

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

Vilson Wronscki Ricardo

**METODOLOGIA DE GERENCIAMENTO DE BASE
DE DADOS PARA ENSINO-APRENDIZAGEM DE
ESTATÍSTICA NA WEB**

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação.

Sílvia Modesto Nassar
Orientadora

Florianópolis, agosto de 2004.

**METODOLOGIA DE GERENCIAMENTO DE BASE DE
DADOS PARA ENSINO-APRENDIZAGEM DE
ESTATÍSTICA NA WEB**

VILSON WRONSKI RICARDO

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciências da Computação Área de Concentração Sistemas de Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciências da Computação.

Prof. Raul Sidnei Wazlawick, Dr.

Banca Examinadora

Prof.^a Sílvia Modesto Nassar, Dra. Eng.
Orientador

Prof. Jorge Coelho, Dr. Eng.

Prof. Marcelo Menezes Reis, Dr. Eng.

Prof. Pedro Alberto Barbetta, Dr. Eng.

“Não podes ensinar nada a um homem;
podes apenas ajudá-lo a encontrar a
resposta dentro dele mesmo”.

(Galileu Galilei)

AGRADECIMENTOS

A meus queridos pais, Antônio e Lídia, por todo o apoio prestado ao longo de toda a minha vida.

À minha esposa Silvana, por me acompanhar nesta caminhada.

À minha orientadora, colega e incentivadora professora Sílvia Modesto Nassar, pelo encorajamento e apoio proporcionado, sem o qual esse trabalho não seria viável.

Ao meu colega, professor Masanao Ohira, cuja sabedoria em Estatística e incentivador constante muito contribuiu para a realização desse trabalho.

A meus colegas do Laboratório de Estatística Aplicada do Departamento de Informática e Estatística da Universidade Federal de Santa Catarina, por todo o auxílio prestado.

Ao Departamento de Informática e Estatística do Centro Tecnológico na pessoa do seu dirigente Luís Fernando Friedrich por todo o apoio e disposição de recursos para o desenvolvimento desse trabalho.

E aos meus filhos Eduardo, Vanessa e André por terem suportado a minha ausência e humor, sem entender claramente todo processo, porém compreendendo a importância disso tudo para mim.

SUMÁRIO

AGRADECIMENTOS	iv
SUMÁRIO	v
LISTA DE FIGURAS	viii
LISTA DE TABELAS	x
SÍMBOLOS, ABREVIATURAS, SIGLAS E CONVENÇÕES	xi
RESUMO	xiii
ABSTRACT	xiv
1 INTRODUÇÃO	1
1.1 Objetivos	4
1.1.1 Objetivo Geral	4
1.1.2 Objetivos Específicos	5
1.2 Estrutura da dissertação	5
2 BANCO DE DADOS	7
2.1 Introdução	7
2.2 Características Operacionais dos SGDB	8
2.3 Arquitetura de um SGBD	9
2.4 Modelo de Dados Relacional	10
2.5 Técnicas de Normalização	12
2.5.1 Primeira Forma Normal	13
2.5.2 Segunda Forma Normal	14
2.5.3 Terceira Forma Normal	15
2.5.4 Relacionamento de Dados	16
2.5.5 Quarta Forma Normal	18
2.5.6 Quinta Forma Normal	19
2.6 SQL (Structured Query Language)	19
2.7 Banco de Dados e a WEB	20
2.8 Considerações finais	22
3 EDUCAÇÃO À DISTÂNCIA	24
3.1 Introdução	24
3.2 Conceituação	26

3.3	A Educação à Distância no Brasil	27
3.4	Regulamentação do EaD no Brasil.....	29
3.5	Considerações finais.....	30
4	SEstatNet.....	31
4.1	Introdução.....	31
4.2	Histórico do desenvolvimento.....	32
4.3	Características do SEstatNet.....	33
4.4	Interface do Sistema	35
4.5	Como funciona o SEstatNet	36
4.6	Considerações Finais	38
5	DESENVOLVIMENTO DA METODOLOGIA	39
5.1	Introdução.....	39
5.2	Trabalhos correlatos	40
5.2.1	CARNEGIE MELLON UNIVERSITY	40
5.2.2	StatCrunch	42
5.2.3	OpenEpi.....	43
5.3	Administração dos dados.....	44
5.3.1	Diagrama Entidade/Relacionamento.....	44
5.3.2	Dicionário de Dados	46
5.4	Esquema lógico do Banco de Dados	51
5.5	Controle Acadêmico.....	53
5.6	Variáveis.....	58
5.6.1	Mensuração.....	59
5.6.2	Transformações de variáveis	59
5.7	SETip.....	62
5.7.1	Processo de Tipificação.....	62
5.7.2	O conhecimento no SETip.....	62
5.7.3	Considerações de implementação.....	63
5.8	Tabela de Dados	64
5.8.1	Características de uso	65
5.8.2	<i>Menu</i> Dados.....	66
5.9	Arquitetura Computacional	66

5.10	Mecanismos de segurança	67
5.11	Considerações finais	69
6	MODELO DE CONTEXTO E DE COMPORTAMENTO	70
6.1	Lista de Eventos do Sistema de Matrícula.....	70
6.2	Diagrama de contexto.....	70
6.3	Modelo de Comportamento	72
6.3.1	Escolher base de dados em Meus dados.....	73
6.3.2	Carregar Dados de Arquivo ou URL.....	74
6.3.3	Amostra de Dados	75
6.3.4	Exportar Dados	76
6.3.5	Função “Apagar”	76
6.3.6	Transformação de Dados	77
6.3.7	Seqüência de Dados.....	78
6.3.8	Classificação de Dados.....	78
6.4	Considerações finais	79
7	CONSIDERAÇÕES FINAIS.....	80
7.1	Conclusão	80
7.2	Sugestões para trabalhos futuros	82
	REFERÊNCIAS	84
	APÊNDICE A: Relação de alguns arquivos fontes do sistema de manipulação de base de dados.....	90

LISTA DE FIGURAS

<i>Figura 2.1: Comando SQL para selecionar o usuário Carlos e suas Urls do exemplo fictício.</i>	18
<i>Figura 2.2: Comando SQL para selecionar informações de usuário e url de todos os alunos do exemplo fictício.</i>	19
<i>Figura 2.3: Fluxo de Informação para acesso de dados pela Web.</i>	21
<i>Figura 4.1: Interface gráfica do SEstatNet.</i>	36
<i>Figura 4.2: Parte de uma base de dados fixa denominada PAP.</i>	37
<i>Figura 5.1: Estrutura do Curso Carnegie Mellon.</i>	41
<i>Figura 5.2: Diagrama E/R da modelagem proposta para ensino-aprendizagem de estatística na Web.</i>	45
<i>Figura 5.3: Esquema lógico do Banco de Dados.</i>	51
<i>Figura 5.4: Comandos SQL para criação das tabelas Cadastro de Usuário e Base de Dados.</i>	52
<i>Figura 5.5: Interface gráfica para registro de um novo usuário do sistema.</i>	53
<i>Figura 5.6: Parte da tela de Certificação de Matrícula do Aluno.</i>	54
<i>Figura 5.7: Funções linearizáveis e suas respectivas transformações.</i>	60
<i>Figura 5.8: Funções cujo padrão de regressão é linearizável e suas respectivas transformações.</i>	60
<i>Figura 5.9: Tabela de Dados.</i>	65
<i>Figura 6.1: Diagrama de Contexto: Aluno – Sistema e Tutor – Sistema.</i>	71
<i>Figura 6.2: Tela do Menu Dados para seleção de operações.</i>	72
<i>Figura 6.3: O aluno não possui nenhuma base de dados armazenada em seu repositório.</i>	73
<i>Figura 6.4: O aluno possui a base de dados PAP armazenada em seu repositório.</i>	73
<i>Figura 6.5: Tela de seleção da opção de carregar dados de arquivo.</i>	74
<i>Figura 6.6: Tela para carregar dados de arquivo.</i>	74
<i>Figura 6.7: Tela para carregar dados de uma URL.</i>	75
<i>Figura 6.8: Amostras de dados disponíveis para uso no SEstatNet.</i>	75
<i>Figura 6.9: Tela utilizada para a exportação de dados.</i>	76
<i>Figura 6.10: Tela para transformação de dados.</i>	77

<i>Figura 6.11: Tela de definição da seqüência a ser gerada.</i>	78
<i>Figura 6.12: Tela de definição da classificação da coluna.</i>	79

LISTA DE TABELAS

<i>Tabela 2.1: Tabela de dados fictícios na Forma Zero. Exemplo de aplicação.....</i>	12
<i>Tabela 2.2: Tabela de dados fictícios na Primeira Forma Normal. Exemplo de aplicação.....</i>	13
<i>Tabela 2.3: Tabelas de dados fictícios após a aplicação da Segunda Forma Normal..</i>	14
<i>Tabela 2.4: Tabelas de dados fictícios após a aplicação da Terceira Forma Normal..</i>	15
<i>Tabela 2.5: Alteração nas estruturas das Tabelas do exemplo fictício para o relacionamento muito-para-muitos.</i>	17
<i>Tabela 5.1: Atributos da Entidade PESSOA.</i>	47
<i>Tabela 5.2: Atributos da Entidade Base de Dados.</i>	47
<i>Tabela 5.3: Atributos de Entidade MATRÍCULA.....</i>	48
<i>Tabela 5.4: Atributos da Entidade CURSO.....</i>	48
<i>Tabela 5.5: Atributo do Relacionamento Coordena.</i>	49
<i>Tabela 5.6: Atributo do Relacionamento Tutoria.</i>	50
<i>Tabela 5.7: Atributo do Relacionamento Participa de.</i>	50
<i>Tabela 6.1: Descrição do contexto: Aluno – Sistema.</i>	71
<i>Tabela 6.2: Descrição do contexto: Tutor – Sistema.</i>	71

SÍMBOLOS, ABREVIATURAS, SIGLAS E CONVENÇÕES

ANSI	American National Standard Institute
BD	Base de Dados
CEFET	Centro Federal de Educação Tecnológica
CTC	Centro Tecnológico
CNPq	Conselho Nacional de Pesquisa
CPAN	Comprehensive Perl Archive Network
DCL	Data Control Language
DDL	Data Definition Language
DML	Data Manipulation Language
E/R	Entidade / Relacionamento
EaD	Educação à Distância
FAQ	Frequently Asked Questions
HPL	Plataforma para Web para definir classes em Perl
HTML	Hyper Text Markup Language
IA	Inteligência Artificial
IBM	International Business Machines
INE	Departamento de Informática e Estatística
IPX/SPX	Internetwork Packet Exchange/Sequenced Packet Exchange
J2EE™	Java 2 Enterprise Edition
JSP	Java Server Pages
LDB	Lei de Diretrizes e Bases da Educação Nacional
LEA	Laboratório de Estatística Aplicada
MEC	Ministério de Educação e Cultura
MySQL	SGBD Relacional obtido em (http://www.mysql.org)
NetBEUI	NetBios Extended User Interface
NIED	Núcleo de Informática Aplicada à Educação
NT	Windows NT (New Technology)
PAP	Programa de Alimentação Popular
PIBIC	Programa Institucional de Bolsas de Iniciação Científica
SENAC	Serviço Nacional de Aprendizagem Comercial

SGBD	Sistema de Gerência de Banco de Dados
SocInfo	Sociedade da Informação no Brasil
SQL	Standard Query Language
TCP/IP	Transmission Control Protocol/Internet Protocol
UFPE	Universidade Federal de Pernambuco
UFSC	Universidade Federal de Santa Catarina
UNESCO	United Nations Educational, Scientific and Cultural Organization
Unicamp	Universidade Estadual de Campinas
URL	Uniform Resource Locator
WWW	World Wide Web
XML	Extensible Markup Language

RESUMO

Um dos meios de comunicação mais emergentes atualmente é a Web, que se torna a cada dia mais comum e acessível. E o ensino à distância é uma dos segmentos que mais cresce na Internet. Hoje existem incontáveis cursos disponíveis e ambientes para a sua criação. Todo o controle e administração dos cursos *on-line* são feitos com o uso de Banco de Dados.

Uma característica muito comum nos *softwares* que utilizam banco de dados é a situação em que o usuário final só pode realizar recuperações. As atualizações com novas informações são limitadas a especialistas do domínio do banco de dados, que supervisionam e analisam os novos dados que estão sendo inseridos. O desafio computacional é mudar essas características.

Em torno dessa idéia surge o objetivo do trabalho que é desenvolver uma metodologia de gerenciamento de base de dados que permita ao usuário alterar suas próprias bases de dados e em tempo real. Dessa forma criou-se uma interface gráfica para a apresentação dos dados aos usuários que disponibiliza recursos como importação e exportação de dados, operações para definição de novas variáveis e transformações por meio de funções embutidas na linguagem SQL. Tendo em vista a sua aplicação em procedimentos de análise estatística de dados destacam-se entre essas funções de transformação: a função inversa, exponencial, raiz quadrada, logarítmica e a padronização pela normal. Considerando ainda que as operações são realizadas em tempo real, foi desenvolvido um módulo de matrícula para cadastro de usuários.

O ambiente de ensino-aprendizagem que a metodologia foi testada é o SEstatNet (Sistema Especialista para Apoio ao Ensino de Estatística) e com ela o *software* apoiará a aprendizagem de estatística via Web, trabalhando simultaneamente com diferentes usuários, que possuem diferentes bases de dados e realizam diferentes análises estatísticas.

A aplicação dessa metodologia resolveu o grande problema de acesso à base de dados que existia nas versões anteriores do SEstatNet, tornando o processo de ensino-aprendizagem de estatística mais dinâmico e eficaz.

Palavras-Chave:

Banco de dados, Educação à distância, Análise Estatística.

ABSTRACT

One of the fastest-growing means of communication today is the *Internet*, which becomes more common and accessible every day. Distance-learning is one of the fastest-growing segments of the Internet. Today countless online courses exist, as well as environments to create such courses. The control and administration of these courses use databases.

A common characteristic of the software that uses database servers is that the end-user can only make database queries. The ability to update databases with new information is limited to specialists in the domain of database administration, who supervise and analyze the data being inserted. Improving this situation is an important problem in computer science.

Based on that idea, the objective of this work is to develop a methodology of database administration that allows users to alter their own databases in real time. To this end, a graphical interface was created for the presentation of data to users. This interface also for the importation and exportation of data, operations to create new variables, and transformation through built-in functions in the SQL language. Some of the outstanding transformational procedures, which can be applied to analytical and statistical procedures, include: inverse, exponential, square root, logarithm, and standardization to the normal. Since the system works in real time, a registration module was developed so that users may register with the system.

The teaching-learning environment in which this methodology will be tested is SestatNet (Specialist System for Supporting the Teaching of Statistics). The methodology will support the teaching of statistics on the web, working simultaneously with different users, who possess different databases and perform different statistical operations.

The application of this methodology solved the problem of access that had existed in previous versions of SestatNet, and made the teaching and learning of statistics more dynamic and effective.

Keywords:

Database, Distance Learning, Statistical Analysis

1 INTRODUÇÃO

O que precisa ser feito...

A análise estatística de observações de uma realidade de interesse necessita de uma estrutura formal dos dados em uma base de dados.

Normalmente representam aspectos do mundo real. Assim sendo, uma base de dados pode ter qualquer tamanho e complexidade, deve possuir uma fonte da qual os dados são derivados, algum grau de interação com eventos do mundo real e um público que está ativamente interessado no seu conteúdo. As bases de dados normalmente são definidas e gerenciadas em um servidor de banco de dados.

Uma característica muito comum nos *softwares* que utilizam banco de dados é a natureza específica dos dados em cada domínio específico da aplicação. Além disso, por sua natureza, apresentam uma característica “estática”, isto é, situação em que o usuário final só pode realizar recuperações no banco de dados. As atualizações com novas informações são limitadas a especialistas do domínio do banco de dados, que supervisionam e analisam os novos dados que estão sendo inseridos.

A metodologia apresentada neste trabalho busca mudar essas características, permitindo que o usuário possa realizar suas próprias atualizações nas suas bases de dados.

Como será feito...

Para que isso possa ser feito será desenvolvido um Módulo de Gerenciamento de Base de Dados que consiste numa coleção de programas que possibilitam aos usuários a criação e manutenção de suas bases de dados, isto é, facilitam o processo de definição, construção e manipulação da base de dados para várias aplicações.

Essas bases de dados têm que ser armazenadas em um servidor de banco de dados. Construir um banco de dados é o processo de armazenar os referidos dados em algum meio de armazenamento que seja controlado pelo Sistema Gerenciador de Banco de Dados (SGBD). A definição desse banco de dados envolve especificar os tipos de dados, as estruturas e as restrições para os dados que serão armazenados. E manipular o banco de dados inclui funções como: fazer consultas para recuperação de dados

específicos, atualizações para refletir alterações no mundo real e gerar relatórios a partir dos dados armazenados.

Por que precisa ser feito...

O Departamento de Informática e de Estatística (INE) da Universidade Federal de Santa Catarina (UFSC) em seu Laboratório de Estatística Aplicada (LEA) vem desenvolvendo um *software*, com técnicas de Inteligência Artificial, SEstat (Sistema Especialista de Apoio ao Ensino-Aprendizagem de Estatística).

A aprovação da Portaria nº 2.253 pelo MEC, que possibilita que até 20% das aulas da graduação sejam ministradas via Ensino à Distância, levou a equipe de desenvolvimento do SEstat a iniciar a adaptação da versão atual de ensino presencial para uma nova versão de ensino à distância (via Web), a qual foi denominada de SEstatNet.

Onde é usado...

O SEstat vem sendo utilizado como ferramenta de apoio ao ensino-aprendizagem de estatística nas disciplinas regulares de cursos de graduação em Engenharia, Ciências da Computação e Sistemas de Informação. Tratam-se de disciplinas de Estatística, oferecida pelo departamento de Informática e de Estatística do Centro Tecnológico da UFSC. As disciplinas abrangem basicamente técnicas de coleta e análise estatística de dados (paramétricas e não paramétricas), à luz da teoria da probabilidade.

Esse *software* está sendo usado desde 1999 para o ensino de estatística nos cursos regulares citados acima. As atividades didático-pedagógicas são realizadas no Laboratório Integrado de Informática do Centro Tecnológico (LIICT), com a presença dos professores da disciplina.

De que maneira é usado...

O papel do professor neste ambiente de ensino é atuar como *suporte* do processo do qual o aluno é o sujeito, agindo quando há uma demanda consciente por parte do aluno, ou quando percebe que é o momento adequado para um suporte teórico. Foi dada grande ênfase à interação entre professor e aluno.

Observou-se, ao longo de vários semestres de utilização do SEstat, que o processo de ensino e aprendizagem de estatística oportuniza: a construção do conhecimento pelo aluno, conforme preconizado por Piaget [PIAGET, 1967] [PIAGET, 1976]; a aprendizagem a partir de uma base de dados de interesse do aluno; a aprendizagem de conceitos de meta-análise de dados; “aprender fazendo” apoiado por um *help* sensível ao contexto [NASSAR et al., 2003].

O que se pensa para o futuro...

Seguindo as novas tendências de educação continuada, com ênfase a educação à distância, a equipe de desenvolvimento do SEstatNet julga ser fundamental iniciar o ensino de estatística via Web para alunos de graduação de nossa Universidade, tendo em vista os recursos tecnológicos disponíveis.

Na modalidade de educação à distância, pelo distanciamento físico de educadores e alunos, é esperado que essa nova maneira de ensinar e aprender se torne cada vez mais presente. A educação à distância pode incorporar várias tecnologias diferentes, as quais podem variar desde uma correspondência comum até os mais elaborados programas computacionais e meios de transmissão. Entretanto, uma das questões mais difíceis é definir quais tecnologias farão parte do ambiente a ser criado e qual a metodologia a ser empregada na integração destas.

O foco da pesquisa...

Uma das aplicações mais importantes nos desenvolvimentos recentes na área de banco de dados é a criação de novos ambientes e funcionalidades em SGBDs, tais como aplicações que suportam acesso universal a bancos de dados da World Wide Web e banco de dados multimídia fornecendo acesso adicional para suportar o armazenamento e o processamento de informações multimídia.

O SEstatNet é uma dessas aplicações. Ele utiliza os conceitos de inteligência artificial distribuída e de sistemas especialistas e se pretende que ele suporte acesso a banco de dados da WWW. O SEstatNet trabalha com diferentes agentes, que possuem diferentes graus de inteligência e que executam tarefas específicas do processo de ensino-aprendizagem e de análise estatística de dados. Tais agentes são capazes, ainda,

de estabelecer comunicação entre si e de decidirem, de forma autônoma, o momento em que devem agir.

O SEstatNet possibilita ao aluno a execução de procedimentos estatísticos e oferece resultados de natureza numérica, gráfica e de raciocínio, utilizados com o objetivo de estimular a percepção objetiva e subjetiva da realidade pesquisada [CATAPAN et al., 2001]. Os procedimentos estatísticos que são ofertados pelo SEstatNet são de Análise Descritiva e Inferência de Dados.

Uma vez que a concepção pedagógica deste *software* permite que o aluno trabalhe com seus próprios dados, o SEstatNet deverá contar com um ambiente que permita gerenciar as bases de dados.

Em torno dessa idéia surge o objetivo do trabalho que é desenvolver uma metodologia de gerenciamento de base de dados que permita ao usuário alterar suas próprias bases de dados e em tempo real.

Dessa forma foi desenvolvida uma interface gráfica para a apresentação dos dados aos usuários que disponibiliza outros recursos como importação e exportação de dados, operações para definição de novas variáveis e transformações por meio de funções embutidas na linguagem SQL.

Tendo em vista a sua aplicação em procedimentos de análise estatística de dados destacam-se entre essas funções de transformação: a função inversa, exponencial, raiz quadrada, logarítmica e a padronização pela normal.

Considerando ainda que as operações sejam realizadas em tempo real, foi desenvolvido um módulo de cadastro de usuários.

O desafio computacional no desenvolvimento desse ambiente é gerenciar processos, executar diferentes algoritmos, armazenar dados e minimizar a transferência de dados entre cliente-servidor, uma vez que todas as operações devem ser realizadas em tempo real.

1.1 Objetivos

1.1.1 Objetivo Geral

Desenvolver uma metodologia de gerenciamento de base de dados que permita ao usuário alterar suas próprias bases de dados em tempo real usando a Internet como meio de comunicação.

1.1.2 Objetivos Específicos

- ✓ Desenvolver um sistema de controle de acesso para o ambiente, que permita a identificação de cada usuário.
- ✓ Desenvolver uma interface gráfica amigável (“Tabela de Dados”).
- ✓ Disponibilizar recursos de importação e exportação de base de dados.
- ✓ Permitir a criação de novas variáveis e transformações (colunas) na Tabela de Dados.
- ✓ Validar essa metodologia no ambiente de ensino de estatística SEstatNet.

1.2 Estrutura da dissertação

O trabalho está estruturado da seguinte maneira:

O **capítulo 2** trata sobre a conceituação de banco de dados descrevendo as características da arquitetura de um SGBD e de um modelo de dados relacional. Sobre os bancos de dados relacionais enfoca o problema de normalização dos dados e a linguagem SQL de comunicação. O capítulo é finalizado descrevendo algumas características de bancos de dados funcionando na Web.

O **capítulo 3** relata sobre a Educação à Distância apresentando os seus objetivos e características adotadas. Descreve como ela está sendo usada no Brasil e qual a regulamentação existente.

O **capítulo 4** descreve o *software* SEstatNet apresentando o histórico e as características adotadas no seu desenvolvimento. Apresenta também a nova interface do sistema e descreve como o *software* funciona.

No **capítulo 5** é desenvolvida a metodologia para gerenciar uma base de dados na Web. Essa metodologia engloba a definição do diagrama Entidade/Relacionamento e o dicionário de dados. Descreve-se como é feito o controle acadêmico do aluno e apresenta a interface que permite o usuário fazer atualizações em sua base de dados. No processo de ensino-aprendizagem de estatística muitas vezes são requeridas algumas transformações de variáveis. Isso é descrito também nesse capítulo, juntamente com a arquitetura utilizada pela metodologia desenvolvida e alguns mecanismos de segurança que devem ser adotados.

O **capítulo 6** descreve as características do desenvolvimento na implantação dessa metodologia apresentando a lista de eventos do sistema de matrícula juntamente com o

diagrama de contexto. A seguir é apresentado o modelo de comportamento do sistema mostrando algumas das telas desenvolvidas para a comunicação e suas funcionalidades.

O trabalho é finalizado apresentando a conclusão e sugestões para trabalhos futuros.

2 BANCO DE DADOS

Esse capítulo apresenta a conceituação de Banco de Dados e as características dos modelos escolhidos para o desenvolvimento dessa metodologia. A seção 2.2 descreve as características operacionais de um Sistema Gerenciador de Banco de Dados e a seção 2.3 apresenta a arquitetura de um Banco de Dados e define o que é um modelo de dado relacional.

O modelo de dado relacional é baseado no conceito de matrizes e é o que mais se aproxima da forma que devem ter as bases de dados para análise estatística. A seção 2.4 descreve detalhadamente esse modelo.

Quando se fala de Banco de Dados relacional logicamente precisa-se falar também de normalização de tabelas e relacionamentos de dados. Esse assunto pode ser visto na seção 2.5 desse capítulo.

A seção 2.6 descreve as características da linguagem SQL para a manipulação de dados em um BD relacional. Ela foi criada na mesma época em que foi desenvolvido o modelo.

Os módulos para acesso ao Banco de Dados e as linguagens para confecção dos *scripts* utilizados na implementação dessa metodologia são descritos na seção 2.7. O capítulo é finalizado com algumas considerações sobre as novas tecnologias de Banco de Dados que estão sendo desenvolvidas.

2.1 Introdução

Um Banco de Dados é antes de tudo uma coleção logicamente coerente de dados com determinada significação intrínseca, cuja principal finalidade é o armazenamento organizado destas informações, visando a otimização dos sistemas, facilitando entrada de dados, alterações, processamentos e consultas [KORTH, 2001].

Um banco de Dados contém os dados dispostos numa ordem pré-determinada em função de um projeto de sistema, sempre para um propósito muito bem definido. Ele representa aspectos do mundo real. Assim, é uma fonte de onde se pode extrair uma vasta gama de informações derivadas e que possui um nível de interação com esses eventos.

Para a criação e manutenção de bancos de dados utiliza-se o Sistema Gerenciador de Banco de Dados (SGBD). Um SGBD é o sistema responsável por gerenciar a forma como os dados/registros são armazenados de forma física no disco rígido do computador e a forma como esses dados são recuperados.

O conjunto formado por um banco de dados e as aplicações que manipulam o mesmo é chamado de Sistema de Banco de Dados [FANDERUFF, 2002].

Cada usuário do sistema SEstatNet deverá ter uma ou mais bases de dados disponíveis para o seu uso. Uma base de dados apresenta uma estrutura semelhante a uma tabela de um Banco de Dados do tipo relacional, onde as colunas e as linhas representam, respectivamente, as variáveis e as observações.

A seguir serão descritos as características operacionais observadas nos Sistemas Gerenciadores de Banco de Dados.

2.2 Características Operacionais dos SGDB

Nos SGBD observam-se sete características operacionais elementares, as quais são descritas a seguir [KIM, 1995]:

- **Controle de Redundâncias** - A redundância consiste no armazenamento de uma mesma informação em locais diferentes, provocando inconsistências. As informações em um Banco de Dados só se encontram armazenadas em um único local, evitando a duplicação descontrolada dos dados. Quando existem replicações dos dados, estas são decorrentes do processo de armazenagem típica do ambiente Cliente-Servidor, totalmente sob controle do Banco de Dados.
- **Compartilhamento dos Dados** - O SGBD deve incluir *software* de controle de concorrência ao acesso dos dados, garantindo em qualquer tipo de situação a escrita/leitura de dados sem erros.
- **Controle de Acesso** - O SGBD deve dispor de recursos que possibilitem selecionar a autoridade de cada usuário. Assim um usuário poderá realizar qualquer tipo de acesso, enquanto outros poderão ler alguns dados e atualizá-los.

Um terceiro grupo de usuários poderão somente acessar um conjunto restrito de dados para escrita e leitura.

- **Interfaceamento** - Um Banco de Dados deverá disponibilizar formas de acesso gráfico, em linguagem natural, em SQL ou ainda via menus de acesso, não sendo uma "caixa-preta" somente sendo passível de ser acessada por aplicações.
- **Esquematisação** - Um Banco de Dados deverá fornecer mecanismos que possibilitem a compreensão do relacionamento existente entre as tabelas e de sua eventual manutenção.
- **Controle de Integridade** - Um Banco de Dados deverá impedir que aplicações ou acessos pelas interfaces possam comprometer a integridade dos dados.
- **Backups** - O SGBD deverá apresentar facilidade para recuperar falha de *hardware* e *software*, através da existência de arquivos de "pré-imagem" ou de outros recursos automáticos, exigindo minimamente a intervenção de pessoal técnico.

Existe a possibilidade de encontrar Banco de Dados que não satisfaz completamente todas as características acima, o que não o invalida como Banco de Dados.

2.3 Arquitetura de um SGBD

Quanto à estrutura de um Banco de Dados pode-se dizer que tem um nível interno, onde é descrita a estrutura física de armazenamento dos dados; um nível intermediário, onde tem a descrição lógica dos dados e um nível externo onde são descritas as visões para grupos de usuários.

Não se pode deixar de lembrar ainda que o Banco de Dados garanta a independência lógica e física dos dados, portanto pode-se alterar o esquema conceitual dos dados, sem alterar as visões dos usuários ou mesmo, alterar o esquema interno sem, contudo alterar seu esquema conceitual [ELMASRI e NAVATHE, 2002].

O modelo de dados é basicamente um conjunto de conceitos utilizados para descrever um Banco de Dados. Equivale à visão conceitual de dados em um SGBD. Todo SGBD deve suportar um modelo que permita uma representação dos dados de uma realidade. O esquema conceitual de uma aplicação é o resultado da adequação dos requisitos de dados dessa aplicação ao modelo de dados do SGBD [KORTH e SILBERSCHATZ, 2001] [HEUSER, 1998].

Não existe uma única forma de representação deste modelo, porém qualquer forma que permita a correta compreensão das estruturas de dados compreendidas no Banco de Dados pode ser considerada adequada.

O modelo relacional é o modelo mais utilizado no momento. Grandes empresas e instituições utilizam gigantes bases relacionais para gerenciar e manter seus dados [ABBEY et al., 2002] [RAMALHO, 1999] [SCHERER et al., 2000].

A seguir será descrito o modelo de dados relacional, pois foi o escolhido para o desenvolvimento desse trabalho. Esse modelo de dados é o que mais se aproxima das características que deve ter a base de dados para análise estatística.

2.4 Modelo de Dados Relacional

O modelo relacional foi formalmente definido por Ted Codd, no Laboratório da IBM em *San Jose* - Califórnia, em 1970. O projeto inicial foi denominado de Sistema R e definia a organização dos dados e linguagens formais para a sua manipulação. Com base nestas linguagens formais, a primeira versão da linguagem SQL (*Structured Query Language*) foi definida. Esta linguagem é, atualmente, um padrão para gerenciamento de dados em SGBDs relacionais. Os SGBDs de grande porte mais famosos são: Oracle, SQL Server, Informix, Sybase e Ingres [FANDERUFF, 2000] [KORTH e SILBERSCHATZ, 2001].

O modelo de dados relacional representa os dados contidos em um Banco de Dados através de “entidade” ou “relação”, que pode ser definida como um objeto do mundo real, concreto ou abstrato e que possui existência independente. Uma entidade equivale ao conceito matemático de conjunto, ou seja, um agrupamento de elementos [SCHERER et al., 2000].

Estas relações contêm informações sobre as entidades representadas e seus relacionamentos. O modelo relacional é claramente baseado no conceito de matrizes,

onde as chamadas linhas (das matrizes) seriam os registros e as colunas (das matrizes) seriam os campos.

Os nomes das tabelas e dos campos são de fundamental importância para a compreensão entre o que está sendo armazenado, onde está armazenado e qual a relação existente entre os dados armazenados.

Cada linha dessa relação será chamada de **tupla** e cada coluna dessa relação será chamada de **atributo**. O conjunto de valores possíveis de serem assumidos por um atributo, será intitulado de **domínio**.

O esquema de uma relação, nada mais é que os campos (colunas) existentes em uma tabela. Já a instância da relação consiste no conjunto de valores que cada atributo assume em um determinado instante. Portanto, os dados armazenados no Banco de Dados, são formados pelas instâncias das relações.

As relações não podem ser duplicadas. No que concerne ao seu tratamento, a ordem de entrada de dados no Banco de Dados não deverá ter qualquer importância para as relações.

Os atributos deverão ser atômicos, isto é, não são passíveis de novas divisões. Chama-se de **chave primária** ao atributo que define um registro, dentre uma coleção de registros. Chaves secundárias, terciárias, etc., são chaves que possibilitam pesquisas ou ordenações alternativas, ou seja, diferentes da ordem criada a partir da chave primária ou da ordenação natural (física) da tabela. Chama-se de **chave composta**, a chave que contém mais de um atributo e **chave estrangeira**, a chave que permite a ligação lógica entre uma tabela (onde ela se encontra) com outra na qual ela é chave primária.

Um dos fatores mais importantes no desenvolvimento de páginas dinâmicas é a definição do banco de dados. Um banco de dados relacional visa manter os dados de forma não redundante (repetição de vários campos em várias tabelas), executar processamento integrado, lidar com relações múltiplas (relacionamentos) e fornecer certo grau de independência dos dados. Diante de um processamento integrado, o banco de dados torna-se simplificado causando melhoria na independência dos dados e, conseqüentemente, obtendo maior desempenho [FANDERUFF, 2000].

Se as tabelas não estão definidas apropriadamente, isso pode dificultar quando for necessária a execução de chamadas SQL no código para extrair os dados desejados. Ao

entender o relacionamento entre os dados e a sua normalização, o programador estará mais bem preparado para começar a desenvolver suas aplicações.

2.5 Técnicas de Normalização

A normalização de dados pode ser vista como um processo para analisar os supostos esquemas de relações com base em suas dependências funcionais e chaves primárias para alcançar as propriedades desejáveis de diminuir redundância e reduzir ao mínimo as anomalias de inclusão, exclusão e atualização [ELMASRI e NAVATHE, 2002].

Quando se trabalha com um SGBD, deve-se conhecer os métodos de normalização do esquema de tabelas no sistema de banco de dados relacional. Eles ajudam a entender o código muito mais facilmente, permitem expandir rapidamente e, em alguns casos, aceleram a sua aplicação.

Basicamente as regras de normalização são reforçadas pela eliminação de redundâncias e dependências inconsistentes no projeto de especificação das tabelas. Explica-se o que isso significa examinando os cinco passos progressivos de normalização de que se deve estar ciente para criar banco de dados funcionais e eficientes. Também se detalham os tipos de relacionamentos que uma estrutura de dados pode utilizar.

Deseja-se criar uma tabela com informações de usuários, onde se armazena o Nome, Curso, E-mail, Departamento, e alguns *Bookmarks* ou *Urls*. Uma estrutura básica com esses dados pode ser vista na Tabela 2.1, onde os dados estão dispostos sem nenhuma preocupação com a normalização da tabela.

Tabela 2.1: Tabela de dados fictícios na Forma Zero. Exemplo de aplicação.

usuários						
usrId	nome	curso	email	depto	url1	url2
1	Carlos	ABC	carlos.inf.ufsc.br	INE	abc.com	xyz.com
2	Flávia	XYZ	flavia.eel.ufsc.br	EEL	abc.com	xyz.com

Esta tabela é dita estar na Forma Zero porque nenhuma das regras de normalização foi aplicada ainda. Note os campos url1 e url2. O que fazer se a aplicação

precisar adicionar mais uma terceira *url*? Quer continuar adicionando colunas em sua tabela e codificar mais este campo *input* no código? Óbvio que não, porque se teria um sistema funcional que poderia crescer com as novas exigências de desenvolvimento.

A solução para esse impasse é aplicar as regras de normalização nessa tabela de dados. As três primeiras formas normais enunciadas em [CODD, 1972], são baseadas nas dependências funcionais entre os atributos de uma relação.

2.5.1 Primeira Forma Normal

A seguir são listadas as regras da Primeira Forma Normal e depois será aplicada nas informações de usuários contidas na Tabela 2.1.

1. *Eliminar grupos repetidos em tabelas individuais.*
2. *Criar tabelas separadas para cada conjunto de dados relacionados.*
3. *Identificar cada conjunto de dados relacionados com uma chave primária.*

A primeira regra está sendo quebrada ao se repetir os campos *url1* e *url2*. E quanto a regra três, chaves primárias? A regra três significa basicamente que se quer definir um tipo de valor inteiro único e auto-incrementado (que aumenta automaticamente) em cada linha (registro). Caso contrário, o que aconteceria se houvesse dois usuários com nomes Flávia e se quer separar um do outro?

Ao se aplicar as regras da Primeira Forma Normal obtém-se a Tabela 2.2 seguinte:

Tabela 2.2: Tabela de dados fictícios na Primeira Forma Normal. Exemplo de aplicação.

usuários					
usrId	nome	curso	email	depto	url
1	Carlos	ABC	carlos.inf.ufsc.br	INE	abc.com
1	Carlos	ABC	carlos.inf.ufsc.br	INE	xyz.com
2	Flávia	XYZ	flavia.eel.ufsc.br	EEL	abc.com
2	Flávia	XYZ	flavia.eel.ufsc.br	EEL	xyz.com

Agora essa tabela é dita estar na Primeira Forma Normal. O problema da limitação de campos de *url* foi solucionado, mas observe as complicações resultantes. Cada vez que se coloca uma nova linha na tabela usuários ocorre a duplicação de dados

do usuário e do curso. O banco de dados não vai apenas ficar muito maior do que é esperado, mas pode facilmente corromper os dados misturando algumas dessas informações redundantes.

2.5.2 Segunda Forma Normal

Deve-se aplicar então as regras da Segunda Forma Normal:

1. *Criar tabelas separadas para conjuntos de valores que se aplicam a registros múltiplos.*
2. *Relacionar estas tabelas com chaves estrangeiras.*

Os valores de *url* devem ser extraídos para uma tabela separada para que se possa adicionar mais *urls* no futuro sem ter que duplicar os dados. Usa-se também um valor de chave primária para relacionar estes campos:

Tabela 2.3: Tabelas de dados fictícios após a aplicação da Segunda Forma Normal.

usuários				
usrId	nome	curso	email	depto
1	Carlos	ABC	carlos.inf.ufsc.br	INE
2	Flávia	XYZ	flavia.eel.ufsc.br	EEL

urls		
urlId	relusrId	url
1	1	abc.com
2	1	xyz.com
3	2	abc.com
4	2	xyz.com

Assim criam-se tabelas separadas e a chave primária na tabela usuários, *usrId*, agora está relacionada à chave estrangeira, *relusrId*, na tabela *urls*. Assim os dados estão numa forma muito melhor agora. Mas o que acontece ao adicionar outro aluno no curso ABC? Ou 200 alunos? Têm-se agora nomes de cursos e de departamentos duplicando-se por todos os lugares. É uma situação perfeita para a duplicação de erros. Logo deve ser aplicada a Terceira Forma Normal.

2.5.3 Terceira Forma Normal

A terceira forma normal é resumida em:

1. *Eliminar campos que não dependem da chave.*

O departamento que oferece o curso e o próprio curso não têm nada a ver com a identificação do usuário. Então eles teriam que ter seus próprios identificadores.

Tabela 2.4: Tabelas de dados fictícios após a aplicação da Terceira Forma Normal.

usuários			
usrId	nome	email	relcurId
1	Carlos	carlos.inf.ufsc.br	1
2	Flávia	flavia.eel.ufsc.br	2

cursos		
curId	curso	depto
1	ABC	INE
2	XYZ	EEL

urls		
urlId	relusrId	url
1	1	abc.com
2	1	xyz.com
3	2	abc.com
4	2	xyz.com

Assim a chave primária curId na tabela de cursos está relacionada à chave estrangeira na tabela de usuários chamada relcurId, e pode-se adicionar 200 usuários enquanto só é inserido o nome "ABC" uma vez. As tabelas de usuários e urls podem crescer o quão grande quiserem sem necessariamente duplicar ou corromper os dados.

A maioria dos desenvolvedores diria que a Terceira Forma Normal já é suficiente, e esse esquema de dados poderia facilmente manipular dados de todos os cursos, e na maioria dos casos eles estariam corretos. Mas observe com atenção o campo de url. Está

ocorrendo a duplicação de dados. Isso é perfeitamente aceitável se esses campos não estivessem sendo pré-definidos.

Se o *input* da página HTML permite que os usuários insiram livremente o seu texto não há nada que se possa fazer a respeito disso, e é apenas uma coincidência que Carlos e Flávia ambos puseram o mesmo *bookmark*. Mas e se tem um menu *drop-down* onde se sabe que somente estas duas urls são permitidas, ou talvez 20 ou mais? Pode-se elevar o nosso esquema de banco de dados para o próximo nível, ou seja, a Quarta Forma Normal.

Muitos desenvolvedores ignoram esta forma porque ela depende de um tipo muito específico de relacionamento, o relacionamento muitos-para-muitos, o qual não foi visto ainda neste exemplo de aplicação.

2.5.4 Relacionamento de Dados

Existem três tipos básicos de relacionamento: um-para-um, um-para-muitos e muitos-para-muitos.

Observe a tabela 2.2 de usuários na primeira forma normal. Por um instante imagina-se que o campo de url foi colocado em uma tabela separada, e toda vez que se insere um registro na tabela de usuários também será inserida uma linha na tabela de urls. Tem-se então um relacionamento um-para-um: cada linha na tabela de usuários teria exatamente uma linha correspondente na tabela de urls. Para os propósitos dessa aplicação fictícia isso não seria nem útil e nem normalizado.

Agora observe as tabelas 2.3 no exemplo da segunda forma normal. Elas permitem a um usuário ter muitas urls associadas a este registro de usuário. Esta é uma relação uma-para-muitos, o tipo mais comum, e até que se alcançasse o dilema apresentado na terceira forma normal, o único tipo que se iria precisar.

O relacionamento muitos-para-muitos, por outro lado, é um pouco mais complexo. Note que no exemplo da terceira forma normal tem-se um usuário relacionado a muitas urls. Como mencionado, deseja-se mudar esta estrutura para permitir que muitos usuários estejam relacionados com muitas urls, dessa forma deve ser aplicado um relacionamento muitos-para-muitos. Observe o que isso faria à estrutura das tabelas.

Tabela 2.5: Alteração nas estruturas das Tabelas do exemplo fictício para o relacionamento muito-para-muitos.

usuarios			
usrId	nome	email	relcurId
1	Carlos	carlos.inf.ufsc.br	1
2	Flávia	flavia.eel.ufsc.br	2

cursos		
curId	curso	depto
1	ABC	INE
2	XYZ	EEL

urls	
urlId	url
1	abc.com
2	xyz.com

url_relacoes		
relacaoid	relacionaurId	relacionausrId
1	1	1
2	1	2
3	2	1
4	2	2

Para diminuir a duplicação de dados e avançar no processo que leva à quarta forma de normalização, cria-se uma tabela somente com chaves primárias e estrangeiras em url_relacoes. Dessa forma as entradas duplicadas na tabela urls foram removidas criando-se a tabela url_relacoes. Agora se pode expressar com exatidão o relacionamento que ambos Carlos e Flávia estão relacionados a cada um, e a ambas urls. Então o que diz exatamente a quarta forma de normalização?

2.5.5 Quarta Forma Normal

As quarta e a quinta formas normais são baseadas respectivamente nos conceitos de dependência multivalorada e de dependência de junção. Muitas das formas normais avançadas são atribuídas a Fagin [FAGIN, 1977] [FAGIN, 1979].

A quarta forma normal pode ser enunciada como:

1. *Num relacionamento muitos-para-muitos, entidades independentes não podem ser armazenadas na mesma tabela.*

Já que isso só se aplica aos relacionamentos do tipo muitos-para-muitos, os desenvolvedores podem corretamente ignorar esta regra. Mas ela torna-se útil em certas situações, como esta apresentada como exemplo de aplicação. Modifica-se com sucesso a tabela de urls para eliminar redundâncias e move-se os relacionamentos em sua própria tabela.

Só para dar um exemplo prático, agora se pode selecionar todas as *urls* de Carlos executando uma chamada SQL como a mostrada na figura 2.1.

```
SELECT
  nome,
  url
FROM
  usuarios,
  urls,
  url_relacoes
WHERE
  url_relacoes.relacionausrld = 1 AND
  usuarios.usrld = 1 AND
  urls.urlld = url_relacoes.relacionaurld
```

Figura 2.1: Comando SQL para selecionar o usuário Carlos e suas Urls do exemplo fictício.

Os comandos SQL mostrados na figura 2.2 são usados para selecionar informações de usuário e *url* de todos.


```
SELECT
nome,
url
FROM
usuarios,
urls,
url_relacoes
WHERE
usuarios.usrld = url_relacoes.relacionausrld AND
urls.urlld = url_relacoes.relacionaurld
```

Figura 2.2: Comando SQL para selecionar informações de usuário e url de todos os alunos do exemplo fictício.

2.5.6 Quinta Forma Normal

Esta é uma forma de normalização que às vezes é aplicada, mas em muitos dos casos não é exigida para se ter a maior funcionalidade a partir da estrutura de dados ou aplicação. Ela sugere que:

1. A tabela original deve ser reconstruída a partir das tabelas que foram geradas por ela.

A vantagem de aplicar esta regra garante que não se criou nenhuma coluna estranha nas tabelas, e que todas as estruturas das tabelas criadas só são tão grandes quanto elas necessitam ser. É de boa prática aplicar esta regra, mas provavelmente não vai precisar dela se o esquema de dados utilizado for muito grande.

2.6 SQL (Structured Query Language)

Quando os bancos de dados relacionais estavam sendo desenvolvidos, foram criadas linguagens destinadas à sua manipulação. O Departamento de Pesquisas da IBM desenvolveu a SQL como forma de interface para o sistema de BD relacional denominado SYSTEM R, no início dos anos 70. Em 1986, o *American National Standard Institute* (ANSI) publicou um padrão SQL. Então, a SQL estabeleceu-se como linguagem padrão de banco de dados relacional [FANDERUFF, 2000] [KORTH e SILBERSCHATZ, 2001] [ABBEY et al., 2002].

A SQL apresenta uma série de comandos que permitem a definição dos dados, chamada de DDL (*Data Definition Language*), composta entre outros pelos comandos *Create*, que é destinado à criação do Banco de Dados, das tabelas que o compõem, além

das relações existentes entre as tabelas. Como exemplos de comandos da classe DDL têm os comandos *Create*, *Alter* e *Drop*.

Os comandos da série DML (*Data Manipulation Language*), são destinados a consulta, inserções, exclusões e alterações em um ou mais registros de uma ou mais tabelas de maneira simultânea. Como exemplos de comandos da classe DML têm os comandos *Select*, *Insert*, *Update* e *Delete*. Uma subclasse de comandos DML, a DCL (*Data Control Language*), dispõe de comandos de controle como *Grant* e *Revoke*.

A Linguagem SQL tem como grande virtude sua capacidade de gerenciar índices, sem a necessidade de controle individualizado de índice corrente, algo muito comum nas linguagens de manipulação de dados do tipo registro a registro. Outra característica muito importante disponível em SQL é sua capacidade de construção de visões, que são formas de se visualizar os dados no formato de listagens independentes das tabelas e organização lógica dos dados [SCHERER et al., 2000].

Outra característica interessante na linguagem SQL é a capacidade de cancelar uma série de atualizações ou de gravar, depois de iniciada uma seqüência de atualizações. Os comandos *Commit* e *Rollback* são responsáveis por estas facilidades.

Nota-se que a linguagem SQL consegue implementar estas soluções, somente pelo fato de estar baseada em Banco de Dados, que garantem por si mesmo a integridade das relações existentes entre as tabelas e seus índices [RAMALHO, 1999].

2.7 Banco de Dados e a WEB

A World Wide Web (WWW ou Web) é um sistema de informação distribuído baseado em hipertextos. Os documentos armazenados na Web podem ser de vários tipos. Os tipos mais comuns são documentos hipertextos formatados de acordo com a *HyperText Markup Language* (HTML), que por sua vez é baseada na *Standard Generalized Markup Language* (SGML). Os textos HTML contêm textos, especificações de fontes e outras instruções de formatação e links. Estes documentos podem ser visualizados com um navegador Web adequado [MORISSEAU-LEROY et al., 2001].

Atualmente, os navegadores Web dão suporte a uma linguagem de programação chamada Java, que permite que os documentos contenham programas executados no

local do usuário, fazendo com que os documentos se tornem ativos ao invés de passivos [ABBEY et al., 2002].

A figura 2.3 apresenta o diagrama de fluxo de informação quando uma página da Web faz acesso a um banco de dados [FANDERUFF, 2002].

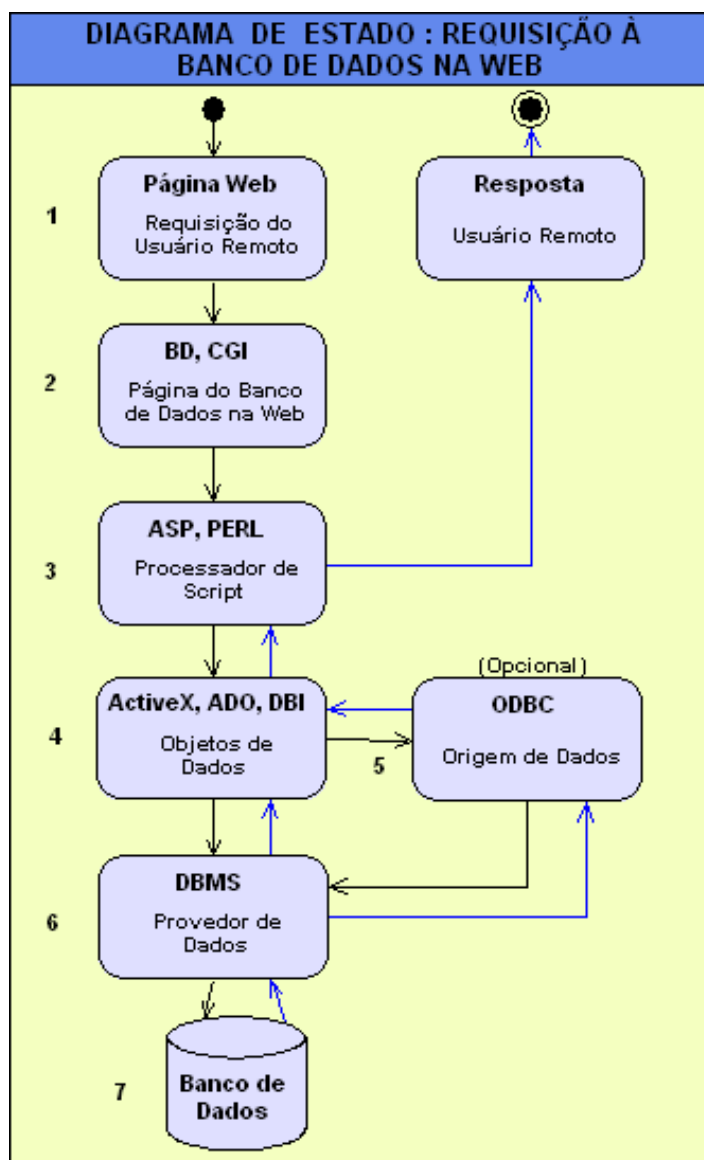


Figura 2.3: Fluxo de Informação para acesso de dados pela Web.

Com o crescimento do comércio eletrônico na Web, os bancos de dados utilizados para processamento de transações precisam estar ligados a Web. Um programa associado a um formulário pode ser executado disparando transações no banco de dados da máquina servidora. Em contrapartida, os resultados da transação podem ser

formatados em HTML e apresentados ao usuário. Pode-se, também, definir documentos HTML em linguagens de macro com consultas SQL embutidas [MORISSEAU-LEROY et al., 2001] [ABBEY et al., 2002].

O protocolo HTTP e a linguagem de marcação HTML têm fornecido estruturas simples e poderosas para transporte e apresentação de dados, o que é a base das interações baseadas na Web.

A figura 2.3 além de mostrar o fluxo que a informação faz ao acessar dados pela Web destaca os ambientes de desenvolvimento da Microsoft e o utilizado no desenvolvimento dessa metodologia.

As páginas de acesso a banco de dados na Web trabalham com CGI (*Common Gateway Interface*). CGI é um protocolo de comunicação (um código) através do qual o servidor HTTP (ou servidor Web) intermedia a transferência de informações entre um programa (no mesmo computador que o servidor) e um cliente HTTP (o seu browser).

A linguagem utilizada no desenvolvimento dos *scripts* é o Perl (*Practical Extraction and Report Language*) cujo principal objetivo é facilitar a extração de dados e emitir relatórios. O Perl é uma linguagem “*interpil*”; isto significa que um programa em Perl é primeiro compilado e em seguida interpretado (não gera código executável) [CPAN, 2004].

O DBI (*Database Interface*) é uma interface genérica para muitos bancos de dados. Isto é, pode-se escrever um script que funciona com diferentes mecanismos de banco de dados sem nenhuma mudança. Para isso é preciso de um *Driver* de Banco de Dados (DBD) definido para cada tipo de banco de dados. Para usar o MySQL este *driver* é chamado DBD::mysql.

Nesse item foi descrito as principais ferramentas de desenvolvimento utilizadas para a implementação da metodologia proposta e a seguir são feitas algumas considerações sobre as novas tecnologias de banco de dados.

2.8 Considerações finais

Os bancos de dados têm sido as ferramentas computacionais mais utilizadas para o armazenamento e a manipulação de grandes quantidades de dados. Para que esta funcionalidade seja alcançada com maior eficiência e primazia, foram desenvolvidas

inúmeras técnicas para armazenamento, indexação e busca mais eficaz, e com maior riqueza de informações, dos dados.

Um meio eficaz de gerenciamento de informações somente é realizado através de um Banco de Dados. A organização de dados passou por um processo de evolução, tendo a modelagem relacional se fixado como um padrão mundial. As tecnologias de banco de dados orientado a objetos, objeto-relacional e a comunicação com bancos de dados através da Web surgiram para atender as novas necessidades de informação.

Nos últimos anos houve um enorme crescimento dos sistemas de informação distribuídos baseados na Internet. As ferramentas que integram a HTML com a SQL facilitam a criação de *front-ends* HTML para os bancos de dados [KORTH e SILBERSCHATZ, 2001]. Grandes empresas, fabricantes de bancos de dados, como a Oracle e a Microsoft, têm aproveitado este potencial para a criação de ferramentas que permitam e facilitem o acesso aos dados na Web.

3 EDUCAÇÃO À DISTÂNCIA

O desenvolvimento da metodologia de gerenciamento de base de dados na web objetiva fundamentalmente o ensino à distância e nesse capítulo ele será analisado sob vários aspectos. Na seção 3.1 é feita uma comparação com o ensino presencial e descrito onde a EaD é usada.

Na seção 3.2 são apresentadas conceituações atuais de educação à distância e os recursos da internet utilizados como suporte à comunicação. A seção 3.3 descreve os objetivos da EaD e a seção 3.4 descreve suas características.

Na seção 3.5 é visto como a EaD está sendo utilizada no Brasil e a sua política de uso. Evidencia também os custos advindos com o seu uso e perspectivas de expansão. As leis que regulamentam a educação à distância no Brasil são apresentadas na seção 3.6. O capítulo é finalizado apresentando as modalidades de ensino e as possibilidades educacionais que surgem com o uso dessas novas tecnologias.

3.1 Introdução

Conceitualmente, sabe-se que educação à distância (EaD) possui características próprias quando comparado ao ensino tradicional. Em alguns aspectos encontram-se posições antagônicas. Por exemplo, as mais evidentes são que na educação à distância o espaço físico e o tempo são totalmente flexíveis, ou seja, o aluno não precisa se deslocar até um determinado local num certo horário para adquirir determinados conceitos. Com a utilização das técnicas de EaD ele pode da sua casa mesmo, caso possua recursos computacionais, adquirir os mesmos conceitos que estavam disponíveis na forma tradicional. Isto, levando-se em conta que ele pode fazê-lo no momento que mais lhe convier. Estas características são fundamentais para tentar levar determinados cursos a regiões as quais dificilmente o ensino tradicional poderia suprir, quer seja por deficiência econômica ou por desinteresse político [SOCINFO, 2000].

É importante observar que a educação à distância não pode ser vista como substitutiva da educação convencional ou presencial. São duas modalidades do mesmo processo. A educação à distância não concorre com a educação convencional, tendo em vista que não é este o seu objetivo, nem poderá ser.

Se a educação à distância apresenta como característica básica a separação física e, principalmente, temporal entre os processos de ensino e aprendizagem, isto significa não somente uma qualidade específica dessa modalidade, mas, essencialmente, um desafio a ser vencido. Isso promove de forma combinada, o avanço na utilização de processos industrializados e cooperativos na produção de materiais com a conquista de novos espaços de socialização do processo educativo [NUNES, 1992].

Esta modalidade de ensino-aprendizagem não pode ser encarada como uma panacéia para todos os males da educação brasileira. Há um esforço muito grande dos educadores e pesquisadores da educação em mostrar que os problemas da educação brasileira não se concentram somente no interior do sistema educacional, mas, antes de tudo, refletem uma situação de desigualdade e polaridade social [NUNES, 1992]. "Certamente que a educação, nas suas mais diversas modalidades, não tem condições de sanear nossos múltiplos problemas nem satisfazer nossas mais variadas necessidades. Ela não salva a sociedade, porém, ao lado de outras instâncias sociais, ela tem um papel fundamental no processo de distanciamento da incultura, da acriticidade e na construção de um processo civilizatório mais digno do que este que vivemos" [LUCHESE, 1989].

Nesse sentido, a educação à distância pode contribuir de forma significativa para o desenvolvimento educacional de um país, notadamente de uma sociedade com as características brasileiras, onde o sistema educacional não consegue desenvolver as múltiplas ações que a cidadania requer.

Do ponto de vista tecnológico, a presença da informática nos processos de capacitação tem gerado grandes avanços nos procedimentos de treinamento à distância ou treinamento independente com ajuda do computador. Caso notório são os procedimentos adotados pelas grandes companhias aéreas e setores das Forças Armadas, com a utilização de simuladores e bancos de dados interativos [NUNES, 1992].

É cada vez maior o número de empresas que descobrem as vantagens do treinamento à distância para a capacitação e atualização de seus funcionários, não somente por conta da redução dos custos, mas principalmente pela possibilidade de envolver um grande número de pessoas ao mesmo tempo e em regiões distantes [RUMBLE e OLIVEIRA, 1992].

3.2 Conceituação

Nas atuais condições de globalização, onde com o desenvolvimento das tecnologias de comunicação, o universal invade o local e vice-versa, modifica a nossa relação com o mundo e com o conhecimento. Segundo *Serres* "antes esta relação era local-local; agora ela é global-local" [SERRES, 1984]. Mudam-se todas as formas de consciência do homem com relação ao mundo como realidade; conseqüentemente emergem novas formas de aprender e de construir o conhecimento a partir de novas competências relacionadas ao trato com a informação.

O ato pedagógico pleno consiste de quatro etapas de aprendizagem que evidenciam diferentes níveis de conhecimento: definição, conceituação, generalização e síntese. Este processo compreende alguns agentes que atuam no sentido da consecução do ato: educador, aluno e os recursos pedagógicos que podem ser utilizados [CATAPAN, 1997] [CATAPAN, 2001].

Tudo isso exige uma transformação nos fundamentos e metodologias de aprendizagem, possibilitando o surgimento de diferentes modalidades de aprender e ensinar.

Os recursos da internet utilizados como suporte à comunicação, tais como as salas de bate-papo, mural eletrônico, quadro compartilhado, fórum, áudio e videoconferência são exemplos de mecanismos disponíveis que permitem ampliar a interação e comunicação em atividades de EaD. Esses recursos ganham relevância na medida em que uma nova maneira de produzir conhecimento vem se instalando com o computador, veiculando a possibilidade de se aprender, fazendo.

No entender de *José Luís García Llamas* "a Educação à Distância é uma estratégia educativa baseada na aplicação da tecnologia à aprendizagem, sem limitação do lugar, tempo, ocupação ou idade dos alunos. Implicam novos papéis para os alunos e para os professores, novas atitudes e novos enfoques metodológicos" [VIANA, 1999].

Segundo *Miguel A. Ramón Martínez*, a Educação à Distância é uma estratégia para operacionalizar os princípios e os fins da educação permanente e aberta, de tal maneira que qualquer pessoa, independentemente do tempo e do espaço, possa converter-se em sujeito protagonista de sua própria aprendizagem, graças ao uso

sistemático de materiais educativos, reforçados por diferentes meios e formas de comunicação.

No que diz respeito às características da EaD, *Anthony Rumble* observa que, a característica geral mais importante da educação à distância é que ela se baseia na comunicação não direta. As conseqüências que esta característica geral da educação à distância traz consigo, podem ser agrupadas em cinco categorias [VIANA, 1999]:

- Democratizar o acesso à educação.
- Propiciar uma aprendizagem autônoma e ligada à experiência.
- Promover um ensino-aprendizagem inovador e de qualidade.
- Incentivar a educação permanente.
- Reduzir os custos.

3.3 A Educação à Distância no Brasil

Durante várias décadas a educação à distância foi utilizada e distribuída pelos mais variados meios como um recurso de ensino capaz de romper as barreiras físicas entre aluno e professor, entretanto, nenhum fato passado provocou tantas transformações sobre esse sistema de ensino quanto a intensificação das tecnologias de comunicação e o enorme aumento do número de usuários da Internet ocorridos nos últimos anos. Esses avanços tecnológicos expandiram as fronteiras da educação à distância e transformaram-na em uma nova dimensão de formação diferenciada do ensino presencial, que passou a conquistar espaço nos meios acadêmicos e empresariais como um tema de caráter estratégico para a democratização do conhecimento e para a consolidação da Sociedade da Informação no Brasil [SOCINFO, 2000].

No Livro Verde, onde o Programa SocInfo está publicado e o Livro Branco que complementa o trabalho de Ciência, Tecnologia e Inovação, considera a educação um elemento-chave na construção de uma sociedade baseada na informação, no conhecimento e no aprendizado. Em relação à EaD, ele informa que a disseminação da Internet tem feito ressurgir um novo interesse, onde a educação à distância pode ser vista como mecanismo complementar, substitutivo ou integrante do ensino presencial [SOCINFO, 2002].

O livro alerta, no entanto, ser necessário o desenvolvimento de metodologias pedagógicas eficientes para o novo meio e de ferramentas adequadas para o estudo

individual ou em grupo. E ressalta ainda que algumas constatações podem ser feitas como resposta à estrutura formal de ensino, e que alguns elementos são fundamentais para introduzir-se uma mudança. A saber: a alfabetização digital, a geração de novos conhecimentos, a aplicação de tecnologias de informação e comunicação e a aplicação destas tecnologias em quaisquer outras áreas.

Percebe-se, portanto, a necessidade de um reposicionamento, onde também é necessária a introdução de novos currículos, ressaltando os Centros Federais de Educação Tecnológica (CEFET) e o SENAC em suas funções de formação técnica e profissional, as graduações e as licenciaturas nas faculdades e universidades e, finalmente, a pós-graduação em tecnologias de informação e comunicação [SOCINFO, 2000].

O Livro Verde recomenda diferentes ações estruturadas que devem ser realizadas. Podemos citar, entre elas, a articulação entre empresas do setor público e empresas privadas; a ampliação do ProInfo; a geração e a difusão de material didático livre; a construção e a distribuição de pacotes tecnológicos de baixo custo; a ampliação da capacidade de formação de recursos humanos; o suporte à graduação e pós-graduação, envolvendo a formação do docente, pesquisadores, entre outros; a experimentação de um novo modelo, integrando vídeo e Internet e, finalmente, o fomento ao desenvolvimento de metodologias de ensino baseadas em tecnologias de informação e comunicação, inclusive leitura e produção de informação no novo meio [SOCINFO, 2000].

É interessante ressaltar que antes de 1984, a produção brasileira de informática já se colocava entre a dos países que mais cresciam no cenário mundial – entre os 10 maiores do mundo – e 60% da indústria nacional trabalhava com equipamentos desenvolvidos no próprio país. Entre 1984 e 1987, o Brasil apresentava maior taxa de crescimento mundial nesta área e, em 1987, tornou-se o sexto maior mercado de microcomputadores, superando a Itália e a Suécia [TAJRA, 1998].

O interesse que o governo brasileiro tinha na reserva de mercado era desenvolver uma política com características independentes, porque é notório que a detenção de conhecimento das áreas tecnológicas, é determinante para o domínio do poder. Quem detém conhecimento, detém poder. Quem detém conhecimento tecnológico, detém ainda mais poder.

Segundo dados divulgados pela UNESCO, até o ano de 2002 o Brasil contava com 60 cursos de graduação e pós-graduação à distância oferecidos, por instituições públicas e privadas, para 84.713 alunos – número que representa apenas 2,56% dos 3,3 milhões de universitários existentes hoje no país. Desse contingente de alunos inscritos em cursos superiores à distância, metade encontrava-se matriculada em cursos de pedagogia e utilizava a Internet como principal recurso tecnológico para o acesso ao material didático disponibilizado. Tais informações comprovam que, apesar dos últimos progressos obtidos na difusão da educação à distância no ensino superior, os cursos à distância no país encontram-se ainda mais disseminados no meio corporativo, onde as instituições costumam oferecer a seus funcionários obtendo bons retornos financeiros.

Entretanto, a empresa de análises e estratégia de mercado IDC (*International Data Corporation*) prevê que o mercado global de educação à distância aumentará cada vez mais nos próximos anos, ultrapassando a marca de U\$ 6,6 bilhões em rendimentos alcançado em 2002 e chegando aos U\$ 23,7 bilhões em 2003. No Brasil, os investimentos no mercado de educação à distância já movimentaram U\$ 80 milhões somente no ano de 2002 e as perspectivas para 2003 apontam uma movimentação de U\$ 100 milhões.

3.4 Regulamentação do EaD no Brasil

As bases legais da educação à distância no Brasil foram estabelecidas pelas seguintes leis e portarias [VIRTUAL, 2004]:

LEI Nº 9.394, DE 20 DE DEZEMBRO DE 1996

Lei de Diretrizes e Bases da Educação Nacional

DECRETO Nº 2.494, DE 10 DE FEVEREIRO DE 1998

Decreto que regulamenta o Art. 80 do LDB nº 939/96 e normatiza a Educação à Distância no Brasil.

PORTARIA Nº 2.253, DE 18 DE OUTUBRO DE 2001

Portaria que normatiza a oferta de disciplinas na modalidade de educação à distância em instituições de ensino superior, com base no art. 81 da Lei nº 9.394, de 1.996.

3.5 Considerações finais

Hoje existem três modalidades de educação:

- ✓ Educação presencial: é a dos cursos regulares, em qualquer nível, onde professores e alunos se encontram sempre num local físico, chamado sala de aula. É o ensino convencional.
- ✓ Semipresencial (parte presencial e parte virtual ou à distância): acontece em parte na sala de aula e outra parte à distância, através do uso de tecnologias.
- ✓ Educação à distância (ou virtual): pode ter ou não momentos presenciais, mas acontece fundamentalmente com professores e alunos separados fisicamente no espaço e/ou no tempo, mas podendo estar presentes através das tecnologias de comunicação.

As tecnologias interativas, sobretudo, vêm evidenciando na educação à distância o que deveria ser o cerne de qualquer processo de educação: a interação e a interlocução entre todos os que estão envolvidos nesse processo.

As possibilidades educacionais que se abrem com a educação à distância são fantásticas. Com o alargamento da banda de transmissão, como acontece na TV a cabo e Internet de banda larga, torna-se mais fácil ver e ouvir à distância. Muitos cursos poderão ser realizados à distância com áudio e vídeo, principalmente na educação continuada.

As possibilidades de interação serão diretamente proporcionais ao número de pessoas envolvidas. As aulas à distância apresentam possibilidade de interação em tempo real (ao vivo) e aulas presenciais terão interação à distância.

4 SEstatNet

Este capítulo apresenta uma descrição da nova versão do *software* SEstatNet, onde a metodologia em desenvolvimento precisa ser validada. A seção 4.1 apresenta vários ambientes de ensino virtuais, atualmente disponíveis no Brasil, para oferecer cursos à distância.

O SEstatNet visa ser um desses ambientes, mas no seu pequeno tempo de vida, aproximadamente 5 anos, já teve várias versões, as quais são descritas na seção 4.2. A seção 4.3 apresenta todas as características implementadas no SEstatNet na sua última versão. A nova interface gráfica do SEstatNet pode ser vista na seção 4.4, que descreve também as características que foram observadas na sua confecção.

A seção 4.5 descreve a funcionalidade do SEstatNet e apresenta também as vantagens e desvantagens de se trabalhar com base de dados flexível e o por quê da construção de um módulo de treinamento. O capítulo é finalizado descrevendo algumas vantagens observadas no SEstatNet.

4.1 Introdução

Diversas universidades e centros de pesquisa vêm unificando esforços no desenvolvimento de novas tecnologias que possam ser utilizadas para dar suporte à educação à distância através da Internet. Tais pesquisas concentram-se principalmente no desenvolvimento de ambientes de ensino virtuais, onde é possível a criação de salas de aula na Internet para o oferecimento de cursos à distância e disponibilizar conteúdos por parte de professores e alunos.

No Brasil, existem algumas iniciativas interessantes que merecem destaque como: a **Escola Virtual Prossiga** (CNPq) que tem como objetivo ministrar cursos à distância através de tecnologias Web, estabelecendo a criação de ambientes virtuais propícios a interações, trocas de informações e trabalhos cooperativos [PROSSIGA, 2004].

Outro projeto interessante é o **Virtus Class** (UFPE). É um sistema de salas de aulas virtuais do Laboratório de Hipermídia da UFPE que conta atualmente com 2.510 salas de aula de diversas áreas do conhecimento e que é disponibilizado com um serviço gratuito na Internet [VIRTUS, 2004].

O projeto **TeleEduc** (Unicamp) é um ambiente para a criação, participação e administração de cursos na Web desenvolvido por pesquisadores do Núcleo de Informática Aplicada à Educação (Nied) da Unicamp [TELEDUC, 2004].

Em relação ao projeto **TeleEduc** faz-se importante ressaltar que o mesmo é um **software livre** sendo possível modificá-lo ou redistribuí-lo sob os termos da **General Public License** [GNU, 2004].

O Laboratório de Estatística Aplicada (LEA), vinculado ao Departamento de Informática e Estatística da Universidade Federal de Santa Catarina desenvolve vários projetos, cursos e programas que exploram a utilização de diversos recursos didáticos e tecnológicos interativos tais como Internet, Intranet, Videoconferência e *softwares*.

Visando facilitar o acesso ao conhecimento estatístico e utilizando-se do conceito de educação à distância (EaD), o LEA vem investindo no projeto **SEstatNet**. O objetivo do SEstatNet é a adição da Web como um ambiente educacional para o ensino-aprendizagem de estatística com a criação de uma transição viável da sala de aula convencional para a sala de aula virtual.

4.2 Histórico do desenvolvimento

O *software* que antecedeu ao SEstatNet é denominado apenas SEstat e seu principal objetivo é oferecer apoio ao ensino-aprendizagem da análise estatística de dados [CECHINEL e AMORIM, 1997] [CECHINEL et al., 1999]. Inicialmente foi utilizado como ambiente na forma de ensino tradicional nos Cursos de Engenharias, Ciência da Computação e Sistemas da Informação da Universidade Federal de Santa Catarina.

Muitas versões foram implementadas. O primeiro grande destaque foi o desenvolvimento de rotinas próprias de cálculos estatísticos [DIAS, 2001].

O primeiro protótipo para um ambiente que permitisse o trabalho colaborativo e à distância entre os usuários do sistema foi desenvolvido em 2002 [NOAL, 2002].

A versão seguinte denominada SEstatNet foi desenvolvida utilizando a plataforma Java 2 [POTTS e FRIEDEL, 1996], mais especificamente utilizando J2EETM. A tecnologia J2EETM tem seu modelo baseado em componentes, simplificando assim seu desenvolvimento [NAKAZAWA e MARAFON, 2003]. Ela controla a infra-estrutura e a interoperabilidade das aplicações, suportando serviços Web com um desenvolvimento

seguro e robusto. Dentre os vários componentes que formam essa tecnologia, utilizaram-se apenas JSP, JavaBeans, JDBC e Java Servlets para o desenvolvimento do SEstatNet [VISUAL, 2004] [DEITEL e DEITEL, 2003]. Essa versão manteve todas as características didático-pedagógicas obtidas no SEstat [DIAS, 2001].

A versão atual e descrita no item a seguir está em fase de desenvolvimento. Esse breve histórico faz-se necessário para mostrar que o ambiente SEstatNet é altamente dinâmico frente as novas tecnologias que se apresentam.

4.3 Características do SEstatNet

No ambiente de ensino-aprendizagem de estatística do SEstatNet o conhecimento é disponibilizado para o aluno através de uma base de conhecimento contendo regras que representam o raciocínio de um especialista em análise estatística de dados. Essa base de conhecimento é gerenciada pelo sistema, e a partir da interação do sistema com o aluno por meio de perguntas e respostas sobre as variáveis de uma base dados (de livre escolha do aluno) e sobre as informações estatísticas que se pretende gerar, é realizada uma análise estatística sobre os dados.

Em outras palavras, a partir das respostas fornecidas pelo aluno, o SEstatNet avalia o atendimento das suposições para a aplicação de cada tipo de método estatístico e realiza a análise propriamente dita, disponibilizando ao aluno as ferramentas e os conteúdos para o entendimento dos resultados gerados.

É importante ressaltar que o aluno pode partir de qualquer ponto de interesse de estudo, sendo capaz de percorrer os caminhos que deseja num movimento de aprendizagem livre. Nesse processo, o aluno interage constantemente com o sistema e adota uma postura crítica perante os conteúdos, tornando-se sujeito ativo no seu processo de ensino-aprendizagem [NASSAR et al., 2003].

No ano de 2003/2004, com recursos do Cnpq/PIBIC, acadêmicos do curso de Sistemas de Informação (entre outras atividades descritas em seus projetos) reestruturaram a arquitetura do SEstatNet, transportando o sistema para a linguagem Perl utilizando a plataforma Web HPL.

Nessa nova estrutura, a base de conhecimento do sistema está sendo disponibilizada em arquivos XML em uma arquitetura de modulação. Esses arquivos

(independentes e externos ao sistema) são manipulados por um objeto que faz o papel da máquina de inferência de um sistema especialista.

Uma vez que a base de conhecimento do sistema está completamente independente do restante do ambiente, tornou-se possível incluir novas regras na base sem a necessidade de interferir na estrutura do sistema, assim como inserir outras bases de conhecimento referentes a outros domínios de aplicação que não necessariamente estejam relacionados com análise estatística de dados.

O SEstatNet foi transportado para um servidor Linux, mais especificamente Red Hat 7.3+. A escolha pelo Linux deve-se à maior compatibilidade e desempenho com todos os recursos utilizados para o desenvolvimento, citados acima, além de ser livre, o que é ideal para o meio acadêmico.

Nota-se que o SEstatNet além de necessitar ter estabilidade, pois deve estar disponível a qualquer hora, deve suportar o acesso simultâneo de em média 40 usuários, situação que ocorre durante uma aula ministrada em laboratórios de ensino do CTC/UFSC. O acesso simultâneo implica em uma grande geração de gráficos e manipulação de dados em um curto espaço de tempo, o que exige total compatibilidade com os recursos utilizados.

As linguagens de programação do SEstatNet são:

- JSP – *JavaServer Pages Technology*. [JAVA, 2004].
- Perl – através da plataforma Web HPL.

O SEstatNet utiliza a biblioteca *R-Project*, que possui melhor controle através do Perl [R, 2004] e para banco de dados é utilizado o MySQL, pois possui um alto desempenho em ambiente Linux.

A implementação da biblioteca *R-Project* incluiu a instalação em ambiente Linux, aprendizagem da manipulação da mesma através da linguagem R, embutida em seu ambiente, integração com os módulos do SEstatNet por meio de pontes entre o ambiente R [R, 2004] e Perl, e desenvolvimento de uma interface Web para a manipulação em tempo real.

Todo o desenvolvimento utiliza ferramentas e tecnologias de código-aberto e os componentes estão disponibilizados no repositório *Comprehensive Perl Archive Network* – CPAN [CPAN, 2004].

4.4 Interface do Sistema

Um *designer* de interface deve observar as seguintes características quando avalia um *software* educacional ou interface gráfica [SHNEIDERMAN, 1997]:

1. *Esforço pela consistência*

Seqüências consistentes de ações devem ser requeridas em situações similares.

2. *Capacitar os usuários freqüentes a usarem de atalhos*

Aumentar a freqüência do uso, também aumenta o desejo de reduzir o número de interações e de aumentar o ritmo da interação.

3. *Fornecer respostas informativas*

Para cada ação do usuário deve haver algum sistema de resposta. Para ações freqüentes e simples, a resposta pode ser modesta, enquanto para ações não tão freqüentes e complexas, a resposta deve ser mais substancial.

4. *Obter homogeneidade no fechamento*

Seqüências de ações devem ser organizadas em grupos com início, meio e fim. A resposta informativa quando da realização de um grupo de ações dá ao usuário a satisfação da realização bem sucedida, o sentimento de alívio.

5. *Oferecer uma manipulação simples dos erros*

Se um erro ocorrer, o sistema deve detectar o erro e o oferecer um mecanismo simples e abrangente para a manipulação do erro.

6. *Permitir a inversão de ações*

Há muitas possibilidades de ações que podem ser reversivas. Unidades com reversibilidade devem ser permitidas para auxiliar na inversão de ações.

7. *Apoiar o controle interno*

O usuário experiente deseja fortemente ter a sensação de que cada ação sua cause mudanças e traga respostas do sistema.

8. *Reduzir a sobrecarga de memória do Usuário*

Reduzir o esforço da memorização do usuário evitando mudança desnecessária das informações da página disponível, pois o ser humano tem uma limitação na sua capacidade de memorização.

Na confecção da interface gráfica do *site* procurou-se implementar todos os tópicos descritos acima. O resultado desse trabalho pode ser observado na figura 4.1, onde se pode identificar a divisão do *browser* em três *frames*.

O frame superior é usado para cadastrar e identificar os usuários que entram na área restrita do portal. Ela contém também data de atualização da página e *links* para contato, suporte e ajuda.



Figura 4.1: Interface gráfica do SEstatNet.

O *frame* da esquerda é ocupado com o menu fixo da página, onde você pode rolar todo o conteúdo da página no *frame* da direita.

O usuário devidamente identificado pode entrar na área restrita do sistema. Quando isso ocorre é aberta uma Tabela de Dados como mostrada na Figura 4.2 e cujas funcionalidades são descritas no próximo capítulo.

4.5 Como funciona o SEstatNet

O *software* SEstatNet foi desenvolvido para trabalhar basicamente como uma página de Internet, onde o usuário faz uso dessa ferramenta navegando em um *browser*. Isso torna dispensável a instalação do programa na máquina do usuário, e ajuda o

professor a monitorar todas as chamadas e interações do aluno com o software, conseguindo assim, fazer uma análise do caminho de aprendizagem do aluno.

Devido ao fato de trabalhar com uma base de dados flexível, o sistema não possui o controle e o conhecimento das características das variáveis com as quais o aluno está lidando, embora alguns módulos sejam planejados com análises internas que possibilitam alertar o usuário para possíveis erros [SANTOS, 2001].

Lista	CASO	LOCAL	PAP	INSTR	TAM	RENDAGE	RENDASE	ALIMBE	APROVBE	APROVSE
1	1	1	0	3	3	9	11	9	3	3
2	2	1	0	3	4	10	5	4	4	1
3	3	1	0	2	5	6	8	6	3	1
4	4	1	1	1	4	7	14	10	4	1
5	5	1	1	2	4	10	3	2	5	1
6	6	1	1	3	6	10	8	7	3	1
7	7	1	1	2	4	5	4	3	4	3
8	8	1	0	3	4	15	11	9	5	1
9	9	1	0	3	3	19	15	11	3	4
10	10	1	0	3	4	5	8	6	4	2
11	11	1	0	2	4	1	12	9	1	5
12	12	1	0	2	5	10	8	6	3	2
13	13	1	0	3	4	13	12	10	2	5
14	14	1	0	2	4	13	14	11	3	1
15	15	1	0	2	8	8	11	8	2	5
16	16	1	1	1	2	2	8	7	2	1
17	17	1	1	2	2	2	12	8	2	1

Figura 4.2: Parte de uma base de dados fixa denominada PAP.

Esta flexibilidade ao domínio pesquisado (“*free context*”) tem a desvantagem de permitir que o usuário responda a perguntas de forma errada sem que haja a intervenção do sistema para a correção da resposta. Em consequência disso, o sistema pode tomar um caminho errado na busca dos resultados estatísticos, fazendo com que o aluno não aprenda a realizar adequadamente a análise estatística de dados.

A solução encontrada para resolver este problema foi a construção de um módulo de treinamento dentro do SStatNet. Este módulo trabalha com uma base de dados fixa (já conhecida pelo sistema), e possibilita ao sistema a identificação de erros nas respostas do usuário assim como avisá-lo sobre eles.

A cada interação do usuário com o SStatNet, o módulo de treinamento irá intervir fazendo observações a respeito da resposta que foi dada, comentando os acertos e fornecendo explicações para os erros.

Desta forma, o usuário que não estiver habituado a utilizar o sistema, utilizará o Módulo de Treinamento para aprender os conceitos estatísticos envolvidos e depois poderá aprimorar esse aprendizado utilizando uma base de dados qualquer escolhida por ele.

A figura 4.2 apresenta uma parte dessa base de dados conhecida pelo sistema denominada PAP (Programa de Alimentação Popular). Procurar-se-á disponibilizar novas bases supervisionadas para poderem ser substituídas conforme o interesse da disciplina ou do treinamento em questão.

4.6 Considerações Finais

O SEstatNet tenta preencher a lacuna atualmente existente no processo de organização e formatação dos conteúdos para cursos de educação à distância pela Internet, uma vez que abre novas perspectivas para a construção e disponibilização do conhecimento, utilizando-se de bases de conhecimento e recursos de Inteligência Artificial.

Outra vantagem observada no SEstatNet é que ele permite ao aluno trabalhar com qualquer base de dados, possibilitando dessa maneira usar vários conjuntos de dados da sua área de conhecimento. Esta característica proporciona ao usuário a generalização do conhecimento estatístico.

É importante salientar que os ambientes de suporte para educação à distância, por mais que ofereçam ferramentas que propiciem a cooperação e interação, não irão conseguir sozinhos que os alunos construam seus conhecimentos se não tiverem uma equipe interdisciplinar que os acompanhe, tanto alunos quanto professores. Pois o acompanhamento é o ponto fundamental para o funcionamento dos ambientes e a construção da aprendizagem.

No capítulo seguinte encontra-se desenvolvido uma metodologia para o gerenciamento de base de dados na Web que permite o ensino-aprendizagem de estatística.

5 DESENVOLVIMENTO DA METODOLOGIA

O desenvolvimento da metodologia será apresentado em dois capítulos, sendo o capítulo 5 dedicado mais a parte conceitual, isto é, concentra-se nas estruturas e restrições durante o projeto do banco de dados. O capítulo 6 será dedicado mais à parte prática, ou seja, a implementação da metodologia.

A seção 5.1 apresenta uma breve introdução sobre modelagem de banco de dados e a seção 5.2 apresenta alguns trabalhos desenvolvidos que disponibilizam facilidades para usar base de dados para análise estatística de dados na Internet.

O modelo Entidade-Relacionamento (ER) é um modelo de dados conceitual de alto nível bastante popular. A seção 5.3 apresenta a modelagem desenvolvida por essa metodologia e o dicionário de dados correspondente.

A seção 5.4 apresenta parte do esquema lógico desenvolvido, Região Base de Dados (*Region_db*) e os comandos SQL usados para a criação das tabelas no Banco de Dados. Nessa seção é apresentada também a interface gráfica para cadastramento dos usuários que queiram usar o sistema.

Se for oferecido algum curso à distância e o aluno quiser fazer, ele tem que efetuar uma matrícula. A seção 5.5 descreve os passos que devem ser seguidos para a efetivação da matrícula do aluno e apresenta a interface gráfica onde o aluno deve fornecer alguns dados complementares.

As seções 5.6 e 5.7 descrevem respectivamente as variáveis envolvidas em uma análise estatística com suas possíveis transformações e o módulo de tipificação de variáveis mostrando como será usado nesse novo ambiente.

A seção 5.8 apresenta as funcionalidades da interface gráfica que exhibe as bases de dados dos usuários.

Os recursos computacionais necessários para a implementação da metodologia e os aspectos de segurança que devem ser considerados são discutidos respectivamente nas seções 5.9 e 5.10.

5.1 Introdução

A modelagem conceitual é uma fase importante no projeto de uma aplicação de banco de dados bem-sucedida. O termo “aplicação de banco de dados” se refere ao

banco de dados propriamente dito e os programas a ele associados que implementam as consultas e atualizações.

Esses programas oferecem interfaces gráficas amigáveis para o usuário (GUIs) utilizando formulários e menus. Assim sendo, parte da aplicação do banco de dados irá precisar do projeto, da implementação e do teste desses programas da aplicação.

Tradicionalmente, o projeto e os testes dos programas da aplicação foram considerados como estando mais na esfera da engenharia de *software* do que sob o domínio do banco de dados. No entanto, cada vez se torna mais claro que existem fatos em comum entre as metodologias de projetos de bancos de dados e as do projeto da engenharia de *software* [ELMASRI e NAVATHE].

5.2 Trabalhos correlatos

Hoje diversas instituições que desenvolvem sistemas de educação à distância, tomam como base padrões e metodologias utilizadas em sistemas de educação à distância famosos e aceitas no meio. Dentre os inúmeros sistemas disponíveis na Web, foram selecionados três sistemas de aprendizagem, exibindo como cada um aborda os mecanismos utilizados no processo de ensino-aprendizagem do aluno à distância.

5.2.1 CARNEGIE MELLON UNIVERSITY

O modelo desenvolvido pela *Carnegie Mellon University*, armazena todo o conteúdo do curso através de informações declarativas e processadas por um sistema genérico baseado em um Banco de Dados (*Oracle V7*).

Pelo fato de ter-se todo o conteúdo representado no BD, o registro dos passos do aluno também estar no BD e a entrega ser feita pelo sistema (Internet), o processo de manipulação de informações fica bem mais fácil. A interface Web do sistema com o banco de dados é feita pelo *Oracle Web Server*. A figura 5.1 representa como os cursos são estruturados com essa plataforma.

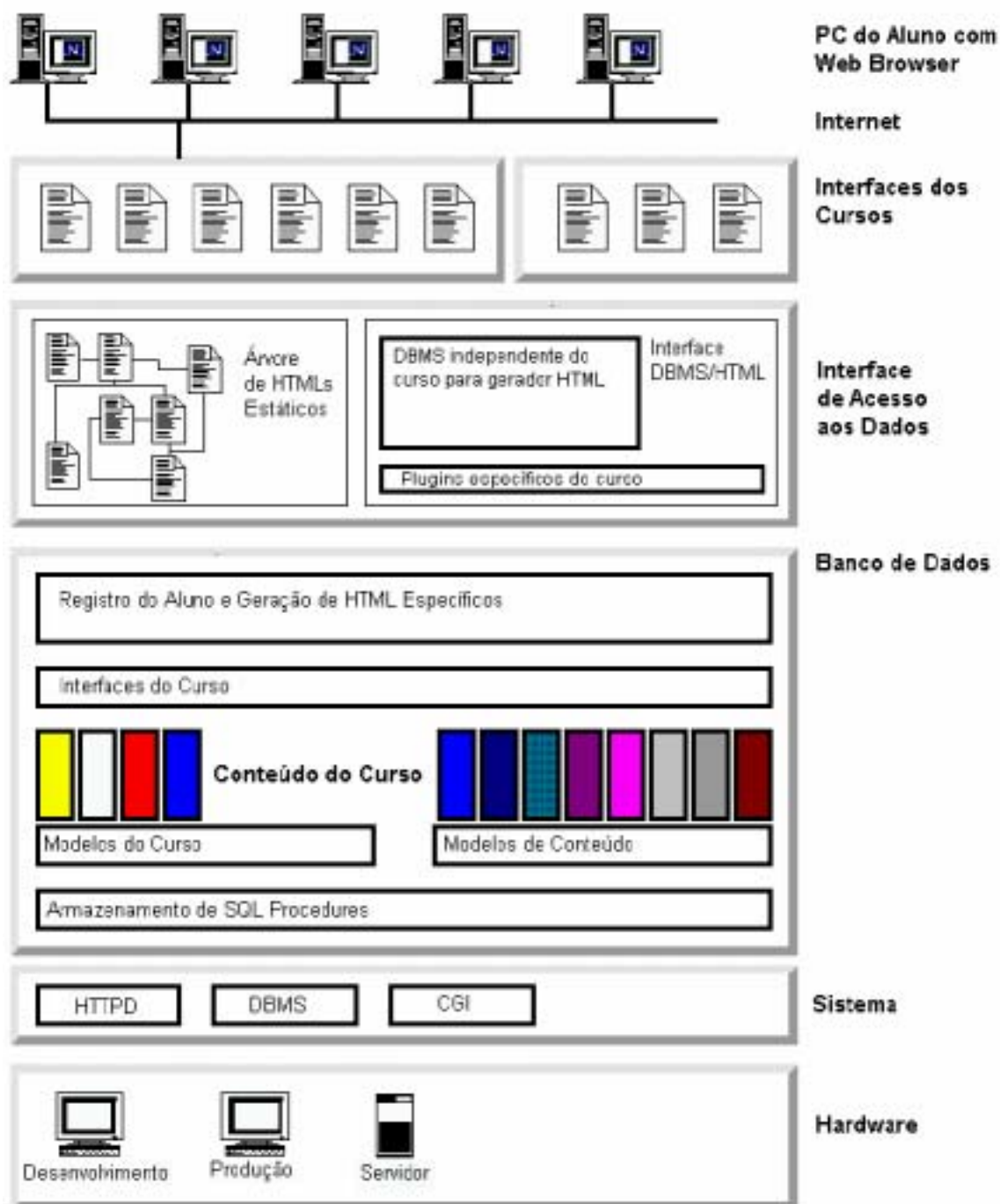


Figura 5.1: Estrutura dos Cursos da Carnegie Mellon University.

Este modelo apresenta um sistema de segurança e autenticação na rede, sendo esta segurança feita com transações SSL Web e a autenticação com *Kerberos login*. O aluno deve possuir conexão à Internet e um *Web browser* instalado em seu computador.

As principais informações contidas no banco de dados deste modelo incluem:

- **Pessoas:** informações dos alunos e professores, independente do curso em que estão.
- **Cursos:** informações do curso, como a sua programação, os alunos inscritos e os professores.
- **Conteúdo:** são as questões, respostas, material de aprendizado, tutoriais, avaliações, podendo ser compartilhado por vários cursos.
- **Modelo do Curso:** descrição dos módulos que farão parte do curso, bem como o conteúdo de cada módulo.
- **Instâncias do Curso:** seleção de um modelo de curso e associação com pessoas e informações de registro.
- **Registro de estudante:** informações sobre o progresso do aluno no curso.
- **Mídia:** arquivos de mídia, HTML, gráficos, programas, usados pelo sistema ou cursos.
- **HTML:** páginas geradas dinamicamente e que serão enviadas aos alunos.

O modelo da *Carnegie Mellon University* possui uma infra-estrutura centrada no estudante, propiciando tanto a entrega do material do curso, como o seu gerenciamento baseado na Web. Além dos tradicionais hipertextos e outras mídias, o sistema gera conteúdos personalizados (*feedback*) para cada aluno, e registra todo o caminho que o aluno fizer durante o curso, obrigando a cumprir as políticas do curso (pré-requisitos, restrições de limite de tempo e momento para exames). O sistema suporta um grande número de alunos e de cursos.

Com relação à utilização do sistema pelo aluno, cada vez que ele acessa o Banco de Dados para uma determinada unidade, registra-se o momento do acesso, bem como a duração deste acesso.

5.2.2 StatCrunch

StatCrunch, também conhecido como WebStat na sua versão mais antiga, é um *software* para análise estatística de dados na *World Wide Web* [WEBSTAT, 2004]. Depois da versão 1 que foi publicada em 1997, lançou-se a versão 2 em 1999 e a versão 3 em 2002 com toda a funcionalidade de um *software* estatístico.

A versão 2 ofereceu uma interface amigável e rotinas estatísticas básicas que satisfazem as necessidades de pesquisas de nível educacional básico. A interface da

página principal oferece uma alternativa de livre acesso a um *software* estatístico comercial pela WWW com a única exigência mínima por parte do usuário que é ter o Java instalado para usar com os *browsers* *Web Internet Explorer* ou *Netscape Navigator*.

O WebStat foi desenvolvido por uma equipe de programadores e estatísticos gerenciados por Webster West do Departamento de Estatística da Universidade de Carolina do Sul. O pacote foi criado inicialmente como tentativa de resolver muitos dos problemas que existem na análise e uso dos atuais *softwares* estatísticos. Gastava-se muito tempo para desenvolver aplicações nas linguagens do Splus, SAS, Minitab, etc. e que são muito específicos para estatísticos. Os estudantes e outros usuários em potencial não tendo acesso a essas linguagens, não podiam usar esses procedimentos. O WebStat deveria suprir essa deficiência usando apenas o Java e a WWW.

Desde então o *software* vem sendo implementado com novos procedimentos de análise estatística e novas funcionalidades. Recentemente foi lançada a versão 4 e trocado o seu nome para StatCrunch.

5.2.3 OpenEpi

O projeto *Interactive Statistical Pages* representa um esforço contínuo para desenvolver e disseminar *software* de análise estatística na forma de páginas Web [DEAN, 2003]. Usando formulários HTML, CGI e *scripts* Perl, Java, JavaScript e outras tecnologias básicas de *browser*, cada página Web contém dentro de si (ou invoca) toda a programação necessária para executar um cálculo específico ou análise.

Essas páginas usam várias técnicas de desenvolvimento:

- Algumas usam **formulários HTML** para enviar seus dados para o servidor Web, que executam *scripts* CGI para realizar os cálculos e enviar o resultados de volta em uma nova página. Pode trabalhar em qualquer *browser* que suporta formulários HTML.
- Outras, incluem tudo isso nas próprias páginas, usando **JavaScript**. Cada página contém dentro de si toda a programação necessária para fazer os cálculos. Esse programa em *JavaScript* pode rodar em qualquer *browser* que suporte a linguagem como o *Netscape V2* e *Internet Explorer V3*, ou versões superiores.

- Ainda outras, invocam **applets Java** para realizar os cálculos. Esse programa em Java pode ser executado em qualquer *browser* que suporte a linguagem como o *Netscape V3* e *Internet Explorer V4*, ou versões superiores.

Estas páginas são criadas por voluntários que contribuem com suas perícias e esforços. Eles residem em computadores de todo o mundo e estão interligados inclusive com o pacote de análise estatística que é livremente acessado pela Internet. Essa é a prática do conceito de *software* livre.

A metodologia ora em desenvolvimento procurou incorporar o que há de melhor dessas aplicações e que satisfazem as necessidades do ambiente SEstatNet conforme será descrito nas seções seguintes.

5.3 Administração dos dados

É importante lembrar que o objetivo principal deste projeto não é a modelagem e construção de um banco de dados “completo” para um sistema de educação à distância. A modelagem proposta prende-se aos dados que auxiliarão na implementação do módulo de matrícula e gerência das bases de dados dos alunos que usam o ambiente de ensino-aprendizagem na Web.

O esquema conceitual é uma descrição concisa dos requisitos de dados feitos pelos usuários e inclui descrições detalhadas dos tipos de entidades, relacionamentos e restrições. Essas são expressas utilizando os conceitos fornecidos pelo modelo de dados de alto nível.

De acordo com os requisitos do ambiente de ensino-aprendizagem de estatística SEstatNet foi feita uma modelagem Entidade/Relacionamento (ER) apropriada. A seguir, este modelo foi utilizado para que um Modelo Relacional de Dados fosse gerado. A seção 5.3.1 contém uma das visões do Banco de Dados do projeto e a seção 5.3.2 contém a descrição das entidades dessa visão.

5.3.1 Diagrama Entidade/Relacionamento

O primeiro passo na criação do banco de dados que será usado consiste na elaboração de um modelo Entidade/Relacionamento. Este modelo abstrai detalhes de implementação, mas abrange todas as necessidades e requisitos do sistema proposto.

A figura 5.2 mostra a Modelagem ER para o módulo a ser desenvolvido e que futuramente será implementado no ambiente SEstatNet. Este módulo será composto por um sistema de matrícula e o controle para criação de base de dados.

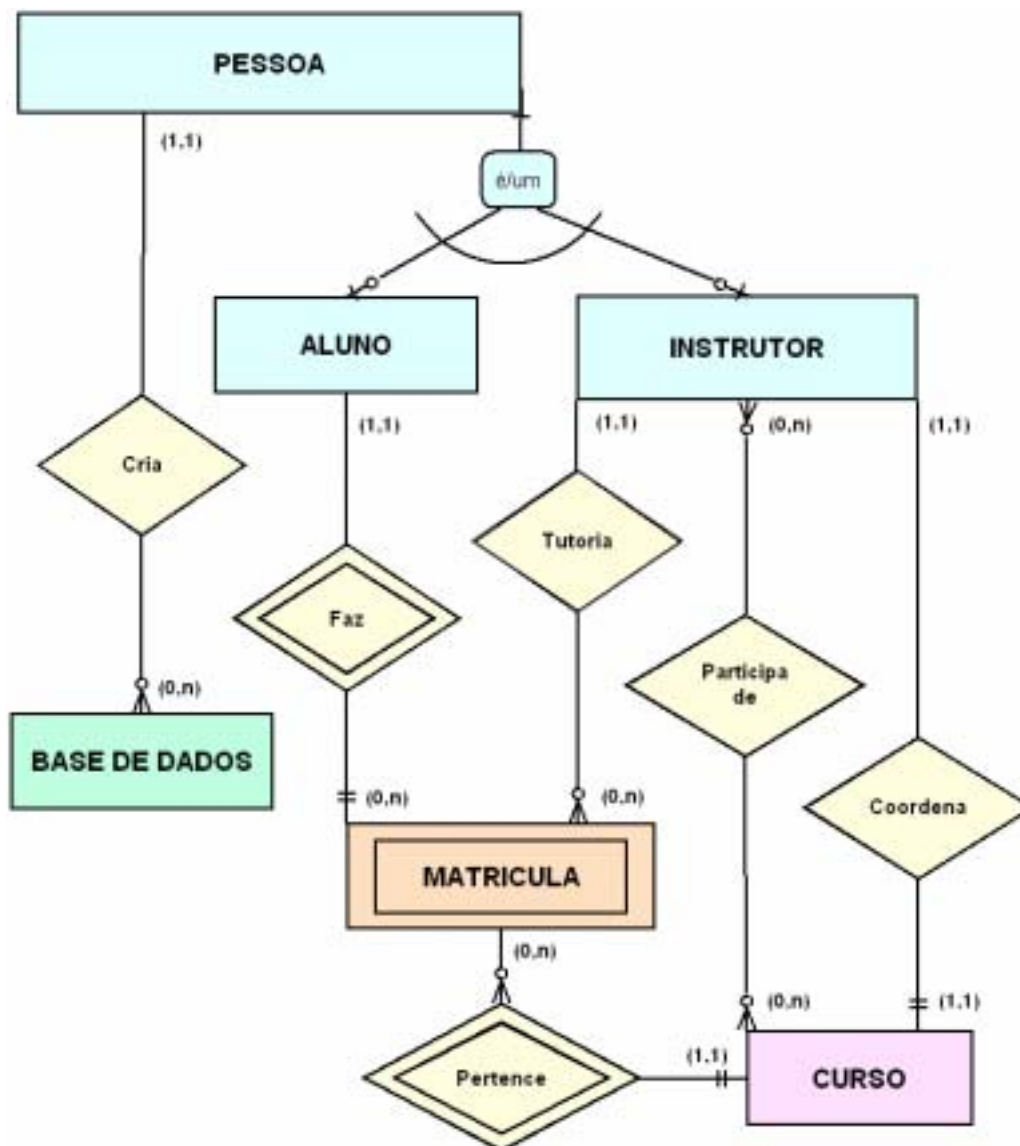


Figura 5.2: Diagrama E/R da modelagem proposta para ensino-aprendizagem de estatística na Web.

O ambiente está projetado do seguinte modo: uma pessoa pode ser um aluno, um tutor ou um profissional. Qualquer pessoa pode criar suas bases de dados. Se for um tutor ele participa de algum curso, sendo um professor orientador dos alunos matriculados no curso. Essa orientação é uma interação onde o tutor corrige exercícios, aplica avaliações e soluciona dúvidas.

Se for um aluno, deve fazer matrícula em um curso que é oferecido. O curso pode ter vários fóruns, cada um com o tema referente ao seu conteúdo. O curso também pode possuir uma FAQ, local onde serão armazenadas as dúvidas mais comuns e respostas a elas.

Também existe a figura do coordenador geral do curso, geralmente o autor do mesmo, que também é um tutor. O coordenador geral do curso pode elaborar provas e listas de exercícios para o curso que coordena.

As seguintes observações podem ser verificadas:

- Todas as pessoas têm que preencher um cadastro inicial para ter acesso ao ambiente.
- Uma pessoa cadastrada pode criar bases de dados e acessar o ambiente SEstatNet.
- Uma matrícula somente recebe material do curso ao qual pertence.

5.3.2 Dicionário de Dados

Este Dicionário de Dados descreve as Entidades e Relacionamentos da modelagem conceitual do sistema proposto e que estão identificados no Banco de Dados. Primeiro aparecem as Entidades e seus respectivos atributos; em seguida são apresentados os Relacionamentos e seus eventuais atributos.

Entidades

Uma Entidade é o objeto básico que o modelo ER representa. É uma representação abstrata de um objeto do mundo real com uma existência independente. Essa existência pode ser física ou conceitual [ELMASRI e NAVATHE].

Cada Entidade possui informações associadas a ela, as quais são chamadas de atributos. Cada atributo tem um nome, uma descrição, um tipo (Tp) que identifica o seu domínio, um indicador de atributo opcional (Op), um indicador de atributo multivalorado (Mv), um indicador de atributo derivado (Dv), um indicador de atributo composto (Cp) e algumas observações sobre componentes, atributos primitivos e domínio (no caso de o atributo não poder assumir todos os valores definidos pelo seu tipo).

Neste dicionário os tipos de dados são indicados por: numérico (N), alfanumérico (A), data e hora (DH) e, somente data (D).

PESSOA

- Descrição: Pessoas que se cadastram no ambiente.
- Sinônimos: Participantes.
- Atributos:

Tabela 5.1: Atributos da Entidade PESSOA.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
IDUSUARIO	Identificação no BD	N					Chave primária, auto-incrementada
LOGIN	Identificação da pessoa	A					
SENHA	Senha para entrada no sistema	A					
E_MAIL	Endereço eletrônico da pessoa	A					
NOME	Nome da pessoa	A					
TIPO_USR	Tipo de usuário	A		X			Valores=(aluno, professor, profissional)
ORGANIZAÇÃO	Empresa que trabalha ou estuda	A					
DATA_CADASTRO	Data de cadastramento	DH					
DATA_ACESSO	Data do último acesso	DH					

ALUNO

- Descrição: Pessoas que estão fazendo algum curso.
- Sinônimos: Discente.

INSTRUTOR

- Descrição: Pessoa que tutora alunos de um curso.
- Sinônimos: Docente, Tutor.
- Observações: Todos os instrutores devem participar de pelo menos um dos relacionamentos: Coordena_curso, É_tutor_em, Participa de.

BASE DE DADOS

- Descrição: Base de Dados criadas pelas Pessoas para analisá-las no ambiente SStatNet.
- Sinônimos: Base de Dados, Amostragem, Dados da pesquisa.
- Atributos:

Tabela 5.2: Atributos da Entidade Base de Dados.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
ID_BASE	Identificador da Base	N					
NOME_BASE	Nome da base de dados	A					
DATA_CRIACAO	Data de criação da base	D					
TAMANHO	Tamanho da base em bytes	N					
N_COL	Número de colunas	N				X	Atributo composto DADOS
NOME_COL	Nome das colunas (variável)	A		X		X	Atributo composto DADOS
TIPO_VAR	Tipo de variável	N		X		X	Atributo composto DADOS
VALORES	Valores das variáveis	N		X		X	Atributo composto DADOS

MATRICULA

- Descrição: Matrícula do aluno em um curso.
- Sinônimos: Inscrição.
- Observação: É uma entidade fraca, pois depende das entidades ALUNO e CURSO.
- Atributos:

Tabela 5.3: Atributos de Entidade MATRÍCULA.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
IDMATR	Identificação da matrícula	N					Chave primária, auto-incrementada
DATA_MATRICULA	Data da matrícula	D					
LOGRADOURO	Rua e número da residência	A				X	Atributo composto ENDERECO
BAIRRO	Bairro onde reside	A				X	Atributo composto ENDERECO
CIDADE	Cidade onde reside	A				X	Atributo composto ENDERECO
UF	Unidade Federativa	A				X	Atributo composto ENDERECO
PAIS	País de origem	A				X	Atributo composto ENDERECO
CEP	Código Postal	N				X	Atributo composto ENDERECO
FONE_RESID	Telefone residencial	N					
FONE_CELULAR	Telefone celular	N	X				
DATA_NASC	Data de nascimento	D					
IDADE	Idade em anos	N			X		
ESCOLARIDADE	Nível de escolaridade	A					Valores = {Grad., Pós-Grad., Especializ.}
NACIONALIDADE	País de origem	A			X		Valores = {Brasileira, Estrangeira}
CPF	Cadastro de Pessoa Física	N					
REGISTRO	Registro geral	N				X	Atributo composto RG
ORGAO_EXP	Órgão expedidor do RG	A				X	Atributo composto RG
DATA_EMISSAO	Data de emissão do registro geral	D				X	Atributo composto RG
CARGO_FUNCAO	Cargo que ocupa na sua Instituição	A	X				
FONE_TRAB	Telefone do trabalho	N					
FONE_FAX	Telefone do FAX	N	X				

CURSO

- Descrição: Cursos oferecidos.
- Atributos:

Tabela 5.4: Atributos da Entidade CURSO.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
ID_CURSO	Identificador do curso	N					
AREA	Área do curso	N					Valores = {Humanas, Biológicas, Exatas}
NOME	Nome do curso	A					
DATA_INSERCAO	Data de inserção da dúvida na FAQ	D				X	Atributo composto FAQ
PERGUNTA	Dúvida freqüente	A				X	Atributo composto FAQ
RESPOSTA	Resposta a uma dúvida freqüente	A				X	Atributo composto FAQ

Relacionamentos

Um relacionamento é uma associação entre entidades. Os relacionamentos possuem nome, cardinalidade máxima e mínima e podem também possuir atributos.

Coordena

- Descrição: Representa um dos relacionamentos entre INSTRUTOR e CURSO. Um instrutor coordena no mínimo zero e no máximo N cursos. Um curso é coordenado por no mínimo um e no máximo N instrutores, em períodos diferentes.
- Observações: Não pode haver interseção entre períodos dos relacionamentos “Coordena_Curso” de uma mesma entidade, ou seja, dois instrutores não podem estar coordenando o mesmo curso em uma mesma data ou período.
- Atributos

Tabela 5.5: Atributo do Relacionamento Coordena.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
PERIODO	Período (início e fim) da coordenação de um curso	p					Chave Parcial

Faz

- Descrição: Representa o relacionamento entre ALUNO e MATRICULA. Um aluno faz no mínimo zero e no máximo N matrículas. Uma matrícula é de um e somente um aluno.

Pertence

- Descrição: Representa o relacionamento entre as Entidades MATRICULA e CURSO. Uma matrícula pertence a um e somente um curso. Um curso possui no mínimo zero e no máximo N alunos matriculados.

Tutoria

- Descrição: Representa o relacionamento entre as Entidades INSTRUTOR e MATRICULA. Um tutor tutoria no mínimo zero e no máximo N matrículas.

Uma matrícula é tutorada por no mínimo 1 e no máximo N tutores, em períodos diferentes.

- Observações:
 1. Um aluno matriculado é tutorado por um e somente um tutor em uma data, ou seja, numa mesma data um aluno não pode ter mais de um tutor;
 2. Para que exista o relacionamento entre instâncias de INSTRUTOR e MATRICULA o valor dos atributos id_curso dos relacionamentos “É tutor em” e “Pertence” devem ser iguais, ou seja, um aluno matriculado em um curso deve ser tutorado por um tutor que seja do mesmo curso.
- Atributos:

Tabela 5.6: Atributo do Relacionamento Tutoria.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
DATA_INICIO	Data do início da tutoria	D					

Participa de

- Descrição: Representa um dos relacionamentos entre INSTRUTOR e a agregação Curso. Um tutor participa de no mínimo zero e no máximo N cursos. Um curso tem como participante no mínimo 1 e no máximo N tutores.
- Atributos:

Tabela 5.7: Atributo do Relacionamento Participa de.

Nome	Descrição	Tp	Op	Mv	Dv	Cp	Observações
PERIODO	Período em que é tutor na oferta de um curso	P					Chave Parcial

Cria

- Descrição: Representa o relacionamento entre as Entidades ALUNO e BASE DE DADOS. Um aluno possui no mínimo zero e no máximo N bases de dados para serem analisadas. Uma Base de Dados pertence a um e somente um aluno.

Uma vez definido o modelo Entidade/Relacionamento do projeto e o seu dicionário de dados, partiu-se para a elaboração do esquema lógico do Banco de Dados.

5.4 Esquema lógico do Banco de Dados

A implementação do esquema lógico de Banco de Dados utilizou o ambiente de desenvolvimento de banco de dados *Fabulous Force Database Designer 4* da KDE/Gnome [DB DESIGNER, 2004]. Este *software* é distribuído sob licença da GNU GPL. A criação das tabelas foi feita usando a Linguagem SQL no sistema de gerenciamento de banco de dados do MySQL.

A figura 5.3 apresenta parte da implementação das tabelas correspondentes ao cadastro inicial dos participantes e as bases de dados.

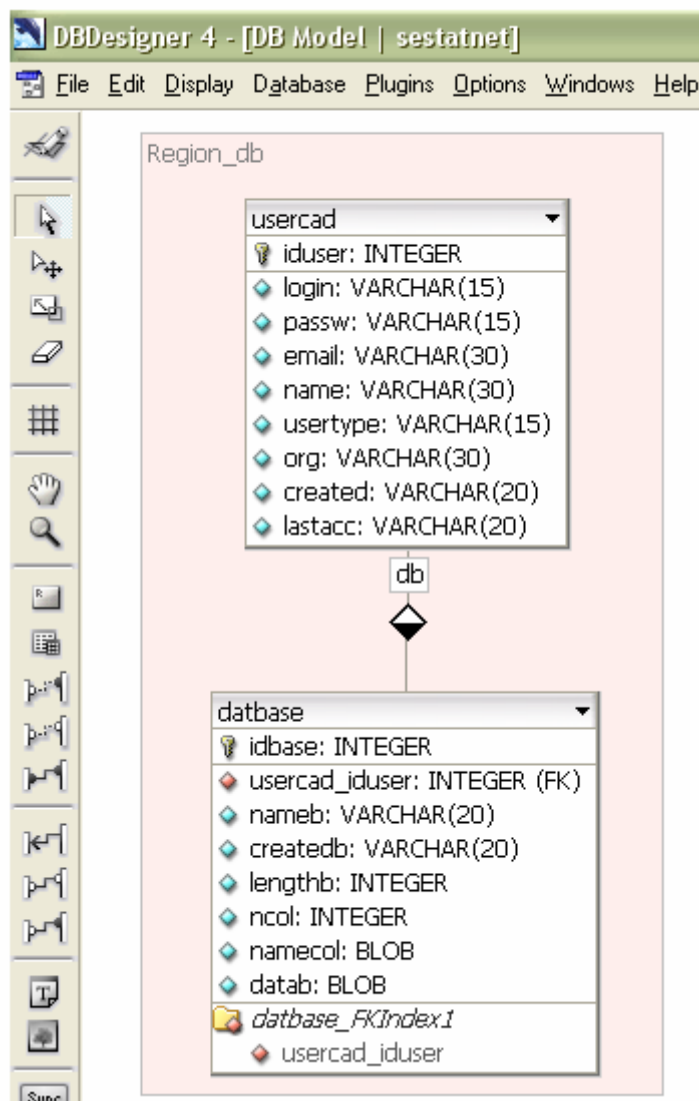


Figura 5.3: Esquema lógico do Banco de Dados.

O *software* permite fazer uma conexão direta com o servidor de banco de dados e cria as tabelas automaticamente com as informações especificadas. Para exemplificar, a figura 5.4 apresenta os comandos SQL necessários para a criação das tabelas acima.

```
CREATE TABLE usercad (
  iduser INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,
  login VARCHAR(15) NOT NULL,
  passw VARCHAR(15) NOT NULL,
  email VARCHAR(30) NOT NULL,
  name VARCHAR(30) NOT NULL,
  usertype VARCHAR(15) NOT NULL,
  org VARCHAR(30) NOT NULL,
  created VARCHAR(20) NOT NULL,
  lastacc VARCHAR(20) NOT NULL,
  PRIMARY KEY(iduser)
);

CREATE TABLE datbase (
  idbase INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,
  usercad_iduser INTEGER UNSIGNED NOT NULL,
  nameb VARCHAR(20) NOT NULL,
  createdb VARCHAR(20) NOT NULL,
  lengthb INTEGER UNSIGNED NOT NULL,
  ncol INTEGER UNSIGNED NULL,
  namecol BLOB NULL,
  datab BLOB NOT NULL,
  PRIMARY KEY(idbase),
  INDEX datbase_FKIndex1(usercad_iduser),
);
```

Figura 5.4: Comandos SQL para criação das tabelas Cadastro de Usuário e Base de Dados.

Para o usuário do sistema essas operações todas ficam subjetivas. Ele precisa apenas acessar na página principal do *site* SEstatNet correspondente a cadastro de usuário. A interface gráfica da figura 5.5 é disponibilizada.

Após o preenchimento dos dados solicitados o sistema faz uma validação dos mesmos e se for aceita cadastra o novo usuário no sistema enviando a senha de acesso para o seu endereço eletrônico (*e-mail*).

Os programas desenvolvidos em Perl e CGI para essa implementação estão disponibilizados no Apêndice A.

Sestat.net Registro de um Novo Usuário

SEstatNet ID: 6-15 caracteres

Senha: 6-15 caracteres

Redigite a senha:

Email:

Nome:

Ocupação:

Empresa/Escola:

Figura 5.5: Interface gráfica para registro de um novo usuário do sistema.

As informações fornecidas na interface gráfica da figura 5.5 ficam registradas na tabela Cadastro de Usuários e são complementadas com mais duas informações do tipo “oculta”: data do registro e data do último acesso feito no ambiente restrito do SEstatNet.

O usuário devidamente cadastrado e de posse de sua identificação de acesso e senha poderá na página principal do ambiente SEstatNet entrar na área restrita, onde são disponibilizados os recursos para acesso e manipulação da sua base de dados. Essas operações são descritas com mais detalhes no próximo capítulo em Modelo de Comportamento.

Um usuário do sistema pode eventualmente participar de algum curso à distância que é oferecido pelo Departamento de Informática e Estatística da UFSC. Nesse caso ele deve proceder conforme está descrito no item seguinte.

5.5 Controle Acadêmico

Para iniciar um curso é necessário fazer a matrícula, preenchendo o respectivo certificado de matrícula disponibilizado na interface geral do sistema. Imediatamente após o preenchimento do certificado de matrícula, o coordenador do curso envia as instruções e os códigos de acesso (*login* e senha de aluno) para o *e-mail*, no endereço apontado pelo mesmo, a fim de garantir a privacidade e autenticidade da operação.

Matricula - Microsoft Internet Explorer

Arquivo Editar Exibir Favoritos Ferramentas Ajuda

Certificado de Matrícula no programa de educação a distância

CENTRO TECNOLÓGICO
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Sestat.net

Dados gerais

Nome da pessoa:

Endereço:

Bairro: Cidade:

CEP: UF: Selezione:

Fone Resid: -

Celular: -

E-mail:

AVISO: Este e-mail será utilizado para correspondências administrativas, inclusive fornecimento de senhas.

CPF:

Data Nasc.: Nacionalidade: Selezione:

R.G.:

Figura 5.6: Parte da tela de Certificação de Matrícula do Aluno.

Na interface apresentada na figura 5.6 o aluno recebe todas as informações pertinentes à realização do seu curso, as quais estão disponíveis no item Regulamentos e Condições Gerais de Uso. Essa regulamentação é descrita a seguir:

REGULAMENTOS E CONDIÇÕES GERAIS DE USO SEstatNet

Apresentação

Os Cursos de Educação a Distância do INE/UFSC foram concebidos com o objetivo de levar a profissionais que atuam na área educacional os mais modernos fundamentos capazes de alavancar o progresso e a modernidade nas unidades educacionais. Utilizando-se de experiências bem sucedidas em renomadas instituições mundialmente consagradas no meio empresarial e contando com o apoio tecnológico do "LEA - Laboratório de Estatística Aplicada", os cursos são subdivididos em diversas áreas de especialização, possibilitando um grande aprofundamento técnico e científico dos participantes. A metodologia de educação a distância permite que sejam atendidas todas

as regiões brasileiras e no exterior, bem como estabelecimentos de quaisquer níveis, democratizando as formas de enriquecimento qualitativo dos gestores educacionais.

Modalidade

Nesta modalidade apresentada de educação à distância, utiliza-se a Internet como ambiente para a realização do curso.

Carga horária

Sistema de crédito

Pré-requisito

Apoio ao aluno

- Orientação inicial para os estudos através da educação à distância e ferramentas encontradas no ambiente privativo de cada curso na Internet.
- Orientação diária ao aluno através de acesso ao Jornal da Educação publicado no website do LEA.
- Plantão de especialistas - via telefone, e-mail ou pessoalmente, no Laboratório de Estatística Aplicada.
- Disponibilização de revistas e periódicos específicos.
- Envio dos textos para aprendizagem e indicação de leituras complementares.
- Apoio para desenvolvimento dos exercícios de aferição de aprendizagem didática.
- Orientação para o trabalho de conclusão do curso (produção científica).
- Interação permanente, sempre que houver necessidade ou dúvidas do aluno.
- Seminários e encontros presenciais, abertos à participação dos discentes.
- Apoio ao estudante após a conclusão do curso através de educação a distância continuada.

Estrutura curricular

O curso está dividido em três unidades didáticas estruturadas como cursos independentes, mas interrelacionados e ainda um trabalho de conclusão de curso que será uma produção científica. Estas unidades serão correspondentes a cada curso especificamente, consoante plano pedagógico previamente estipulado.

Produção científica

Ao final da terceira unidade o aluno deverá fazer seu trabalho de conclusão do curso, sob forma de monografia, permitindo uma visão de conjunto do aprendizado e servindo de base para a difusão do conhecimento.

Certificação

Aspectos gerais

Aspectos específicos

Ao término de cada unidade didática há uma certificação, assim que ocorra o aproveitamento dos estudos.

Concluídos os três módulos e elaborado a produção científica, há a certificação plena do curso.

Sistema de matrícula e desenvolvimento do curso

1. Para iniciar o curso é necessário fazer a matrícula, preenchendo a ficha de matrícula.
2. Imediatamente após o preenchimento do certificado de matrícula, o Instituto envia as instruções e os códigos de acesso a cada curso (*login* e senha de aluno), diretamente ao responsável pelo certificado de matrícula através de *e-mail*, no endereço apontado pelo mesmo, a fim de garantir a privacidade e autenticidade da operação.
3. Por tratar-se de uma matrícula virtual, o sistema foi preparado para aceitar os procedimentos de operacionalização e realização do curso por um prazo provisório de até 10 dias.

4. O aluno, devidamente habilitado a realizar o respectivo curso (tendo conhecimento do seu respectivo *login* e senha), deverá, por sua vez, acessar o ambiente privativo de ensino a distância através da "caixa de *login*" encontrada no *website* do SEstatNet.
5. Utilizando os recursos disponibilizados através da Internet, o aluno estuda em casa, no local de seu trabalho ou onde preferir, acompanhando o conteúdo programático.
6. No final de cada módulo há um conjunto de exercícios que devem ser feitos pelo aluno e enviado ao LEA por *e-mail* (em condições excepcionais, fax ou correio) permitindo a aferição de aprendizagem.
7. Caso não seja alcançado o mínimo exigido a equipe do Instituto encaminhará orientação para estudos complementares, ocorrendo nova(s) avaliação(ões) até que se atinja o aproveitamento previsto.
8. Após a conclusão de todos os módulos, o aluno deverá elaborar uma produção científica (monografia ou trabalho similar). Para tanto, a partir do 2º módulo, já terá a orientação para o desenvolvimento.
9. Ao término de cada unidade há a certificação parcial e ao final do curso, a certificação plena.

IMPORTANTE

- Com relação às normas de utilização no âmbito tecnológico e ambiente da Internet deverão ser observadas as condições de uso e os regulamentos específicos do "LEA", responsável pelas ferramentas de educação à distância na área privativa do *site*. Essas informações devem ser lidas previamente pelo usuário que poderá acessá-las através deste *link*: REGULAMENTO.

Informações Complementares

Sede do LEA:

Universidade Federal de Santa Catarina
Depto de Informática e Estatística
Laboratório de Estatística Aplicada
Cidade Universitária - Trindade

CEP 88040-900 - Florianópolis - Santa Catarina - Brasil

Telefone: (48) 231-7747

<http://www.sestatnet.ufsc.br> - e-mail: sestat@sestatnet.ufsc.br

Se o aluno concordar com a regulamentação exposta ele fica devidamente habilitado a realizar o curso. Assim, tendo conhecimento do seu *login* e senha poderá por sua vez, acessar o ambiente privativo de educação à distância através da "caixa de *login*" encontrada na interface principal do sistema. A partir daí deverá realizar as atividades previstas no curso, desde que respeite as normas estabelecidas.

Utilizando os recursos disponibilizados através da Internet, o aluno estuda em casa, no local de seu trabalho ou onde preferir, acompanhando o conteúdo programático.

Os itens descritos até agora se referem ao controle externo do ambiente: controle de acesso e como os dados serão armazenados no Banco de Dados. Uma base de dados é composta por colunas de dados que são denominadas de variáveis. A manipulação de uma base de dados está diretamente ligada à maneira como se pode trabalhar com as variáveis que a compõem. No item 5.6 serão definidos alguns aspectos que deverão ser levados em conta no desenvolvimento dessa metodologia como o tipo de variável e suas possíveis transformações.

5.6 Variáveis

Pode-se chamar de variáveis às características que podem ser observadas ou medidas nos elemento da população, sob as mesmas condições. Uma variável observada ou medida num elemento da população deve gerar um e apenas um resultado [BARBETTA, 2003].

Algumas variáveis como sexo, educação e estado civil apresentam como possíveis realizações, uma qualidade ou atributo do indivíduo pesquisado, ao passo que outras como número de filhos, salário e estatura apresentam como possíveis realizações, números resultantes de uma contagem ou mensuração. As variáveis do primeiro tipo são chamadas *qualitativas* e as do segundo tipo são chamadas *quantitativas* [BUSSAB, 1997].

Dentre as variáveis *qualitativas*, ainda pode-se fazer uma distinção entre dois tipos:

- Variável qualitativa *nominal*, para a qual não existe nenhuma ordenação nas possíveis realizações.
- Variável qualitativa *ordinal*, para a qual existe certa ordem nos possíveis resultados.

De modo análogo, as variáveis quantitativas podem sofrer uma classificação dicotômica:

- Variáveis quantitativas *discretas*, cujos possíveis valores formam um conjunto finito ou enumerável de números e que resultam freqüentemente de uma contagem, como por exemplo, número de filhos.
- Variáveis quantitativas *contínuas*, cujos valores formam um intervalo contínuo de números reais, como por exemplo, a estatura e o peso de um indivíduo.

5.6.1 Mensuração

Mensuração é o processo que consiste em atribuir valores a objetos ou observações. O grau de mensuração que se atinge é função das regras sob as quais se faz a atribuição de números. As operações e relações utilizadas na obtenção dos conjuntos de valores definem e limitam as manipulações e operações permissíveis ao lidar-se com os valores; tais manipulações e operações devem ser as da estrutura numérica à qual a mensuração é isomorfa [SIEGEL, 1975].

5.6.2 Transformações de variáveis

A análise estatística de dados requer muitas vezes que algumas transformações sejam feitas sobre as variáveis a serem analisadas. Essas transformações dependem do tipo de variável que se está trabalhando.

No caso de variáveis qualitativas nominais pode-se desejar reagrupar algumas categorias. As possíveis transformações para variável qualitativa ordinal seriam a de reagrupar categorias ou definir postos para elas. Quando a variável é quantitativa (contínua ou discreta) pode-se reagrupar, criar postos ou realizar alguma transformação.

Basicamente existem três situações em que a transformação é requerida em uma análise estatística [REBELO, 2004]. A primeira seria quando se deseja linearizar uma função. Nesse caso, conforme podem ser observadas na figura 5.7, as transformações são realizadas sobre a variável independente.

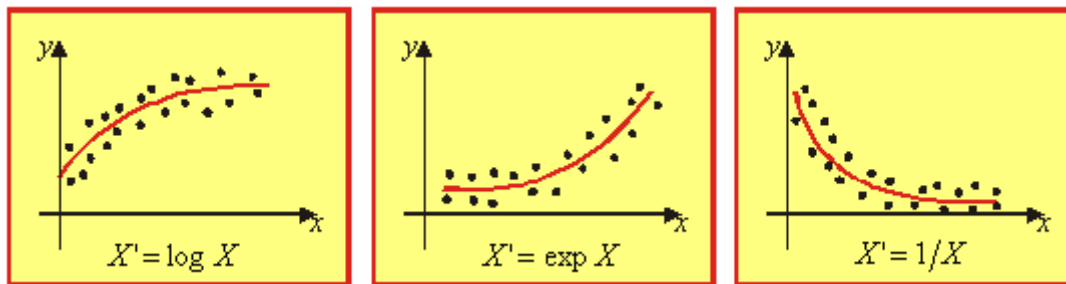


Figura 5.7: Funções linearizáveis e suas respectivas transformações.

O pressuposto para esse tipo de transformação é que os erros tenham uma distribuição aproximadamente normal e com variância constante. Nesse caso a variável dependente Y não é afetada pela transformação.

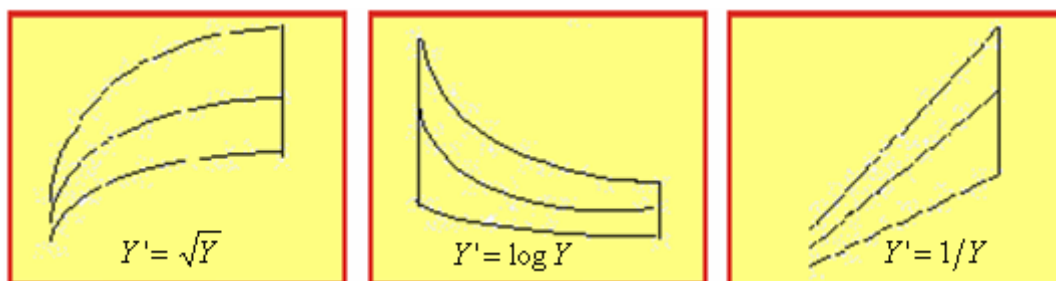


Figura 5.8: Funções cujo padrão de regressão é linearizável e suas respectivas transformações.

No segundo caso é quando aparecem juntas, tanto para a variável dependente e independente, a não normalidade dos erros e variâncias heterogêneas. Nesse caso pode-se observar na figura 5.8, que a transformação será feita sobre a variável dependente Y para modificar a sua forma e dispersão.

Para cada uma das transformações aplicadas é importante analisar os gráficos de dispersão e de resíduos para determinar a transformação ou transformações mais adequadas [NETER et al., 1996].

Um terceiro tipo de transformação é para eliminar o efeito da mensuração sobre as variáveis que estão envolvidas na análise, isto é, torná-las adimensionais. Esse tipo de transformação é denominado de normalização ou padronização pela normal e é realizada sobre a variável dependente e independente utilizando as seguintes fórmulas:

$$\frac{x_i - \bar{x}}{S_x} \quad \text{e} \quad \frac{y_i - \bar{y}}{S_y}$$

onde: \bar{x} = Média aritmética da variável independente x ,
 \bar{y} = Média aritmética da variável dependente y e
 S_x e S_y = Desvios-padrão correspondentes.

Portanto, tornar um valor adimensional é diminuir o valor médio e dividir pelo desvio padrão observado.

Pode-se ver a seguir que as operações envolvidas nessas transformações são basicamente funções que estão disponíveis na linguagem SQL de manipulação de banco de dados:

- **LOG10(X)**: retorna o logaritmo de X na base 10.
 Exemplos: `mysql> SELECT LOG10(2);`
`-> 0.301030`
`mysql> SELECT LOG10(100);`
`-> 2.000000`
`mysql> SELECT LOG10(-100);`
`-> NULL`
- **EXP(X)**: retorna o valor de e (base do logaritmo natural) elevado à potência de X.
 Exemplos: `mysql> SELECT EXP(2);`
`-> 7.389056`
`mysql> SELECT EXP(-2);`
`-> 0.135335`
- **SQRT(X)**: retorna a raiz quadrada de X (para um valor não-negativo).
 Exemplos: `mysql> SELECT SQRT(4);`
`-> 2.000000`
`mysql> SELECT SQRT(20);`
`-> 4.472136`

A Estatística é utilizada para transformar dados em informações sobre determinada realidade para resolver um problema ou tomar uma decisão [NASSAR, 1995]. A aplicação de métodos estatísticos adequados implica em conhecer o tipo de variáveis envolvidas na análise. Então na seção 5.7 será discutida a implementação de um módulo de tipificação de variáveis.

5.7 SETip

O SETip é um sistema especialista para tipificar dados de uma pesquisa e tem como objetivo oferecer a habilidade de tipificar dados de uma base de dados flexível de modo a possibilitar a identificação de informações erradas fornecidas pelos alunos.

Nesse item será dada uma descrição detalhada de como funciona internamente o SETip no processo de tipificação e aprendizagem. Maiores detalhes podem ser consultados em [SANTOS, 2001].

5.7.1 Processo de Tipificação

No processo de tipificação os dados são enviados a um agente que inicialmente verifica se os valores da variável são do tipo *string*. Se forem então esse agente tipifica a variável como sendo qualitativa. Caso contrário verifica se os valores são do tipo *flutuante* e se forem então a tipificação é feita como quantitativa.

Se não ocorreu a tipificação verifica se existe número negativo e se existir então a tipificação é feita como quantitativa. Se não existir negativo verifica se os dados são compostos por apenas dois valores distintos (0 e 1, 1 e 2 ou 100 e 200) e se isso ocorrer, então a tipificação é feita como qualitativa. Caso contrário enviar os dados a outro agente que irá transformar a variável em um *caso*.

Para transformar a variável em um *caso* é necessário encontrar parâmetros que combinados possam dizer se a variável é qualitativa ou quantitativa. Essa etapa é feita buscando algumas características básicas, com auxílio de outro agente, para ajudar na busca dos parâmetros. Quando esses parâmetros são encontrados, então a variável é transformada em um *caso* que será comparado com os existentes na Base de Conhecimento (BC), procurando-se o mais semelhante. Os *casos* da BC são armazenados em tabela com seis colunas, sendo que as cinco primeiras são os parâmetros e a última o tipo da variável – tipo igual a zero a variável é qualitativa e tipo igual a um a variável é quantitativa.

5.7.2 O conhecimento no SETip

Para que o SETip fosse capaz de reconhecer uma variável, quanto ao seu tipo (qualitativa ou quantitativa), foi necessário descobrir alguns padrões apresentados por várias variáveis. Para descobrir esses padrões, coletou-se um conjunto de variáveis

originadas dos diversos tipos de pesquisa – pesquisa social, agropecuária, médica e financeira.

Pela análise dos dados, foram detectadas algumas características básicas que diferenciam um tipo de variável da outra, como por exemplo, valores da variável sendo do tipo *string*, *flutuante*, negativos ou compostos por duas categorias apenas. De posse desse conhecimento, o sistema está apto a reconhecer algumas variáveis.

Para verificar se os valores de uma variável são do tipo *string* ou *flutuante*, o sistema usa funções da própria linguagem de programação. Para verificar negativos, basta fazer um teste do tipo, valor menor que zero. E para verificar se há apenas duas categorias, foi implementada uma operação que testa a igualdade entre as equações mostradas abaixo, ou verifica se o valor mínimo é igual a zero e o valor máximo é igual a um.

$$A = \frac{Min + Max}{2} \quad B = 1.5 * Min$$

onde: Min = Valor mínimo do conjunto de dados e

Max = Valor máximo do conjunto de dados.

Como esse conhecimento não era suficiente para tipificar um número satisfatório de variáveis, foi necessário descobrir padrões que tornassem isso possível. Com o uso de heurísticas, ferramentas de *Data Mining* e *softwares* estatísticos, chegou-se a cinco parâmetros que combinados, podem determinar, com certa precisão, se uma variável é qualitativa ou quantitativa.

5.7.3 Considerações de implementação

Essa base de conhecimento foi gravada como uma tabela do banco de dados. Ela é acessada pelo *software* SStatNet, que por sua vez verifica se as respostas dos alunos estão de acordo com as características das variáveis envolvidas na análise.

Outra opção de uso disponibilizada com o desenvolvimento dessa metodologia é realizar a tipificação das variáveis antes de acessar o *software* SStatNet. Para usar essa opção foi disponibilizado um *menu* que realiza a tipificação automática das variáveis da base de dados que o aluno está trabalhando. Se o aluno não concordar com a tipificação feita ele pode refiná-la utilizando a caixa de seleção existente no topo de cada coluna de dados. Na figura 5.9 essas características podem ser observadas.

No processo de aprendizagem os dados são enviados a um agente que irá transformar a variável em um *caso* e enviará a outro agente que irá colocar o novo conhecimento na BC. A alteração da BC só poderá ser feita por um especialista ou alguém que tenha autorização para isso. O acesso só é permitido através de uma senha que foi colocada para que a sua integridade seja mantida.

Essa é mais uma interface restrita do ambiente e só é disponibilizada para os usuários com autorização.

O SETip tem um bom desempenho no sentido de classificar variáveis quanto ao seu tipo (qualitativa ou quantitativa). Ele pode ser melhorado à medida que seja utilizado, pois foi planejado e implementado com a capacidade de aprender com os seus erros.

A seção 5.8 descreve sobre a interface gráfica que mostra a base de dados dos usuários do sistema e suas funcionalidades.

5.8 Tabela de Dados

A tabela de dados, como pode ser observada na figura 5.9, se refere ao conjunto de células usadas para entrar e editar dados. A primeira linha da tabela de dados apresenta o nome das colunas. Pode acontecer que alguns valores de dados fornecidos nesta linha não serão usados em análises subsequentes.

A coluna mais a esquerda (primeira coluna) serve como identificação das linhas na tabela de dados.

A tabela de dados apresenta barras de rolagem vertical na direita e horizontal em baixo para realizar o movimento dentro da tabela. As teclas “*Page Down*” e “*Page Up*” podem ser utilizadas também para a rolagem da tabela.

The screenshot shows a web browser window titled "Wilson - Microsoft Internet Explorer" displaying the SEstatNet application. The application has a menu bar with "SEstatNet", "Dados", "Stat", "Gráficos", "Ajuda", and "Tipificação". Below the menu is a row of dropdown menus labeled "Tipo de Variável". The main area contains a data table with 17 rows and 7 columns. The columns are labeled "var1" through "var7". The first row is highlighted in blue.

Tipo de Variável							
Linha	var1	var2	var3	var4	var5	var6	var7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
14							
15							
16							
17							

Figura 5.9: Tabela de Dados.

O número de células apresentadas na tabela não pode ser incrementado pelo uso desses botões. Isso pode ser feito usando ações de teclado ou pelo redimensionamento da janela principal que contém a tabela de dados.

5.8.1 Características de uso

A célula que está destacada recebe entrada do teclado e algumas críticas são realizadas. A interface tenta carregar o número com precisão dupla, mas se isso não é possível ele armazena a entrada como uma string. Se for digitado "38sestat" será considerado apenas "38". Em outras palavras, strings que iniciam com números serão truncados depois do número. Ela usa também a notação científica sendo que "0,000001" pode ser mostrado como "1E-6". Todos os números são mostrados com precisão de ponto flutuante, mas eles são armazenados com precisão dupla.

Sempre se deve pensar como o cursor ocupando a posição à direita da célula destacada. Nesse momento, o cursor não pode ser posicionado manualmente dentro da célula destacada. A tecla "Backspace" pode ser usada para apagar valores digitados um a um, começando com o último valor digitado. A chave "Delete" pode ser usada para remover todos os valores em uma célula destacada.

Quando a célula perde o destaque, os valores digitados na célula podem ser formatados para uma melhor apresentação. No caso em que o número apresenta o valor “0,00000000012345” pode ser representado como “1,2345E-10”.

A largura de uma coluna será ajustada automaticamente se a informação de entrada excede o seu tamanho. Quando trabalhamos com grandes arquivos de dados essa característica de auto-ajuste deve ser desativada para melhorar o tempo de resposta.

Para alterar a célula que está destacada, ou clique o “*Mouse*” dentro de uma nova célula ou use as teclas “*Tab*”, “*Enter*” ou de movimentos. Se o usuário tentar mover para uma célula que não está completamente visível, a tabela de dados se posicionará de modo que ela fique totalmente visível.

A tecla de tabulação “*Tab*” move o destaque para a célula à direita e a tecla de retorno “*Enter*” move o destaque para a célula de baixo. As teclas de movimento (para cima, para baixo, para a direita e para a esquerda) mudam o destaque para a célula na direção selecionada.

Todas essas teclas podem ser usadas para aumentar o número de células exibidas na tabela. Por exemplo, se a tabela de dados está atualmente mostrando 5 colunas e o usuário aperta a tecla de tabulação enquanto a célula de destaque é a quinta coluna, a interface gráfica irá mostrar a sexta coluna com a célula de destaque na sexta coluna.

5.8.2 Menu Dados

O *menu* Dados disponibiliza todas as possíveis formas de manipulação dos dados necessárias para a implementação dessa metodologia. Entre elas podemos destacar a aquisição de dados, operações possíveis de serem feitas na tabela de dados e as que podem ser feitas nas colunas de dados.

O detalhamento dessas operações será apresentado na seção 6.4 do próximo capítulo em Modelo de Comportamento.

A seguir apresenta-se a arquitetura computacional necessária para a implementação dessa metodologia.

5.9 Arquitetura Computacional

A arquitetura computacional para a implementação do módulo de gerenciamento de base de dados do ambiente para ensino-aprendizagem de estatística necessita de dois servidores principais e uma interface Web.

Os servidores são: o servidor de banco de dados e o servidor de páginas *HTML* com *scripts*. De um lado estão estes servidores, e do outro está o cliente com um *browser* que suporta *HTML*.

Como se observa, a arquitetura se divide em três níveis distintos: o nível em que se encontra o servidor de páginas *HTML* com *script*, o nível do servidor de banco de dados, que fica somente encarregado do tratamento das consultas e atualizações nas tabelas, e o nível do cliente, que interpreta somente código *HTML*.

Através da linguagem *script* e de expressões em linguagem SQL utilizadas nas páginas, o servidor de páginas consegue a interação necessária com o servidor de banco de dados para a realização da funcionalidade do sistema. Todo *script* de acesso ao banco de dados é, portanto, executado no servidor de páginas. O cliente não tem qualquer conhecimento deste código *script* e nem da existência do servidor de banco de dados.

Os servidores que fazem parte do módulo de gerência de base de dados do ambiente SStatNet estão instalados no servidor de aplicações do Laboratório de Estatística Aplicada. Trata-se de uma máquina com sistema operacional linux [LINUX, 2004], onde o servidor de Banco de Dados utilizado é o MySQL [MySQL, 2004] e o servidor de páginas é o Apache [APACHE, 2004].

A segurança é uma das principais preocupações dos administradores de rede hoje em dia. As transações pela Internet são aceitas pelos seus usuários somente quando é oferecida uma grande margem de segurança, e como a Internet é um dos meios mais promissores para a comunicação dos tempos de hoje, vários estudos estão sendo feitos para cada vez mais aumentar a segurança dos dados na rede. A seção seguinte apresenta alguns aspectos de segurança que devem ser levados em consideração na implementação dessa metodologia.

5.10 Mecanismos de segurança

Conectar um servidor ou uma rede local de uma corporação diretamente à Internet é uma ação que deve ser cuidadosamente planejada, pois os riscos inerentes são consideráveis; temos exemplos diários em jornais e em notícias pela Internet toda sobre o que *hackers* mal intencionados podem fazer se eles encontrarem uma abertura. Para melhorar este panorama, é necessário minimizar riscos de ataque.

Vários fatores precisam ser considerados ao se definir os procedimentos de segurança contra as ameaças em potencial. Deve-se definir exatamente o que se está tentando proteger, de quem e do que é preciso protegê-lo. A metodologia de segurança deve ser direcionada para [PEDATELLA, 2000]:

- ✓ Evitar danos e prejuízos intencionais a arquivos e sistemas.
- ✓ Evitar danos e prejuízos acidentais a arquivos e sistemas.
- ✓ Limitar os prejuízos que a remoção ou danos a arquivos possam causar.
- ✓ Proteger a integridade e a privacidade dos dados.
- ✓ Evitar os acessos não autorizados ao sistema.
- ✓ Propiciar sistemas de recuperação de desastres.

Deve-se ter em mente que a segurança também deve ser implementada a nível físico, como a certificação de que o servidor está fisicamente protegido, utilização de dispositivos de proteção da rede elétrica, tolerância à falhas no servidor, *backups*, e outros.

Para a proteção da rede, além do nível físico, podem ser implementadas algumas técnicas de isolamento. O isolamento de protocolo faz com que apenas uma parte da rede se comunique com a Internet usando o TCP/IP e o resto da rede se comunique em outro protocolo, como o IPX/SPX ou o NetBEUI.

Outro mecanismo de segurança que está sendo utilizado no ambiente SEstatNet é o *firewall*, que é tanto um dispositivo de *hardware* (um roteador) como um pacote de *software* rodando em um computador especialmente configurado que fica entre uma rede segura e uma rede insegura. O *firewall* realiza várias tarefas importantes, como evitar o acesso não autorizado de pessoas na rede, limitar o tráfego de entrada e de saída, autenticar os usuários, registrar as informações sobre tráfego e produzir relatórios. Ele monitora todo o tráfego que flui entre duas redes e pode bloquear completamente certos tipos de tráfego; se o *firewall* for bem implementado, um intruso não poderá chegar até a rede interna e ela ficará protegida.

Para o projeto SEstatNet, a segurança é fundamental, pois existe o controle de acesso e informações importantes que são passadas pela rede. Para os cursos de ensino à distância criados nesta plataforma serão tomadas outras medidas de segurança como encriptação, expiração para a mudança de senhas e áreas de acesso restrito.

5.11 Considerações finais

A metodologia aqui desenvolvida disponibiliza as funcionalidades de um ambiente de educação à distância e contando com um sistema de matrículas de alunos (*login/senha*), recursos para manipulação de base de dados via Web, além de outros recursos como repositório de documentos e lista de *e-mails* de usuários envolvidos no processo de ensino-aprendizagem.

Isso permite que alunos e professores troquem informações sobre os conceitos de estatística abordados ao longo do curso de uma forma semelhante ao que ocorre no ensino presencial [NOAL, 2002].

O capítulo seguinte apresentará a implementação da metodologia descrevendo o modelo de contexto e modelo de comportamento.

6 MODELO DE CONTEXTO E DE COMPORTAMENTO

Nesse capítulo apresenta-se o modelo de contexto do sistema de matrícula e o modelo de comportamento para o módulo de gerenciamento de base de dados.

O modelo do contexto é a descrição das principais atividades que devem ser atendidas pelo sistema. É constituído da lista de eventos e do diagrama do contexto. A seção 6.1 descreve a lista de eventos para o sistema de matrícula dos alunos que usam o ambiente de ensino à distância e a seção 6.2 apresenta o diagrama do contexto correspondente a essa lista de eventos.

O modelo de comportamento apresentado na seção 6.3 descreve as principais operações do sistema e o seu comportamento de acordo com as ações que são executadas.

6.1 Lista de Eventos do Sistema de Matrícula

A lista de eventos é elaborada partindo da lista das necessidades relacionadas com o aplicativo. Um evento corresponde a uma entidade externa. Ele gera um estímulo que por sua vez dispara uma atividade dentro do Sistema. A atividade é executada e retorna uma resposta planejada para a entidade externa que gerou o evento [VIRGILIO, 2000].

Abaixo se enumera a lista de eventos do sistema de matrícula:

1. O tutor necessita da relação de seus alunos.
2. O tutor necessita enviar mensagem para seus alunos.
3. O tutor necessita liberar matrícula de seus alunos.
4. O tutor necessita cancelar matrícula de alunos.
5. O aluno necessita efetuar a matrícula.
6. O aluno necessita receber confirmação de matrícula (*login* e senha).

6.2 Diagrama de contexto

O diagrama de contexto mostra a entidade responsável por um evento externo com um retângulo e os fluxos como setas unidirecionais que partem da entidade externa [KERN, 1994].

A figura 6.1 mostra as interações que ocorrem entre cada um dos personagens externos (aluno e tutor) no sistema de matrícula.

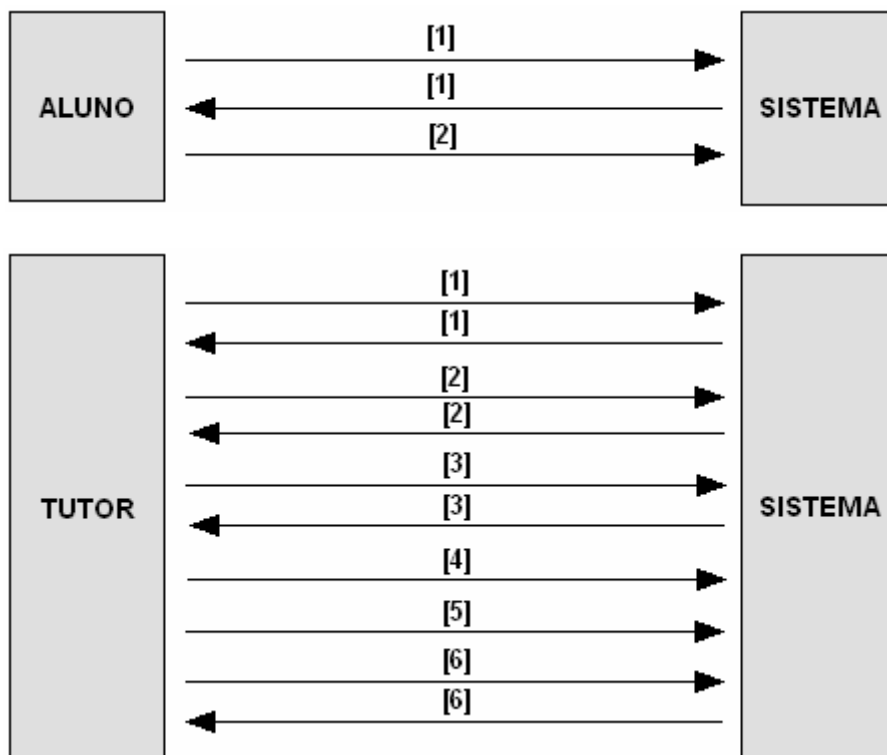


Figura 6.1: Diagrama de Contexto: Aluno – Sistema e Tutor – Sistema.

No diagrama de contexto acima, as entidades aluno e tutor e o sistema de matrícula em que essas entidades interagem estão representadas por um retângulo. A descrição dos eventos externos (estímulo e resposta) está nas tabelas 6.1 e 6.2 correspondentes ao diagrama de contexto criado.

Tabela 6.1: Descrição do contexto: Aluno – Sistema.

Número	Estímulo	Resposta
1	Solicita matrícula	Matrícula
2	Envia matrícula	

Tabela 6.2: Descrição do contexto: Tutor – Sistema.

Número	Estímulo	Resposta
1	Solicita relação de alunos.	Relação de Alunos do tutor.
2	Solicita acessos de alunos.	Relação de Acessos dos Alunos.
3	Solicita liberação de matrícula	Libera matrícula para o aluno
4	Envia mensagem	
5	Excluir alunos do curso	
6	Solicita estatística de tempo conectado de alunos.	Relação com dados estatísticos sobre tempo conectado.

6.3 Modelo de Comportamento

Este modelo, como já foi dito, descreve as principais operações do sistema e o seu comportamento de acordo com as ações que são executadas. Para proporcionar melhor entendimento do módulo de manipulação de dados, algumas das telas que compõem o sistema são apresentadas aqui para que essas operações possam ser mais bem visualizadas e entendidas.

O *menu* DADOS na Tabela de Dados do ambiente proposto disponibiliza todas as possíveis formas de manipulação dos dados como pode ser visto na figura 6.2. Entre elas podemos citar:

- ✓ Aquisição de dados.
- ✓ Tabela de dados.
- ✓ Criação de colunas.

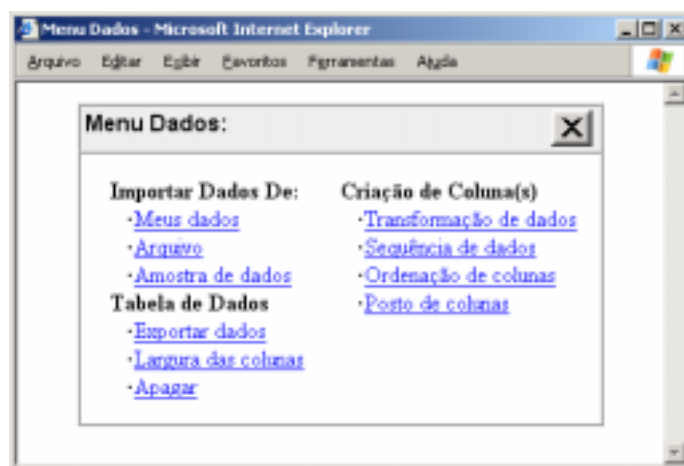


Figura 6.2: Tela do Menu Dados para seleção de operações.

Os dados podem ser importados para o ambiente de uma das maneiras seguintes:

- ✓ Pegar o arquivo importado previamente da lista do repositório Meus dados.
- ✓ Importar o arquivo de uma máquina local ou da rede.
- ✓ Escolher um exemplo de aplicação disponível em Amostra de dados.

O ambiente de gerenciamento de base de dados oferece vários procedimentos para adicionar novas colunas de dados ou executa uma variedade de transformações de dados

em colunas já existentes. Esses recursos serão apresentados a partir do item 6.3.6 e suprem as transformações de dados apresentadas no desenvolvimento da metodologia.

6.3.1 Escolher base de dados em Meus dados

Cada vez que se faz uma importação de dados para o ambiente de ensino-aprendizagem de estatística pela segunda opção de importação “Arquivo”, a base de dados será armazenada no repositório Meus dados.

A figura 6.3 mostra a situação em que não existe base de dados no repositório do aluno.

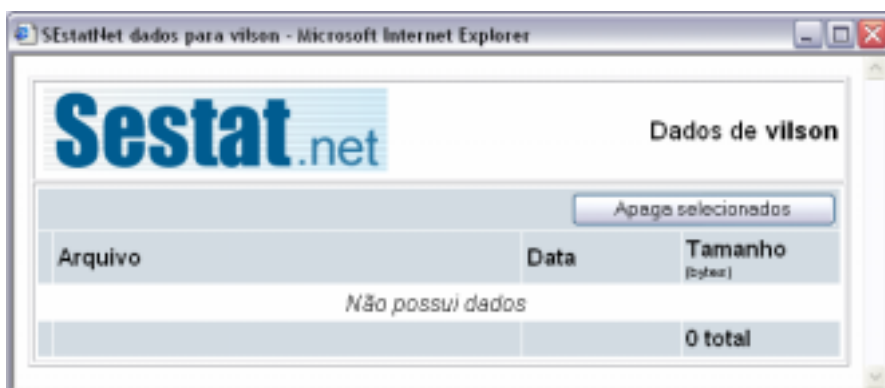


Figura 6.3: Aluno não possui base de dados armazenada em seu repositório.

E a figura 6.4 mostra a situação em que existe a base de dados PAP armazenada no repositório do aluno.

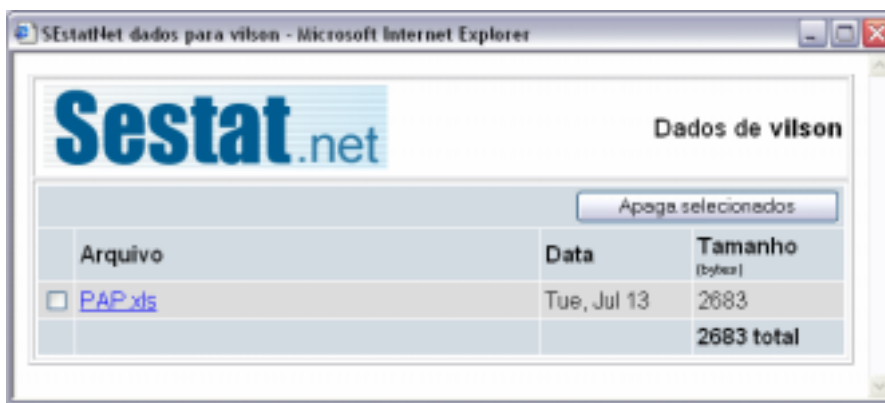


Figura 6.4: Aluno possui a base de dados PAP armazenada em seu repositório.

Quando existe arquivo na lista do aluno, para carregá-lo basta clicar no *link* do nome do arquivo, que ele será colocado na Tabela de Dados. Se o aluno deseja apagar

algum desses arquivos de sua lista ele deve selecioná-los marcando a caixa de seleção a esquerda do nome do arquivo e acionar o botão Apaga selecionados.

6.3.2 Carregar Dados de Arquivo ou URL

Os arquivos de dados podem ser importados de um computador local ou de um endereço www e devem ter o formato texto ou Excel (com extensão .xls).

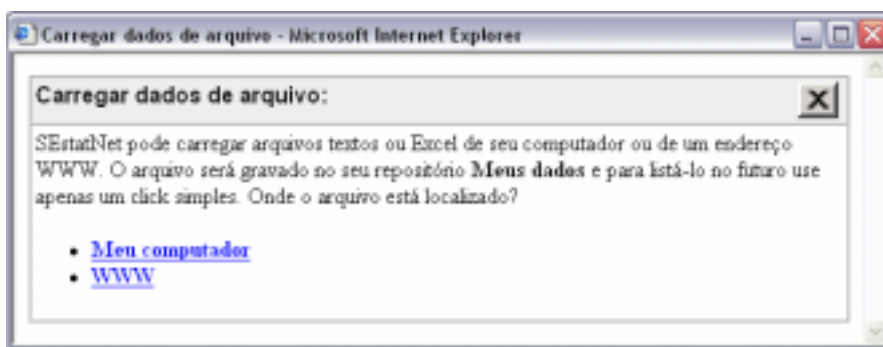


Figura 6.5: Tela de seleção da opção de carregar dados de arquivo.

Se na tela apresentada na Figura 6.5 o aluno escolher a opção de carregar o arquivo de uma máquina local, o ambiente disponibiliza a tela apresentada na Figura 6.6 onde o aluno deve informar o nome do arquivo, especificar se os nomes das variáveis estão na primeira linha do arquivo e selecionar o delimitador se o arquivo é do tipo texto.

Caso o aluno desconheça a localização do arquivo pode acionar o botão Procurar que a janela de procura padrão do sistema operacional utilizado será disponibilizada para o aluno especificar o local no seu computador onde o arquivo reside.

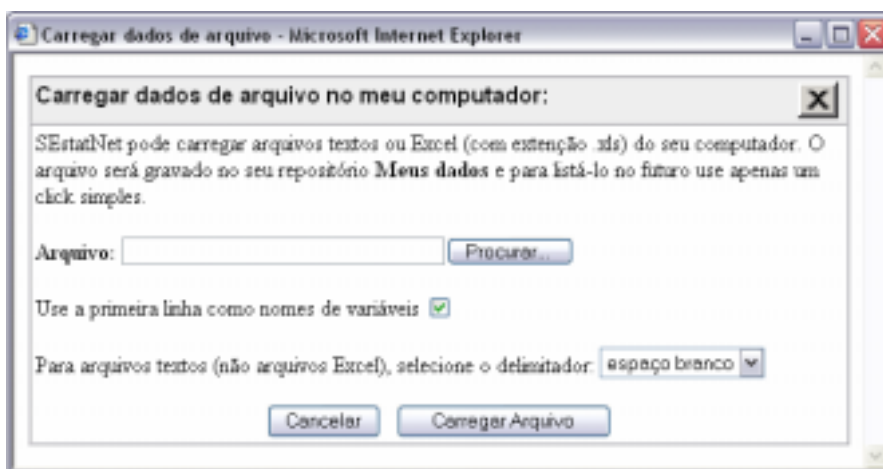


Figura 6.6: Tela para carregar dados de arquivo.

Após especificar os dados necessários o aluno deve acionar o botão Carregar Arquivo para efetuar a importação.

A tela da figura 6.7 apresenta os campos que o aluno deve preencher para carregar o arquivo de uma URL (*Uniform Resource Locator*).

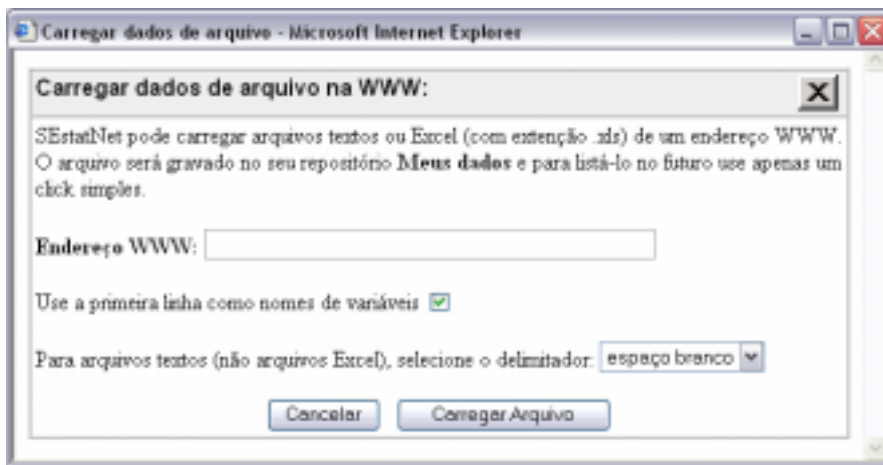


Figura 6.7: Tela para carregar dados de uma URL.

O aluno deve digitar o local da Internet na caixa de entrada “Endereço www”, como por exemplo: <http://www.sestatnet.ufsc.br/teste/meusdados.xls>. Se a primeira linha do arquivo contém o nome das variáveis, marque a opção “Use a primeira linha como nome das variáveis”. Clique no botão “Carregar Arquivo” e em pouco tempo, os dados irão aparecer na tabela de dados.

6.3.3 Amostra de Dados

Um grande número de amostras de dados será disponibilizado pelo Laboratório de Estatística Aplicada para que os usuários experimentem o ambiente SEstatNet.



Figura 6.8: Amostras de dados disponíveis para uso no SEstatNet.

As instruções de uso estão descritas na tela da figura 6.8 e as informações sobre os dados são apresentadas quando o aluno clicar no *link* info situado na linha que contém o nome do arquivo. Alguns arquivos de dados não são aplicáveis a certas análises ou gráficos. Deve-se ter a cautela para testar análise ou gráficos com dados apropriados.

A forma de apresentação da Tabela de Dados foi discutida no item 5.8. A seguir mostra-se outras operações que podem ser realizadas na Tabela de Dados que dão uma maior flexibilidade no seu uso.

6.3.4 Exportar Dados

Os dados armazenados em uma tabela de dados podem ser exportados no formato texto ou no formato HTML.

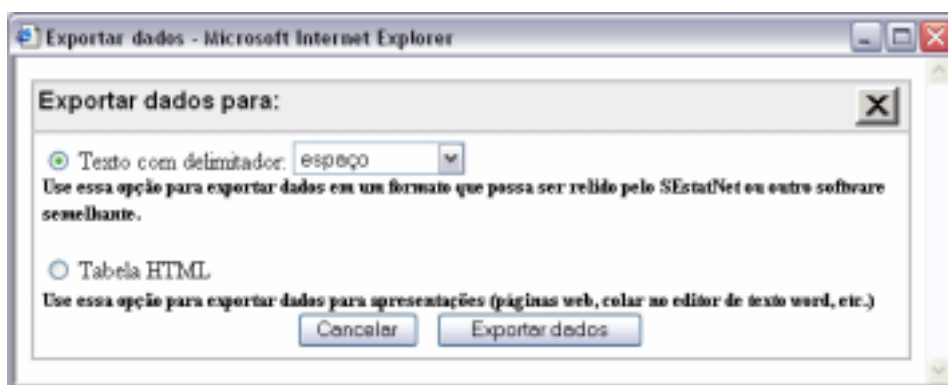


Figura 6.9: Tela utilizada para a exportação de dados.

Conforme a figura 6.9 para exportar dados com a opção texto, deve-se especificar o delimitador de dados a ser usado (o valor padrão será o espaçamento).

A opção texto deve ser usada para exportar dados no formato que possa ser lido por outros programas semelhantes ao SStatNet ou pelo próprio em futuros trabalhos. A opção HTML é usada para exportar dados com o propósito de publicação (páginas web, colar em programas de edição de texto, etc.).

6.3.5 Função “Apagar”

Essa função criará uma nova Tabela de Dados vazia. Todos os resultados de análise e gráficos não serão afetados porque são apresentados em janelas separadas.

6.3.6 Transformação de Dados

Com esta opção é oferecida muita flexibilidade para fazer transformações de dados. O usuário pode digitar uma expressão diretamente na caixa de transformação usando operações matemáticas básicas e/ou funções como logaritmo, média, soma, desvio padrão, etc. A figura 6.10 mostra a tela que apresenta essas opções de transformação de variáveis.

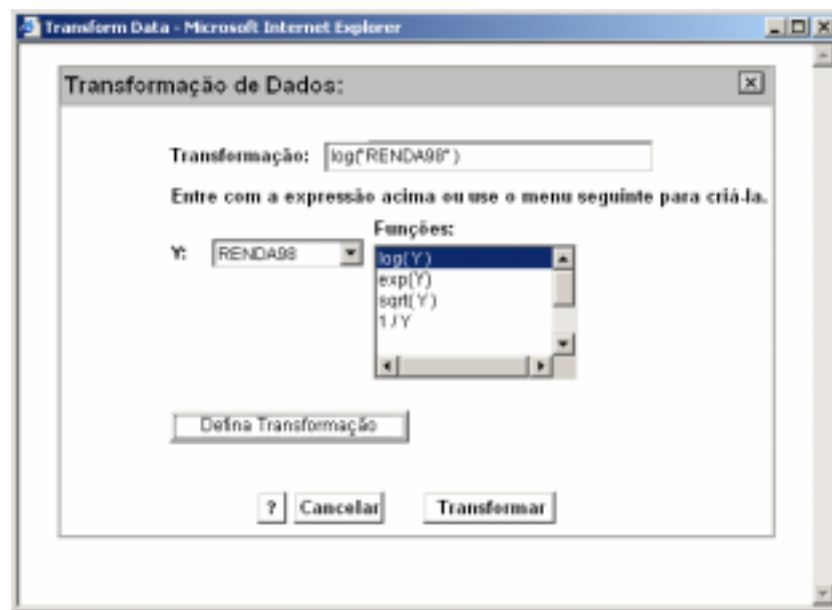


Figura 6.10: Tela para transformação de dados.

Pode-se também construir uma expressão de um “Menu” de transformações simples. A opção de digitar dá mais recursos para o usuário, enquanto que o menu de opções permite que usuários menos familiarizados com expressões matemáticas façam uso delas.

A lista de transformações simples inclui: $\log(Y)$, \sqrt{Y} , $1/Y$ e $\exp(Y)$. Em cada caso, a coluna necessária (Y) deve ser selecionada.

Quando o botão “Faça transformação” for acionado, a sintaxe matemática formal será colocada na caixa de transformação e clicando o botão “Transformação” serão computados os valores para a nova coluna transformada. Essa coluna será adicionada à direita das colunas existentes na Tabela de Dados.

6.3.7 Seqüência de Dados

Com essa opção o usuário tem a facilidade de criar uma nova coluna de dados seqüenciais personalizados. A figura 6.11 mostra a tela de definição dos dados da seqüência a ser criada.



Figura 6.11: Tela de definição da seqüência a ser gerada.

Os parâmetros de definição da seqüência numérica são os seguintes:

- ✓ Na caixa de entrada DE, especifique o valor numérico do início da sucessão.
- ✓ Na caixa de entrada ATE, especifique o valor numérico do fim da seqüência.
- ✓ Na caixa de entrada POR, especifique o valor numérico do incremento.
- ✓ Especifique o número de vezes para repetir cada valor na sucessão.
- ✓ Especifique o número de vezes para repetir a sucessão.

Clique o botão “Criar sucessão” para gerar uma nova coluna na Tabela de Dados contendo os valores da sucessão.

6.3.8 Classificação de Dados

Esta opção permite classificar colunas numéricas ou textuais em ordem ascendente ou descendente. A figura 6.12 mostra a tela de entrada dos dados de classificação de coluna.

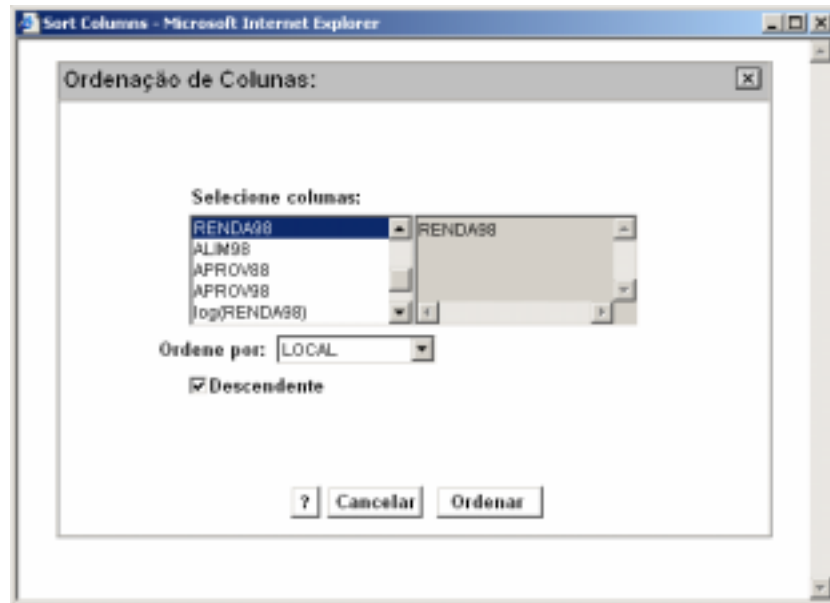


Figura 6.12: Tela de definição da classificação da coluna.

Para classificar uma coluna deve-se:

- ✓ Selecionar a coluna a ser classificada.
- ✓ Selecionar a coluna a ser usada para definir a nova ordenação em “Ordene por”. Essa definição é opcional.
- ✓ O padrão é classificar em ordem ascendente. Para mudar para ordem descendente deve-se marcar a caixa “Descendente”.

Clicando o botão “Ordenar” cria-se uma nova coluna para cada coluna selecionada.

6.4 Considerações finais

Nesse capítulo foi apresentada a implementação da metodologia de gerenciamento de base de dados para a análise estatística.

Outros recursos podem ser acrescentados aos atuais para que o usuário tenha maior flexibilidade no manuseio de sua base de dados. Alguns desses recursos serão descritos em sugestões para trabalhos futuros.

7 CONSIDERAÇÕES FINAIS

O principal objetivo deste projeto foi desenvolver uma metodologia de gerenciamento de base de dados, permitindo que usuários alterem suas próprias bases de dados na Web.

Neste sentido, foi feito o projeto de administração dos dados necessários para a implementação de um sistema de gerenciamento de base de dados. Todos os aspectos da modelagem e evolução do projeto até a implementação do sistema tiveram um planejamento e um embasamento teórico auxiliar.

7.1 Conclusão

As formas de interatividade entre educador/aluno foram planejadas de modo cuidadoso, para que o módulo de gerenciamento de base de dados utilizado tivesse um *feedback* esperado. A proposta adotada nesse desenvolvimento foi a implementação em ambiente *Unix*, usando a interface CGI em conjunto com a linguagem Perl. A linguagem Java também se mostrou bastante vantajosa para algumas tarefas realizadas, mas os programas feitos em Java executam mais lentamente, enquanto Perl é bastante rápido.

O desenvolvimento do sistema de matrícula busca permitir o controle de acesso dos usuários no ambiente de ensino-aprendizagem. Essa página fica protegida dentro do sistema e só é apresentada ao usuário mediante a confirmação da identificação e senha do usuário. Se o usuário é novo, ele deve se registrar no sistema usando o *link* existente na página principal.

A maioria dos usuários de computador está familiarizada com o uso de uma planilha eletrônica. Então a metodologia procurou desenvolver uma interface gráfica semelhante a uma planilha que oferece recursos adicionais de gerência de dados pela Web. A interface apresenta na primeira coluna a identificação das linhas e, na primeira linha, a identificação das colunas. As colunas são nomeadas com nomes padrões ou com nomes de variáveis dadas pelos usuários. Na idealização dessas interfaces foram considerados todos os conceitos estudados e a tendência atual no que diz respeito a *sites* educacionais.

Para realizar uma análise estatística os dados dos usuários têm que estar armazenados no banco de dados. Essa tarefa é realizada com recursos de importação implementadas no sistema. Os dados podem estar no computador local ou no repositório do usuário. Os dados podem também ser exportados no formato texto para serem usados por outro programa ou, no formato Web, para serem colados em documentos para apresentação.

Dentre os recursos adicionais implementados na planilha de dados, destacam-se a tipificação automática de variáveis (SETip) e facilidades de criação e transformação de colunas de dados com funções logarítmica, exponencial, inversa, raiz quadrada e a padronização pela normal, que são operações essenciais no ensino-aprendizagem de estatística.

A tipificação de dados permite reconhecer o tipo de variável, classificando-a como qualitativa ou quantitativa, de acordo com alguns padrões descobertos nos dados. Apesar de não ser o único aspecto para se decidir sobre o método estatístico a ser empregado na análise dos dados, a identificação do tipo de variável é um aspecto importante para essa decisão.

As transformações apresentadas são realizadas através de recursos disponíveis na linguagem SQL para banco de dados. O que o módulo faz é permitir a criação e execução do comando SQL com as características selecionadas e aplicá-lo na base de dados do usuário. Nessa primeira etapa do trabalho, foram implementadas somente as transformações básicas requeridas em uma análise estatística de dados, mas que podem ser ampliadas muito facilmente com outros tipos de transformações.

Essa metodologia foi validada subjetivamente por especialistas em análise de dados. No entanto, necessita ser validada objetivamente no ambiente de ensino-aprendizagem de estatística do SEstatNet. Isso só será possível de ser realizado com a conclusão do desenvolvimento da nova versão do *software* SEstatNet para a Web e, conseqüente, aplicação nos próximos semestres letivos.

A seguir destacam-se algumas contribuições que essa metodologia ao ser implementada pode trazer:

- Para usuários iniciantes em análise estatística de dados poderão realizar a sua análise em um ambiente muito amigável e com acompanhamento através de uma Ajuda em tempo real sobre os métodos que esteja usando.

- Para usuários experientes o módulo apresenta uma série de ferramentas que poderão ser utilizadas sem gasto de tempo em programação como na maioria dos *softwares* comerciais.
- A maioria dos recursos disponibilizados com essa metodologia tem como objetivo facilitar a entrada e manipulação de dados dos usuários para serem usados em um programa de análise estatística. Os dados estando no mesmo ambiente de análise farão com que o tempo gasto no ensino-aprendizagem de estatística seja otimizado.
- Embora essa metodologia esteja sendo validada em um curso de ensino-aprendizagem de estatística, pode ser facilmente implementada em cursos técnicos de outras áreas de aplicação.

7.2 Sugestões para trabalhos futuros

Durante a fase de implementação desta metodologia, observou-se que alguns aspectos importantes em um módulo de gerenciamento de base de dados não foram levados em consideração, mas tal fato se dá devido à complexidade de um projeto como esse. Para esgotar todos os aspectos necessários seria preciso um maior tempo de trabalho, mas futuros alunos poderão dar continuidade complementando o modelo que foi definido e utilizando todas as informações contidas neste documento.

Novas funcionalidades podem ser implementadas na Tabela de Dados do ambiente de gerência de base de dados. Um tipo de manipulação muito útil para o ambiente seria a Seleção de Linhas. Os valores em uma linha da Tabela de Dados poderiam ser destacados clicando o mouse dentro da célula que contém a identificação da linha. Clicando novamente na célula de identificação da linha, a mesma perderia a seleção. Todas as linhas destacadas poderiam ser liberadas clicando-se no rótulo Linha na linha de identificação das colunas.

Um dos agentes que o SStatNet apresenta é o controle de *log* que é um mecanismo usado pelo tutor para o acompanhamento dos caminhos percorridos por cada aluno e que possibilita a avaliação do seu desempenho e a identificação das principais dificuldades enfrentadas no seu processo de ensino-aprendizagem. Assim esse agente precisa ser implementado no módulo de manipulação de base de dados com esse mesmo objetivo.

No sistema de gerenciamento de base de dados algumas funcionalidades semelhantes ao registro de base de dados dos usuários poderiam ser implementadas como, por exemplo, guardar as modificações que os alunos realizam em sua base de dados. Essas modificações englobariam a geração de colunas automáticas pelo sistema e transformações realizadas nas colunas da base de dados.

Os sistemas de gerenciamentos de banco de dados estão em constante evolução em função das novas perspectivas de uso que se apresentam. Novas tecnologias como banco de dados orientado a objetos, objeto-relacional e a comunicação com banco de dados através da Web têm surgido para atender as novas necessidades de informação. Grandes empresas fabricantes de banco de dados, como o *Oracle* e a *Microsoft*, estão criando novas ferramentas que permitem e facilitam o acesso aos dados na Web. O uso dessas novas ferramentas é essencial para o bom desempenho e aprimoramento do sistema.

REFERÊNCIAS

- ABBEY, Michael; COREY, Mike; ABRAMSON, Ian. **Oracle9i – Guia Introdotório – Aprenda os fundamentos do Oracle 9i**. Editora Campus, Rio de Janeiro – RJ, 2002.
- APACHE – **Servidor de HTTP**: <http://www.apache.org/> . Acesso: Julho de 2004.
- BARBETTA, Pedro A. **Estatística Aplicada às Ciências Sociais**. 5ª ed. Editora da UFSC, Florianópolis, 2003.
- BUSSAB, Wilton O., MORETTIN, Pedro A. **Estatística Básica**. Editora Saraiva, 5ª. Edição. São Paulo, 2003.
- CATAPAN, Araci H. **O Ato Pedagógico: a Construção do Conceito**. DOIS PONTOS, vol. 5, nº 35. Nov. e Dez., p. 67 – 69, 1997.
- CATAPAN, Araci H.; FIALHO, Francisco A. P. **Tertium: o novo modo do ser, do saber e do aprender (construindo uma taxionomia para medição pedagógica em tecnologia de comunicação digital)**. [240 f.] Tese (Doutorado em Engenharia de Produção). Curso de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina. Florianópolis, 2001.
- CECHINEL, Cristian; AMORIM, Lidiane. **Sistema especialista de apoio ao Ensino de Estatística**. Monografia (Bacharel em Ciências da Computação). Curso de Bacharelado em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 1997.
- CECHINEL, Cristian et al. **Concepção e Implementação de um Ambiente de Ensino de Estatística**. In: PRESTA – Conferência Internacional Experiências e Perspectivas do Ensino da Estatística: desafios para o século XXI. Florianópolis, 1999.

- CODD, Edgar. F. **Further Normalization of the Data Base Relational Model**, in Rustin, 1972.
- CPAN – **Comprehensive Perl Archive Network**. <http://www.cpan.org>. Acesso: Maio de 2004.
- DB DESIGNER, Fabulous Force. **FabFORCE.net**. Versão 4.0.5.6 Beta. <http://fabforce.net>. Acesso: Junho de 2004.
- DEAN, Andrew G. **Open Source Software for Public Health Epidemiology**. Última revisão: Janeiro de 2003. Última revisão na Internet: Fevereiro de 2004. <http://www.openepi.com/Jan2004/menu/OpenEpiMenu.htm> . Acesso: Julho de 2004.
- DEITEL, Harvey. M.; DEITEL, Paul J. **JAVA Como Programar**. Bookman, 4ª ed., São Paulo, 2003.
- DIAS, Kirliam M. **Sistema especialista para auxílio ao Ensino de Estatística**. (168 f.) Monografia (Bacharel em Ciências da Computação). Curso de Bacharelado em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2001.
- ELMASRI, Ramez ; NAVATHE, Shamkant B. **Sistemas de Banco de Dados – Fundamentos e Aplicações**. Editora LTC, 3ª ed., Rio de Janeiro, 2002.
- FAGIN, Ron. **Multivalued Dependencies and a New Normal Form for Relational Databases**. TODS, 2:3, 1977.
- FAGIN, Ron. **Normal Forms and Relational Database Operators**. In SIGMOD, 1979.
- FANDERUFF, Damaris. **Oracle 8i - Utilizando SQL*Plus e PL/SQL**. Makron Books, 1ª ed., São Paulo, 2000.

FANDERUFF, Damaris. **Solução de Data Warehouse para possibilitar a análise estratégica integrada e distribuída dos indicadores de desempenho (IDS) do Estado de SC.** Dissertação (Mestrado em Engenharia de Produção). Curso de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina. Florianópolis, 2002.

GNU – **General Public License.** <http://www.gnu.org/>. Acesso: Janeiro de 2004.

HEUSER, Carlos A. **Projeto de Banco de Dados.** Publicação de comunicação de intenção de publicação, 1998.

JAVA – **JavaServer Pages Technology** <http://www.sun.com/software/learnabout/java/> . Acesso: Janeiro 2004.

KERN, Vinícius M. **Banco de Dados Relacionais: Teoria e prática de projeto.** Editora Érica, São Paulo, 1994.

KIM, Worley. **Modern Database Systems: The Object Model, Interoperability, and Beyond.** ACM Press, Addison-Wesley, 1995.

KORTH, Henry F.; SILBERSCHATZ, Abraham. **Sistema de Banco de Dados.** Makron Books, 3ª ed., São Paulo, 2001.

LINUX – **Sistema Operacional Linux.** <http://www.linux.org/> . Acesso: Julho de 2004.

LUCHESE, Cipriano C. **Democratização da educação: ensino à distância como alternativa.** Tecnologia Educacional números 89/90/91 jul/dez, ABT, Rio de Janeiro, 1989.

MORISSEAU-LEROY, Nirva; SOLOMON, Martin K.; BASU, Julie. **ORACLE8i – Programação de Componentes Java com EJB, CORBA e JSP.** Editora Campus, Rio de Janeiro, 2001.

MySQL – **Servidor de BD: MySQL.** [http://www.mysql.org.](http://www.mysql.org/) Acesso: Julho 2004.

- NAKAZAWA, Carlos A.; MARAFON, Márcio J. **SEstatNet – Ensino de Estatística mediado por Computador**. Monografia (Bacharel em Ciências da Computação). Curso de Bacharelado em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2003.
- NASSAR, Silvia M. **Sistema Estatístico Inteligente para Apoio a Pesquisas Médicas**. Tese (Doutorado em Engenharia Elétrica). Curso de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina. Florianópolis, 1995.
- NASSAR, Silvia M.; CATAPAN, Araci H.; BASTOS, Rogério C.; REIS, Marcelo. M., WRONSCKI, Vilson R.; GAUTHIER, Francisco O. **SEstaNet Sistema de Ensino-Aprendizagem de Estatística On-Line**. II Congresso Brasileiro de Educação Superior à Distância – Unirede. Brasília, 2003.
- NETER, John; KUTNER, Michael H.; MACHTSHEIM, Christopher J.; WASSERMAN, William. **Applied Linear Regression Models**. 3rd ed. United States, 1996.
- NOAL, Renato B. **Ambiente colaborativo para ensino de estatística com o SESTAT**. Dissertação (Mestrado em Ciências da Computação). Curso de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2002.
- NUNES, Ivônio B. **Pequena Introdução à Educação a Distância**. Educação à Distância número 1, INED. Brasília, 1992.
- PEDATELLA, Rodrigo N. UFG Virtual – Estudo e Implantação de Serviços de Rede. Universidade Federal de Goiás, 2000.
- PIAGET, Jean. **A Psicologia da Inteligência**. Fundo de Cultura AS, Lisboa, 1967.
- PIAGET, Jean. **A Equilibração das Estruturas Cognitivistas – Problema Central do Desenvolvimento**. Editora Zahar, Rio de Janeiro, 1976.

- POTTS, Anthony; FRIEDEL Jr., David H. **Java: Programming Language Handbook**. Coriolis Group Books, 1^a ed., Scottsdale , 1996.
- PROSSIGA – **Escola Virtual Prossiga**. Brasília: Ministério da Ciência e Tecnologia, Programa Prossiga. <http://www.prossiga.br/escolavirtual/>. Acesso: Maio de 2004.
- R – **Publications related to R**: <http://www.r-project.org/doc/bib/R-publications.html>. Acesso: Abril de 2004.
- RAMALHO, José A. **Oracle8i**. Editora Berkeley Brasil, São Paulo – SP, 1999.
- REBELO, Raquel A. **Planejamento de uma Ferramenta Computacional de Ensino-Aprendizagem de Análise de Regressão**. Dissertação (Mestrado em Ciências da Computação). Curso de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2004.
- RUMBLE, Greville.; OLIVEIRA, João **Vocational Education at a Distance**. International perspectives, Kogan Page, London, 1992.
- SANTOS, José G. **SETip – Sistema Especialista para Tipificar Dados de uma Pesquisa: Variáveis Qualitativas e Quantitativas**. Dissertação (Mestrado em Ciências da Computação). Curso de Pós-Graduação em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2001.
- SCHERER, Douglas; GAYNOR William Jr.; VALENTINSEN, Arlene; CURSETJEE, Xerxes. **Oracle 8i: Dicas e Técnicas**. Editora Campus, Rio de Janeiro – RJ, 2000.
- SERRES, Michel. **Hermès. Vol 1. La Communication**. Editions de Minuit, Paris, 1984.
- SHNEIDERMAN, Bem. **Designing the User Interface - Strategies for Effective Human-Computer Interaction**. Addison-Wesley Pub Co, 1997.

SIEGEL, Sidney. **Estatística não-Paramétrica para as Ciências do Comportamento**. Makron Books do Brasil Editora Ltda., São Paulo, 1975.

SOCINFO – **Sociedade da Informação no Brasil: livro verde**. Ministério da Ciência e Tecnologia, Programa Sociedade da Informação. Brasília, 2000.

SOCINFO – **Sociedade da Informação no Brasil: livro branco – Ciência, Tecnologia e Inovação**. Ministério da Ciência e Tecnologia, Programa Sociedade da Informação. Brasília, 2002.

TAJRA, Samya F. **Informática na Educação: Professor na Atualidade**. Editora Érica, São Paulo, 1998.

TELEDUC – **Teleduc-Ambiente de Ensino à Distância**. Universidade de Campinas: Núcleo de Informática Aplicada à Educação. <http://teleduc.nied.unicamp.br/teleduc/>. Acesso: Abril de 2004.

VIRTUAL – **UNIT VIRTUAL – Cursos de ensino à distância**. Universidade Tiradentes. Aracaju, Estância, Itabaiana e Sergipe, <http://www.nead.unit.br/ead/legislacao.php> . Acesso: Junho de 2004.

VIANA, R. C. **Um Modelo Computacional aplicado ao ensino à Distância**, Universidade Federal de Goiás, 1999.

VIRGILIO, Almeida; MENASCÉ, Daniel. **Scaling for E-Business: Technologies, Models, Performance, and Capacity Planning**. Prentice Hall Inc., USA, 2000.

VIRTUS – **Sistema de salas de aulas virtuais do Virtus**. Universidade Federal de Pernambuco, Laboratório de Hiperídia. <http://www.virtus.ufpe.br/>. Acesso: Abril de 2004.

VISUAL – **Visualbuilder.com – JSP Tutorial**. Disponível em: <http://visualbuilder.com>. Acesso: Maio de 2004.

WEBSTAT – **Statistical software for data analysis on the Web**. Disponível em: <http://www.webstatsoftware.com> . Acesso: Julho de 2004.

APÊNDICE A: Relação de alguns arquivos fontes do sistema de manipulação de base de dados

Este apêndice contém o código fonte de alguns dos programas que compõem o sistema de registro de usuários e manipulação de base de dados. Foi feita apenas uma breve descrição sobre o funcionamento dos programas. Não foram anexados todos os códigos fontes devido a grande quantidade dos mesmos.

Os programas foram implementados na linguagem Perl e Java e fazem uso da interface CGI para manipular os dados fornecidos pelo usuário. Esses programas fazem conexão com um servidor de Banco de Dados MySQL e utilizam a linguagem SQL para as consultas no BD.

Para executar estes programas é preciso ter um *login* e senha de aluno ou tutor do ambiente SEstatNet. O acesso é feito através da seguinte URL <http://www.sestatnet.ufsc.br>.

Cadastro de Usuários:

A interface que a CGI gera para entrada dos dados é a mostrada na figura 5.5 e a interface de validação apresenta o mesmo formato. As mensagens de erros, se ocorrerem, aparecem à direita das caixas de textos.

```
#!/usr/bin/perl
#####
# by Vilson Wronscki #
# Last Modified: 23/08/2004 #
# Program Adduser (newlogin, password1, password2, newemail, newname, #
# newusertype, neworg , newdate) #
#####
#
# include the DBI and CGI modules in the script
#use strict;
use CGI qw(:all);
use CGI::Carp qw( fatalToBrowser );

use DBI; # Here's how to include the DBI module

$this_cgi = 'adduser.cgi';
```



```

if ($ENV{'REQUEST_METHOD'} eq "POST"){

# Get the input
read(STDIN, $buffer, $ENV{'CONTENT_LENGTH'});

# Split the name-value pairs
@pairs = split(/&/, $buffer);

foreach $pair (@pairs) {
    ($name, $value) = split(/=/, $pair);

    # Un-Webify plus signs and %-encoding
    $value =~ tr/+/ /;
    $value =~ s/%([a-fA-F0-9][a-fA-F0-9])/pack("C", hex($1))/eg;
    $value =~ s/<!--(.\n)*-->//g;

    $FORM{$name} = $value;
}

#####
# DATA VALIDATION #
#####

$obrigatorio = "Campo obrigatório";
for ($i=0;$i<7;$i++) {
    @erro[$i] = "OK";
}

# newlogin
if (!$FORM{'newlogin'}) {
    @erro[0] = $obrigatorio;
}
elseif (length($FORM{'newlogin'}) < 6) {
    @erro[0] = "Deve ter no mínimo 6 caracteres";
}

# password1
if (!$FORM{'password1'}) {
    @erro[1] = $obrigatorio;
}
elseif (length($FORM{'password1'}) < 6) {
    @erro[1] = "Deve ter no mínimo 6 caracteres";
}
}

```

```

# password2
if (!${FORM{'password2'}}) {
    @erro[2] = $obrigatorio;
}
elseif (length(${FORM{'password2'}}) < 6) {
    @erro[2] = "Deve ter no mínimo 6 caracteres";
}
elseif (${FORM{'password1'}} ne ${FORM{'password2'}}) {
    @erro[2] = "Deve ser igual a Senha";
}

# newemail
@nn = split(/\@/, ${FORM{'newemail'}});
$n = @nn;
# string vazia
if (!${FORM{'newemail'}}) {
    @erro[3] = $obrigatorio;
}
# @ no inicio e no fim da string
elseif (($FORM{'newemail'} =~ /^\@/) or (${FORM{'newemail'}} =~ /\@$/)) {
    @erro[3] = "Email Inválido";
}
# mais de um @ no meio da string
elseif ($n != 2) {
    @erro[3] = "Email Inválido";
}

# newname
# substitui branco/escape por underscore
${FORM{'newname'}} =~ s/\s/_/gi;
if (!${FORM{'newname'}}) {
    @erro[4] = $obrigatorio;
}
elseif (length(${FORM{'newname'}}) < 8) {
    @erro[4] = "Deve ter no mínimo 8 caracteres";
}

# newusertype
if (!${FORM{'newusertype'}}) {
    @erro[5] = $obrigatorio;
}

# neworg
if (!${FORM{'neworg'}}) {

```

```

        @erro[6] = $obrigatorio;
    }
    elsif (length($FORM{'neworg'})< 2) {
        @erro[6] = "Deve ter no mínimo 2 caracteres";
    }

#####
# FORM OF VALIDATION #
#####

foreach $ok (@erro) {
    if ($ok !~ "OK") {

        print "Content-type: text/html\n\n";

        print <<"VAL";
    }
}

<html>
<head>
<title>Validação do Registro - Erro!</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>

<body bgcolor="#ffffff" text="#000000">

<table bgcolor="white" border="1" align="center" cellpadding="6"
cellspacing="0">
    <tr>
        <td>
            <table border="0" align="center" cellpadding="3" cellspacing="0">
                <script>
                    if(window.opener != null)
                        document.write('<tr><td colspan="2" align="left" valign="middle"
nowrap bgcolor="#eeeeee"><a href="javascript:window.close();"></a></td></tr>');
                </script>

                <tr>
                    <td colspan="3" align="center" bgcolor="#eeeeee">
                        <center></center>
                        <b><font face="Arial, Helvetica, sans-serif" size="+1"
color="660000">Validação do novo registro SEstatNet:</font></B>
                    </td>
                </tr>
            </table>
        </td>
    </tr>
</table>

```

```
<tr>
  <td align="right"><b>SEstatNet ID:</b></td>
  <td align="left">${FORM{'newlogin'}}</td>
  <td align="left">@erro[0]</td>
</tr>

<tr>
  <td align="right"><b>Senha:</b></td>
  <td align="left"><B>.....</B></td>
  <td align="left">@erro[1]</td>
</tr>

<tr>
  <td align="right"><b>Redigite a senha:</b></td>
  <td align="left"><B>.....</B></td>
  <td align="left">@erro[2]</td>
</tr>

<tr>
  <td align="right"><b>Email:</b></td>
  <td align="left">${FORM{'newemail'}}</td>
  <td align="left">@erro[3]</td>
</tr>

<tr>
  <td align="right"><b>Nome:</b></td>
  <td align="left">${FORM{'newname'}}</td>
  <td align="left">@erro[4]</td>
</tr>

<tr>
  <td align="right"><b>Ocupação:</b></td>
  <td align="left">${FORM{'newusertype'}}</td>
  <td align="left">@erro[5]</td>
</tr>

<tr>
  <td align="right"><b>Empresa/Escola:</b></td>
  <td align="left">${FORM{'neworg'}}</td>
  <td align="left">@erro[6]</td>
</tr>

<tr>
```

```

        <td colspan="3" align="center" bgcolor="#eeeeee">
            <center><A href="http://localhost/cgi-bin/adduser.cgi">Tente
novamente</A></center>
            <center><A href="http://localhost/sestatnet/index.html">Voltar para
Página inicial</A></center>
        </td>
    </tr>

```

```

</table>

```

```

</td>

```

```

</tr>

```

```

</table>

```

```

</body></html>

```

```

VAL

```

```

    exit 0;
}
}

```

```

#####
# OPEN DATABASE #
#####

```

```

# variable for the database handle
my $dbh;

```

```

# open the DBI connection to the database
$dbh = open_dbh();

```

```

# login para selecao
$auteursch = '';
$auteursch .= $FORM{'newlogin'};
$auteursch .= '';

```

```

$sth = $dbh->prepare("select count(*) from usercad where login = $auteursch");

```

```

$sth->execute;

```

```

$num = $sth->fetchrow();

```

```

# data
($dateh, $datec) = date_current();

```

```

# login duplicado
if ($num)
{
    &already_there;

    close_dbi($dbh);

    exit 0;

}

# O login pode ser gravado na base
else
{
    $sth = $dbh->prepare("INSERT INTO
usercad(iduser,login,password,email,name,usertype,org,created,lastacc)
VALUES(?,?,?,?,?,?,?,?,?)");

    $sth->execute(0,$FORM{'newlogin'},$FORM{'password1'},$FORM{'newemail'},
$FORM{'newname'},$FORM{'newusertype'},$FORM{'neworg'},$datec,$dateh);

    close_dbi($dbh);

    &new_user;

#####
# ENIVAR LOGIN E SENHA POR E-MAIL #
#####

# destino da mensagem
# $mailprog = "/usr/lib/sendmail";
$sendto = "$FORM{'newemail'} ($FORM{'newname'})";
$subject = "Registro no SEstatNet";

# enviar saída para sendmail
open(MAIL, "|$mailprog -t") or err_trap("Não pode conectar o servidor de
Email");
#
# cabeçalho da mensagem
print MAIL "To: $sendto\n";
print MAIL "From: adm\@sestatnet.ufsc.br \n";
print MAIL "Subject: $subject\n\n";

```

```

# conteúdo do descritor MAIL
print MAIL <<"EOM";
O usuário $FORM{'newname'} foi cadastrado no SEstatNet em $dateh .
A seguir segue os dados para acesso restrito:
Login: $FORM{'newlogin'} e
Senha: $FORM{'password1'}.

Esperamos que tenha um bom aprendizado na utilização do sistema
e que obtenha bons resultados.
EOM

print MAIL "Obrigado pelo seu interesse.\n";
close(MAIL);

    exit 0;
}

}else{

#####
# FORM OF INPUT DATA #
#####

    print "Content-type: text/html\n\n";

    print <<"REG";
<html>
<head>
<title>Registro no SEstatNet</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>

<body bgcolor="#ffffff" text="#000000">

<form name="register" method="POST" action="">

<table bgcolor="white" border="1" cellpadding="6" cellspacing="0"
align="center">
    <tr>
        <td>
            <table border="0" align="center" cellpadding="2" cellspacing="0">
                <tr>
                    <td>

```

```

<script language="JavaScript">
    if(window.opener != null)
        document.write('<tr><td colspan="2" align="right" valign="middle"
nowrap bgcolor="#eeeeee"><a href="javascript:window.close();"></a></td></tr>');
</script>
</td>
</tr>
<tr>
    <td colspan="2" align="center" bgcolor="#eeeeee">
        <center></center>
        <b><font face="Arial, Helvetica, sans-serif" size="+1
color="660000">Registro de um novo usuário SEstatNet:</font></B>
        </td>
</tr>

<tr>
    <td align="left"><b>SEstatNet ID:</b></td>
    <td align="left"><input type="text" name="newlogin" value=""
maxlength="15" size="15"> 6-15 caracteres</td>
</tr>

<tr>
    <td align="left"><b>Senha:</b></td>
    <td align="left"><input type="password" name="password1" value=""
maxlength="15" size="15"> 6-15 caracteres</td>
</tr>

<tr>
    <td align="left"><b>Redigite a senha:</b></td>
    <td align="left"><input type="password" name="password2" value=""
maxlength="15" size="15"></td>
</tr>

<tr>
    <td align="left"><b>Email:</b></td>
    <td align="left"><input type="text" name="newemail" value=""
maxlength="30" size="35"></td>
</tr>

<tr>
    <td align="left"><b>Nome:</b></td>

```



```

        <td align="left"><input type="text" name="newname" value="" maxlength="30"
size="35"></td>
    </tr>

    <tr>
        <td align="left"><b>Ocupação:</b></td>
        <td align="left">
            <select name="newusertype">
                <option value="">Selecione</option>
                <option value="Estudante">Estudante</option>
                <option value="Professor">Professor</option>
                <option value="Profissional">Profissional/Outro</option>
            </select>
        </td>
    </tr>

    <tr>
        <td align="left"><b>Empresa/Escola:</b></td>
        <td align="left"><input type="text" name="neworg" value="" maxlength="30"
size="35"></td>
    </tr>

    <tr>
        <td width="36%" height="2">
            <div align="left">&nbsp;</div>
        </td>
        <td width="64%" height="2">
            <div align="left"><input type="submit" value="Registrar" width="81"
height="22" border="0" name="make"></div>
        </td>
    </tr>

    <tr>
        <td width="36%" height="2">
            <div align="left">&nbsp;</div>
        </td>
        <td width="64%" height="2">
            <div align="left"><A href="http://localhost/sestatnet/index.html">Voltar
para Página inicial</A></div>
        </td>
    </tr>
</table>

</td>

```

```

</tr>

</table>
</form>
</body>
</html>

REG
}

exit;

#####
# OPEN_DBI #
#####

sub open_dbi
{
    # Declare and initialize variables
    my $host = 'localhost';
    my $db = 'sestatnet';
    my $db_user = 'vilson';
    my $db_password = '*****';

    # Connect to the requested server

    my $dbh = DBI->connect("dbi:mysql:$db:$host", "$db_user", "$db_password",
        {RaiseError => 0, PrintError => 0} )
        or err_trap("Falha na conexão da base de dados");
    return $dbh;
}#end: open_dbi

#####
# ERR_TRAP #
#####

sub err_trap
{
    my $error_message = shift(@_);
    die "$error_message\n ERROR: $DBI::err ($DBI::errstr)\n";
}#end: err_trap

#####
# CLOSE_DBI #

```

```
#####
```

```
sub close_dbi
{
    my $dbh = shift(@_);

    $dbh->disconnect or err_trap("Não pode desconectar a base de dados");
}#end: close_dbi
```

```
#####
```

```
# ALREADY_THERE #
```

```
#####
```

```
sub already_there {

    print "Content-type: text/html\n\n";

print <<EOT;
    <html>
    <head>
    <title>Novo Usuário - Erro!</title>
    </head>
    <body bgcolor="#FFFFFF">
    <center>
    <h1>ERRO: Usuário Duplicado.</h1>
    <hr><P>
    <font size=5>Já consta em nosso banco de dados o usuário<b>
$FORM{'newlogin'}</b>
    </font><P>
    <font size=4>Você precisa selecionar outro "User Name" ou uma variação
deste.<P>
    </font></center><hr>
    </body></html>
EOT

} # End sub already_there
```

```
#####
```

```
# DATE_CURRENT #
```

```
#####
```

```
sub date_current
```

```

{
    $time = time() ;

    ($sec,$min,$hour,$mday,$mon,$year,$yday,$isdst) = localtime($time);
## or just localtime( time() )

## Let's trate the return of localtime():

## $mon start from 0, let's add 1:
$mon += 1 ;

## year 2000 is 100 on $year, let's add 1900:
$year += 1900 ;

## get the last two digits of the year
#$year = sprintf("%02d", $year % 100);

## adding the 0 before d:
if ($sec < 10) { $sec = "0$sec" ;}
if ($min < 10) { $min = "0$min" ;}
if ($hour < 10) { $hour = "0$hour" ;}
if ($mday < 10) { $mday = "0$mday" ;}
if ($mon < 10) { $mon = "0$mon" ;}

## The @DATE:
$DATE1 = "$mday/$mon/$year $hour:$min:$sec" ;
$DATE2 = "$year-$mon-$mday $hour:$min:$sec";
@DATE = ($DATE1, $DATE2);

return @DATE;

}#end: date_current

#####
# NEW_USER #
#####

sub new_user {

    print "Content-type: text/html\n\n";

print <<EOC;
    <html>

```

```

    <head>
<title>Novo Usuário - Ok!</title>
    </head>
    <body bgcolor="#FFFFFF">
    <center>
    <h1>AVISO: Usuário Registrado.</h1>
    <hr><P>
    <font size=4>O seu Login e Senha estão sendo enviados para<BR>
    o email <b> $FORM{'newemail'}</b> que você informou<BR>
    no cadastramento efetuado em: <b> $dateh</b></font><P>
    <A href="http://localhost/sestatnet/index.html">Voltar para Página
inicial</A>
    </font></center><hr>
    </body></html>
EOC

```

```
} # End sub new_user
```

Validação de Usuários:

Na interface geral do sistema (Figura 4.1) o usuário precisa digitar a sua identificação e senha para acessar a área restrita do ambiente SEstatNet. O programa seguinte faz a validação do usuário e se tiver permissão acessa esse ambiente disponibilizando todos os recursos de análise e manipulação de dados. Quando isso ocorre o programa atualiza a base de dados com a data atual de acesso do usuário.

As mensagens de erros, se ocorrerem, são mostradas no *browser*. As rotinas de acesso à base de dados e data atual são idênticas as do programa anterior.

```

#!c:/Perl/bin/Perl.exe
#!/usr/bin/perl
#####
# by Vilson Wronscki #
# Last Modified: 06/09/2004 #
# Program Login (username, password) #
#####
#
# include the DBI and CGI modules in the script
#use strict;
use CGI qw(:all);
use CGI::Carp qw( fatalsToBrowser );

```

```

use DBI; # Here's how to include the DBI module

# Nome do arquivo que voce deseja proteger. (TXT)
$txt_file = 'c:/data/secure/secure.html';

# Pagina inicial
$login_page = 'http://localhost/sestatnet/index.html';

$this_cgi = 'login.cgi';

# Get the input
read(STDIN, $buffer, $ENV{'CONTENT_LENGTH'});

# Split the name-value pairs
@pairs = split(/&/, $buffer);

foreach $pair (@pairs) {
    ($name, $value) = split(/=/, $pair);

# Un-Webify plus signs and %-encoding
    $value =~ tr/+/ /;
    $value =~ s/%([a-fA-F0-9][a-fA-F0-9])/pack("C", hex($1))/eg;
    $value =~ s/<!--(.\n)*-->/g;

    $FORM{$name} = $value;
}

# Falta de dados
&whoareyou unless $FORM{'password'};
&whoareyou unless $FORM{'username'};

#####
# OPEN DATABASE #
#####

# variable for the database handle
my $dbh;

# open the DBI connection to the database
$dbh = open_dbh();

# login para selecao
$auteursch = '';
$auteursch .= $FORM{'username'};

```

```

$autsrch .= '';

($dateh, $datec) = date_current();

$sth = $dbh->prepare("SELECT iduser,password,lastacc FROM usercad WHERE login
= $autsrch");

$sth->execute;

($num, $passw, $lastdate) = $sth->fetchrow();

# Not an authorized user
if (!$num)
{
    &authorized_user;

    close_dbh($dbh);

    exit 0;

}

# Passwords don't match
elsif ($passw ne $FORM{'password'})
{
    &match_password;

    close_dbh($dbh);

    exit 0;

}

#####
# RESTRICT AREA #
#####
else
{

$today = '';
$today .= $dateh;
$today .= '';

# Update the access date

```

```

    $sth = $dbh->prepare("UPDATE usercad SET lastacc=$today WHERE
iduser=$num");

    $sth->execute;

    close_dbh($dbh);

# This section opens the htm file and prints it
    print "Content-type: text/html\n\n";

    open(FILE, "$txt_file");
    @filein = <FILE>;
    close(FILE);

    print @filein;

}

exit 0;

#####
# WHOAREYOU #
#####

sub whoareyou {

    print "Content-type: text/html\n\n";

print <<"EOT";
    <html>
    <head>
<title>Dados para Acesso - Erro!</title>
    </head>
    <body bgcolor="#FFFFFF">
    <center>
    <h1>ERRO: Área Restrita.</h1>
    <hr><P>
<font size=4>Você precisa de um Nome de Usuário e de uma Senha<BR>
para acessar o ambiente restrito <b>SEstatNet</b>.<BR>
<P><a href="\$login_page\">Tente novamente.</a>
    </font>
    </center><hr></body>
    </html>

EOT

```



```

exit 0;
}#end: whoareyou

#####
# AUTHORIZED_USER #
#####

sub authorized_user {

    print "Content-type: text/html\n\n";

print <<"USR";
    <html>
    <head>
    <title>Usuário não autorizado - Erro!</title>
    </head>
    <body bgcolor="#FFFFFF">
    <center>
    <h1>ERRO: Usuário não autorizado.</h1>
    <hr><P>
    <font size=5>Usuário: <b>${FORM{'username'}}</b> não existe.<P>
    </font>
    <font size=4>Você precisa de um Nome de Usuário e de uma Senha<BR>
    para acessar o ambiente restrito <b>SEstatNet</b>.<P>
    <a href="\${login_page}">Tente novamente.</a><p>
    </font>
    </center><hr></body>
    </html>
USR

}# end: authorized_user

#####
# MATCH_PASSWORD #
#####

sub match_password {

    print "Content-type: text/html\n\n";

print <<"PSW";
    <html>
    <head>

```

```
<title>Acesso negado - Erro!</title>
</head>
<body bgcolor="#FFFFFF">
<center>
<h1>ERRO: Acesso negado.</h1>
<hr><p>
<font size=5>A Senha do Usuário: <b>${FORM{'username'}}</b> não confere.<p>
$achou</font>
<font size=4>Você precisa de um Nome de Usuário e de uma Senha<BR>
para acessar o ambiente restrito <b>SEstatNet</b>.<p>
<a href="\${login_page}">Tente novamente.</a><p>
Se você esqueceu a Senha acesse o link na página principal.
</font><p>
</center><hr></body>
</html>
```

PSW

```
}# end: match_password
```