes inteligentes, raciocínio baseado em casos e outros.

Segundo [BIS1999], árvores de decisão expressam uma forma simples de lógica condicional buscando a representação de uma série de questões que estão escondidas sobre a base da dados. Em uma árvore de decisão existem dois tipos de atributos, o decisivo, que é aquele que contém o resultado ao qual se quer chegar, e os não-decisivos, que contêm os valores que conduzem a uma decisão.

Diante do exposto acima, foi desenvolvido um protótipo de um Sistema de Informação Gerencial aplicado a CIA & CIA, da Pró-Reitoria de Ensino da FURB, utilizando *Data Mining*. Através deste protótipo, a pessoa responsável pela divulgação dos cursos da Universidade poderá ter acesso a vários dados que possibilitarão a obtenção de diversas informações como: comparativo entre resultados obtidos por meio de questionários dos anos de 1999 e 2000; perfil da preferência dos alunos pelas instituições de ensino superior da nossa região; perfil da preferência pelas áreas de ensino oferecidas por essas instituições; fatores de maior importância para os alunos na infra-estrutura de um curso.

## 1.1 OBJETIVO

O objetivo principal deste trabalho é o desenvolvimento de um protótipo de um Sistema de Informação Gerencial aplicado a CIA & CIA, da Pró-Reitoria de Ensino da FURB, utilizando *Data Mining*.

## 1.2 ORGANIZAÇÃO DO TEXTO

O primeiro capítulo define o objetivo do trabalho, apresentando a justificativa para seu desenvolvimento.

O segundo capítulo apresenta uma visão geral sobre o SI, que o trabalho propõe-se a utilizar, mostrando conceitos, tipos, problemas e suas utilidades.

O terceiro capítulo enfatiza os conceitos, técnicas e aplicações de *Data Mining*.

O quarto capítulo aborda sobre a estrutura e o funcionamento da Pró-Reitoria de Ensino da FURB e da CIA & CIA.

O quinto capítulo apresenta a análise, as características, o desenvolvimento e a utilização do modelo criado.

O sexto capítulo completa o trabalho, apresentando as conclusões, limitações e sugestões para serem implementadas e aprimoradas.

## **2 SISTEMAS DE INFORMAÇÃO**

### **2.1 CONCEITOS**

Nos dias atuais, a utilização dos recursos de informação torna-se fator indispensável na reestruturação de novas organizações, quanto ao uso eficaz dos recursos de informações tradicionais, tais como os equipamentos, as instalações, os recursos humanos e os financeiros. À medida que se avança numa época de interação moderna e de melhor oferta, quanto maior for o número de informações, melhor será a sua utilização na busca de novas soluções para as pessoas que tomam decisões importantes dentro das organizações. Já que a busca destas informações vem crescendo cada vez mais, as organizações empresariais investiram em sua busca por meio da tecnologia de software e hardware, para aprimorar o desenvolvimento dos sistemas de informação.

De acordo com [DAL2000], a não-utilização das informações como recursos estratégicos leva o executivo, na maioria das vezes, a administrar impulsivamente ou baseado em modismo. Hoje, o fenômeno da moda são os sistemas de informação. A utilização de um Sistema de Informação pode vir a facilitar o processo decisório com a obtenção de dados estrategicamente escolhidos e de conteúdos relevantes para qualquer nível e tamanho de organização.

### **2.2 DADOS**

De acordo com [OLI1992], dados são quaisquer elementos identificados em sua forma bruta, que por si só não conduzem a uma compreensão de determinado fato ou situação. Toda organização utiliza-se de dados. Por dados entende-se áreas de conhecimento, cursos da Universidade, turno em que os alunos estudam, entre outros. Porém, esses dados em sua forma bruta pouco contribuem para a busca de uma visão mais integrada de uma determinada situação. Para isto, são utilizados dados transformados, que podem-se classificar como informação.

"A informação é o resultado do tratamento de dados existentes a cerca de alguém ou de alguma coisa. A informação aumenta a consistência e o conteúdo cognoscível dos dados", [CRU1998].

## 2.3 INFORMAÇÃO

O uso eficaz da informação nas organizações passa a ser um patrimônio, que é considerado um fator-chave para o sucesso das organizações. Este fator torna-se mais expressivo quando as organizações se defrontam com mudanças de mercado e avanços das tecnologias. A informação é resultante da organização e análise dos dados, sendo que a qualidade desta informação é muito mais importante do que a quantidade de informação. Para que a qualidade seja um fator de decisão na organização, é preciso estabelecer algumas regras básicas ([DAL2000]):

- a) a informação não deve ser demasiada;
- b) a informação não deve ser escassa;
- c) a sobrecarga de informação é de pouca utilidade; e
- d) deve haver o reaproveitamento e reciclagem das informações.

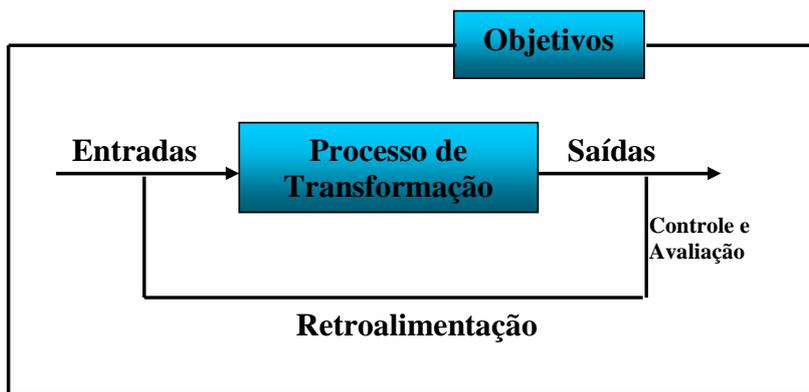
A informação é um processo pelo qual a organização informa-se sobre ela própria e seu ambiente e por ele informa ao seu ambiente sobre ela mesma, por meio da criação, comunicação, tratamento e memorização das informações nas formas mais diversas.

## 2.4 SISTEMA

Define-se sistema como sendo a disposição das partes de um todo, que de forma coordenada formam estrutura organizada, com a finalidade de executar uma ou mais atividades [CRU1998].

Conforme [OLI1996], os sistemas são compostos por seis componentes: os objetivos do sistema, as entradas do sistema, o processo de transformação do sistema, as saídas do sistema, os controles e avaliações do sistema e *feedback* ou retroalimentação ou realimentação do sistema (Figura 1).

**Figura 1 – Componentes de um Sistema**



*Fonte: [OLI1996]*

Os objetivos do sistema são a razão de sua existência. Os objetivos referem-se tanto aos objetivos do usuário do sistema quanto aos objetivos do próprio sistema.

As entradas são as fontes que alimentam o sistema. As entradas fornecidas ao sistema irão gerar as saídas do sistema, alinhadas aos objetivos deste.

O processo de transformação do sistema é definido como a transformação de um insumo (entrada) em um produto, serviço ou resultado (saída). É a maneira pela qual os elementos componentes do sistema interagem para que sejam produzidas as saídas desejadas.

As saídas do sistemas correspondem aos resultados obtidos do processo de transformação. Devem ser coerentes com os objetivos do sistema e quantificáveis de acordo com critérios e parâmetros previamente fixados.

Os controles e avaliações verificam se as saídas produzidas estão coerentes com os objetivos estabelecidos. Os controles e avaliações são realizados mediante o estabelecimento de um padrão, como uma medida de desempenho do sistema.

A retroalimentação ou *feedback* do sistema pode ser considerado como a reintrodução de uma saída fornecida sob a forma de informação. O *feedback* serve para regular as

informações realimentadas, resultantes das divergências verificadas entre os objetivos definidos e as saídas produzidas.

## 2.5 ELEMENTOS DO SISTEMA DE INFORMAÇÃO

De acordo com [BIN1994], o Sistema de Informação (SI) “é o requisito básico para a decisão automatizada, pois o processo decisório apoia-se na malha de sistemas de informação da empresa”.

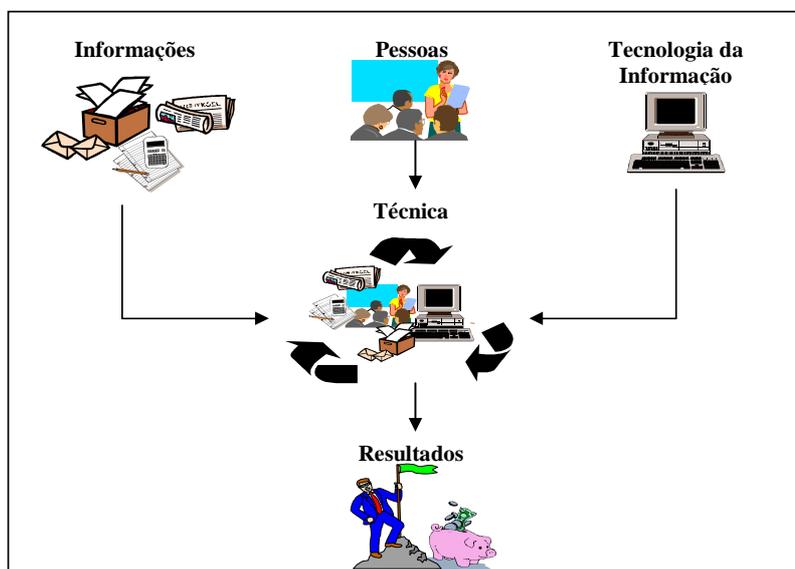
Um Sistema de Informação é um tipo especializado de sistema e pode ser definido de inúmeros modos. Um modo é dizer que sistemas de informação são conjuntos de elementos ou componentes inter-relacionados que coletam (entrada), manipulam e armazenam (processo), disseminam (saída) os dados e informações e fornecem um mecanismo de *feedback*. A entrada é a atividade de captar e reunir novos dados, o processamento envolve a conversão ou transformação dos dados em saídas úteis, e a saída envolve a produção de informação útil. O *feedback* é a saída que é usada para fazer ajustes ou modificações nas atividades de entrada ou processamento [STA1998].

Os Sistemas de Informação designam a logística indispensável à realização do processo de informação, a qual não se reduz somente à informática, como poderia parecer inicialmente. São o conjunto interdependente das pessoas, das estruturas da organização, das tecnologias da informação (hardware e software), dos procedimentos e métodos que deveriam permitir às organizações disporem, no tempo desejado, das informações que necessitam (ou necessitarão) para seu funcionamento atual e para sua evolução [VIC1994].

De acordo com [PRA1994], Sistemas de Informação são formados pela combinação estruturada de vários elementos, organizados da melhor maneira possível, visando atingir os objetivos da organização. São integrantes dos Sistemas de Informação: a informação (dados formatados, textos livres, imagens e sons), os recursos humanos (pessoas que coletam, armazenam, recuperam, processam, disseminam e utilizam as informações), as tecnologias de informação (o hardware e o software usados no suporte aos Sistemas de Informação), e as

práticas de trabalho (métodos utilizados pelas pessoas no desempenho de suas atividades). Pode-se observar estes elementos na Figura 2.

**Figura 2 - Elementos de um Sistema de Informação**



*Fonte: [PRA1994]*

De acordo com DALFOVO ([DAL2000]), os Sistemas de Informação podem ser divididos em quatro níveis, conforme Tabela 1:

**Tabela 1 - Níveis dos Sistemas de Informação**

Nível	Funcionalidade	Propósito
Operacional	Monitoram as atividades elementares e transacionais da organização.	Responder a questões de rotina e fluxo de transações (ex.: vendas, recibos, folha).
Conhecimento	São SI de suporte aos funcionários especializados e de dados em uma organização.	Ajudar a empresa a integrar novos conhecimentos ao negócio e controlar fluxo de papéis.
Gerencial	Suportam monitoramento, controle, tomada de decisão e atividades administrativas.	Controlar e prover informações de rotina para a direção setorial.
Estratégico	Suportam as atividades de planejamento de longo prazo dos administradores seniores.	Compatibilizar mudanças no ambiente externo com as capacidades organizacionais existentes.

## 2.6 TIPOS DE SISTEMAS DE INFORMAÇÃO

De acordo com [STA1998] e [ALT1992], a seguir são apresentados os tipos de SI mais relevantes:

- a) Sistema de Processamento de Transações (SPT): coletam e armazenam dados sobre transações e às vezes controlam decisões que são executadas como parte de uma transação. Uma transação é um evento empresarial que pode gerar ou modificar dados armazenados num Sistema de Informação. Ele foi o primeiro Sistema de Informação que surgiu e é frequentemente encontrado. Por exemplo, quando pagamos uma conta com o Cartão de Crédito é o SPT que efetua a transação com a Central e valida o cartão. Enfim, ele grava as informações e assegura que as mesmas estejam consistentes e disponíveis;
- b) Sistema de Automação de Escritório (SAE): ajuda as pessoas a processar documentos e fornece ferramentas que tornam o trabalho no escritório mais eficiente e eficaz. Também pode definir a forma e o método para executar as tarefas diárias e dificilmente afeta as informações em si. Exemplos deste tipo de Sistema são editores de texto, planilhas de cálculo, softwares para correio eletrônico e outros. Todas as pessoas que têm em sua função tarefas como redigir textos, enviar mensagens, criar apresentações, são usuárias de Sistemas de Automação de Escritórios;
- c) Sistema de Informação Gerencial (SIG): converte os dados de uma transação do SPT em informação para gerenciar a organização e monitorar o seu desempenho. Ele enfatiza a monitoração do desempenho da empresa para efetuar as devidas comparações com as suas metas. As pessoas que o utilizam são os gerentes e as que precisam monitorar seu próprio trabalho. Um exemplo disto são os relatórios que são tirados diariamente para acompanhar o Faturamento da empresa;
- d) Sistemas Especialistas (SE): tornam o conhecimento de especialistas disponível para outros, e ajudam a resolver problemas de áreas em que o conhecimento de especialistas é necessário. Eles podem guiar o processo de decisão e assegurar que

os fatores-chave serão considerados, e também pode ajudar uma empresa a tomar decisões consistentes. As pessoas que usam estes sistemas são aquelas que efetuam tarefas em que deveria existir um especialista. Um sistema especialista pode ser, por exemplo, um sistema no qual médicos dizem os sintomas e são pesquisados, em uma base de conhecimento, os possíveis diagnósticos;

- e) Sistema de Apoio à Decisão (SAD): ajuda as pessoas a tomar decisões, provendo informações, padrões, ou ferramentas para análise de informações. Ele pode prover métodos e formatos para porções de um processo de decisão. Os maiores usuários são os analistas, gerentes e outros profissionais. Os sistemas que disponibilizam gráficos 3D para comparativos são exemplos;
- f) Sistema de Informações Executivas (SIE): fornece informações aos executivos de uma forma rápida e acessível, sem os forçar a pedir ajuda a especialistas em Análises de Informações. É utilizado para estruturar o planejamento da organização e o controle de processos, e pode, eventualmente, também ser utilizado para monitorar o desempenho da empresa. Um exemplo são os sistemas que fornecem comparativos simples e fáceis de Vendas x Estoque x Produção.

## **2.7 SISTEMA DE INFORMAÇÃO GERENCIAL**

“Sistema de Informação Gerencial é o conjunto de tecnologias que disponibiliza os meios necessários à operação do processo decisório em qualquer organização por meio do processamento dos dados disponíveis”[CRU1998].

Segundo [CRU1998], o SIG é desenvolvido para garantir a administração eficiente a qualquer tipo de empresa. É esse sistema que vai garantir que os dados operacionais utilizados para manter a empresa produzindo serão traduzidos em informações passando a todos que tiverem funções executivas.

De acordo com [OLI1992], um Sistema de Informação Gerencial (SIG) é o processo de transformação de dados em informações que são utilizadas na estrutura decisória da

empresa, bem como proporcionam a sustentação administrativa para otimizar os resultados esperados. Esse tipo de sistema é orientado para a tomada de decisões estruturadas. Os dados são coletados internamente na organização, baseando-se somente nos dados corporativos existentes e no fluxo de dados. Os aspectos para a otimização do desenvolvimento e a implantação do SIG nas empresas são: metodologia de elaboração, suas partes integrantes, sua estruturação, sua implementação e avaliação, bem como as características básicas do executivo administrador do SIG.

Para OLIVEIRA [OLI1996], Sistema de Informações Gerenciais é o processo de transformação de dados em informações que são utilizadas na estrutura decisória da empresa, bem como proporcionam a sustentação administrativa para otimizar os resultados esperados. São voltados aos gerentes de empresas que acompanham os resultados das organizações semanalmente, mensalmente e anualmente, estando preocupados com os resultados diários. Os dados são coletados internamente na organização, baseando-se somente nos dados corporativos existentes e no fluxo de dados. A característica dos Sistemas de Informação Gerenciais é utilizar somente dados estruturados, que também são úteis para o planejamento de metas estratégicas.

### **2.7.1 COMPONENTES DO SISTEMA DE INFORMAÇÃO GERENCIAL**

Conforme [STA1998] um SIG é composto de uma coleção integrada de sistemas de informação funcionais. Cada um desses sistemas trabalha dando suporte às áreas funcionais específicas dentro da organização. Cada área funcional utiliza seu próprio conjunto de subsistemas, os quais interagem, de alguma forma, com o SIG.

Um SIG é composto por diversos componentes, todos trabalhando de forma independente e ao mesmo tempo integrada, objetivando um fim comum, que é o de fornecer informações ao SIG e este, por sua vez, fornecerá informações úteis à tomada de decisões. Partindo-se de uma análise funcional, o SIG é composto basicamente de um conjunto de subsistemas que trabalham de maneira integrada para tornar mais fácil o compartilhamento de informações dentro da organização, aumentando, assim, a eficiência. Partindo-se de uma

análise mais voltada para o processo administrativo, o SIG é composto de elementos e atividades inerentes ao processo decisório de uma organização.

Conforme [OLI1996], o SIG é composto, de uma forma mais abrangente, pelos elementos mostrados na Figura 3.

**Figura 3 – Componentes do SIG**



*Fonte: Adaptado de [OLI1996]*

## 2.7.2 FASES DO DESENVOLVIMENTO E IMPLANTAÇÃO DO SIG

Segundo [OLI1996], a administração das empresas está cada vez mais necessitando de informações bem depuradas. A eficácia empresarial está sendo muito prejudicada pela utilização de sistemas que fornecem informações incompreensíveis e inúteis para a tomada de decisão. Uma empresa deve desenvolver um sistema estruturado de informações gerenciais que alimente o processo decisório e o desenvolvimento, implementação e avaliação das decisões e ações relacionadas. Todas as fases de desenvolvimento e implantação de um SIG devem ser debatidas pelo gerente que deve, acima de tudo, conhecer profundamente a organização da empresa. Antes de desenvolver e operacionalizar um SIG, o gerente deve identificar as finalidades para as quais o SIG foi ou deve ser estruturado, pois esta atividade evita o desenvolvimento de um sistema que forneça dados e informações irrelevantes para o processo decisório da organização.

Segundo [OLI1996], o desenvolvimento de um SIG está baseado em quatro grandes fases. O objetivo principal deste modelo de desenvolvimento é fazer com que o gerente possa efetuar-lo respeitando a realidade da organização, bem como os resultados a serem alcançados. O SIG deve atender a determinados aspectos na sua operacionalização, como: administração, geração/arquivamento, controle/avaliação, disseminação, utilização e retroalimentação. Estes aspectos são fundamentais para o delineamento das quatro grandes fases de desenvolvimento e aplicação do SIG.

Segundo [OLI1996], as quatro grandes fases metodológicas para elaboração e desenvolvimento do Sistema de Informação Gerencial são:

- a) fase I - Conceituação do SIG
- b) fase II - Levantamento e Análise do SIG
- c) fase III - Estruturação do SIG
- d) fase IV - Implantação e Avaliação do SIG

### **2.7.2.1 FASE I - CONCEITUAÇÃO DO SIG**

A primeira fase do desenvolvimento de um SIG tem como objetivo obter uma idéia preliminar e geral do volume e complexidade do projeto. Nesta fase de desenvolvimento do SIG é que se deve identificar as informações e dados necessários e confiáveis ao desenvolvimento.

Esta fase caracteriza-se pela realização de reuniões e entrevistas para que se possa avaliar a situação atual da organização, ou seja, os problemas existentes, os sistemas atuais e até que ponto estes correspondem aos objetivos da organização e quais as necessidades e restrições dos usuários envolvidos. A fase de conceituação do SIG deve fornecer condições de efetuar-se um planejamento adequado do SIG. Nesta fase, o gerente catalisador do SIG deve considerar o desenvolvimento de um sistema de informações gerenciais que objetive auxiliar os vários gerentes da organização na tomada de decisão, através de informações que sejam confiáveis, na quantidade e época adequadas e com custo compatível com o volume e o nível da qualidade das informações.

### **2.7.2.2 FASE II - LEVANTAMENTO E ANÁLISE DO SIG**

A segunda fase do desenvolvimento de um SIG é a fase em que é necessário identificar as informações relacionadas às atividades do processo de tomada de decisões, avaliar estas informações, estudar e desenvolver novas informações, e implementar e avaliar as novas informações dentro do contexto decisório da organização. É através do SIG que fluem as informações, permitindo o funcionamento otimizado da organização, que depende destas para sua sobrevivência. A fase de levantamento e análise do SIG é caracterizada pela obtenção das informações que dele deverão fazer parte e seus fluxos. Nesta fase é que se verifica, após a obtenção das informações relevantes, a viabilidade do projeto do SIG.

Para que seja possível identificar os focos de decisões, é necessário saber com quais documentos a organização efetua o tratamento das informações dentro do processo decisório. No levantamento e análise do SIG devem ser respondidas perguntas pertinentes à análise das

entradas, processamento e saídas de tal forma que o resultado final seja uma combinação custo/benefício aceitável para a organização.

### **2.7.2.3 FASE III - ESTRUTURAÇÃO DO SIG**

A terceira fase do desenvolvimento de um SIG pode ser efetuada visando aos relatórios gerenciais, que representam os resumos consolidados e estruturados das informações necessárias ao processo decisório. Estas informações devem estar em nível otimizado de qualidade, ou seja, a satisfação e manutenção do usuário da informação (o gerente decisor).

Na fase de estruturação do SIG é necessário que se definam alternativas de soluções que operacionalizem o objetivo principal do SIG. Estas alternativas devem considerar equipamentos e abordar formas diferentes de desenvolver e implementar o SIG. Nesta fase deve-se: completar o fluxo geral do sistema de informações, os componentes das informações e as suas iterações; identificar o processo de tratamento de arquivos; determinar os arranjos físicos (*layouts*); especificar a formatação dos documentos e relatórios de entrada; definir a necessidade de relatórios; desenvolver a estrutura lógica geral do sistema de informações; determinar procedimentos e momentos de controle e avaliação; estabelecer a estimativa de custo do sistema de informações; elaborar um plano detalhado para a implantação; documentar todos os aspectos desta fase do projeto ao coordenador do sistema e aos usuários; e estabelecer a decomposição do sistema em subsistemas para facilitar o seu desenvolvimento e implementação. Ao final desta fase deve-se ter consolidado um relatório escrito, pois esta situação possui as vantagens da apresentação completa de todos os fatos importantes ao desenvolvimento do SIG.

O SIG deve ser estruturado respeitando a filosofia de atuação da organização, considerando sua postura frente ao risco organizacional. Para facilitar a estruturação é válido subdividir o projeto a longo prazo em projetos menores, ser coerente com o potencial humano disponível, alocar todos os custos necessários, estabelecer plano de carga de trabalho e fazer gráficos de desempenho.

#### **2.7.2.4 FASE IV - IMPLANTAÇÃO E AVALIAÇÃO DO SIG**

Segundo [OLI1996], a quarta fase é considerada a fase mais problemática do desenvolvimento do SIG, pois envolve intensivamente aspectos comportamentais e que devem ser tratados pela equipe responsável. Nesta fase deve-se preparar a documentação informativa necessária aos usuários, treinar estes usuários, supervisionar a implementação das diversas partes do SIG e acompanhar a implementação do SIG, consolidando um processo adequado de avaliação. É na fase de implantação e avaliação do SIG que se verifica como e onde o SIG pode ser melhorado, comparar com os objetivos originais e analisar todas as qualidades ou defeitos do SIG.

Nesta fase o gerente da organização deve preparar a documentação informativa necessária para os diversos usuários, treinar todos os usuários do SIG, supervisionar a implementação das diversas partes do sistema de informações gerenciais e acompanhar a implementação do SIG consolidando um adequado processo de avaliação, tendo em vista a sua otimização ao longo do tempo.

#### **2.7.3 PROTOTIPAÇÃO**

Segundo [MEL1990], a prototipação representa uma boa solução para a maioria dos problemas desta área. Para ele, a criação de modelos ou protótipos se constitui numa grande solução para a área de desenvolvimento de sistemas, pois:

- não possui uma seqüência rigorosa das etapas de desenvolvimento, como ocorre nas metodologias tradicionais, mas não deve ser considerada como uma forma desorganizada de trabalho;
- antecipa ao usuário final uma versão ou modelo do sistema, para que ele possa avaliar e identificar erros através da utilização;
- como qualquer metodologia de desenvolvimento, a prototipação possui etapas em que são elaboradas as parcelas do produto final pretendido. À medida que uma etapa é concluída, um conjunto de especificações e detalhes técnicos é produzido, devendo ser documentado para que se possa evitar uma situação de desinformação e conseqüente descontrole.

O modelo de prototipação de sistema proposto por [MEL1990] é a metodologia genérica de prototipação de sistemas descrito a seguir:

- a) etapa de exame de viabilidade do projeto: na primeira etapa da metodologia de prototipação, devem ser examinados e avaliados todos os aspectos relativos ao perfil técnico dos analistas e usuários, padrões mínimos de desenvolvimento, bem como os aspectos comportamentais do ambiente organizacional. É uma etapa presente em qualquer metodologia de desenvolvimento de sistemas, pois é nela que se identifica quais as condições para o desenvolvimento ou não de um sistema de informações. Esta etapa caracteriza-se pela análise de três pontos fundamentais:
  - a seleção e alocação de recursos de software e hardware;
  - o exame dos fatos geradores do sistema, ou seja, analisar o “por quê?” do desenvolvimento do sistema e “quais” suas características em termos de complexidade e incerteza; e
  - o planejamento de informática, que indica o caminho e as linhas gerais de fixação do ambiente das informações da empresa;
- b) etapa de identificação de necessidades e requisitos do sistema: serve de base para a elaboração do modelo preliminar do protótipo. É nela que se identifica e define os objetivos do sistema a ser prototipado, os dados gerados e requeridos para o alcance dos objetivos, os relacionamentos entre estes dados e as funções administrativas que os mantêm atualizados. Nesta etapa deve-se analisar a essencialidade do sistema, para só depois partir para o detalhamento dos dados e funções que o compõe e determinar também o tempo de desenvolvimento e a qualidade do protótipo em termos de atendimento das necessidades e requisitos do usuário;
- c) etapa de desenvolvimento do modelo vivo de trabalho: é a etapa que depende quase que totalmente do analista. É com base nos modelos de dados e funções que deve ser construído o modelo de operacionalização do protótipo. É uma etapa que deve ser realizada num curto espaço de tempo para que o usuário não perca o interesse, a motivação e até mesmo a confiança no projeto. Nesta etapa devem ser utilizadas, para o desenvolvimento do modelo vivo, técnicas de construção de

diálogos/interatividade; de elaboração e geração de telas; de elaboração e geração de relatórios; e de geração de estrutura e de relacionamento de dados;

- d) etapa de demonstração e uso do modelo: é a demonstração do protótipo que definirá a aprovação ou a rejeição do projeto. Na etapa de demonstração do modelo devem participar todos os usuários envolvidos com o sistema. É na demonstração que se verifica se o projeto atende às necessidades de informação dos usuários, e se realiza uma série de atividades como: identificação de omissões de dados ou funções, explicação de cada componente do protótipo, identificação de extensões que permitiriam melhorar ainda mais o contexto funcional do usuário, entre outras;
- e) etapa de revisão e melhoramentos: é uma etapa que depende exclusivamente da fase anterior. Com base na etapa anterior podem ser descartados ou incluídos alguns componentes funcionais no protótipo. Devem ser efetuadas nesta etapa todas as alterações requisitadas pelos usuários, correções de erros e acréscimo de novas rotinas nos programas;
- f) etapa de utilização do protótipo: após a realização da revisão e melhoramento dos componentes do protótipo ele deve ser realmente utilizado. Este protótipo final é que servirá unicamente de base para o desenvolvimento do sistema definitivo. Nesta etapa podem ocorrer, ainda, sérias críticas de alguns usuários em relação à estética do protótipo, visto que, como o próprio nome diz é apenas um protótipo, os embelezamentos devem ser feitos somente no desenvolvimento do sistema definitivo.

### 3 DATA MINING

A tecnologia tem tornado fácil o agrupamento e armazenamento de grandes quantidades de dados brutos, mas a análise deste material tende a ser lenta e dispendiosa. Dados armazenados não analisados podem esconder informações úteis. Isso conduz à urgente necessidade do desenvolvimento de métodos semi-automáticos para descobrir este tipo de informação escondida.

Atualmente, muitas revistas de informática e de negócios têm publicado artigos sobre *Data Mining*. Contudo, há poucos anos, muito pouca gente tinha ouvido falar a respeito. Apesar dessa tecnologia ter uma longa evolução de sua história, o termo, como o conhecemos hoje, só foi introduzido recentemente, nos anos 90.

Segundo [GRO1997], *Data Mining* (ou mineração de dados) é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais. O *Data Mining* vai muito além da simples consulta a um banco de dados, no sentido de que permite aos usuários explorar e inferir informação útil a partir dos dados, descobrindo relacionamentos escondidos no banco de dados. Pode ser considerada uma forma de descobrimento de conhecimento em bancos de dados (*Knowledge Discovery in Databases - KDD*), área de pesquisa de bastante evidência no momento, envolvendo Banco de Dados.

As expressões *Data Mining*, mineração de dados ou garimpagem de dados referem-se ao processo de extrair dados potencialmente úteis a partir de dados brutos que estão armazenados em bancos de dados dos diversos sistemas implantados nas organizações. A tecnologia utilizada no *Data Mining* utiliza-se da procura em grandes quantidades de dados armazenados procurando extrair padrões e relacionamentos que podem ser fundamentais para os negócios da organização. O *Data Mining* utiliza-se de um conjunto de técnicas avançadas para identificar os padrões e associações que os dados refletem, com isso oferecendo conclusões que podem trazer valiosas vantagens em nível de mercado para a organização [COM1999].

*Data Mining* é um conjunto de técnicas que envolve métodos matemáticos, algoritmos e heurísticas para descobrir padrões e regularidades em grandes conjuntos de dados [WES1998]. Vários fatores levaram as grandes organizações a armazenar grandes quantidades de dados nos últimos anos. A disponibilidade de computadores a baixo custo pode ser vista como a principal causa do surgimento destas enormes bases de dados. Assim, o mundo das organizações mantém, a um baixo custo, os dados relativos às mais variadas áreas.

Muitas organizações têm investido em tecnologias de informação para ajudar na gerência dos negócios. É nesse contexto que *Data Mining* está cada vez mais sendo utilizado. Ele pode ser visto como uma forma de selecionar, explorar e modelar grandes conjuntos de dados para detectar padrões de comportamento dos consumidores. Desse modo, é possível direcionar melhor uma campanha de *marketing* de uma organização ou adaptar os negócios para que alcancem um maior número de clientes de forma mais personalizada.

Segundo [BIS1999], o setor de *marketing* também está se revolucionando com o uso de *Data Mining*. Em vez de realizar imensas e caras campanhas de âmbito geral, essas organizações descobriram que, dividindo o público-alvo em categorias, é possível realizar campanhas mais direcionadas, mais baratas e com um retorno muito maior. Para dividir o público-alvo em categorias, é necessário conhecer esse público, e o *Data Mining* tem sido imprescindível nesse sentido.

De acordo com [BER1997], o objetivo do *Data Mining* é descobrir o conhecimento, extraí-lo implicitamente sem que seja necessário conhecer a estrutura das informações do banco de dados sobre ele aplicado; este processo é denominado de *Knowledge Discovery in Databases* – KDD, que será detalhado no próximo item.

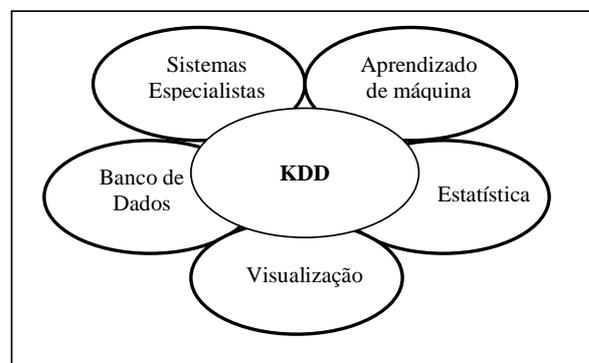
### **3.1 PROSPECÇÃO DE CONHECIMENTO**

Prospecção de conhecimento em bases de dados (*Knowledge Discovery in Databases* - KDD) é um processo que envolve a automação da identificação e do reconhecimento de padrões em um banco de dados. Trata-se de uma pesquisa de fronteira, que começou a se expandir mais rapidamente nos últimos cinco anos. Sua principal característica é a extração

não-trivial de informações a partir de uma base de dados de grande porte. Essas informações são necessariamente implícitas, previamente desconhecidas, e potencialmente úteis [FIG1998].

O processo KDD é interdisciplinar e envolve áreas relativas a aprendizado de máquina, reconhecimento de padrões, bases de dados, estatística e matemática, aquisição de conhecimento para sistemas especialistas e visualização de dados. Este processo utiliza métodos, algoritmos e técnicas oriundos destas diversas áreas, com o objetivo principal de extrair conhecimento a partir de grandes bases de dados. A interdisciplinaridade de áreas no processo KDD pode ser visualizada pela Figura 4 :

**Figura 4 - Processo KDD**



Devido a essas características incomuns, todo o processo de KDD depende de uma nova geração de ferramentas e técnicas de análise de dados, e envolve diversas etapas. A principal, que forma o núcleo do processo, e que muitas vezes se confunde com ele, chama-se *Data Mining*, ou Mineração de Dados, também conhecido como processamento de padrões de dados, arqueologia de dados, ou colheita de informação (*information harvesting*).

O KDD compreende todo o processo de descoberta de dados, enquanto o *Data Mining* refere-se a aplicação de algoritmos para extração de padrões de dados, sem os passos adicionais do KDD e da análise dos resultados [AVI1998].

## 3.2 AS ETAPAS DO PROCESSO DE KDD

O processo de KDD (Figura 5) começa com o entendimento do domínio da aplicação e a relevância do conhecimento em relação às metas a serem atingidas. Em seguida, é feita a seleção dos conjuntos de dados a serem utilizados durante o processo do KDD, isto é, um agrupamento organizado de dados, que será o alvo da prospecção. A etapa da limpeza dos dados (*data cleaning*) vem a seguir, por meio de um pré-processamento dos dados, visando a adequá-los aos algoritmos. Isso se faz pela integração de dados heterogêneos, eliminação de incompletude dos dados, repetição de tuplas, problemas de tipagem, etc. Essa etapa pode tomar até 80% do tempo necessário para todo o processo, devido às bem conhecidas dificuldades de integração de bases de dados heterogêneas [FAY1996].

**Figura 5 - As etapas do processo de KDD**



Fonte: [FIG1998]

Os dados pré-processados devem, ainda, passar por uma transformação que os armazena adequadamente, visando facilitar o uso das técnicas de *Data Mining*.

Prosseguindo no processo, chega-se à fase de *Data Mining* especificamente, que começa com a escolha dos algoritmos a serem aplicados. Essa escolha depende fundamentalmente do objetivo do processo de KDD: classificação, segmentação, agrupamento por afinidades, estimativas, árvores de decisão, etc. De modo geral, na fase de *Data Mining*, ferramentas especializadas procuram padrões nos dados. Essa busca pode ser efetuada automaticamente pelo sistema ou interativamente com um analista responsável pela geração de hipóteses. Diversas ferramentas distintas, como redes neurais, indução de árvores de decisão, sistemas baseados em regras e programas estatísticos, tanto isoladamente quanto em combinação, podem ser então aplicadas ao problema. Em geral, o processo de busca é iterativo, de forma que os analistas revêm o resultado, formam um novo conjunto de

questões para refinar a busca em um dado aspecto das descobertas, e realimentam o sistema com novos parâmetros. Ao final do processo, o sistema de *Data Mining* gera um relatório das descobertas, que passa então a ser interpretado pelos analistas de mineração. Somente após a interpretação das informações obtidas encontra-se o conhecimento.

Uma diferença significativa entre *Data Mining* e outras ferramentas de análise está na maneira como exploram as inter-relações entre os dados. As diversas ferramentas de análise disponíveis dispõem de um método baseado na verificação, isto é, o usuário constrói hipóteses sobre inter-relações específicas e então verifica ou refuta, através do sistema. Esse modelo torna-se dependente da intuição e habilidade do analista em propor hipóteses interessantes, em manipular a complexidade do espaço de atributos, e em refinar a análise baseado nos resultados de consultas ao banco de dados potencialmente complexas. Já o processo de *Data Mining* fica responsável pela geração de hipóteses, garantindo mais rapidez, acurácia e completude aos resultados [KRE1999].

### 3.3 FUNÇÕES DO DATA MINING

O *Data Mining* pode desempenhar uma série limitada de tarefas, dependendo das circunstâncias. Cada classe de aplicação em *Data Mining* tem como base um conjunto de algoritmos que serão usados na extração de relações relevantes dentro de uma massa de dados [HAR1988]:

- a) classificação;
- b) estimativa;
- c) agrupamento por afinidade;
- d) previsão;
- e) segmentação.

Cada uma destas propostas difere quanto à classe de problemas que o algoritmo será capaz de resolver.

### **3.3.1 CLASSIFICAÇÃO**

Classificação é uma técnica que consiste no mapeamento ou pré-classificação de um conjunto pré-definido de classes. Em geral, algoritmos de classificação incluem árvores de decisão ou redes neurais. Os algoritmos classificadores utilizam-se de exemplos para determinar um conjunto de parâmetros, codificados em um modelo, que será mais tarde utilizado para a discriminação do restante dos dados. Uma vez que o algoritmo classificador foi desenvolvido de forma eficiente, ele será usado de forma preditiva para classificar novos registros naquelas mesmas classes pré-definidas.

### **3.3.2 ESTIMATIVA**

Uma variação do problema de classificação envolve a geração de valores ao longo das dimensões dos dados: são os chamados algoritmos de estimativa. A estimativa lida com resultados contínuos, ao contrário da classificação que lida com resultados discretos. Fornecidos alguns dados, usa-se a estimativa para estipular um valor para alguma variável contínua desconhecida como receita, altura ou saldo de cartão de crédito. Em vez de um classificador binário determinar um risco “positivo” ou “negativo”, a técnica gera valores de “escore”, dentro de uma determinada margem. A abordagem de estimativa tem a grande vantagem de que os registros individuais podem ser agora ordenados por classificação, e as redes neurais são adequadas a esta tarefa.

### **3.3.3 AGRUPAMENTO POR AFINIDADE**

Trata-se de um algoritmo tipicamente endereçado à análise de mercado, em que o objetivo é encontrar tendências dentro de um grande número de registros de compras, por exemplo, expressas como transações. Essas tendências podem ajudar a entender e explorar padrões de compra naturais, e podem ser usadas para ajustar mostruários, modificar prateleiras ou propagandas, e introduzir atividades promocionais específicas. Um exemplo mais distinto, em que essa mesma técnica pode ser utilizada, é o caso de um banco de dados escolar, relacionando alunos e disciplinas. Uma regra do tipo “85% dos alunos inscritos em ‘Programação I’ também estão inscritos em ‘Teoria da Computação’” pode ser usada pela

direção ou secretaria para planejar o currículo anual, ou alocar recursos como salas de aula e professores.

### 3.3.4 PREVISÃO

A previsão é o mesmo que classificação ou estimativa, exceto pelo fato de que os registros são classificados de acordo com alguma atitude futura prevista. Em um trabalho de previsão, o único modo de confirmar a precisão da classificação é esperar para ver. Essa tarefa é uma variante do problema de agrupamento por afinidades, no qual as regras encontradas entre as relações podem ser usadas para identificar seqüências interessantes, que serão utilizadas para prever acontecimentos subseqüentes. Nesse caso, não apenas a coexistência de itens dentro de cada transação é importante, mas também a ordem em que aparecem, e o intervalo entre elas. Seqüências podem ser úteis para identificar padrões temporais, por exemplo, entre compras em uma loja, ou utilização de cartões de crédito, ou ainda tratamentos médicos.

### 3.3.5 SEGMENTAÇÃO

A segmentação é um processo de agrupamento de uma população heterogênea em vários subgrupos ou *clusters* mais homogêneos. O que a distingue da classificação é que segmentação não depende de classes pré-determinadas. Essa segmentação é realizada automaticamente por algoritmos que identificam características em comum e particionam o espaço n-dimensional definido pelos atributos. Os registros são agrupados de acordo com a semelhança e depende do usuário determinar qual o significado de cada segmento, caso exista algum. Muitas vezes a segmentação é uma das primeiras etapas dentro de um processo de *Data Mining*, já que identifica grupos de registros correlatos, que serão usados como ponto de partida para futuras explorações. O exemplo clássico é o de segmentação demográfica, que serve de início para uma determinação das características de um grupo social, visando desde hábitos de compras até utilização de meios de transporte.

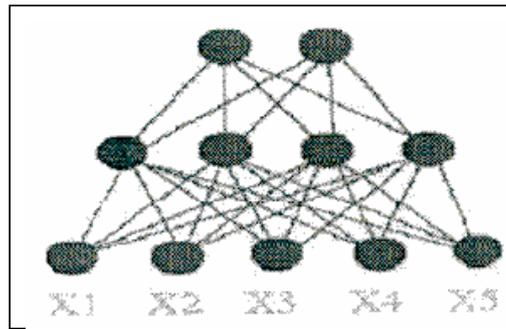
## 3.4 TÉCNICAS DE DATA MINING

As técnicas utilizadas atualmente em *Data Mining* são extensões de métodos analíticos já conhecidos há algum tempo. O que as diferencia, é que as técnicas aqui apresentadas são mais voltadas a encontrar estratégias para os negócios. Isto acontece graças ao aumento do desempenho dos computadores. As técnicas apresentadas a seguir são extraídas de Bispo (1999).

### 3.4.1 REDES NEURAS ARTIFICIAIS

As redes neurais artificiais trabalham de forma semelhante às redes neurais biológicas. São coleções de nós conectados que possuem uma camada de entrada e outra de saída e que realizam diversos processamentos em seus nós. Uma rede neural artificial é capaz de aprender sozinha, por meio de um treinamento apropriado. Em cada passo do treinamento, o conjunto de dados de saída é comparado com um resultado já conhecido; se for diferente, uma correção é calculada e processada automaticamente nos nós da rede. Esses passos são repetidos até que uma condição de parada seja atingida, ou quando um número de correções estipulado é alcançado.

Redes neurais artificiais são um processo opaco; quais são os meios para se obter um resultado não tem uma interpretação clara. Normalmente esse processo é aplicado sem se entender a razão por trás de seus resultados. Essa falta de explicações inibe a confiança, aceitação e aplicação dos resultados. Alguns produtos de redes neurais artificiais têm algoritmos que podem traduzir o processo utilizado em um conjunto de regras que podem ajudar a compreender o que ela está fazendo. Dependendo da aplicação, uma rede neural artificial pode consumir um tempo muito longo de aprendizagem. É muito utilizada em aplicações que envolvem classificação. A estrutura de uma rede neural artificial pode ser descrita na Figura 6:

**Figura 6 – Estrutura de Rede Neural Artificial**

Fonte: [AVI1998]

### 3.4.2 ALGORITMOS GENÉTICOS

É uma técnica de otimização que usa diferentes processos, tais como combinação genética, mutação e seleção natural, baseando-se em conceitos de evolução. Os algoritmos genéticos aplicam a mecânica da genética e seleção natural à pesquisa usada para encontrar os melhores conjuntos de parâmetros que descrevem uma função de previsão. Eles são utilizados no *Data Mining* dirigido e são semelhantes à estatística, em que a forma do modelo precisa ser conhecida em profundidade. Os algoritmos genéticos usam os operadores seleção, cruzamento e mutação para desenvolver sucessivas gerações de soluções. Com a evolução do algoritmo, somente os mais previsíveis sobrevivem, até as funções convergirem em uma solução ideal [BER1997].

### 3.4.3 INDUÇÃO DE REGRAS

A Indução de Regras (IR) ou *Rule Induction*, se refere à detecção de tendências dentro de grupos de dados, ou de “regras” sobre o dado. As regras são, então, apresentadas aos usuários como uma lista “não encomendada”. Vários algoritmos e índices são colocados para executar esse processo, sendo que a maioria do processo é feita pela máquina, e uma pequena parte é feita pelo usuário.

Por exemplo, a tradução das regras para dentro de um modelo aproveitável é feito pelo usuário, ou por uma interface de árvores de decisão. Do ponto de vista do usuário, o maior problema com as regras é que o programa de *Data Mining* não faz o *ranking* das regras por sua importância. O analista de negócio é então forçado a encarregar-se de criar um manual de análise para todas as regras relatadas a fim de determinar aquelas que são mais importantes no modelo de *Data Mining*, e para os assuntos de negócio envolvidos. E isso pode ser um processo tedioso.

### 3.4.4 ANÁLISE ESTATÍSTICA DE SÉRIES TEMPORAIS

A estatística é a mais antiga tecnologia em *Data Mining*, e é parte da fundamentação básica de todas as outras tecnologias. Ela incorpora um envolvimento muito forte do usuário, exigindo engenheiros experientes, para construir modelos que descrevem o comportamento do dado através dos métodos clássicos de matemática. Interpretar os resultados dos modelos requer conhecimento especializado. O uso de técnicas de estatística também requer um trabalho muito forte de máquinas/engenheiros.

A análise de séries temporais é um exemplo disso, apesar de frequentemente ser confundida como um gênero mais simples de *Data Mining* chamado *forecasting* (previsão).

Enquanto a análise de séries temporais é um ramo altamente especializado da estatística, o *forecasting* é de fato uma disciplina muito menos rigorosa, que pode ser satisfeita, embora com menos segurança, pela maioria das outras técnicas de *Data Mining*.

### 3.4.5 VISUALIZAÇÃO

As técnicas de Visualização são um pouco mais difíceis de definir, porque muitas pessoas as definem como “complexas ferramentas de visualização”, enquanto outras como simplesmente a capacidade de geração de gráficos.

Nos dois casos, a Visualização mapeia o dado, sendo minerado de acordo com dimensões especificadas. Nenhuma análise é executada pelo programa de *Data Mining* além de manipulação estatística básica. O usuário, então, interpreta o dado enquanto olha para o

monitor. O analista pode pesquisar a ferramenta depois, para obter diferentes visões ou outras dimensões.

### 3.4.6 ÁRVORES DE DECISÃO

Segundo [BIS1999], árvores de decisão dividem os dados em subgrupos com base nos valores das variáveis. O resultado é uma hierarquia de declarações do tipo “Se...então...”, que são utilizadas principalmente para classificar dados. Há uma onda de interesse em produtos baseados em árvore de decisão, principalmente porque eles são mais rápidos que as redes neurais para alguns problemas organizacionais, e também é mais fácil compreender o seu funcionamento.

Árvores de decisão expressam uma forma simples de lógica condicional buscando a representação de uma série de questões que estão escondidas sob a base da dados. Em uma árvore de decisão existem dois tipo de atributos: o decisivo, que é aquele que contém o resultado ao qual queremos chegar; e os não-decisivos que contêm os valores que conduzem a uma decisão.

Por uma fórmula matemática denominada entropia, são realizados cálculos sobre os atributos não-decisivos, denominados classes, nos quais é escolhido um nó inicial também chamado raiz. A partir deste nó, será realizada uma série de novos cálculos com o objetivo de decidir a estrutura de formação da árvore a ser gerada. Este processo é repetido até que todos os atributos a serem processados estejam perfeitamente classificadas ou já se tenham processado todos os atributos.

Os três principais algoritmos conhecidos que implementam árvores de decisão, são ID3, C4.5 e PERT, sendo que os algoritmo C4.5 e PERT são um aperfeiçoamentos do algoritmo ID3, com alguns conceitos avançados de podagem (técnica de cortar nós da árvore que não são potencialmente úteis) e preocupação com o seu desempenho em relação ao tempo de processamento.

O objetivo do algoritmo ID3 é gerar os valores categóricos de um atributo chamado *classe*, para isso utilizando-se de um método de classificação que tem o objetivo de realizar

testes que são introduzidos na árvore, separando os casos de treino em subconjuntos. Cada subconjunto deve consistir de exemplos de uma única classe.

A distribuição de classes pode ser representada em forma de uma lista de probabilidades  $p(c_1) \dots p(c_n)$ , em que cada  $p$  indica a probabilidade de um exemplo pertencer a uma classe.

De acordo com [BIS1999], os valores das funções que calculam essas probabilidades representam a informação necessária para classificar um caso e são chamados de entropia e *gain*, sendo calculados com as seguintes fórmulas demonstradas na Figura 7.

**Figura 7 - Fórmulas para calcular entropia e *gain***

Entropia(S) =  $\sum -p(I) \log_2 p(I)$  onde

Log2 é o logaritmo de número com base 2

$p(I)$  é quantidade de ocorrências cada valor possível de uma classe dividido pela quantidade total da classe.

Gain (S,A) = Entropia(S) -  $\sum ((|S_v| / |S|) * \text{Entropia}(S_v))$  onde

$\sum$  é cada valor possível de todos os valores do atributo A

$S_v$  é a quantidade de ocorrências de cada atributo definido por A

$|S_v|$  é o número total de elementos definido por  $S_v$

$|S|$  é o número total de elementos da coleção.

[B1] Comentário:

*Fonte: adaptado de [BIS1999]*

## 4 PRÓ-REITORIA DE ENSINO - PROEN

A Universidade Regional de Blumenau – FURB é composta de quatro Pró-Reitorias:

- a) Pró-Reitoria de Administração – PROAD
- b) Pró-Reitoria de Extensão e Relações Comunitárias – PROERC
- c) Pró-Reitoria de Pesquisa e Pós-Graduação – PROPEP
- d) Pró-Reitoria de Ensino de Graduação – PROEN

À Pró-Reitoria de Ensino de Graduação (PROEN) compete a execução das tarefas relacionadas à organização, coordenação e superintendência de todas as atividades acadêmicas docentes e discentes da Instituição, em nível de graduação. São atribuições da Pró-Reitoria de Ensino de Graduação:

- a) Superintender os cursos de graduação, baixando normas para sua execução, supervisionando e orientando as atividades dos Colegiados de Curso.
- b) Definir as políticas gerais de aprimoramento dos docentes e da qualidade do Ensino de Graduação.
- c) Estabelecer, em conjunto com os Departamentos e Colegiados de Cursos, as linhas gerais do Projeto Político Pedagógico da Universidade.
- d) Estimular projetos orientados para a melhoria da qualidade dos cursos de graduação.

A Pró-Reitoria de Ensino conta com a seguinte estrutura organizacional, de acordo com a Resolução 25/97, que aprova o Regimento Geral da Reitoria da Universidade Regional de Blumenau:

- 1. Divisão de Registros Acadêmicos - DRA
  - 1.1. Seção de Registros de Discentes da Graduação
  - 1.2. Seção de Diplomas
- 2. Divisão de Administração do Ensino da Graduação - DIAEG
  - 2.1. Seção de Registros de Docentes - SRD
  - 2.2. Seção de Apoio Pedagógico - SAP
  - 2.3. Seção de Apoio Técnico ao Ensino - SATE

- 2.4. Seção de Apoio ao Ensino Informatizado
- 2.5. Seção de Audiovisuais
- 2.6. Comissão de Avaliação Institucional - COMAVI
- 2.7. Central de Informação aos Alunos – CIA & CIA

## **4.1 CENTRAL DE INFORMAÇÃO AOS ALUNOS – CIA & CIA**

Neste item será apresentada a CIA & CIA, a qual é a área em que será aplicado este trabalho. A definição de outras áreas podem ser encontradas no Regimento Geral da Reitoria da Universidade Regional de Blumenau.

A CIA & CIA é o setor responsável pela divulgação dos cursos da Universidade Regional de Blumenau. Dentre as suas principais atividades destacam-se:

1. fornecer aos alunos do ensino médio que estão cursando o 3º ano, ou já concluíram, informações sobre os cursos oferecidos pela Universidade. São visitadas 63 escolas das cidades de Jaraguá do Sul, Brusque, Joinville, São Bento do Sul, Rio Negrinho, Guaramirim, Massaranduba, Timbó, Benedito Novo, Indaial, Rio dos Cedros, Dr.Pedrinho, Ascurra, Apiúna, Pomerode e Rio do Sul, efetuando o convite para participarem do Fórum de Informação dos Cursos de Graduação (FIC-FURB), pessoalmente (sala por sala). O objetivo desta atividade é colocar adolescentes e adultos com dificuldades na escolha Universitária no curso certo.
2. fornecer informações sobre os cursos de graduação da FURB e sobre o vestibular do SUPRA via fone, fax e e-mail e enviar correspondência aos alunos, anexando folhetos dos cursos, provas anteriores, índice candidato/vaga, etc. Por meio de um tratamento diferencial aos futuros alunos com informações precisas sobre a Universidade, pode-se obter maior número de inscritos no SUPRA.
3. demonstrar ao aluno a preocupação/organização da Universidade com seus futuros alunos, enviando, por mala direta, um manual de inscrição do SUPRA para cada aluno. Para esta atividade, exige-se um banco de dados (dados sobre alunos e escolas) bastante atualizado.

4. realizar, com os adolescentes com dificuldade de efetuar sua escolha Universitária, Técnicas de Orientação Profissional, sendo que ao final desta atividade o aluno tenha no máximo três opções de escolha para os cursos da FURB.
5. mostrar ao aluno a preocupação da Universidade em relação ao mercado de trabalho, proporcionando uma real visualização regional, já que os futuros acadêmicos sentem a necessidade de saber mais sobre o mercado de trabalho e suas possibilidades de ingresso. Por uma interação entre indústria/comércio da região, redefinindo as necessidades de empregabilidade, entrando em contato com todas as empresas, apresentando a Universidade, além de verificar as necessidades de cada uma (emprego), é possível formar uma parceria FURB/INDÚSTRIA/COMÉRCIO, proporcionando, ainda, a possibilidade de participação dos funcionários no FIC-FURB.
6. tornar a Universidade mais conhecida para alunos de escolas distantes da cidade de Blumenau (Chapecó, Criciúma, Tubarão, Curitiba, Joaçaba, São Miguel do Oeste, etc), por meio de palestras sobre os cursos da Universidade.
7. participar de feiras e eventos com objetivo de divulgação da Universidade, atendendo aos alunos, entregando material informativo, esclarecendo dúvidas e informando quais os cursos oferecidos pela Universidade.
8. formar parcerias com Bancos ou outras empresas que possam patrocinar o material a ser distribuído na matrícula dos “calouros”; realizar palestras de treinamento aos monitores envolvidos na matrícula e organizar os *kits* entregues aos alunos; organizar, num só local todas as atividades relativas à matrícula de “calouros” (Ginásio de Esportes), proporcionando, ainda, serviços adicionais, foto, xerox, DAE, café, etc. O objetivo desta atividade é proporcionar aos calouros a possibilidade de efetuar a matrícula num único local, tornando este momento numa ocasião prazerosa ao aluno/pais com atendimento personalizado, diferenciando-se das demais Instituições por explicar o funcionamento da Universidade, além de dar orientações sobre seu curso, sua sala, as vantagens inerentes ao fato de serem alunos da Instituição.
9. conseguir ampliar o número de inscritos no SUPRA, com alunos que estudam em outras cidades e que normalmente fazem uma segunda opção de vestibular (além da Federal), treinando alunos da FURB que têm sua família em uma destas

idades, fornecendo-lhes passagens até o local, seguro de vida e um crédito financeiro para cada cursinho ou escola particular visitada.

A CIA & CIA, com o evento do Fórum de Informação dos Cursos de Graduação (FIC – FURB), realizado todos os anos e reunindo mais de 3 mil alunos, entre os que estão estudando ou já concluíram o 3º ano do ensino médio, vários dados são coletados através da aplicação de questionários. Existe, hoje, uma necessidade grande de gerenciamento desses dados, de forma a serem transformados em informações úteis, visto ser um processo que abrange investimentos altos, tanto em recursos financeiros, quanto em recursos humanos.

No Anexo 1 é apresentado o questionário aplicado durante o evento do FIC-FURB, para Avaliação Qualitativa de Demanda por Cursos, objeto de estudo deste trabalho.

## 5 TECNOLOGIAS E FERRAMENTAS UTILIZADAS

No presente capítulo serão explicadas as principais tecnologias e ferramentas utilizadas neste trabalho: análise estruturada, diagrama de fluxo de dados, modelo entidade-relacionamento, dicionário de dados, ferramentas CASE, *Power Designer*, ferramenta de programação – ambiente visual Delphi e trabalhos correlatos.

### 5.1 ANÁLISE ESTRUTURADA

O desenvolvimento deste protótipo foi baseado na metodologia de desenvolvimento de sistemas Análise Estruturada. De acordo com [YOU1990], a análise é uma fase crítica do desenvolvimento de sistemas; com isso, afeta todas as fases seguintes do desenvolvimento. A Análise Estruturada tem como objetivo resolver essas dificuldades fornecendo uma abordagem sistemática, para desenvolver, inicialmente, a análise, e posteriormente produzir uma especificação de sistema.

Segundo [YOU1990], a análise estruturada é uma metodologia na qual tanto os analistas quanto os usuários sabem que o produto final da prototipação será o próprio sistema, já na sua forma aperfeiçoada.

#### 5.1.1 CONCEITOS

De acordo com [JOA1993], a análise estruturada é um tipo de análise de sistemas que tem como objetivo resolver as dificuldades encontradas na fase de análise no desenvolvimento de sistemas e programas de software. As dificuldades da fase de análise podem ser representadas por problemas de comunicação, mudanças nos requisitos do sistema e técnicas inadequadas de avaliação. A análise estruturada fornece uma abordagem sistemática, etapa por etapa, para se desenvolver a análise e produzir uma especificação de sistema nova e melhorada, centralizando-se em uma comunicação clara e concisa.

A análise estruturada tem por objetivo a modelagem funcional dos sistemas por meio da especificação dos processos de transformação de dados. Utiliza-se de ferramentas gráficas para a visualização dos fluxos de informação e suas transformações, e funciona por meio da decomposição funcional, por uma abordagem *top-down*, e por refinamentos sucessivos. Consiste basicamente em diagrama de contexto, diagrama de fluxo de dados (DFD), modelo entidade-relacionamento (MER), dicionário de dados, e ferramentas para a descrição lógica dos processos.

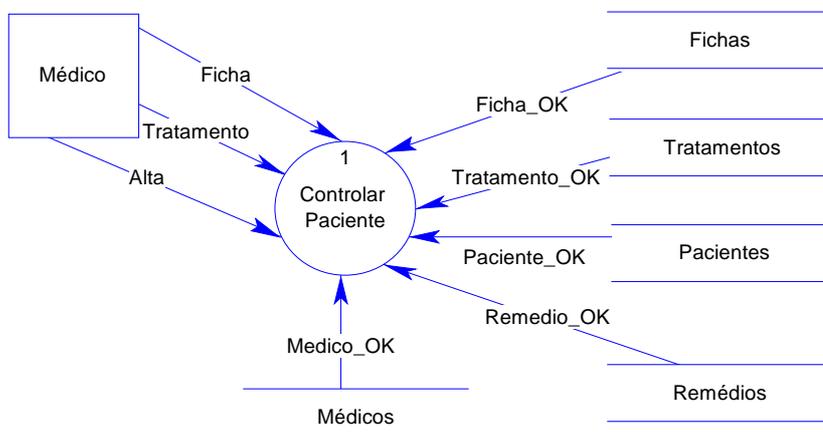
### 5.1.2 DIAGRAMA DE FLUXO DE DADOS

Conforme [JOA1993], o Diagrama de Fluxo de Dados (DFD) é usado como o primeiro passo em um projeto estruturado. O DFD apresenta os processos e o fluxo de dados global entre esses processos em um sistema. O DFD é uma ferramenta de análise de sistemas, para desenhar os componentes procedurais básicos e os dados que passam entre eles.

Segundo [YOU1990], o DFD ao nível lógico é a ferramenta principal para entendimento e manipulação de um sistema de qualquer complexidade, juntamente com o refinamento desta notação para uso em análise.

Na Figura 8 está representado um exemplo de Diagrama de Fluxo de Dados.

**Figura 8 - Diagrama de Fluxo de Dados**



### 5.1.3 MODELO ENTIDADE-RELACIONAMENTO

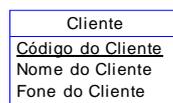
De acordo com [JOA1993], os analistas de dados precisam, para o levantamento e compreensão dos dados, da ajuda das pessoas envolvidas com o sistema e de ferramentas que permitam a diagramação de dados de forma compreensível.

O diagrama de entidade-relacionamento ou modelo de entidade-relacionamento (MER) é a maneira de se obter esta diagramação. O MER determina os relacionamentos entre as entidades que fazem parte da administração da empresa. O MER pode ser decomposto em modelos de dados detalhados.

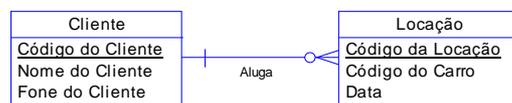
O objetivo da construção de MER e modelos de dados é criar uma descrição da semântica dos dados da realidade e suas necessidades informacionais, ou seja, aprender a realidade e transmitir informações precisas sobre a mesma.

O modelo entidade-relacionamento é composto basicamente pelas seguintes estruturas:

- a) **Entidade** - são componentes reais ou abstratos, a respeito dos quais são armazenados dados. Um tipo de entidade refere-se a uma classe de entidades que mantêm os mesmos atributos. Como exemplo, pode-se citar a entidade Cliente.



- b) **Atributo** - é a representação de propriedades de uma entidade, ou seja, é uma única peça de representação de uma entidade. Como exemplo, pode-se citar Nome do Cliente.
- c) **Ocorrência** - é o conjunto de atributos de uma entidade. Como exemplo, pode-se citar Código do Cliente + Nome do Cliente + Fone do Cliente.
- d) **Relacionamento** - é uma associação entre dois tipos de entidades que mostra como elas são relacionadas. Como exemplo, pode-se citar Aluga.



- e) **Identificador** - é um ou mais atributos que determinam de modo único uma ocorrência de entidade. Como exemplo, pode-se citar Código do Cliente.
- f) **Grau de relacionamento** - é o número de entidades que participam de uma associação.
- g) **Classe de relacionamento ou cardinalidade** - quantas ocorrências de cada entidade são envolvidas no relacionamento.

#### 5.1.4 DICIONÁRIO DE DADOS

Segundo [JOA1993], o dicionário de dados contém definições de todos os dados do MER, informação física sobre os dados, tais como dispositivos de armazenamento e métodos de acesso aos dados.

De acordo com [YOU1990], o dicionário de dados é o local estruturado no qual são mantidos os conteúdos dos fluxos de dados, os conteúdos dos depósitos de dados e dos processos.

## 5.2 FERRAMENTAS CASE

Com o intuito dos analistas de sistemas desenvolverem seus projetos de maneira mais rápida, mais abrangente e mais facilmente modificável, surge, então, a necessidade de se utilizar ferramentas automatizadas de apoio ao desenvolvimento de sistemas. Esta necessidade ocasionou, então, o surgimento da automação do desenvolvimento de software, pela técnica denominada *Computer Aided Software Engineering* – Engenharia de Software Apoiada por Computador (CASE). O surgimento da tecnologia CASE é considerado por alguns especialistas, a mais profunda transformação ocorrida na comunidade de software.

De acordo com [JOA1993], dentre as várias funções das ferramentas CASE podemos citar:

- utilização da diagramação (MER, DFD, Fluxogramas);
- prototipação;
- geração automática de códigos;

- documentação automática;
- dicionário de dados; e
- informações sobre o projeto.

De uma maneira mais genérica, pode-se citar como benefícios do uso uma ferramenta CASE os seguintes pontos:

- tornar prático o uso das técnicas estruturadas;
- melhorar a qualidade do software desenvolvido;
- simplificar e reduzir custos de manutenção;
- acelerar o processo de desenvolvimento;
- aumentar a produtividade da equipe de desenvolvimento; e
- promover a comunicação eficiente entre os analistas.

### **5.2.1 POWER DESIGNER**

É uma ferramenta que pode ser usada tanto para a criação de diagramas de fluxo de dados como para a criação de diagramas entidade-relacionamento. O Power Designer 6.1.0 é composto de três módulos, dentre os quais pode-se citar: Power Designer Process Analyst, para a criação de modelos de fluxos de dados (DFD's); e Power Designer Data Architect, para a criação de modelos entidade-relacionamento (MER). Apenas estes dois componentes foram utilizados na análise deste trabalho.

## **5.3 FERRAMENTA DE PROGRAMAÇÃO – AMBIENTE VISUAL DELPHI**

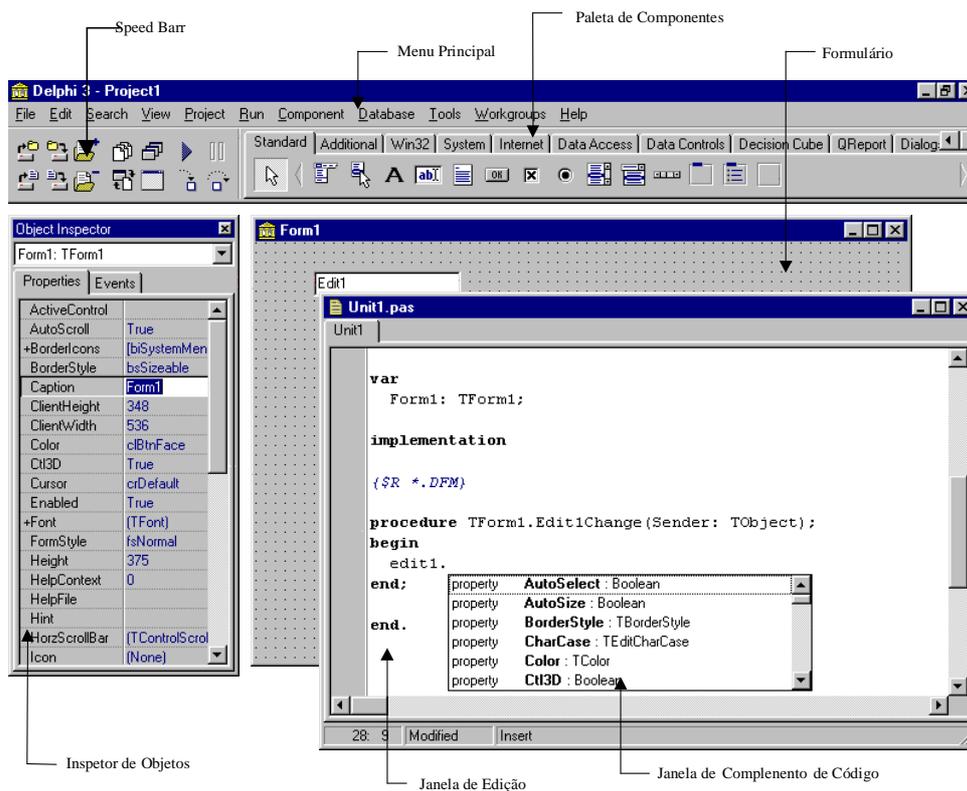
O Delphi na versão 5.0, para plataforma Windows 95, da empresa Borland, e o Banco de Dados Paradox é usado na implementação do protótipo deste trabalho.

Segundo [CAN2000], Delphi é uma versão de desenvolvimento rápido de aplicativos do Turbo Pascal para Windows. O Delphi oferece uma interface melhorada e muitos recursos que facilitam o desenvolvimento de aplicativos. O Delphi oferece ao desenvolvedor de

aplicativos vários diferenciais, tais como a combinação de uma barra de atalho e de vários auxiliares de programação, como o inspetor de objetos.

O Delphi é baseado em projetos. Um projeto é, essencialmente, uma aplicação em Delphi e deve-se em primeiro lugar determinar qual o tipo de interface de usuário será utilizada. O Delphi permite a manipulação dos componentes no programa através de suas propriedades e métodos, dispensando quase todo o acesso de baixo nível do Windows. Na Figura 9 é apresentada a área de trabalho do Delphi.

**Figura 9 - Área de trabalho do Delphi**



## 5.4 TRABALHOS CORRELATOS

A seguir, serão apresentados alguns trabalhos já desenvolvidos, entre os quais, os Trabalhos de Conclusão de Curso dos acadêmicos Ricardo Kremer ([KRE1999]) e Geandro Compolt ([COM1999]).

O objetivo principal do trabalho de Ricardo Kremer era auxiliar na tomada de decisões por meio de um Sistema de Apoio à Decisão utilizando técnicas de *Data Mining*, mais especificamente para efetuar previsões genéricas. Foi implementado um protótipo que permitia ao usuário definir um modelo de previsão, no qual ele poderia ser treinado para responder às variáveis de previsão com flexibilidade. Foram analisadas características de Sistemas de Informação, *Data Warehouses*, técnicas de *Data Mining*, Inteligência Artificial e Redes Neurais.

No trabalho de Geandro Compolt, o objetivo principal era gerar um modelo de classificação de dados utilizando técnicas de *Data Mining*, mais especificamente árvores de decisão. Foi implementado um protótipo que permitia ao usuário definir um valor-prioridade para cada atributo que fazia parte do modelo de classificação. Foram analisadas características de Sistemas de Informação, técnicas de *Data Mining* e montada uma base de dados fictícia, com informações de condições que conduziam à concessão de crédito a fornecedores.

Neste trabalho, o objetivo principal é gerar um modelo de classificação de dados utilizando técnicas de *Data Mining*, mais especificamente árvores de decisão. Para auxiliar esta tarefa, é implementado um protótipo que permite ao usuário definir um valor-prioridade para cada atributo que faz parte do modelo de classificação. Para a elaboração do protótipo são analisadas as características de Sistemas de Informação e técnicas de *Data Mining*, e montada uma base de dados fornecida pela CIA & CIA, que é aplicada à classificação.

## 6 DESENVOLVIMENTO DO PROTÓTIPO

Levando em conta os objetivos propostos por este trabalho, construiu-se um Sistema de Informação Gerencial que fosse flexível e de fácil utilização. Utilizaram-se as fases do SIG proposto por [OLI1996]. Aproveitando a flexibilidade da linguagem escolhida, resolveu-se utilizar a filosofia *Data Mining* com as etapas do processo KDD, e a técnica de árvores de decisão. Pelo uso da análise estruturada, conseguiu-se desenvolver um protótipo com as informações do questionário de Avaliação Qualitativa de Demanda por Cursos, objeto de prioridade no contexto atual para a CIA & CIA.

A seguir será apresentada a implementação da metodologia SIG, especificação e apresentação das telas.

### 6.1 METODOLOGIA SIG

Na primeira fase do SIG, conforme item 2.7.2.1, obteve-se uma idéia preliminar e geral do volume e complexidade do projeto por meio da realização de reuniões e entrevistas com o responsável pela Central de Informação aos Alunos da Universidade. Com isto, verificou-se que existem problemas no processamento dos dados coletados pela CIA & CIA, que se utiliza do banco de dados do Sphinx Léxica para armazenar todos os seus dados, sendo este um software que não satisfaz completamente os objetivos da organização, pois somente emite um relatório com resultados em percentuais para as perguntas do questionário de Avaliação Qualitativa de Demanda por Cursos, não demonstrando a situação real entre a relação dos dados armazenados e as situações consideradas ideais para a Universidade, como por exemplo se a FURB está realmente em primeiro lugar na preferência dos alunos, se os cursos em que se investe mais estão realmente sendo mais procurados, entre outros.

Na Fase II do SIG, descrita no item 2.7.2.2, foram identificadas as informações relacionadas às atividades do processo de tomada de decisões e implementadas e avaliadas as novas informações dentro do contexto decisório da CIA & CIA, através do Administrador,

que indicou quais as questões de maior relevância da Avaliação Qualitativa de Demanda dos Cursos (Anexo 1):

- a) resultado das perguntas dos questionários aplicados no evento do FIC – FURB;
- b) comparativo de algumas questões entre os anos de 1999 e 2000 (Questões 9, 10, 11, 12, 14, 15, 16, 18 e 19) demonstrando os percentuais de cada resposta ;
- c) preferência dos alunos pelas instituições de ensino superior da região;
- d) preferência pelas diferentes áreas de ensino;
- e) preferência dos alunos em itens que compõem a infra-estrutura de um curso;

Nas Fases III e IV do SIG, descritas respectivamente nos itens 2.7.2.3 e 2.7.2.4, foram escolhidas as técnicas e ferramentas utilizadas neste trabalho. Optou-se pela utilização da Mineração de Dados por ser um conjunto de técnicas que permite selecionar os dados de maior relevância para o Administrador, transformando-os em informações úteis para o processo decisório. Após a especificação de prioridades dos atributos previamente definidas pelo Administrador apresentadas nas tabelas 2, 3, 4, 5, 6, 7 e 8 (Anexo 2), a técnica de árvore de decisão foi aplicada para verificar a relação entre os dados armazenados e os resultados esperados pelo Administrador. Os dados dos questionários foram importados do Banco de Dados do Sphinx Léxica para o Banco de Dados Paradox do ambiente visual Delphi. Foi desenvolvida a estrutura lógica geral e completado o fluxo geral do sistema de informações para então, ser desenvolvido o protótipo.

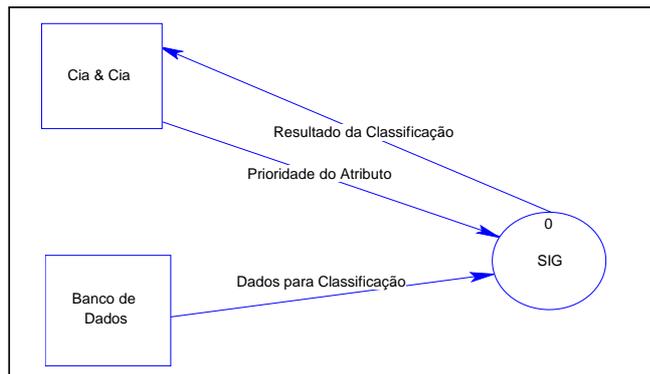
## **6.2 ESPECIFICAÇÃO**

A seguir serão mostrados o Diagrama de Contexto, Diagrama de Fluxo de Dados e Modelo Entidade-Relacionamento do Sistema de Informação Gerencial CIA & CIA.

### **6.2.1 DIAGRAMA DE CONTEXTO**

Para o desenvolvimento do SIG deste trabalho o Diagrama de Contexto está representado na Figura 10.

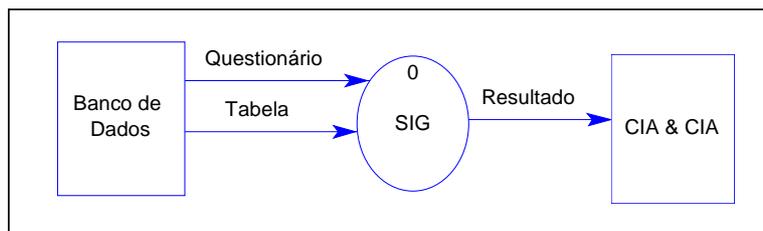
**Figura 10 - Diagrama de Contexto do Sistema de Informação Gerencial**



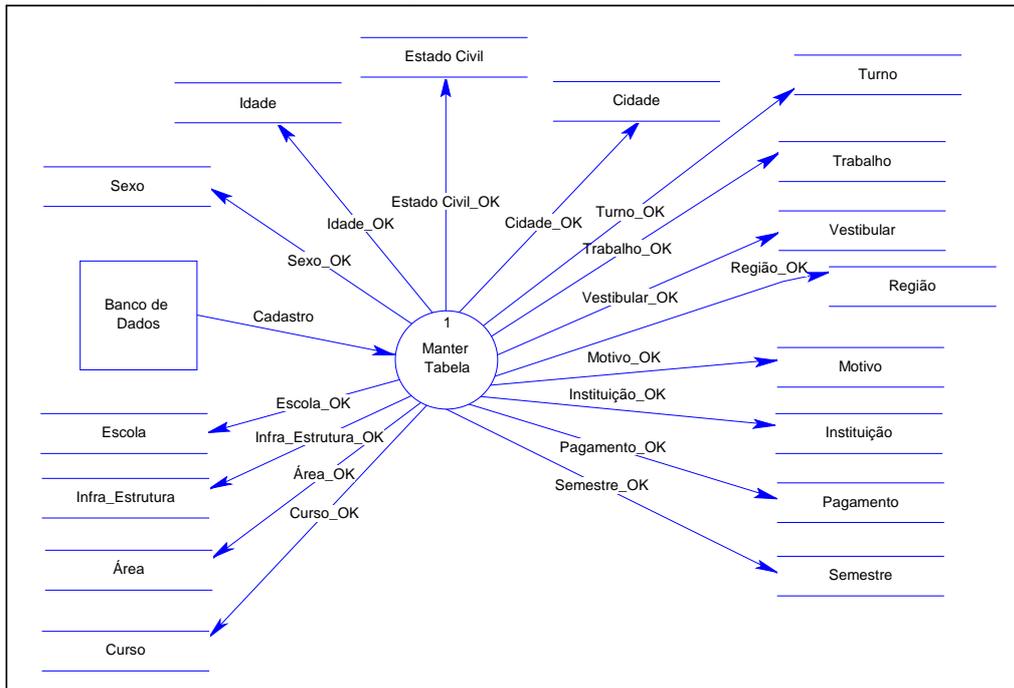
## 6.2.2 DIAGRAMA DE FLUXO DE DADOS

O Diagrama de Fluxo de Dados do Sistema de Informação Gerencial CIA & CIA está representado no nível 0 na Figura 11. Nas Figuras 12, 13 e 14 estão representados os níveis 1, 2 e 3 respectivamente.

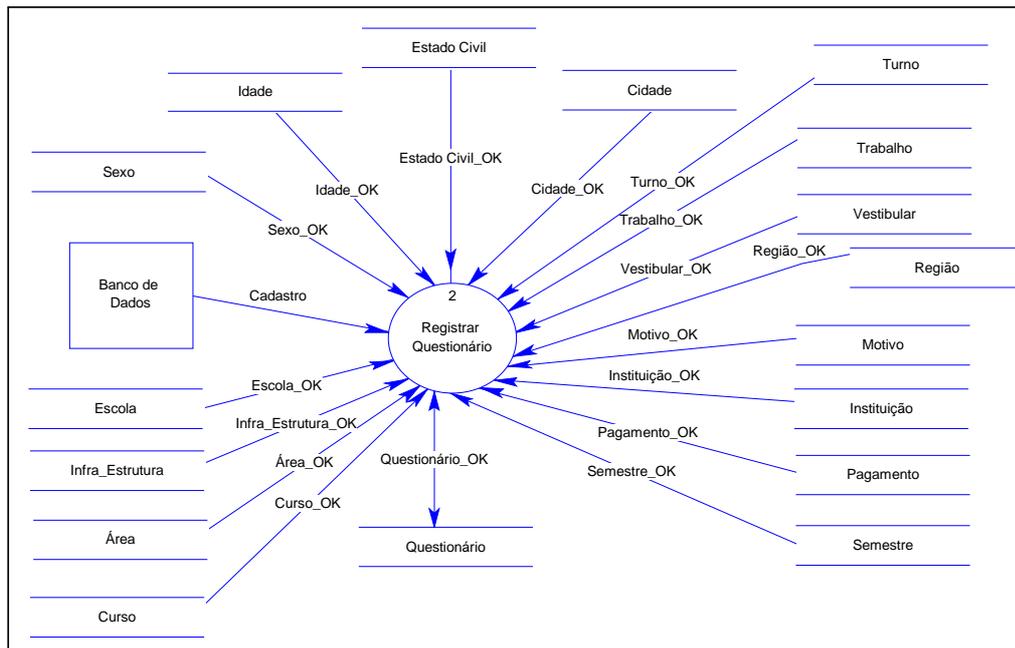
**Figura 11 - Diagrama de Fluxo de Dados (Nível 0)**



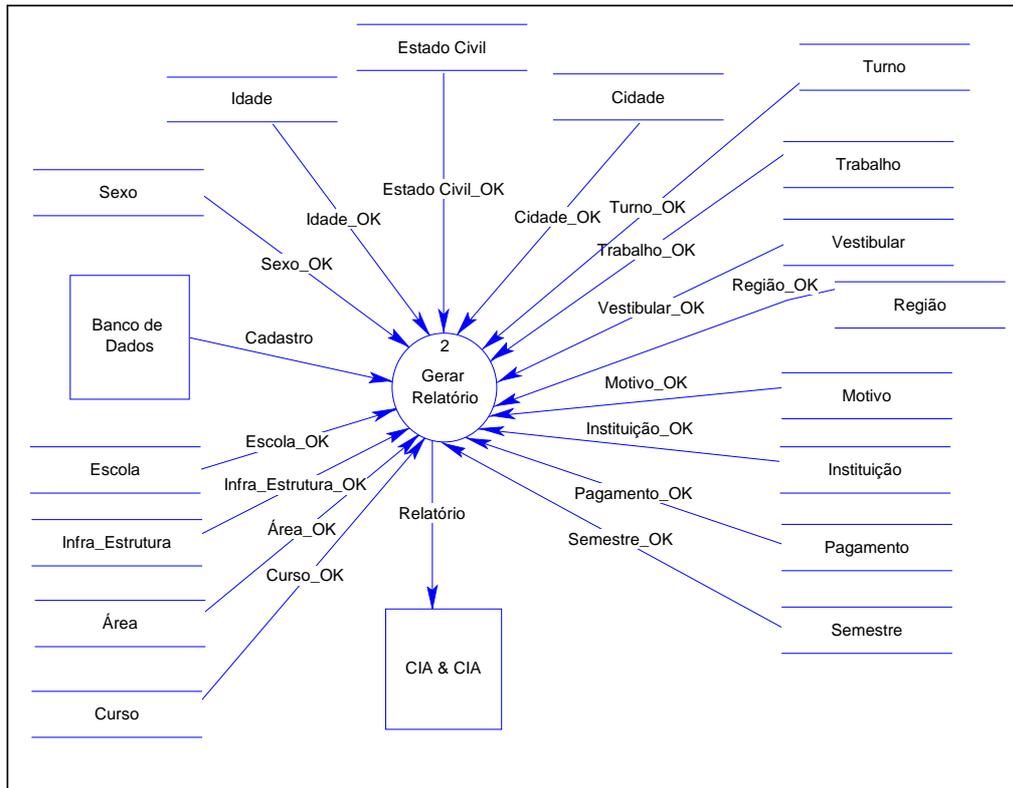
**Figura 12 - Diagrama de Fluxo de Dados (Nível 1)**



**Figura 13 - Diagrama de Fluxo de Dados (Nível 2)**



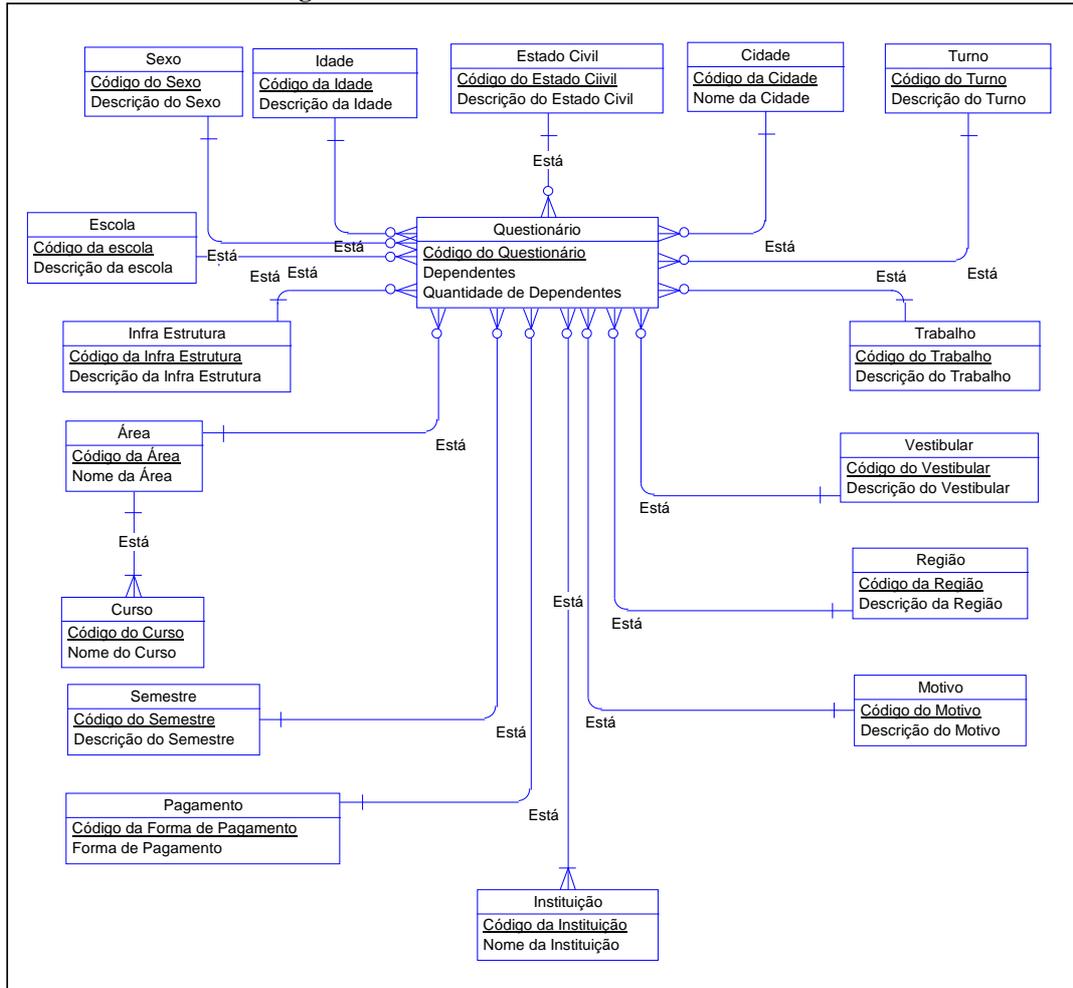
**Figura 14 - Diagrama de Fluxo de Dados (Nível 3)**



### 6.2.3 MODELO ENTIDADE-RELACIONAMENTO

O modelo entidade-relacionamento do Sistema de Informação Gerencial CIA & CIA é representado na Figura 15, na qual estão os atributos de cada uma das questões da Avaliação Qualitativa de Demanda por Cursos, questionário aplicado pela CIA & CIA, para coleta de dados estratégicos para o SIG.

**Figura 15 - Modelo Entidade-Relacionamento**



## 6.2.4 DICIONÁRIO DE DADOS

Em seguida é apresentado o relatório com o Dicionário de Dados do Sistema de Informação Gerencial CIA & CIA (Tabela 2), gerado pelo Power Designer a partir do modelo entidade-relacionamento (Figura 15). Na tabela estão presentes o nome das entidades (Attribute List), o nome dos atributos para o modelo físico (Name) e lógico (Code), o tipo (Type), se é identificador (I) e se é obrigatório (M).

**Tabela 2 - Dicionário de Dados**

## Área

## Attribute List

Name	Code	Type	I	M
Código da Área	CD_AREA	A1	Yes	Yes
Nome da Área	NM_AREA	A40	No	Yes

## Cidade

## Attribute List

Name	Code	Type	I	M
Código da Cidade	CD_CIDADE	I	Yes	Yes
Nome da Cidade	NM_CIDADE	A40	No	Yes

## Curso

## Attribute List

Name	Code	Type	I	M
Código do Curso	CD_CURSO	I	Yes	Yes
Nome do Curso	NM_CURSO	A40	No	Yes

## Escola

## Attribute List

Name	Code	Type	I	M
Código da escola	CD_ESCOLA	A1	Yes	Yes
Descrição da escola	DS_ESCOLA	A10	No	Yes

## Estado Civil

## Attribute List

Name	Code	Type	I	M
Código do Estado Civil	CD_CIVIL	A1	Yes	Yes
Descrição do Estado Civil	DS_CIVIL	A25	No	Yes

## Idade

## Attribute List

Name	Code	Type	I	M
Código da Idade	CD_IDADE	A1	Yes	Yes
Descrição da Idade	DS_IDADE	A20	No	Yes

## Infra Estrutura

## Attribute List

Name	Code	Type	I	M
Código da Infra Estrutura	CD_INFRA_ESTR	A2	Yes	Yes
Descrição da Infra Estrutura	DS_INFRA_ESTR	LA200	No	Yes

## Instituição

## Attribute List

Name	Code	Type	I	M
Código da Instituição	CD_INST	A1	Yes	Yes
Nome da Instituição	NM_INST	A15	No	No

## Motivo

## Attribute List

Name	Code	Type	I	M
Código do Motivo	CD_MOTIVO	A1	Yes	Yes
Descrição do Motivo	DS_MOTIVO	A40	No	Yes

## Pagamento

## Attribute List

Name	Code	Type	I	M
Código da Forma de Pagamento	CD_PAGTO	A1	Yes	Yes
Forma de Pagamento	DS_PAGTO	A30	No	Yes

## Questionário

## Attribute List

Name	Code	Type	I	M
Código do Questionário	CD_QUESTIONARIO	I	Yes	Yes
Dependentes	DS_DEPENDENTES	BL	No	Yes
Quantidade de Dependentes	QT_DEPENDENTES	I	No	Yes

## Região

## Attribute List

Name	Code	Type	I	M
Código da Região	CD_REGIAO	A1	Yes	Yes
Descrição da Região	DS_REGIAO	A10	No	Yes

## Semestre

## Attribute List

Name	Code	Type	I	M
Código do Semestre	CD_SEMESTRE	A1	Yes	Yes
Descrição do Semestre	DS_SEMESTRE	A10	No	Yes

## Sexo

## Attribute List

Name	Code	Type	I	M
Código do Sexo	CD_SEXO	A1	Yes	Yes
Descrição do Sexo	DS_SEXO	A10	No	Yes

## Trabalho

## Attribute List

Name	Code	Type	I	M
Código do Trabalho	CD_TRABALHO	A1	Yes	Yes
Descrição do Trabalho	DS_TRABALHO	A30	No	Yes

## Turno

## Attribute List

Name	Code	Type	I	M
Código do Turno	CD_TURNO	A1	Yes	Yes
Descrição do Turno	DS_TURNO	A10	No	Yes

## Vestibular

## Attribute List

Name	Code	Type	I	M
Código do Vestibular	CD_VESTIBULAR	A1	Yes	Yes
Descrição do Vestibular	DS_VESTIBULAR	A10	No	Yes

## 6.3 APRESENTAÇÃO DAS TELAS

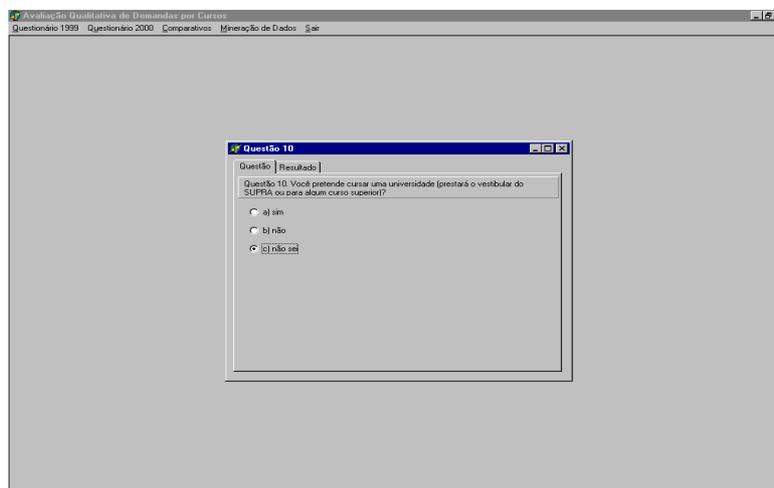
A seguir serão mostradas as telas do protótipo. Na Figura 16 é apresentada a tela de abertura do protótipo.

**Figura 16 - Tela de Abertura do Protótipo**



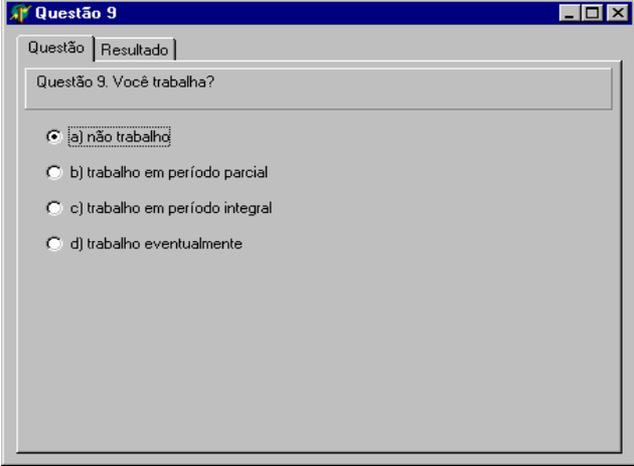
A Figura 17 mostra a tela principal na qual constam os módulos das questões da Avaliação Qualitativa de Demanda por Cursos, realizada nos anos de 1999 e 2000. Aparece, também, o módulo de Comparativos, no qual se tem a comparação dos resultados entre 1999 e 2000 para as questões mais relevantes para o Administrador. No módulo Mineração de Dados aparecem as questões nas quais foi aplicada a técnica de árvores de decisão.

**Figura 17 - Tela Principal**



A Figura 18 mostra a tela em que aparece a questão 9 dos módulos Questionário 1999 e Questionário 2000. Nesta opção o Administrador pode verificar o conteúdo das questões 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18 e 19, bem como os resultados em percentuais de cada uma das respostas escolhidas pelos participantes do FIC-FURB, tanto para o ano de 1999, quanto para 2000 (Figura 19).

**Figura 18 - Questão 9 dos Módulos Questionários**

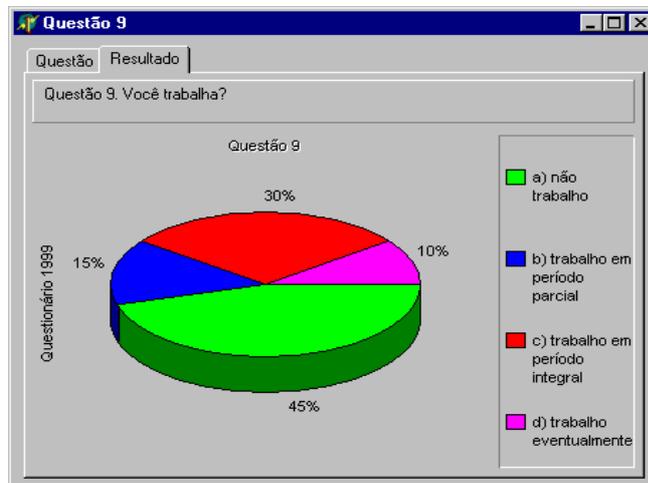


Questão 9

Questão 9. Você trabalha?

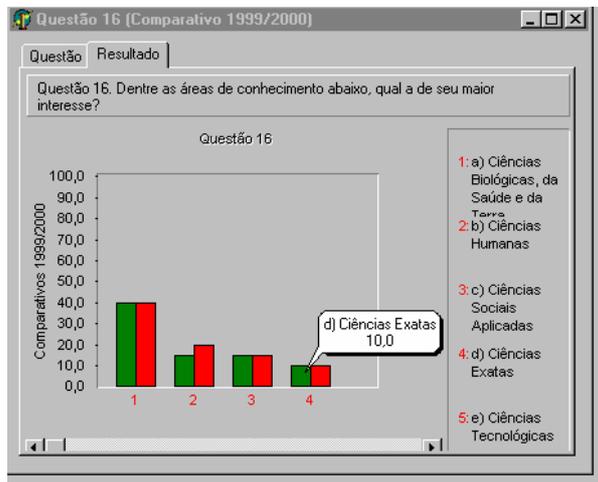
- a) não trabalho
- b) trabalho em período parcial
- c) trabalho em período integral
- d) trabalho eventualmente

**Figura 19 - Resultado da Questão**



Na Figura 20 é apresentada a tela da questão 16 do módulo Comparativos, que mostra a comparação entre os percentuais de cada resposta para os anos de 1999 e 2000. Com as informações dos resultados, o Administrador pode verificar se houve aumento ou decréscimo na procura por uma área de conhecimento e então tomar decisões estratégicas a fim de aumentar a procura pelos cursos da FURB.

**Figura 20 - Questão 16 do Módulo Comparativos**



Na Figura 21 é apresentado o algoritmo da árvore de decisão, conforme item 3.4.6. utilizado para calcular a Entropia e o Gain das questões 13, 17 e 20, conforme item 3.4.6.

**Figura 21 - Algoritmo da Árvore de Decisão**

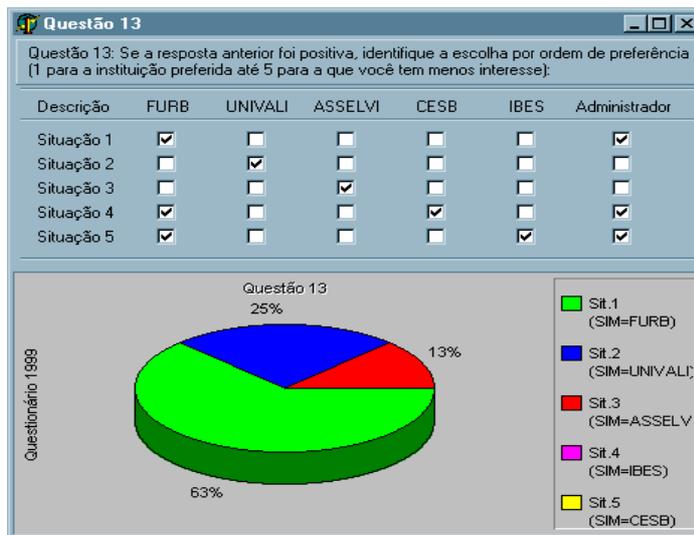
```

function TfrmTcc.ArvoreDecisao(S, xSim, xNao: Integer): Double;
function Entropia(S, xSim, xNao: Integer): Double;
begin
  result := -((xSim / S) * log2(xSim/S)) -
    ((xNao / S) * log2(xNao/S));
end;
function Gain(S, xSim, xNao: Integer): Double;
begin
  result := Entropia(S, xSim, xNao) - ((xSim+xNao) / S)* Entropia(S, xSim, xNao);
end;
begin
  Result := Gain(S, xSim, xNao);
end;
function TfrmTcc.ArvoreDecisao2(S, xSituacao: Integer): Double;
function Entropia2(S, xSituacao): Double;
begin
  result := -((xSituacao / S) * log2(xSituacao/S));
end;
function Gain2(S, xSituacao: Integer): Double;
begin
  result := Entropia2(S, xSituacao) - ((xSituacao) / S)* Entropia2(S, xSituacao);
end;
begin
  Result := Gain2(S, xSituacao);
end;
end.

```

A Figura 22 mostra a questão 13 do módulo de Mineração de Dados, na qual foi aplicada a técnica de árvore de decisão com a fórmula da entropia descrita na Figura 21, para os anos de 1999 e 2000, onde aparece a preferência dos alunos pelas instituições de ensino superior da região. Para esta questão, o Administrador atribuiu pesos para as situações 1 a 5 descritas pela Tabela 3 (Anexo 2), em que a situação ideal é verificada na situação 1. Nas situações 2 e 3, verifica-se um alto risco para a FURB, pois a UNIVALI e a ASSELVI são grandes concorrentes. Nas situações 5 e 6, existe um risco médio, pois atualmente o CESB e o IBES não representam grandes riscos para a FURB.

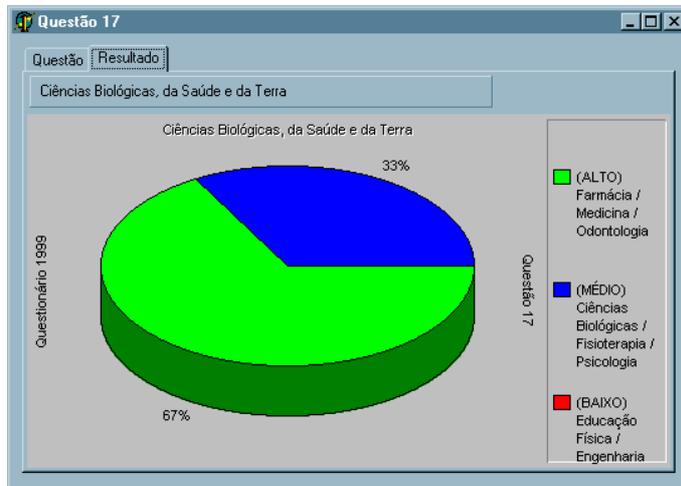
**Figura 22 - Preferência dos Alunos pelas Instituições de Ensino Superior (Questão 13)**



Nas Tabelas 4, 5, 6, 7 e 8 (Anexo 2) são descritos os pesos atribuídos pelo Administrador para a questão 17a, 17b, 17c, 17d e 17e, respectivamente, em que constam os Níveis Alto, Médio e Baixo, que representam a preferência pelos cursos de graduação das diferentes áreas de conhecimento de maior relevância para o Administrador. O nível alto representa a situação ideal, o nível médio a situação mediana, e o nível baixo a situação de menor importância.

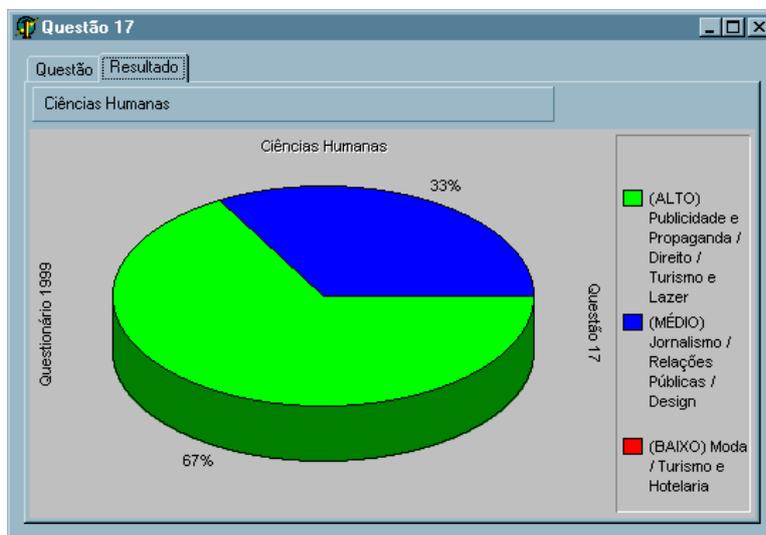
A Figura 23 mostra a preferência dos alunos na área de Ciências Biológicas, da Saúde e da Terra (Questão 17a).

**Figura 23 - Preferência dos Alunos na Área de Ciências Biológicas, da Saúde e da Terra (Questão 17a)**



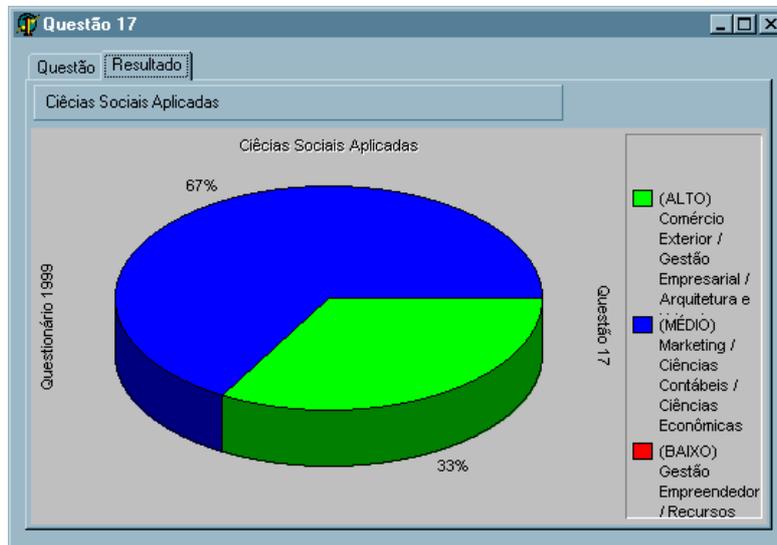
A Figura 24 mostra a preferência dos alunos na área de Ciências Humanas (Questão 17b).

**Figura 24 - Preferência dos Alunos na Área de Ciências Humanas (Questão 17b)**



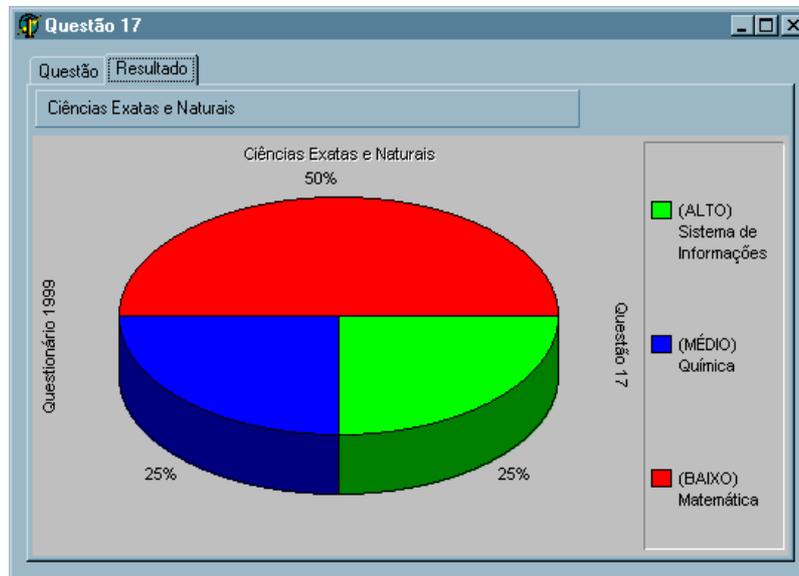
A Figura 25 mostra a preferência dos alunos na área de Ciências Sociais Aplicadas (Questão 17c).

**Figura 25 - Preferência dos Alunos na Área de Ciências Sociais Aplicadas (Questão 17c)**



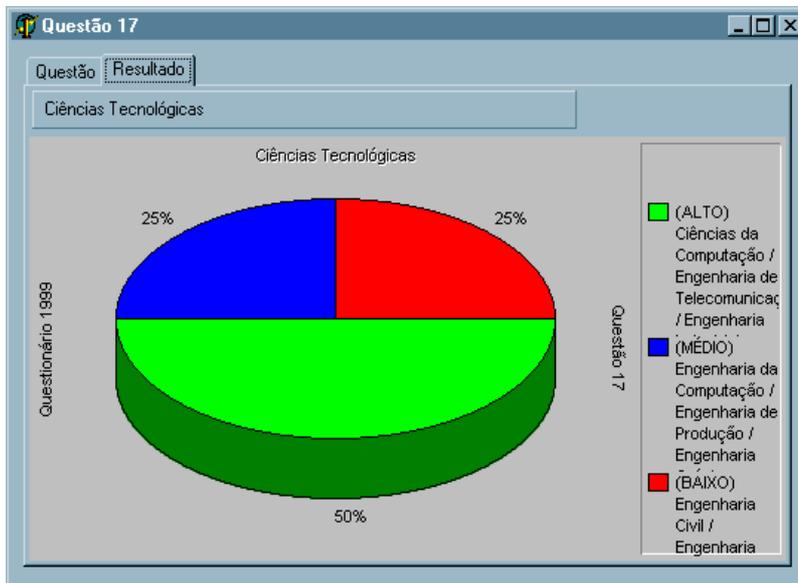
A Figura 26 mostra a preferência dos alunos na área de Ciências Exatas (Questão 17d).

**Figura 26 - Preferência dos Alunos na Área de Ciências Exatas (Questão 17d)**



A Figura 27 mostra a preferência dos alunos na área de Ciências Tecnológicas (Questão 17e).

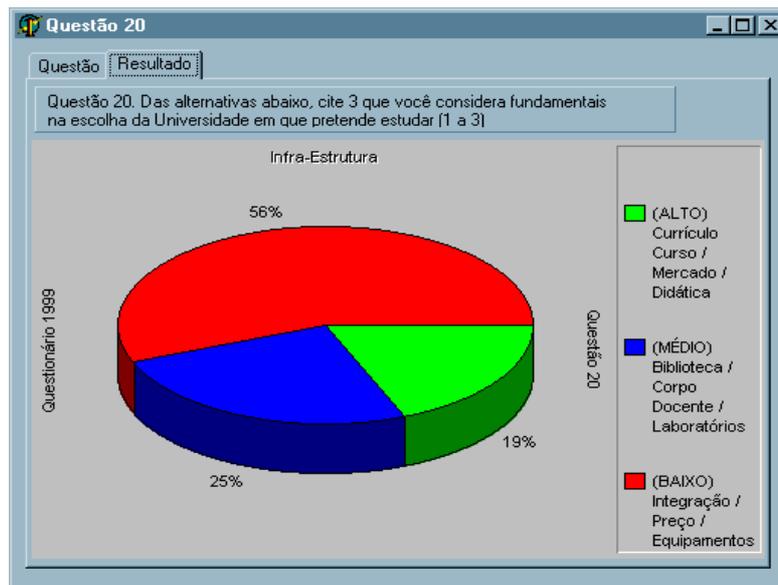
**Figura 27 - Preferência dos Alunos na Área de Ciências Tecnológicas (Questão 17e)**



A Tabela 9 (Anexo 2) também mostra os pesos atribuídos pelo Administrador para a questão 20, em que constam os Níveis Alto, Médio e Baixo, que representam a preferência por alguns itens que compõem a Infra-Estrutura de um curso com maior relevância para o Administrador. O nível alto representa a situação ideal, o nível médio a situação mediana, e o nível baixo a situação de menor importância.

A Figura 28 mostra a preferência dos alunos em itens que compõem a Infra-Estrutura de um Curso (Questão 20).

**Figura 28 - Preferência dos Alunos na Infra-Estrutura de um Curso (Questão 20)**



Por meio do módulo Mineração de Dados, o Administrador pode tomar decisões estratégicas pela comparação entre uma prioridade pré-definida e a realidade verificada nas respostas do questionário de Avaliação Qualitativa de Demanda por Cursos, no qual os dados-alvo foram selecionados, pré-processados e transformados, de forma a contribuir para sua interpretação, gerando o conhecimento almejado e presente na última etapa do KDD.

## 7 CONCLUSÕES E SUGESTÕES

Este capítulo apresenta as conclusões, limitações e sugestões referentes ao trabalho desenvolvido.

### 7.1 CONCLUSÃO

Partindo da necessidade de se extrair conhecimento por meio da interpretação de dados, foi estudada a tecnologia de *Data Mining*. Foram estudadas suas funções, suas técnicas, e as etapas que levem à descoberta do conhecimento que é o objetivo principal do *Data Mining*.

Neste trabalho foi enfatizado o uso do *Data Mining* com Árvores de Decisão, empregado em um Sistema de Informação Gerencial para modelos de classificação e segmentação de dados. Tendo isso como base, foi possível verificar que a utilização do *Data Mining*, juntamente com as etapas de KDD, mostrou-se bastante eficiente.

Os resultados obtidos com o *Data Mining* melhoram os negócios em uma organização já próspera. Ele não necessariamente proporciona mudanças revolucionárias, mas é uma poderosa ferramenta de descoberta para organizações como a FURB, que deseja conhecer melhor os seus clientes/futuros alunos e possui uma visão a longo prazo. Essa tecnologia está consolidando a informação como um dos recursos naturais mais críticos das organizações, senão o mais importante.

Foram realizados testes com o modelo de dados construído para a execução do processo de *Data Mining* nos quais o protótipo mostrou-se eficiente para a definição de modelos de classificação e segmentação de dados.

Durante a construção do modelo, foram utilizadas algumas etapas/fases da metodologia de análise estruturada, as quais auxiliaram em muito no desenvolvimento do

projeto. As ferramentas OLAP ajudou muito pela facilidade de aprendizado que proporciona e pelo fácil acesso aos dados, de forma *On-line*.

Considera-se que o objetivo principal do trabalho - o desenvolvimento de um SIG para efetuar classificações e segmentações de dados utilizando *Data Mining* - foi atingido.

## 7.2 LIMITAÇÕES

O protótipo construído apresenta as seguintes limitações:

- a) a fonte de dados que o protótipo utiliza para processamento é fixa, desta forma não permitindo ao usuário mudar a fonte de dados ou alterar o conjunto de atributos a serem processados;
- b) os atributos envolvidos no processo de classificação possuem domínio fixo.

## 7.3 SUGESTÕES

Sugere-se o estudo do *Data Mining* aplicando outras tarefas e técnicas para a tomada de decisões, como o uso de outras técnicas.

Em relação aos módulos, poderiam ser criados outros, para contemplar novas características relevantes para o Administrador no futuro. Também poderiam ser gerados módulos automáticos para os questionários dos anos seguintes.

Sugere-se, também, a implementação de outros protótipos para os questionários utilizados na matrícula dos calouros e para o questionário sócio-cultural aplicado na primeira semana de aula dos alunos ingressantes.

Mais um item importante é o desenvolvimento de SIG para outros setores da Pró-Reitoria de Ensino e das outras Pró-Reitorias, que apresentam grande volume de dados a serem processados.

## ANEXO 1 – QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA DE DEMANDA POR CURSOS

1. Sexo: a ( ) masculino b ( ) feminino
2. Idade: a ( ) até 18 anos b ( ) de 18 a 24 anos c ( ) mais de 24 anos
3. Estado Civil: a ( ) solteiro(a) b ( ) casado(a) c ( ) viúvo(a)  
d ( ) separado(a)/divorciado(a)
4. Tem dependentes: a ( ) sim b ( ) não
5. Quantos: \_\_\_\_\_
6. Cidade em que reside: \_\_\_\_\_
7. Está cursando o ensino médio em escola: a ( ) pública b ( ) privada
8. Em qual turno você estuda? a ( ) matutino b ( ) vespertino c ( ) noturno
9. Você trabalha? a ( ) não trabalho b ( ) trabalho em período parcial  
c ( ) trabalho em período integral d ( ) trabalho eventualmente
10. Você pretende cursar uma universidade (prestará o vestibular do SUPRA ou para algum curso superior)? a ( ) sim b ( ) não c ( ) não sei
11. Você pretende fazer o curso superior em Blumenau ou região?  
a ( ) sim b ( ) não c ( ) não sei
12. Se a resposta anterior foi negativa, diga o motivo:  
a ( ) localização b ( ) problemas financeiros c ( ) baixa qualificação  
d ( ) falta de interesse e ( ) o curso desejado só existe em outra região
13. Se a resposta anterior foi positiva, identifique a escolha por ordem de preferência  
(1 para a instituição preferida; até 5 para a que você tem menos interesse):  
a ( ) FURB b ( ) UNIVALI c ( ) ASSELVI d ( ) CESB  
e ( ) IBES f ( ) UNIVILLE g ( ) não sei h ( ) outra
14. Quando você pretende prestar vestibular?  
a ( ) 1º sem/2000 b ( ) 2º sem/2000 c ( ) 1º sem/2001 d ( ) 2º sem/2001  
e ( ) 1º sem/2002 f ( ) 2º sem/2002 g ( ) não sei
15. Como pretende manter-se durante o curso universitário?  
a ( ) trabalhando b ( ) com bolsa de estudos c ( ) crédito educativo  
d ( ) com recursos dos pais e ( ) ainda não sei

16. Dentre as áreas de conhecimento abaixo, qual a de seu maior interesse?

- a ( ) Ciências Biológicas, da Saúde e da Terra      b ( ) Ciências Humanas  
 c ( ) Ciências Sociais Aplicadas                      d ( ) Ciências Exatas  
 e ( ) Ciências Tecnológicas

17. Da área que você tem maior interesse, identifique três cursos, dando notas de 1 a 3, sendo 1 para o curso com o qual você mais se identifica:

a) área de Ciências Biológicas, da Saúde e da Terra:

- |                         |  |                          |
|-------------------------|--|--------------------------|
| ( ) Ciências Biológicas | ( ) Ciências Biológicas –<br>Biotecnologia | ( ) Educação Física      |
| ( ) Enfermagem          | ( ) Engenharia Ambiental                   | ( ) Engenharia Florestal |
| ( ) Farmácia            | ( ) Fisioterapia                           | ( ) Fonoaudiologia       |
| ( ) Medicina            | ( ) Nutrição                               | ( ) Oceanografia         |
| ( ) Odontologia         | ( ) Psicologia                             |                          |

b) área de Ciências Humanas:

- |   |  |   |
|---|--|---|
| ( ) Ciência Política                          | ( ) Ciências Sociais                   | ( ) Ciências Sociais – Ênfase em<br>Desenvolvimento Sustentável |
| ( ) Ciências da Religião                      | ( ) Comunicação Social –<br>Jornalismo | ( ) Comunicação Social –<br>Publicidade e Propaganda            |
| ( ) Comunicação Social –<br>Relações Públicas | ( ) Design                             | ( ) Direito   |
| ( ) Educação Artística                        | ( ) Filosofia                          | ( ) Gastronomia   |
| ( ) Geografia                                 | ( ) Gestão do Lazer e Eventos          | ( ) História  |
| ( ) Letras                                    | ( ) Moda                               | ( ) Pedagogia   |
| ( ) Relações Internacionais                   | ( ) Secretariado Executivo Bilíngue    | ( ) Serviço Social  |
| ( ) Turismo e Hotelaria                       | ( ) Turismo e Lazer                    |   |

c) área de Ciências Sociais Aplicadas:

- |   |  |   |
|---|--|---|
| ( ) Administração –<br>Comércio Exterior    | ( ) Administração – Ênfase em<br>Gestão de Negócios Imobiliários | ( ) Administração – Finanças              |
| ( ) Administração – Gestão<br>Empreendedora | ( ) Administração – Gestão<br>Empresarial                        | ( ) Administração – Gestão de<br>Negócios |
| ( ) Administração –<br>Marketing            | ( ) Administração – Recursos<br>Humanos                          | ( ) Administração – Serviços              |
| ( ) Arquitetura e                           | ( ) Automação de Escritórios e<br>Secretariado                   | ( ) Ciências Contábeis                    |
| ( ) Ciências Econômicas                     | ( ) Desenho Industrial – Design                                  | ( ) Logística                             |

d) área de Ciências Exatas:

- |                |             |                            |
|----------------|-------------|----------------------------|
| ( ) Matemática | ( ) Química | ( ) Sistemas de Informação |
|----------------|-------------|----------------------------|

## e) área Tecnológica:

- |   |  |   |
|---|--|---|
| <input type="checkbox"/> Ciências da Computação                           | <input type="checkbox"/> Engenharia Civil                  | <input type="checkbox"/> Engenharia da Computação   |
| <input type="checkbox"/> Engenharia de Produção –<br>Tecnologias + Limpas | <input type="checkbox"/> Engenharia de<br>Telecomunicações | <input type="checkbox"/> Engenharia Industrial Elétrica                                       |
| <input type="checkbox"/> Engenharia Industrial -<br>Mecânica              | <input type="checkbox"/> Engenharia Química                | <input type="checkbox"/> Tecnólogo em Processos<br>Industriais – Modalidade<br>Eletromecânica |

18. Em qual turno você prefere cursar a universidade?

- a ( ) matutino      b ( ) vespertino      c ( ) noturno      d ( ) integral

19. Em qual(is) turno(s) você está impossibilitado de estudar?

- a ( ) matutino      b ( ) vespertino      c ( ) noturno      d ( ) integral

20. Das alternativas abaixo, dê notas de 1 a 3 às que você considera fundamentais na escolha da Universidade em que pretende estudar. A nota 1 deve ser atribuída à alternativa que você considere mais importante:

- a ( ) Biblioteca com qualidade e quantidade de livros adequados para o curso.
- b ( ) Corpo docente qualificado, predominantemente com mestrado e/ou doutorado.
- c ( ) Currículo do curso atualizado e coerente com as exigências do mercado.
- d ( ) Infra-estrutura de apoio (banco, restaurante, livraria, fotocopiadora).
- e ( ) Laboratórios capazes de atender às necessidades do curso.
- f ( ) Oportunidades de futura colocação no mercado.
- g ( ) Oportunidades de integração dos alunos, através de participação em atividades de pesquisa e/ou relacionadas com a comunidade.
- h ( ) Preço inferior ao praticado pela concorrência.
- i ( ) Procedimentos didáticos dos professores adequados às necessidades de formação dos alunos.
- j ( ) Quantidade adequada de alunos por sala de aula.
- l ( ) Salas de aula com mobiliário que proporciona conforto aos alunos  
( ar-condicionado, cadeira estofada, cortina).
- m ( ) Salas de aula equipadas com projetores multimídia, internet, retroprojetores e outros equipamentos desta natureza.

## ANEXO 2 – TABELAS DE PRIORIDADES DOS ATRIBUTOS

**Tabela 3 - Atributos da Questão 13**

Descrição	FURB	UNIVALI	ASSELVI	CESB	IBES	Administrador
Situação 1	Sim	Não	Não	Não	Não	Sim
Situação 2	Não	Sim	Não	Não	Não	Não
Situação 3	Não	Não	Sim	Não	Não	Não
Situação 4	Sim	Não	Não	Sim	Não	Sim
Situação 5	Sim	Não	Não	Não	Sim	Sim

**Tabela 4 - Atributos da Questão 17a**

Ciências Biológicas, da Saúde e da Terra	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	AD
Alto	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Não	Sim
Médio	Sim	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim	Não
Baixo	Não	Não	Sim	Não	Não	Sim	Não								

A1 – Ciências Biológicas

A2 – Ciências Biológicas – Biotecnologia

A3 – Educação Física

A4 – Enfermagem

A5 – Engenharia Ambiental

A6 – Engenharia Florestal

A7 – Farmácia

A8 – Fisioterapia

A9 – Fonoaudiologia

A10 – Medicina

A11 – Nutrição

A12 – Oceanografia

A13 – Odontologia

A14 – Psicologia

AD – Administrador

**Tabela 5 - Atributos da Questão 17b**

Ciências Humanas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	AD
Alto	Não	Não	Não	Não	Não	Sim	Não	Não	Sim	Não	Sim	Sim												
Médio	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não															
Baixo	Não	Sim	Não	Não	Não	Não	Sim	Não	Não															

A1 – Ciência Política

A2 – Ciências Sociais

A3 – Ciências Sociais – Des. Sust.

A4 – Ciências da Religião

A5 – Jornalismo

A6 – Publicidade e Propaganda

A7 – Relações Públicas

A8 – Design

A9 – Direito

A10 – Educação Artística

A11 – Filosofia

A12 – Gastronomia

A13 – Geografia

A14 – Gestão do Lazer e Eventos

A15 – História

A16 – Letras

A17 – Moda

A18 – Pedagogia

A19 – Relações Internacionais

A20 – Secretariado Executivo Bilíngüe

A21 – Serviço Social

A22 – Turismo e Hotelaria

A23 – Turismo e Lazer

AD - Administrador

**Tabela 6 - Atributos da Questão 17c**

Ciências Sociais Aplicadas	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	AD
Alto	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim
Médio	Não	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Não	Sim	Sim	Não	Não	Não
Baixo	Não	Não	Não	Sim	Não	Não	Não	Sim	Não	Não	Não	Não	Não	Sim	Não	Não

A1 - Comércio Exterior

A2 – Negócios Imobiliários

A3 – Finanças

A4 – Gestão Empreendedora

A5 – Gestão Empresarial

A6 – Gestão de Negócios

A7 – Marketing

A8 – Recursos Humanos

A9 – Serviços

A10 – Arquitetura e Urbanismo

A11 – Automação de Escritórios

A12 – Ciências Contábeis

A13 – Ciências Econômicas

A14 – Desenho Industrial

A15 - Logística

AD - Administrador

**Tabela 7 - Atributos da Questão 17d**

Ciências Exatas e Naturais	Matemática	Química	Sistemas de Informações	Administrador
Alto	Não	Não	Sim	Sim
Médio	Não	Sim	Não	Não
Baixo	Sim	Não	Não	Não

**Tabela 8 - Atributos da Questão 17e**

Ciências Tecnológicas	A1	A2	A3	A4	A5	A6	A7	A8	A9	Administrador
Alto	Sim	Não	Não	Não	Sim	Sim	Não	Não	Não	Sim
Médio	Não	Não	Sim	Sim	Não	Não	Não	Sim	Não	Não
Baixo	Não	Sim	Não	Não	Não	Não	Sim	Não	Sim	Não

A1 – Ciências da Computação

A2 – Engenharia Civil

A3 – Engenharia de Computação

A4 – Engenharia de Produção

A5 – Engenharia de Telecomunicações

A6 – Engenharia Industrial Elétrica

A7 – Engenharia Industrial Mecânica

A8 – Engenharia Química

A9 – Eletromecânica

**Tabela 9 - Atributos da Questão 20**

Infra-Estrutura	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	AD
Alto	Não	Não	Sim	Não	Não	Sim	Não	Não	Sim	Não	Não	Não	Sim
Médio	Sim	Sim	Não	Não	Sim	Não							
Baixo	Não	Não	Não	Não	Não	Não	Sim	Sim	Não	Não	Não	Sim	Não

A1 – Biblioteca

A2 – Corpo Docente

A3 – Currículo Curso

A4 – Infra-Estrutura Apoio

A5 – Laboratórios

A6 – Mercado

A7 – Integração

A8 – Preço

A9 – Didática

A10 – Qtde Alunos

A11 – Mobiliário

A12 – Equipamentos

AD – Administrador

## Referências Bibliográficas

- [ALT1992] ALTER, Steven. **Information systems: a management perspective**. USA: Addison-Wesley Publishing, 1992.
- [AVI1998] ÁVILA, Bráulio Coelho. **Data Mining**. VI Escola de Informática da SBC – Região Sul. Blumenau, 1998. p. 87–106.
- [BER1997] BERRY, Michael J.A.; LINOFF, Gordon. **Data Mining techniques**. USA: Wiley Computer Publishing, 1997.
- [BIN1994] BINDER, Fábio Vinício. **Sistema de apoio à decisão**. São Paulo: Érica, 1994.
- [BIS1999] BISPO, Carlos Alberto.; CAZARINI, Edson Walmir. Transformando dados em informações via data mining. **Revista Developers Magazine**, Rio de Janeiro: ano 3, n. 29, p. 36-38, jan.1999.
- [CAN2000] CANTU, Marco. **Dominando o Delphi 5: a Bíblia**. São Paulo: Makron Books, 2000.
- [COM1999] COMPOLT, Geandro Luis. **Sistemas de Informação Executiva Baseado em um Data Mining Utilizando a Técnica de Árvores de Decisão**. Blumenau: FURB, 1999. Trabalho de Conclusão de Curso.
- [CRU1998] CRUZ, Tadeu. **Sistemas de Informações Gerenciais: tecnologia da informação e a empresa do século XXI**. São Paulo: Atlas, 1998.
- [DAL2000] DALFOVO, Oscar; AMORIN, Sammy Newton. **Quem tem informação é mais competitivo**. Blumenau: Acadêmica, 2000.
- [FAY1996] FAYYAD, Usama M...[et all]. **Advances in knowledge discovery and Data Mining**. Menlo Park: AAAI: MIT, 1996.
- [FIG1998] FIGUEIRA, Rafael Medeiros Andrade. **Miner: um software de inferência de dependências funcionais**. Rio de Janeiro, 1998. Trabalho de Conclusão de Curso – Instituto de Matemática, Universidade Federal do Rio de Janeiro.

- [GRO1997] GROTH, Robert. **Data mining**: a hands-on approach for business professionals. New Jersey: Prentice Hall, 1997.
- [HAR1988] HARMON, Paul; King, David. **Sistemas Especialistas**. Rio de Janeiro: Campus, 1988.
- [JOA1993] JOAO, Belmiro N. **Metodologias de Desenvolvimento de Sistemas**. São Paulo: Érica, 1993.
- [KRE1999] KREMER, Ricardo. **Sistema de Apoio à Decisão para Previsões Genéricas Utilizando Data Mining**. Blumenau: FURB, 1999. Trabalho de Conclusão de Curso.
- [MEL1990] MELENDEZ, Rubem Filho. **Prototipação de sistemas de informações**: fundamentos, técnicas e metodologias. São Paulo: Livros Técnicos e Científicos, 1990.
- [OLI1992] OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**. São Paulo: Atlas, 1992.
- [OLI1996] OLIVEIRA, Djalma. **Sistemas de informações gerenciais**: estratégicas, táticas, operacionais. São Paulo: Atlas, 1996.
- [PRA1994] PRATES, Maurício. Conceituação de sistemas de informação do ponto de vista do gerenciamento. **Revista do Instituto de Informática**, São Paulo, PUC-CAMP, mar/set 1994.
- [STA1998] STAIR, Ralph M. **Princípios de sistemas de informação**: uma abordagem gerencial. Rio de Janeiro: LTC, 1998.
- [VIC1994] VICO MANÃS, Antonio. **Administração da informática**. São Paulo: Érica, 1994.
- [WES1998] WESTPHAL, Christopher; BLAXTON, Teresa. **Data mining solutions**. Canadá : John Wiley & Sons Inc, 1998.

[YOU1990] YOURDON, Edward. **Análise Estruturada Moderna**. Rio de Janeiro: Campus, 1990.