

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIAS DA COMPUTAÇÃO

**PROTÓTIPO DE SISTEMA DE INFORMAÇÃO APLICADO A
CIA & CIA. (CENTRAL DE INFORMAÇÃO AOS ALUNOS)
DA PRÓ-REITORIA DE ENSINO DA FURB UTILIZANDO
DATA MINING**

Acadêmica: Biancca Nardelli

Orientador: Oscar Dalfovo

2000/2-11

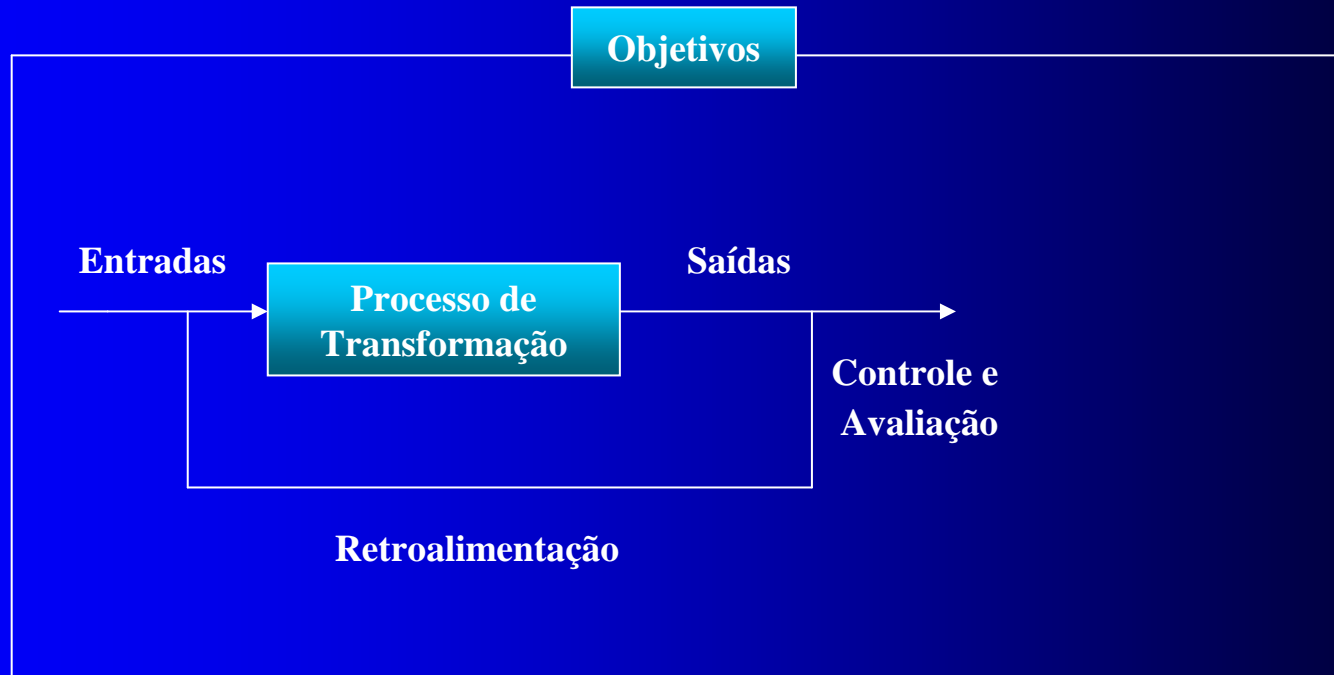
APRESENTAÇÃO

- **Introdução**
- **Sistemas de Informação**
 - **Sistemas de Informação Gerencial (SIG)**
 - **Fases do SIG**
- **Data Mining**
 - **Prospecção de Conhecimento - KDD**
 - **Árvores de Decisão**
- **Pró-Reitoria de Ensino**
 - **CIA & CIA.**
- **Tecnologias e Ferramentas Utilizadas**
 - **Análise Estruturada, Ferramenta CASE, Banco de Dados, Ambiente Visual Delphi, OLAP**
- **Desenvolvimento do Protótipo**
 - **Metodologia SIG**
 - **Especificação do Protótipo (Diagrama de Contexto, Diagrama de Fluxo de Dados, Modelo Entidade-Relacionamento)**
 - **Apresentação das Telas**
- **Conclusões e Sugestões**
 - **Limitações, Sugestões**
- **Bibliografia**

INTRODUÇÃO

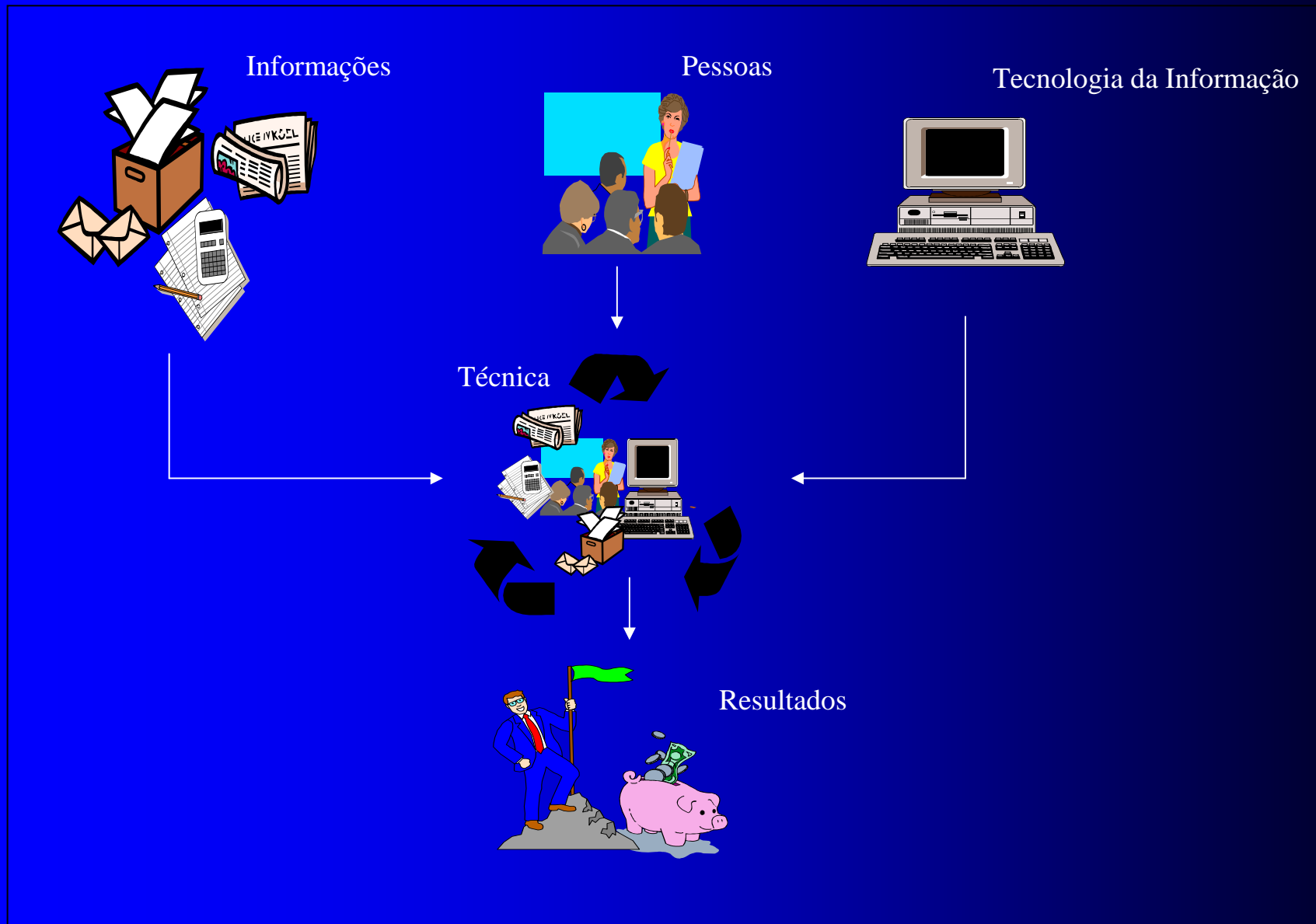
- Necessidade de divulgação dos cursos da Universidade Regional de Blumenau-FURB.
- Necessidade de gerenciamento dos dados, transformando em informações úteis.
- De acordo com Dalfovo ([DAL2000]), um Sistema de Informação facilita o processo decisório com a obtenção de dados estrategicamente escolhidos e de conteúdos relevantes para qualquer nível e tamanho de organização.

SISTEMAS DE INFORMAÇÃO



Fonte: [OLI1996]

SISTEMAS DE INFORMAÇÃO



Fonte: adaptado de [PRA1994]

SISTEMA DE INFORMAÇÃO GERENCIAL - SIG

Para OLIVEIRA [OLI1996], Sistema de Informação Gerencial 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.

Fases do SIG

- Fase I - Conceituação do SIG
- Fase II - Levantamento e Análise do SIG
- Fase III - Estruturação do SIG
- Fase IV - Implantação e Avaliação do SIG

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.

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.

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).

Fase IV - Implantação e Avaliação do SIG

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.

Data Mining

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.

Prospecção de Conhecimento KDD

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.



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.

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.

Fonte: adaptado de [BIS1999]

Pró-Reitoria de Ensino

São atribuições da Pró-Reitoria de Ensino de Graduação:

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

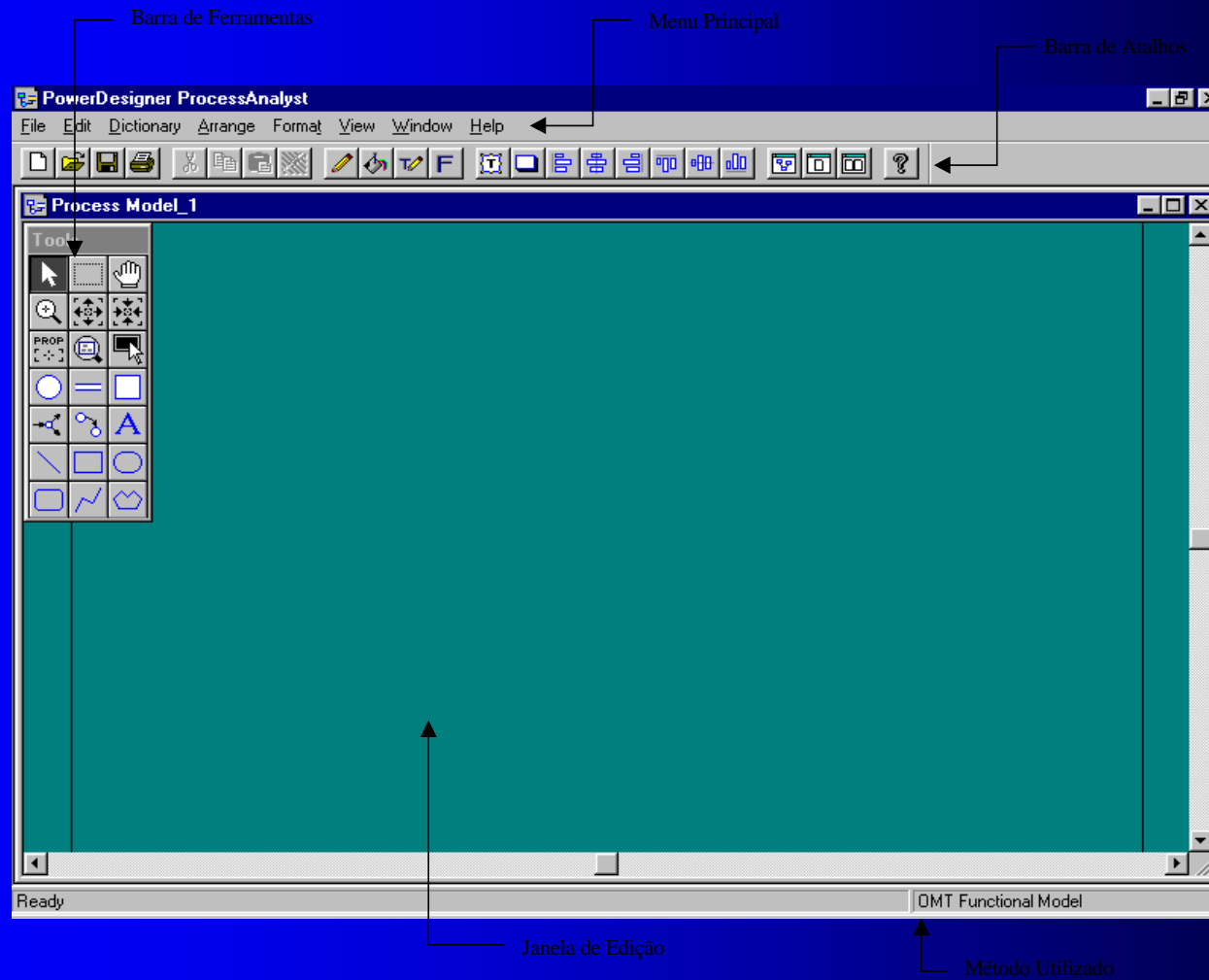
CIA & CIA. (Central de Informação aos Alunos)

- Divulgação de Cursos da Universidade;
- FIC - FURB;
- Eventos e Feiras;
- Parcerias;
- Matrículas;

Tecnologias e Ferramentas Utilizadas - Análise Estruturada

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 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.

FERRAMENTAS UTILIZADAS POWER DESIGNER



Banco de Dados

O banco de dados possibilita a realização de diversas operações, como:

- a adição de novos (vazios) arquivos ao banco de dados;
- a inserção de novos dados aos arquivos existentes;
- a recuperação de dados dos arquivos existentes;
- a atualização de dados nos arquivos existentes;
- a eliminação de dados nos arquivos existentes; e
- a renovação permanente de arquivos existentes (vazios ou outros) do banco de dados.

Neste trabalho foi utilizado o Banco de Dados Paradox, do Ambiente Visual Delphi.

FERRAMENTAS USADAS DELPHI

The image shows the Delphi 3 IDE interface with several components labeled:

- Speed Barr**: Points to the toolbar at the top of the main window.
- Menu Principal**: Points to the menu bar at the top of the main window.
- Paleta de Componentes**: Points to the component palette at the top of the main window.
- Formulário**: Points to the design-time form window (Form1) in the center.
- Inspector de Objetos**: Points to the Object Inspector window on the left, which displays the properties of the selected form.
- Janela de Edição**: Points to the code editor window (Unit1.pas) at the bottom.
- Janela de Complemento de Código**: Points to the code completion popup window in the code editor.

The Object Inspector window shows the following properties for Form1: TForm1

Property	Value
ActiveControl	
AutoScroll	True
+BorderIcons	[biSystemMen
BorderStyle	bsSizeable
Caption	Form1
ClientHeight	348
ClientWidth	536
Color	clBtnFace
Ctl3D	True
Cursor	crDefault
Enabled	True
+Font	(TFont)
FormStyle	fsNormal
Height	375
HelpContext	0
HelpFile	
Hint	
HorzScrollBar	(TControlScrol
Icon	(None)

The code editor window (Unit1.pas) shows the following code:

```
Unit1  
  
var  
    Form1: TForm1;  
  
implementation  
  
    {$R *.DFM}  
  
    procedure TForm1.Edit1Change(Sender: TObject);  
    begin  
        edit1.  
    end;  
end.
```

The code completion popup window shows the following properties for TForm1:

property	AutoSelect : Boolean
property	AutoSize : Boolean
property	BorderStyle : TBorderStyle
property	CharCase : TEditCharCase
property	Color : TColor
property	Ctl3D : Boolean

ON-LINE ANALYTICAL PROCESSING **OLAP**

Permite que analistas, gerentes e executivos obtenham, de maneira rápida, consistente e interativa, acesso a uma variedade de visualizações possíveis de informação que foi transformada de dados puros para refletir a dimensão real do empreendimento do ponto de vista do usuário, ou seja, um conjunto de funcionalidades que tenta facilitar a análise multidimensional.

Metodologia SIG

- Fase I: Coleta dos dados pela CIA & CIA no FIC-FURB e importação para o Paradox;
- Fase II: Definição, pelo Administrador, das questões de maior relevância;
- Fases III e IV: Escolha das técnicas a serem aplicadas e especificação das prioridades dos atributos.

Atributos da Questão 13

Descrição	FURB	UNIVALI	ASSELM	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

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	Não	Não	Não	Não	Não	Não	Não	Não

A1 – Ciências Biológicas

A2 – Ciências Biológicas – Biotecnologia

A3 – Educação Física

A4 – Enfermagem

A5 – Engenharia Ambiental

A6 – Engenharia Florestal

A7 – Farmácia

A8 – Fisioterapia

A9 – Foncaudiologia

A10 – Medicina

A11 – Nutrição

A12 – Oceanografia

A13 – Odontologia

A14 – Psicologia

AD – Administrador

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	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Sim	Sim
Médio	Não	Não	Não	Não	Sim	Não	Sim	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
Baixo	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não	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

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

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

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

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	Não	Não	Não	Não	Não	Não	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

ESPECIFICAÇÃO DO PROTÓTIPO

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

Diagrama de Contexto do Sistema de Informação Gerencial

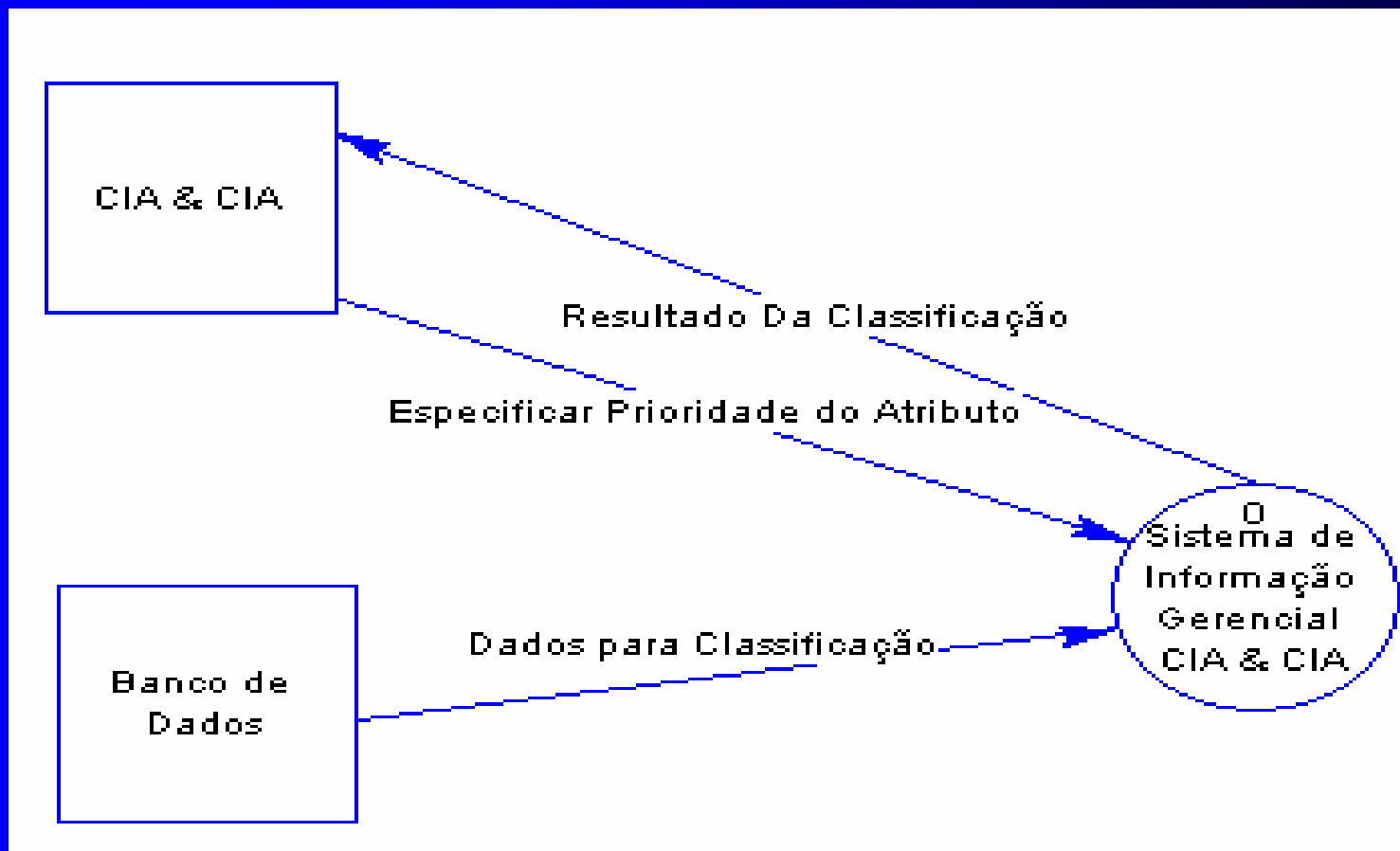


Diagrama de Fluxo de Dados (Nível 0)

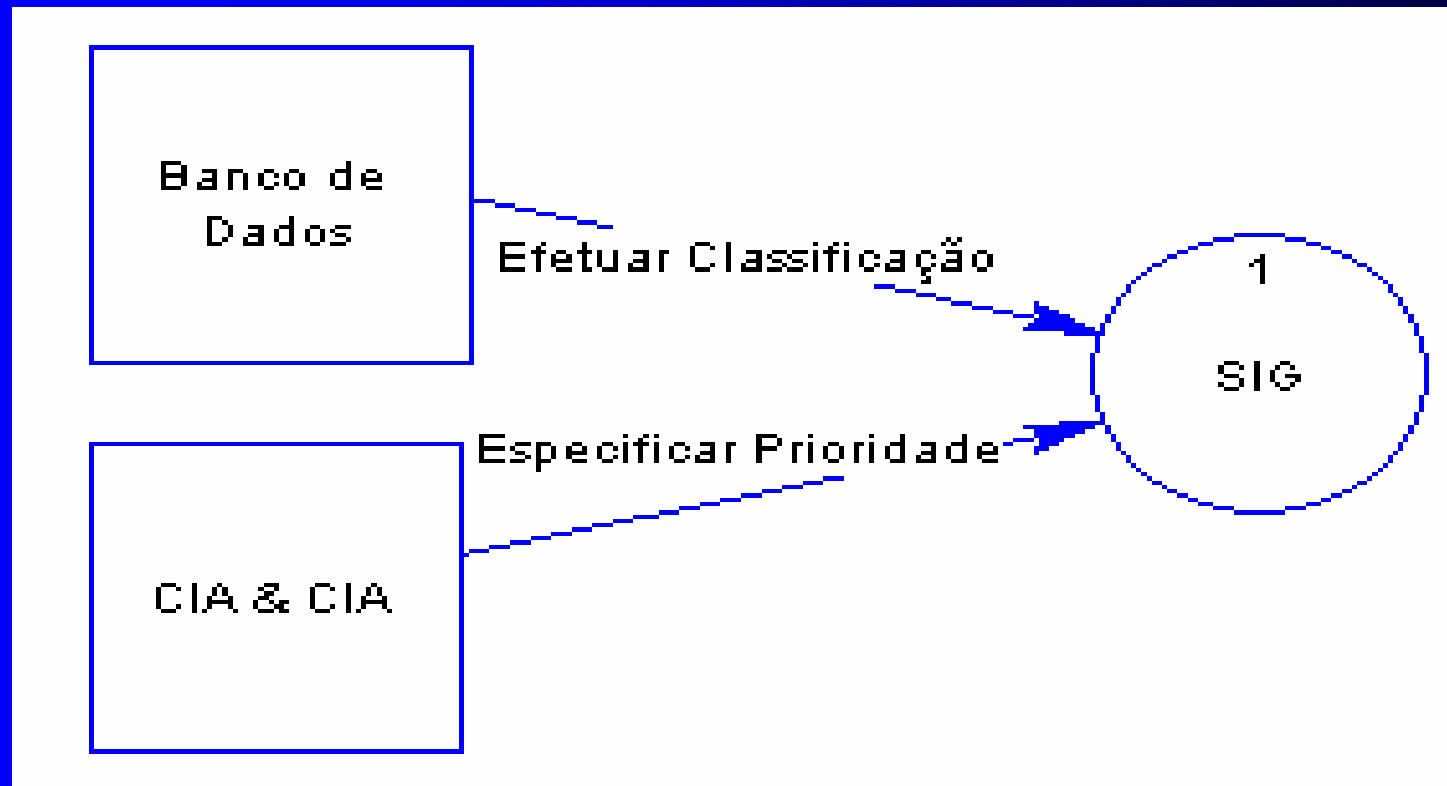


Diagrama de Fluxo de Dados do Questionário (Nível 1)

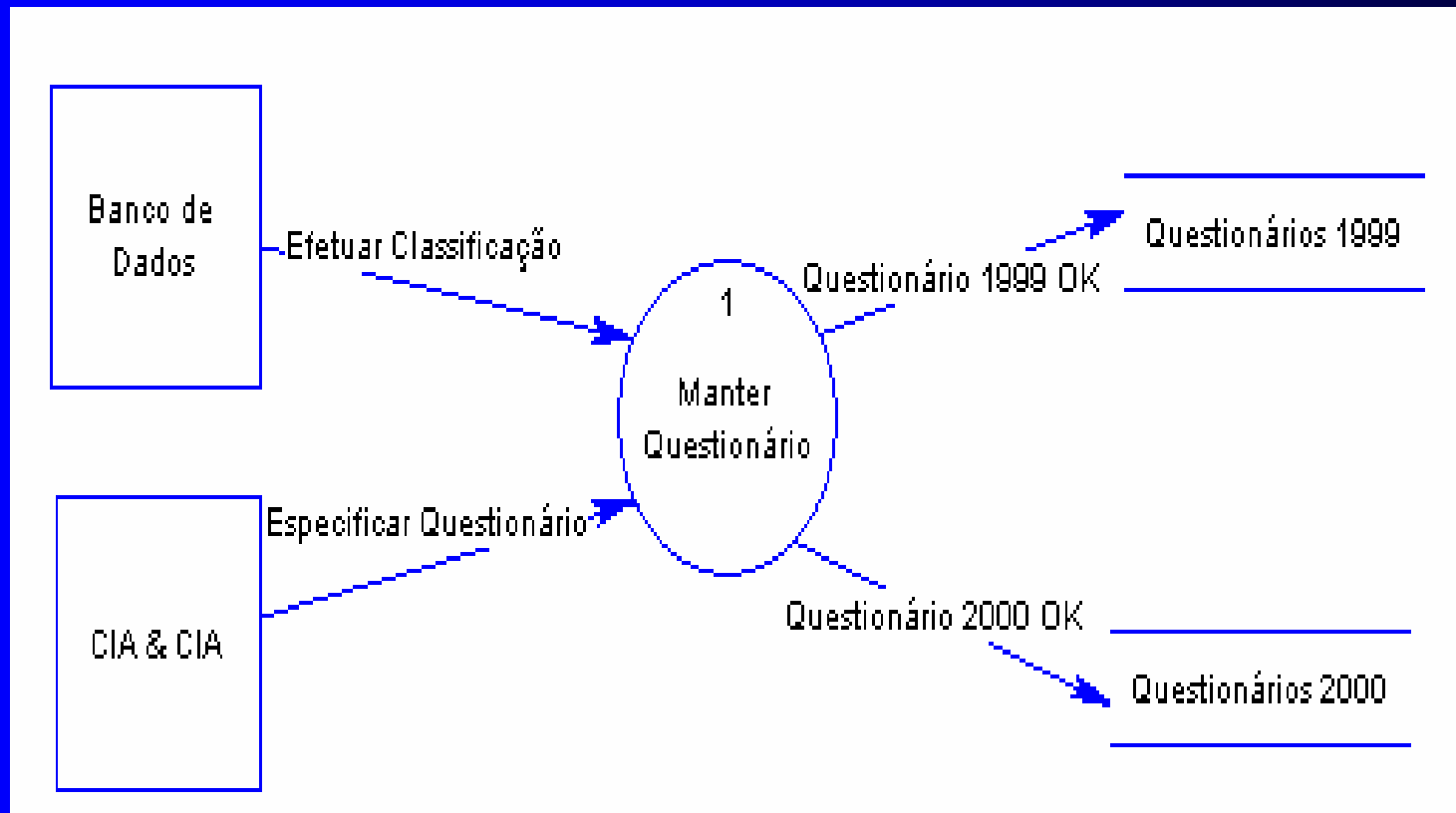


Diagrama de Fluxo de Dados do Comparativo (Nível 1)

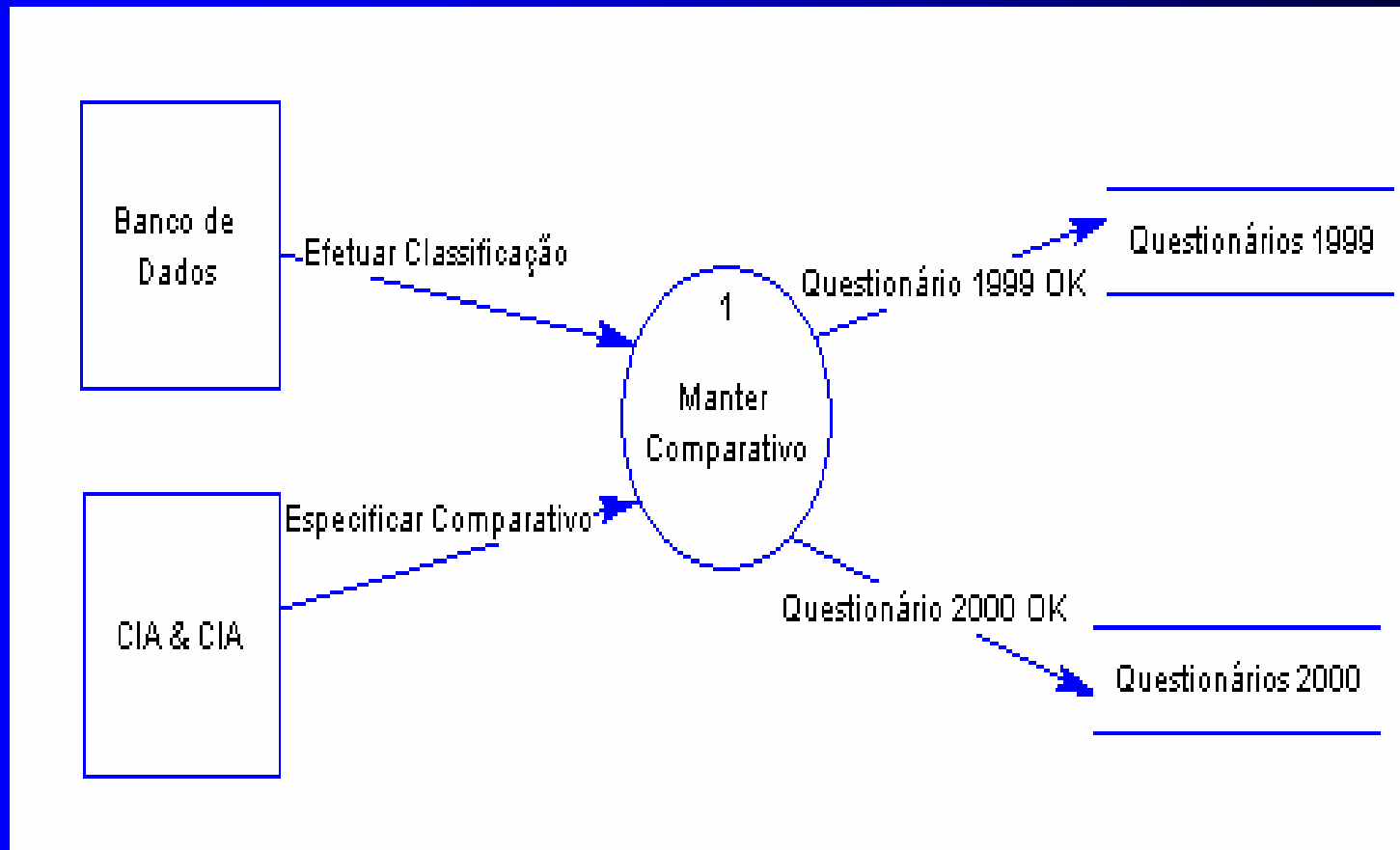
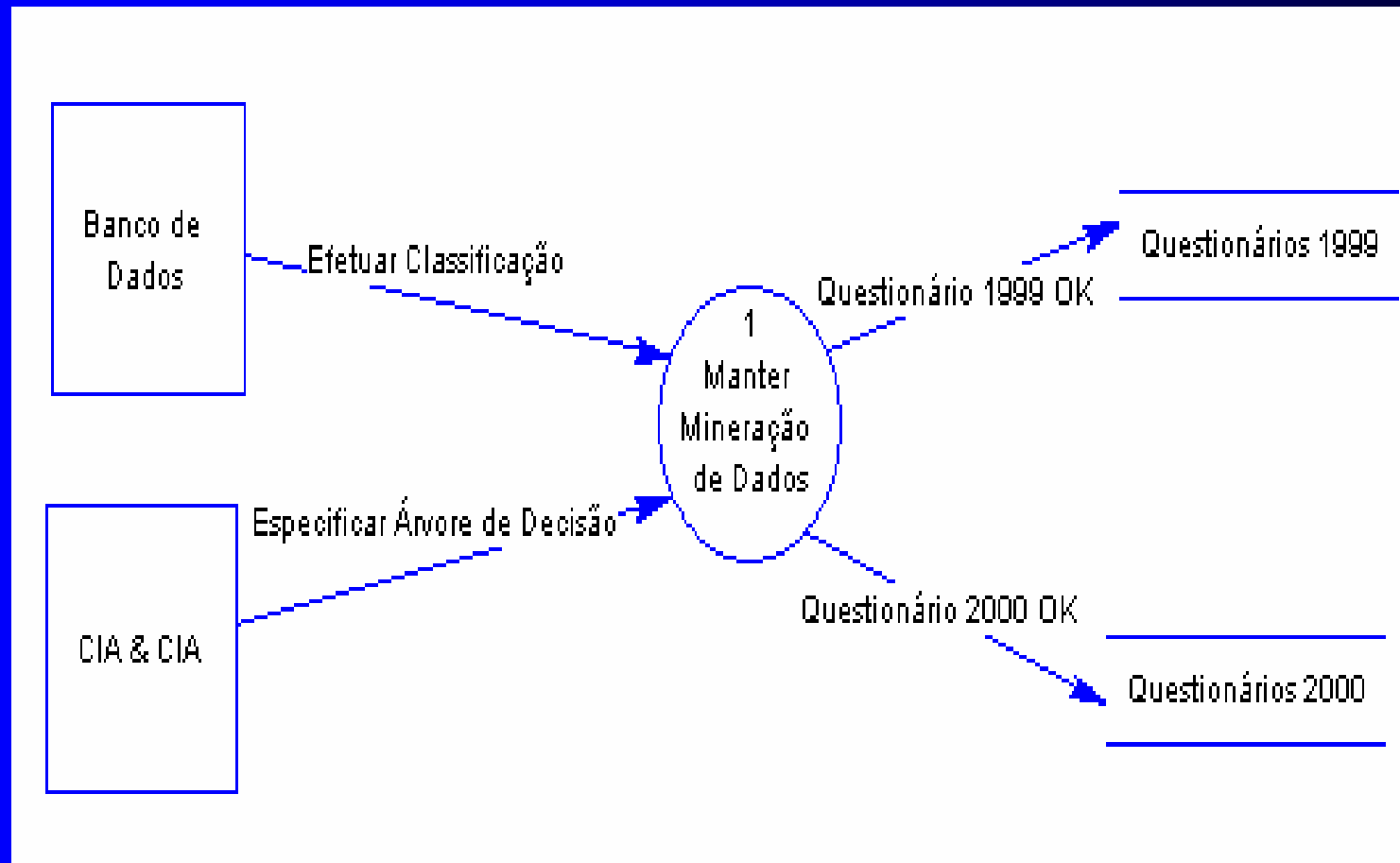
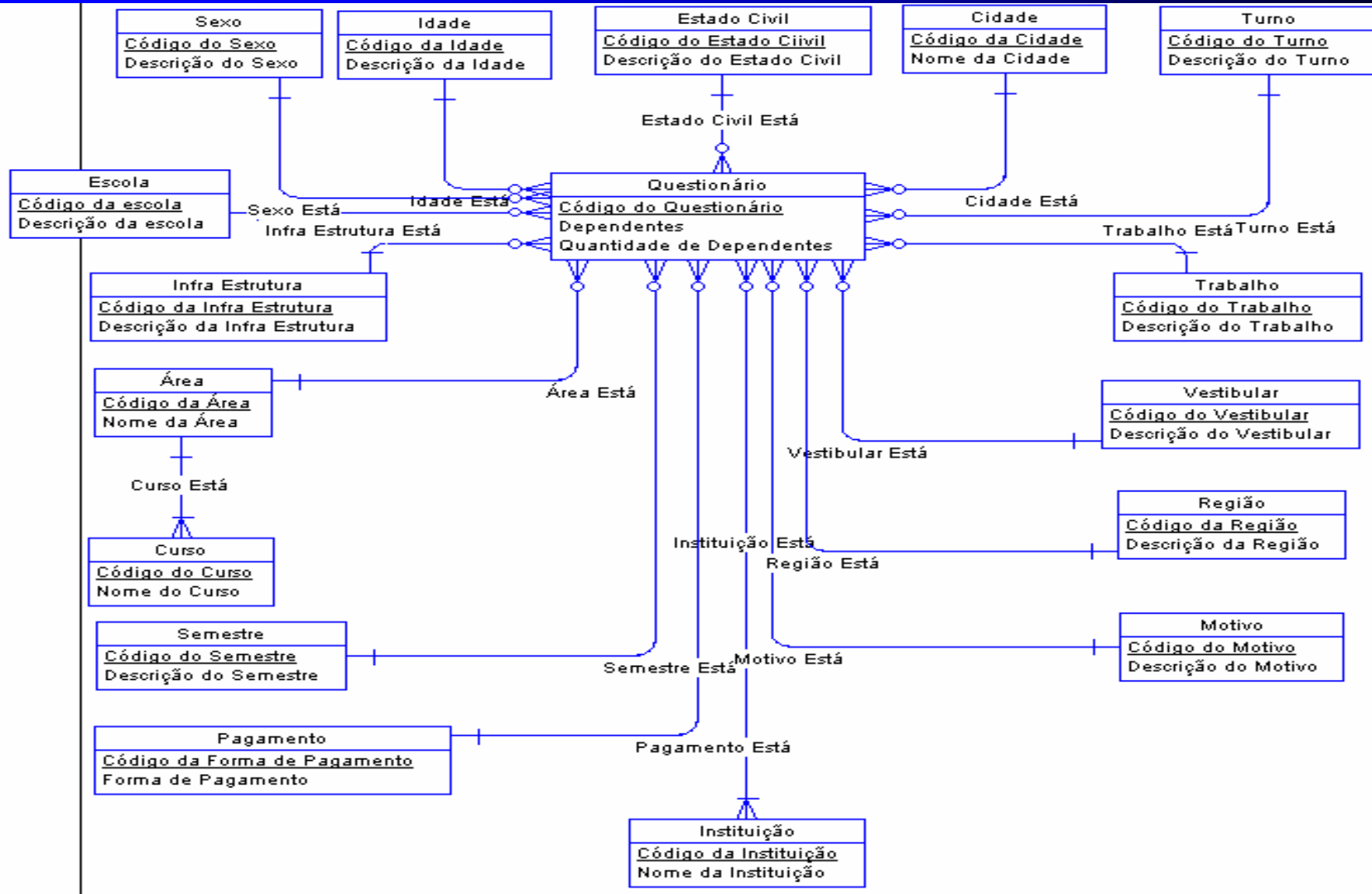


Diagrama de Fluxo de Dados da Mineração de Dados (Nível 1)



Modelo Entidade-Relacionamento



IMPLEMENTAÇÃO - Tela de Abertura



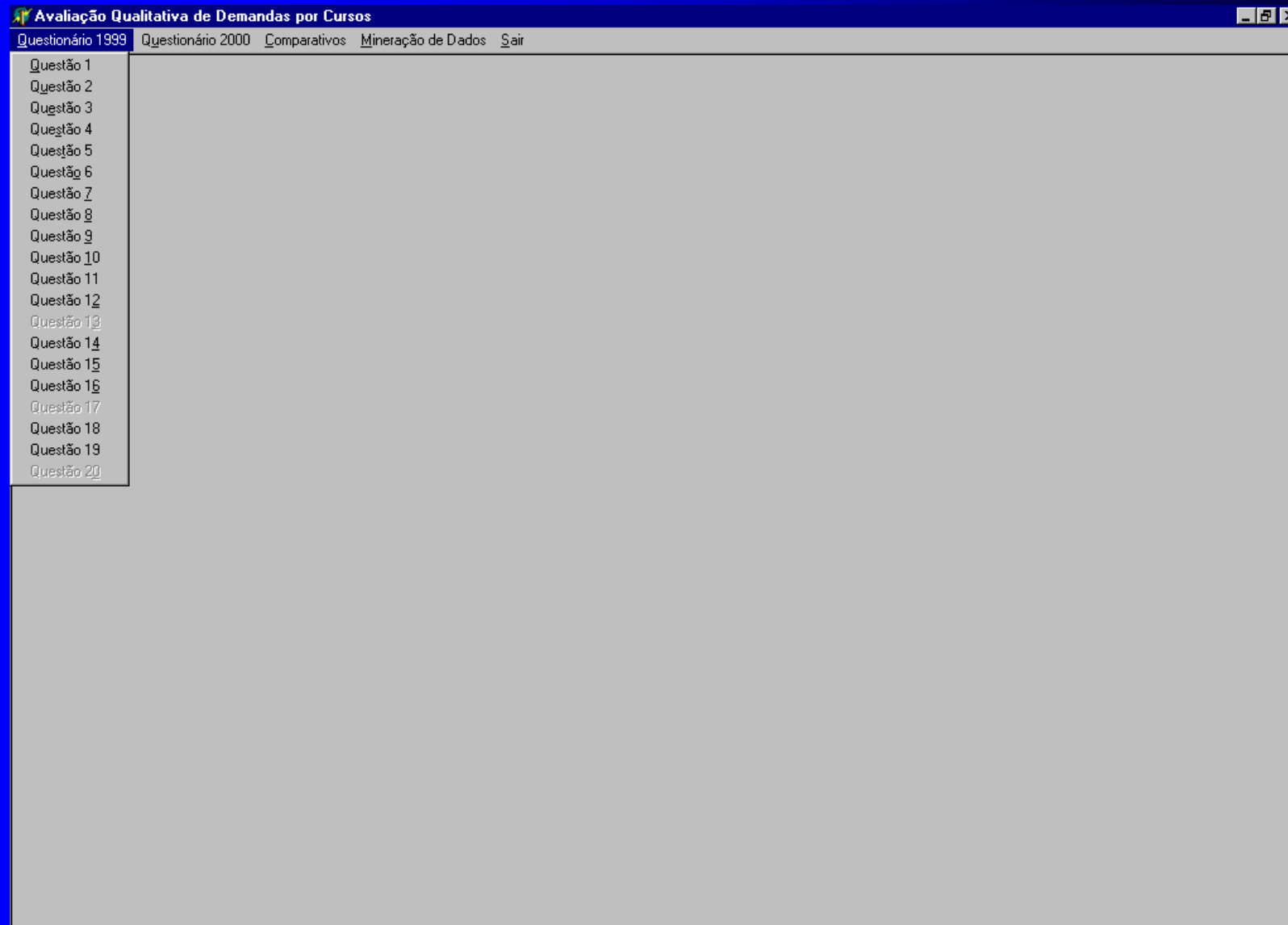
Universidade Regional de Blumenau

Protótipo de Sistemas de Informação aplicado
a CIA & CIA (Central de Informações aos
Alunos) da Pró-Reitoria de Ensino da FURB
utilizando Data Mining

Aluna: Bianca Nardelli

Professor: Oscar Dalfovo

IMPLEMENTAÇÃO - Tela Principal



IMPLEMENTAÇÃO - Questionário

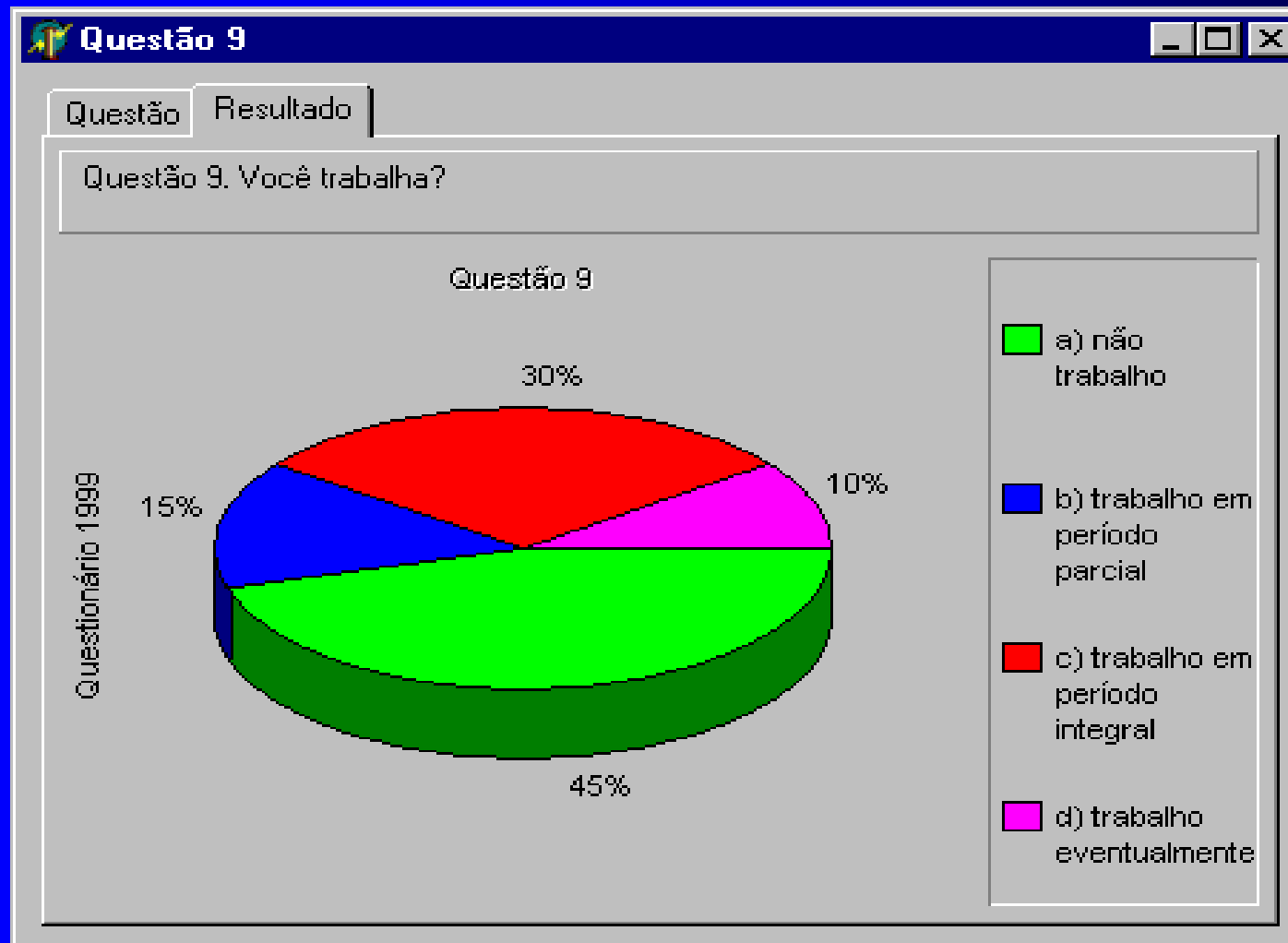
Questão 9

Questão | Resultado

Questão 9. Você trabalha?

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

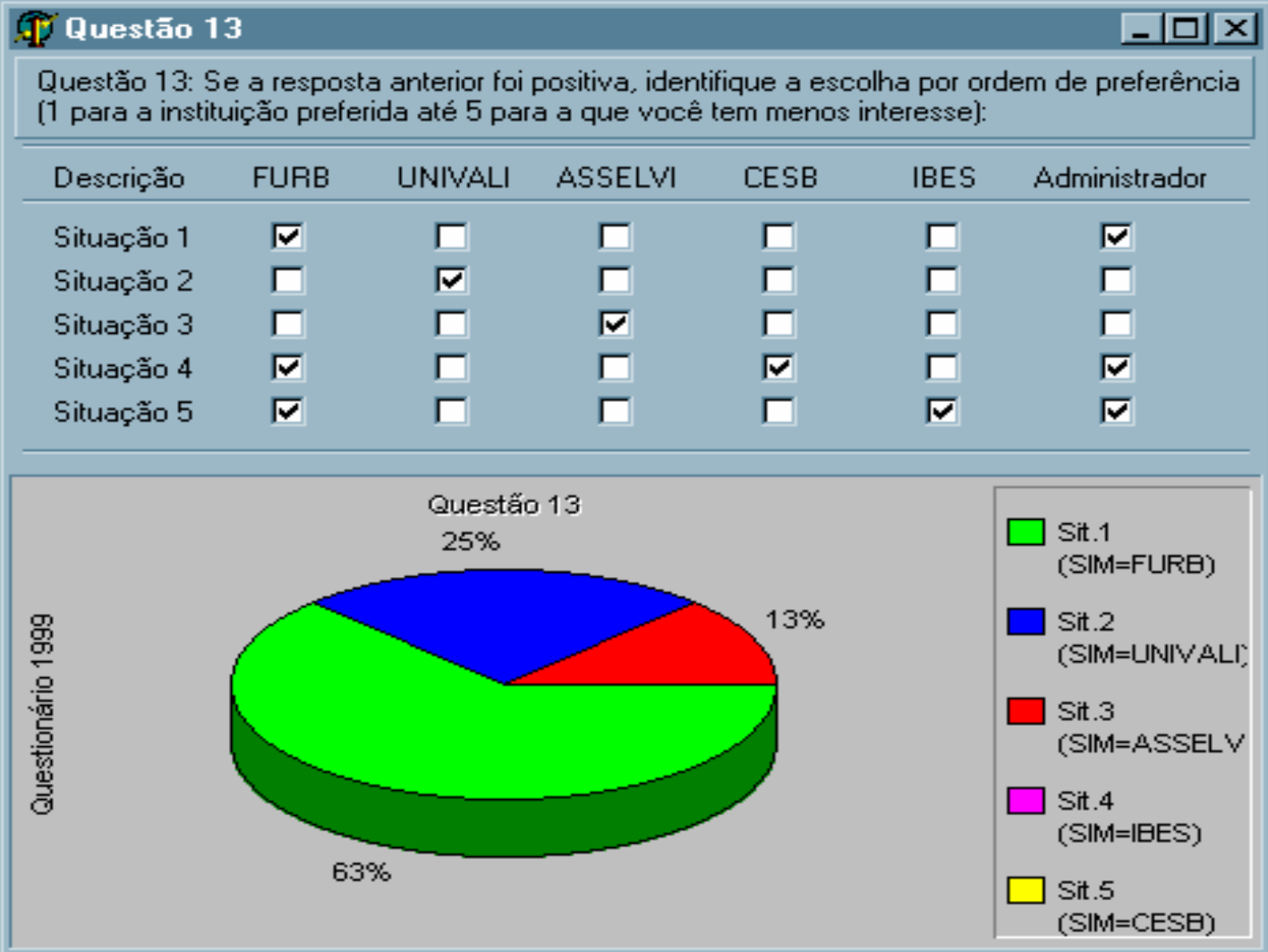
IMPLEMENTAÇÃO - Resultados



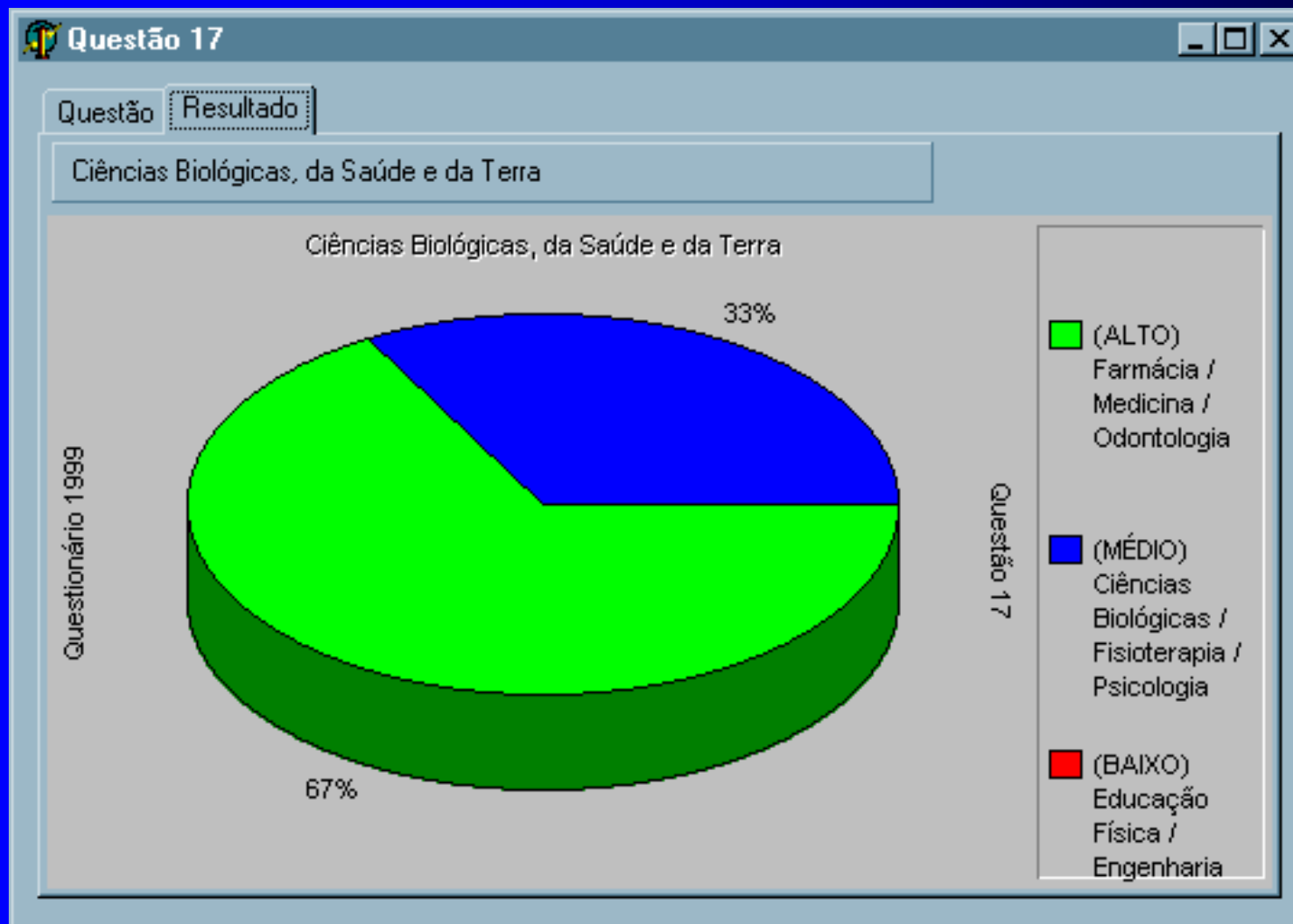
IMPLEMENTAÇÃO - Comparativos



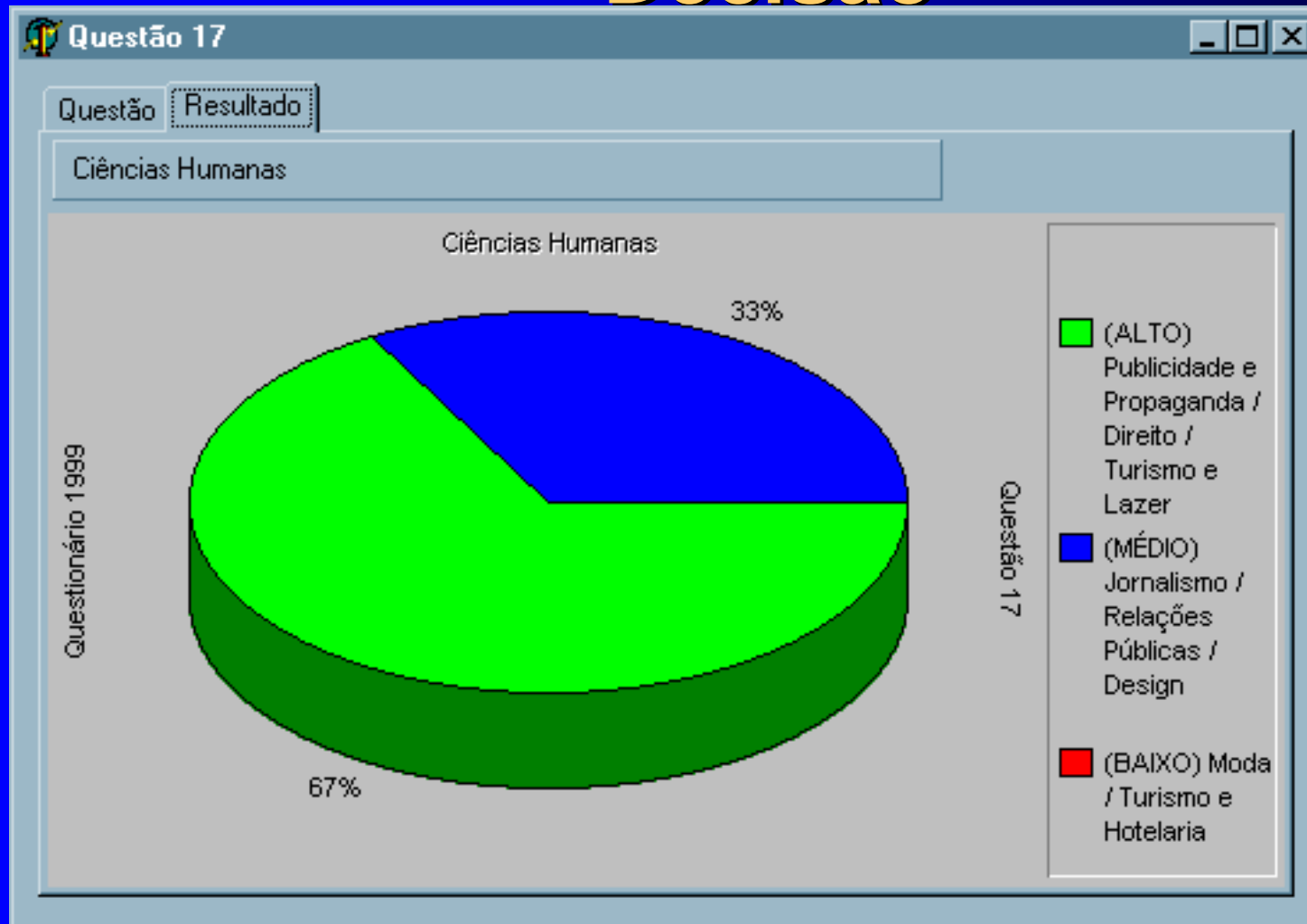
IMPLEMENTAÇÃO - Árvore de Decisão



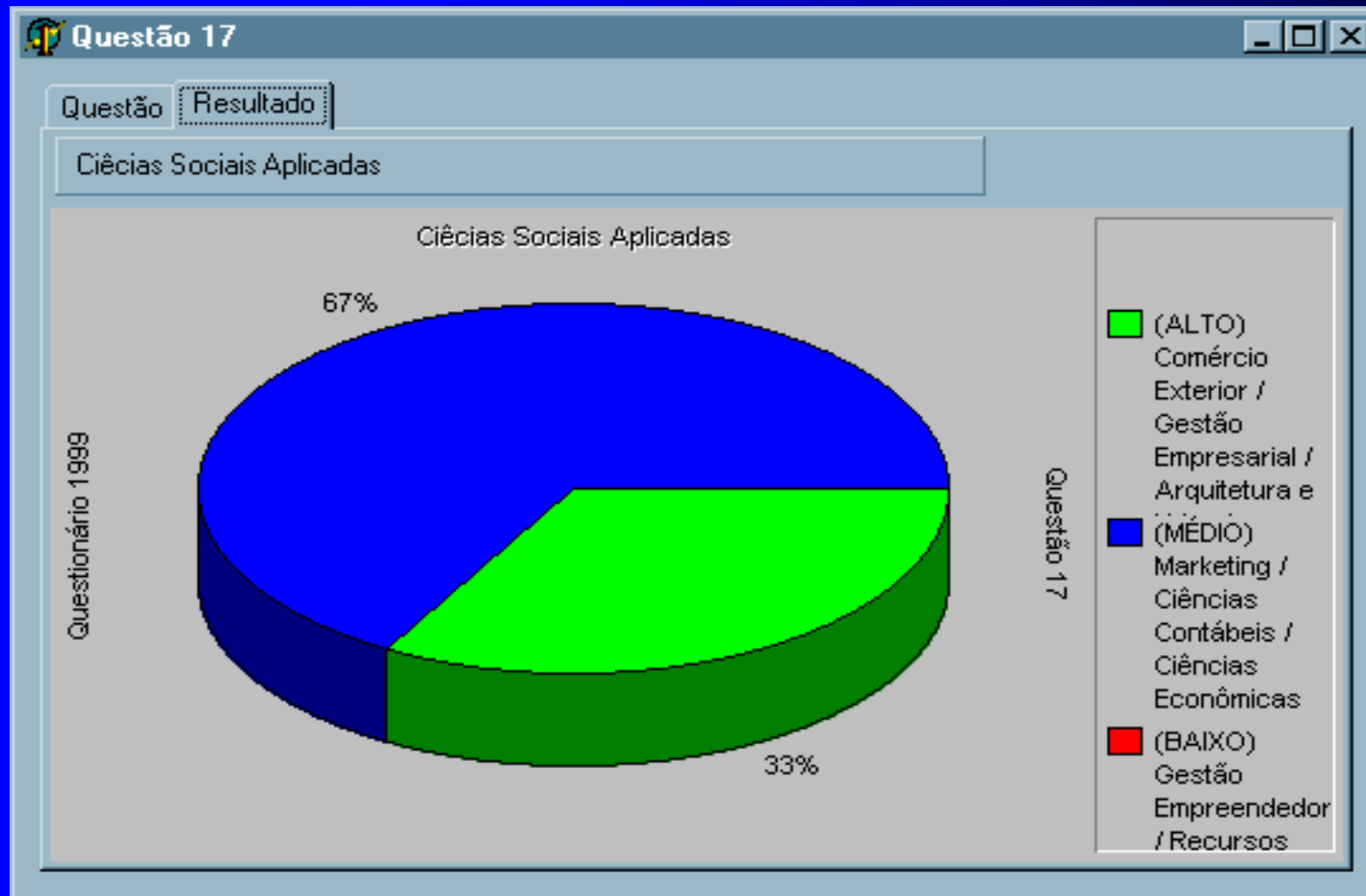
IMPLEMENTAÇÃO - Árvore de Decisão



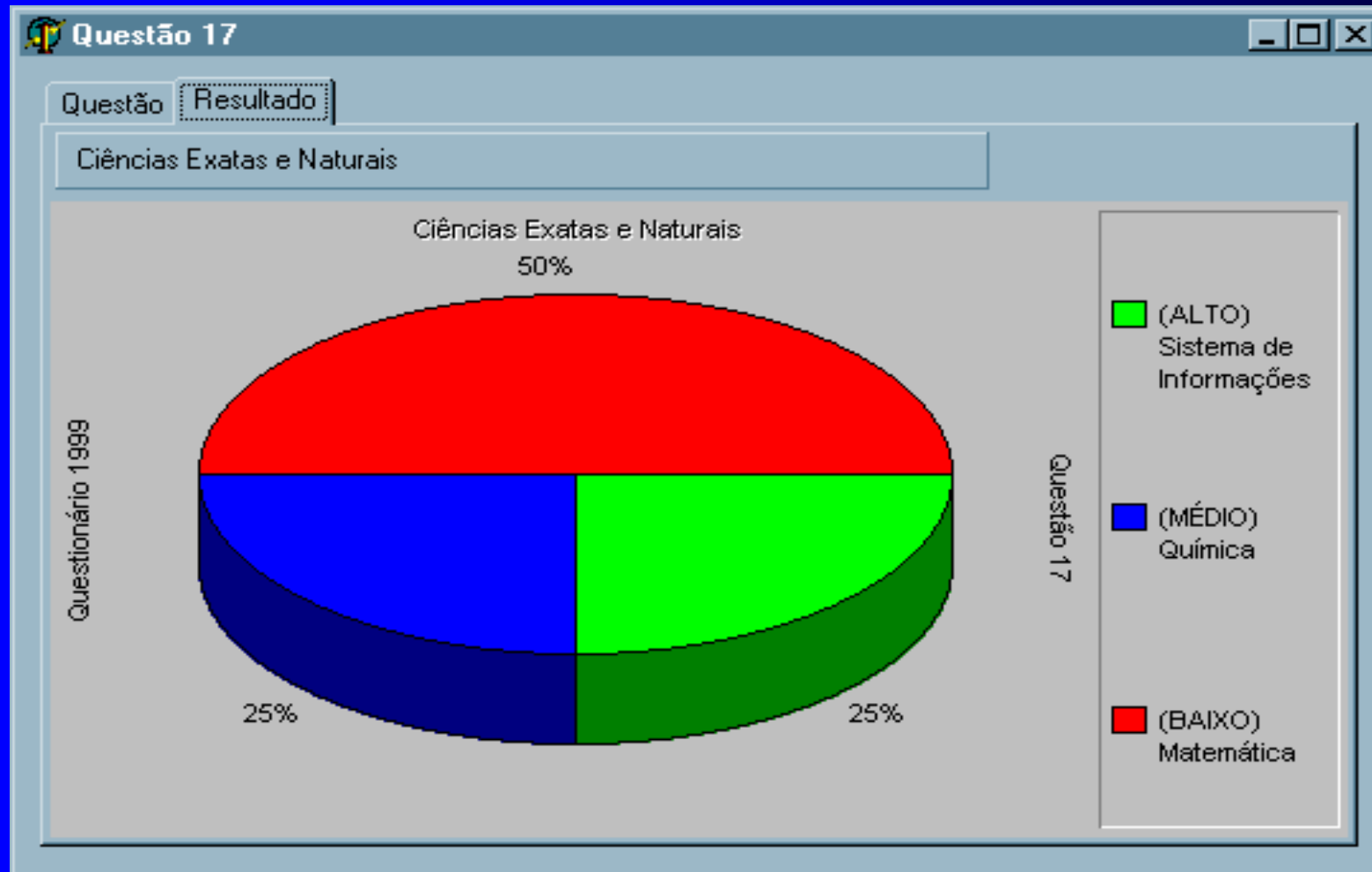
IMPLEMENTAÇÃO - Árvore de Decisão



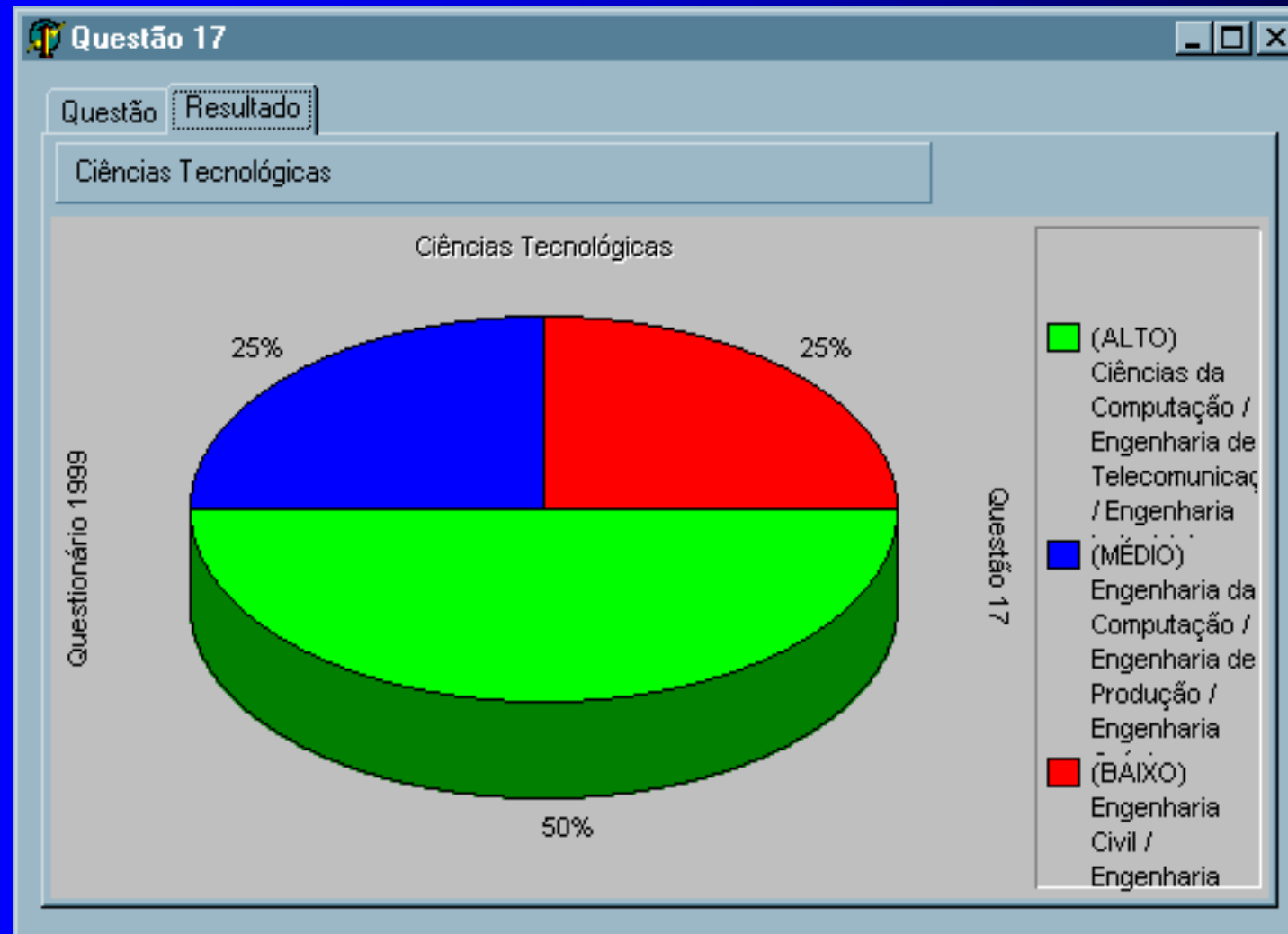
IMPLEMENTAÇÃO - Árvore de Decisão



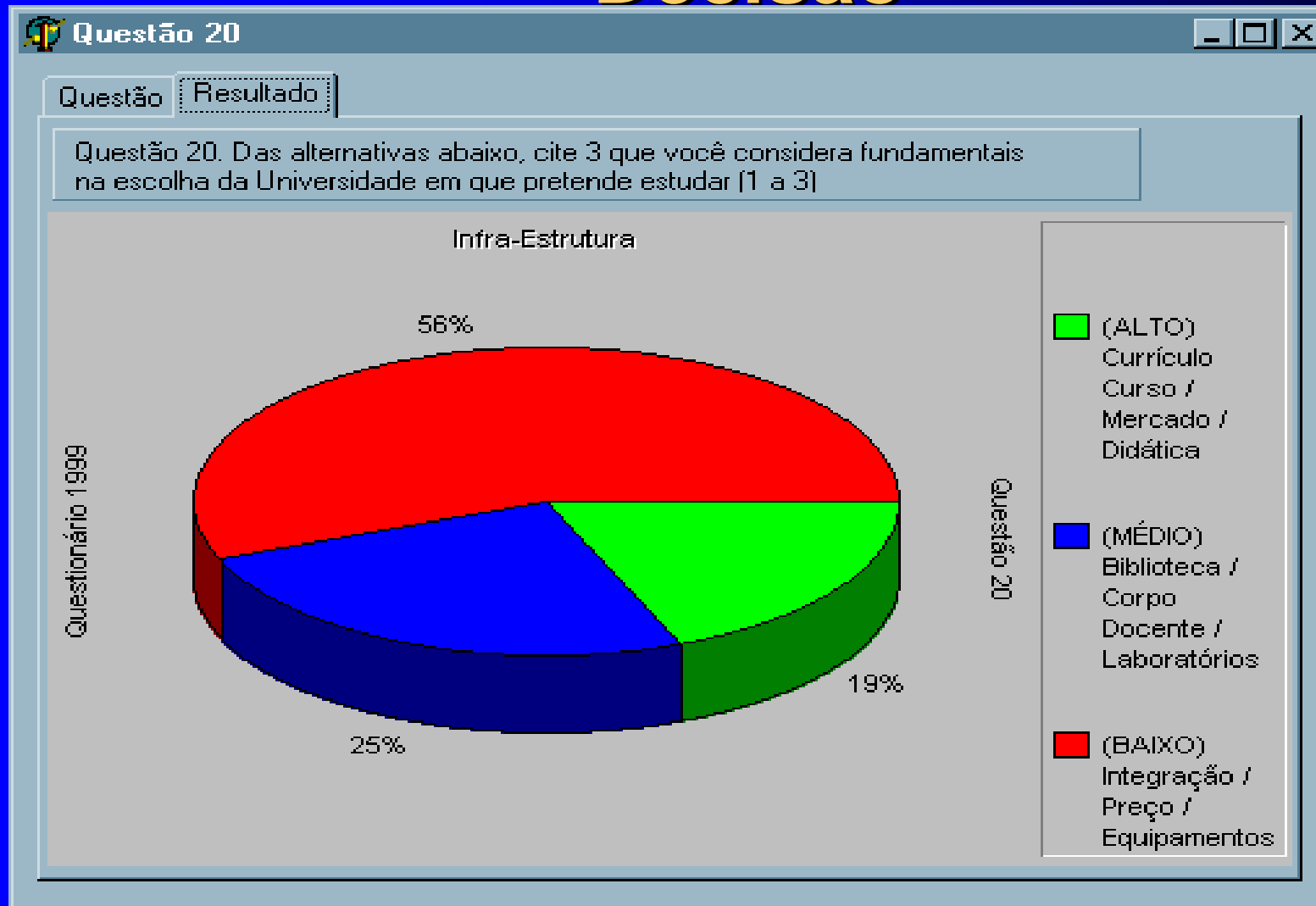
IMPLEMENTAÇÃO - Árvore de Decisão



IMPLEMENTAÇÃO - Árvore de Decisão



IMPLEMENTAÇÃO - Árvore de Decisão



CONCLUSÃO

- 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*.

LIMITAÇÕES

O protótipo construído apresenta as seguintes limitações:

- 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;
- os atributos envolvidos no processo de classificação possuem domínio fixo.

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.

BIBLIOGRAFIA

- [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.
- [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.
- [DAL2000] DALFOVO, Oscar; AMORIN, Sammy Newton. Quem tem informação é mais competitivo. Blumenau: Acadêmica, 2000.
- [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.
- [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.
- [YOU1990] YOURDON, Edward. Análise Estruturada Moderna. Rio de Janeiro: Campus, 1990.