

---

# COMPORTAMENTO VERBAL

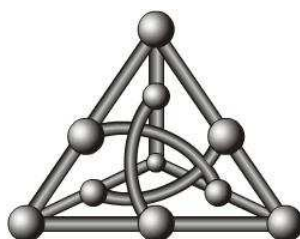
## Um Repositório Temático

Trabalho de Conclusão de Curso  
Bacharelado em Análise de Sistemas

Douglas Nantes Gualberto  
Fernando Levandoski

---

Faculdade de Computação  
Universidade Federal de Mato Grosso do Sul



**Orientadora: Prof. Dra. Luciana Montera Cheung**

Campo Grande, 28 de Janeiro de 2011

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Descrição do Problema . . . . .	3
1.2	Objetivo Geral . . . . .	3
1.3	Objetivos Específicos . . . . .	3
1.4	Justificativa e Metodologia . . . . .	3
<b>2</b>	<b>Requisitos do Sistema</b>	<b>5</b>
2.1	Visão geral do sistema . . . . .	5
2.2	Requisitos funcionais do sistema . . . . .	5
2.3	Requisitos não funcionais do sistema . . . . .	7
2.4	Funcionalidades do Sistema . . . . .	7
2.5	Diagrama de caso de uso . . . . .	12
<b>3</b>	<b>O Banco de Dados</b>	<b>14</b>
3.1	Diagrama Entidade Relacionamento . . . . .	14
3.2	O Modelo Relacional . . . . .	16
<b>4</b>	<b>O Aplicativo Web</b>	<b>19</b>
4.1	Interfaces do Sistema . . . . .	19
<b>5</b>	<b>Data Mining</b>	<b>24</b>
5.1	Motivação . . . . .	24
5.2	Classificação Bayesiana . . . . .	24
5.3	Processo de Classificação . . . . .	25
5.3.1	Cálculos das probabilidades das classes . . . . .	25

5.3.2	Cálculo da probabilidade de ocorrência de cada atributo da amostra desconhecida . . . . .	25
5.3.3	Cálculo da probabilidade total da amostra desconhecida . . . . .	25
5.4	Aplicação do Algoritmo de Bayes . . . . .	27
5.5	Implementação do Algoritmo . . . . .	30
5.5.1	Resultados obtidos com Weka . . . . .	31
<b>6</b>	<b>Conclusão</b>	<b>33</b>
	<b>Referências Bibliográficas</b>	<b>34</b>

# Capítulo 1

## Introdução

Em 1957 o norte-americano Burrhus Frederic Skinner (1904-1990) – autor da obra de maior expressão do behaviorismo [3], definiu comportamento verbal como: “*o comportamento estabelecido e mantido por consequências mediadas por outras pessoas*”, ou de outra forma, “*o comportamento dos indivíduos pode ser influenciado bem como direcionado por estímulos*”.

O comportamento verbal está sujeito aos mesmos princípios que governam o comportamento não verbal, a única característica que o difere significativamente e o faz merecer uma análise separada é a natureza do reforço que o estabelece e o mantém e que requer a mediação de outra pessoa.

Segundo Skinner (1957) o ouvinte responde aos estímulos verbais produzidos pelo falante. O indivíduo é falante ao comportar-se verbalmente perante o outro e torna-se um ouvinte ao comportar-se funcionalmente a estímulos verbais produzidos por outros indivíduos.

Em 2007 o Professor Dr. Lucas Cordoba Ferraz, do Departamento de Psicologia da Universidade Federal de Mato Grosso do Sul, iniciou uma pesquisa com o intuito de verificar se os autores pós Skinner mantiveram, ou não, a essência da definição de comportamento verbal dada por seu autor. Para a realização da pesquisa, serão avaliados artigos da área publicados em duas revistas americanas Jaba<sup>1</sup> e Jeab<sup>2</sup>. O grande volume de dados a ser armazenado e avaliado fez com que o professor e coordenador da pesquisa procurasse uma forma segura e eficiente de organizar o conteúdo da mesma.

Frente às necessidades expostas pelo professor, foi sugerido a implantação de um Repositório Digital. Repositórios Digitais são coleções de informações digitais, que podem ser construídas de diferentes formas e com diferentes propósitos [2]. Para o problema em questão será utilizado um Repositório Temático que corresponde a um repositório digital especializado em armazenar dados de uma determinada área do conhecimento, no caso, comportamento verbal. Um Repositório Digital pode ser também entendido como um Banco de Dados (BD) e, por isso, utilizaremos os dois termos indiscriminadamente.

---

<sup>1</sup><http://seab.envmed.rochester.edu/jaba/index.html>

<sup>2</sup><http://seab.envmed.rochester.edu/jeab/index.html>

Os dados dos artigos selecionados serão coletados e triados pelos acadêmicos que fazem parte da pesquisa e posteriormente irão alimentar o repositório digital. A implantação do repositório irá permitir aos membros da pesquisa avaliar/analisar informações de forma rápida e confiável e assim dinamizar a pesquisa que está sendo realizada.

## 1.1 Descrição do Problema

O prof. Dr. Lucas e sua equipe têm realizado, até o presente momento, todo o trabalho de busca, cadastro e avaliação dos artigos de interesse de forma manual, trabalho este que demanda grande esforço e não oferece um meio de consulta ágil e confiável. Diante deste contexto, foi proposto a criação de um banco de dados a fim de automatizar e, assim, agilizar algumas das tarefas envolvidas nesta pesquisa. Para tornar a manipulação do banco de dados uma tarefa completamente independente do conhecimento de uma linguagem específica, como a SQL, por exemplo, um aplicativo web será desenvolvido.

## 1.2 Objetivo Geral

O objetivo geral é desenvolver uma ferramenta computacional para automatizar o processo de cadastro (armazenamento) e consulta de artigos, tornando a realização destas tarefas mais rápida e eficiente. Espera-se também colaborar com a pesquisa sobre “*Comportamento Verbal*” no sentido de torná-la mais robusta, isto é, menos suscetível à falhas humanas.

## 1.3 Objetivos Específicos

Dentre os objetivos específicos deste trabalho estão:

- O projeto e a implementação de um banco de dados para armazenar dados referentes à artigos científicos.
- O desenvolvimento de uma interface web amigável para a manipulação do banco de dados.
- Implementação de consultas estáticas ao banco de dados.
- Implementação de algoritmo de aprendizagem de máquina para *Data Mining*.

## 1.4 Justificativa e Metodologia

A necessidade do professor Dr. Lucas em informatizar etapas do seu trabalho de pesquisa é evidente. Após a leitura e análise dos artigos relativos à Comportamento

Verbal, os dados importantes são escritos em rascunhos ou salvos em planilhas do Excel. Não existe nenhuma forma de organização desses dados e toda informação que se queira recuperar tem que ser garimpada em um misto de anotações e planilhas.

Um sistema computacional trará benefícios no âmbito de armazenamento (banco de dados), automação de consultas, bem como na extração de novos conhecimentos, por meio de *Data Mining*.

O desenvolvimento do projeto é também uma oportunidade para os acadêmicos que o desenvolvem colocarem em prática os conhecimentos adquiridos nas disciplinas de Programação de Computadores, Engenharia de Software, Banco de Dados e Inteligência Artificial cursadas durante a graduação.

Para que o desenvolvimento do trabalho fosse possível, foi preciso um bom entendimento sobre o tema Comportamento Verbal para identificar características importantes sobre o sistema, bem como requisitos funcionais e não funcionais. Além das leituras sobre o tema, diversas reuniões com o professor Dr. Lucas, coordenador da pesquisa, bem como com outros participantes do grupo, foram realizadas.

Para o desenvolvimento do sistema foi escolhida a linguagem PHP, por ser uma linguagem de programação Web que oferece forte portabilidade e também por ser uma linguagem de uso comum. Para o banco de dados decidiu-se pela utilização do banco *MySQL* e da ferramenta *PHPMyAdmin*, a qual possibilita o gerenciamento eficaz do banco de dados. As duas escolhas foram motivadas pela licença pública (GNU *General Public License*) das mesmas, além de serem de fácil uso.

# Capítulo 2

## Requisitos do Sistema

A referência [10] foi utilizada na elaboração deste capítulo. Os requisitos de um sistema constituem um conjunto de tarefas que este deve ser capaz de executar para atender às necessidades do cliente. As técnicas de engenharia de requisitos descritas no capítulo 7 da referência citada foram amplamente utilizadas para se determinar os requisitos do sistema proposto.

Há no mercado aplicativos como o *Google Scholar* que é uma ferramenta de pesquisa do *Google* que permite pesquisar em trabalhos acadêmicos, literatura escolar, jornais de universidades e artigos variados, porém nenhum desses aplicativos atende completamente às necessidades da pesquisa em questão. Apesar de ter propósito parecido com softwares existentes no mercado o SIAP se propõe a cadastrar, fornecer consulta e classificação dos artigos interessantes à pesquisa. O SIAP possui formulários de cadastro que se moldam perfeitamente para atender o objetivo da pesquisa.

### 2.1 Visão geral do sistema

O sistema deve ter um conjunto de formulários que recebam todos os dados interessantes para a pesquisa, deve armazenar esses dados a fim de fornece-los rapidamente quando solicitados em alguma pesquisa do usuário e ainda deve ser capaz de classificar os artigos de acordo com as informações fornecidas pelo usuário.

### 2.2 Requisitos funcionais do sistema

#### RF01 - Cadastro de Usuário

O sistema deve permitir o cadastro das pessoas que fazem parte do grupo de pesquisa. As informações a serem cadastradas são: nome do acadêmico ou coordenador, login, senha, telefone, e-mail, instituição de ensino, data de ingresso no projeto de pesquisa e data de saída do projeto de pesquisa.

sistema deve permitir ao coordenador da pesquisa, e somente a ele, incluir/excluir usuários bem como listar/editar os dados dos usuários cadastrados no projeto de pesquisa.

#### **RF02 - Consulta de Usuário**

O sistema deve permitir ao coordenador, e somente a ele, consultar os dados dos acadêmicos cadastrados na pesquisa.

#### **RF03 - Cadastro de Artigo**

O sistema deve permitir que todo usuário possa efetuar o cadastro de algum artigo, porém o referido cadastro deve ser, posteriormente, confirmado pelo coordenador da pesquisa.

O sistema deve, para cada artigo, permitir o cadastro dos seguintes dados: título, ano de publicação, autores, no máximo 5 (cinco) palavras-chave, resumo, comentário, nome do participante da pesquisa responsável pelo cadastro, nome da revista fonte da publicação, tipo de sujeito sendo analisado, tipo da análise, delineamento, objeto de estudo/categoria, status do artigo e se ele discute ou não a sua aplicação. O status representa a situação de um artigo em relação à confirmação do cadastro realizada pelo coordenador da pesquisa. O sistema deve registrar a data de cadastro do artigo no banco de dados.

O sistema deve aguardar confirmação do coordenador da pesquisa antes que um artigo tenha seu cadastrado confirmado.

O sistema deve permitir que após a confirmação (pelo coordenador da pesquisa) de um cadastro, somente este possa modificar os dados deste artigo.

#### **RF04 - Consulta a Artigos**

O sistema deve permitir a todo usuário consultar os dados dos artigos cadastrados.

O sistema deve ser capaz de efetuar consultas baseadas em qualquer um dos dados cadastrados para os artigos.

#### **RF05 - Cadastrar Novas Categorias**

O sistema deve permitir que sejam cadastradas novas categorias para os artigos conforme estas sejam identificadas.

#### **RF06 - Modificar Cadastro**

Com exceção do nome do participante da pesquisa responsável pelo cadastro do artigo e da data de cadastro do artigo, o sistema deve permitir que os dados já cadastrados possam ser modificados pelo coordenador da pesquisa.

#### **RF07 - Excluir Cadastro**

O sistema deve permitir ao coordenador, e somente a ele, excluir cadastro com status de já confirmados.

#### **RF08 - Emissão de relatórios**

O sistema deve permitir a emissão de relatórios com os dados dos usuários bem como a emissão de relatórios com os dados dos artigos selecionados por pesquisas realizadas no banco de dados.



## 2.3 Requisitos não funcionais do sistema

### RNF01 - Usabilidade

O sistema deve possuir uma interface para login e outras que possibilitem a entrada dos dados de interesse mas que sejam simples e de fácil uso.

### RNF02 - Hardware

O sistema deve ser alocado em um servidor exclusivo, quando estiver totalmente operacional.

## 2.4 Funcionalidades do Sistema

Os casos de uso apresentados a seguir visam a descrição dos cenários de interação entre o usuário e o sistema. Os casos de uso servirão como direcionamento para a posterior implementação do sistema.

<b>Caso de uso:</b>	<b>Cadastrar Usuário</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	O professor deve ter os dados do aluno a ser cadastrado.
<b>Pós-Condições:</b>	É liberado ao usuário o acesso ao sistema.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O professor cadastra os dados do novo usuário com uma senha e nome de acesso padrão.</li><li>2. O sistema libera o acesso ao novo usuário.</li></ol>

---

<b>Caso de uso:</b>	<b>Excluir Usuário</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	O usuário deve estar cadastrado.
<b>Pós-Condições:</b>	Acesso bloqueado ao usuário excluído.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O professor exclui o usuário em questão.</li><li>2. O sistema bloqueia o acesso ao usuário excluído.</li></ol>

**Tratamento de exceções:** 1a. O usuário possui cadastro de artigos a serem confirmados.  
1a.1 O coordenador decide pela exclusão ou confirmação dos cadastros pendentes e retorna ao item 1 do fluxo principal.

---

**Caso de uso:** **Editar cadastro do usuário**

**Interessados:** Professor coordenador da pesquisa.

**Pré-Condições:** O usuário deve estar cadastrado.

**Pós-Condições:** Dados do cadastro do usuário alterados.

**Fluxo Principal:**

1. O professor seleciona o usuário que deseja.
2. O professor altera os dados que deseja do usuário selecionado.
3. O professor salva o cadastro do usuário já modificado.

---

**Caso de uso:** **Alterar senha**

**Interessados:** Professor coordenador da pesquisa e acadêmicos já cadastrados.

**Pré-Condições:** O usuário deve estar cadastrado.

**Pós-Condições:** Senha antiga substituída pela nova senha.

**Fluxo Principal:**

1. O usuário altera a sua senha.
2. O sistema registra a nova senha em substituição à senha antiga.

**Tratamento de exceções:**

- 1a. O usuário esqueceu a senha.  
1a.1 O usuário envia uma mensagem ao professor solicitando uma nova senha.  
1a.2 O professor efetua login com a senha de administrador e cadastra nova senha (padrão) para o usuário, volta-se ao item 1 do Fluxo principal.

---

<b>Caso de uso:</b>	<b>Cadastrar artigo</b>
<b>Interessados:</b>	Professor coordenador da pesquisa e acadêmicos cadastrados.
<b>Pré-Condições:</b>	O usuário deve estar cadastrado e os dados do artigo devidamente coletados.
<b>Pós-Condições:</b>	Dados do arquivo armazenados no banco de dados.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O usuário registra os dados do artigo conforme o item <b>cadastro de artigo</b> dos requisitos funcionais no capítulo 2.</li><li>2. Os Dados do arquivo ficam armazenados provisoriamente em um pré-cadastro aguardando a confirmação do coordenador da pesquisa.</li><li>3. O sistema armazena no banco de dados os itens do artigo pré-cadastrado.</li></ol>

---

<b>Caso de uso:</b>	<b>Confirmar cadastro de artigo</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	Artigo pré-cadastrado.
<b>Pós-Condições:</b>	Dados do artigo armazenados definitivamente no BD.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O professor analisa os dados do artigo pré-cadastrado.</li><li>2. O sistema armazena os dados do artigo que teve o pré-cadastro aceito.</li></ol>
<b>Tratamento de exceções:</b>	<ol style="list-style-type: none"><li>1a. O professor não aceita o pré-cadastro.<ol style="list-style-type: none"><li>1a.1 O pré-cadastro não é confirmado e o acadêmico responsável é informado no seu próximo login, através de mensagem digitada pelo coordenador, sobre as correções necessárias.</li></ol></li></ol>

---

<b>Caso de uso:</b>	<b>Excluir cadastro de artigo</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	O artigo deve estar cadastrado e com status de confirmado.
<b>Pós-Condições:</b>	Os dados do artigo são excluídos do sistema.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O professor efetua a busca dos artigos que deseja excluir.</li><li>2. O sistema pesquisa todos os artigos solicitados pelo professor.</li><li>3. O professor seleciona os artigos desejados resultantes da busca e confirma a exclusão.</li></ol>

---

<b>Caso de uso:</b>	<b>Pesquisar artigo</b>
<b>Interessados:</b>	Professor coordenador da pesquisa e acadêmicos cadastrados.
<b>Pré-Condições:</b>	O usuário deve estar cadastrado.
<b>Pós-Condições:</b>	Os artigos que atendem às condições impostas na consulta feita pelo usuário são exibidos.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"><li>1. O usuário realiza a pesquisa.</li><li>2. O sistema faz a triagem e exibição dos artigos que atendem aos atributos da pesquisa.</li></ol>

---

<b>Caso de uso:</b>	<b>Modificar dados de artigo cadastrado</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	O artigo deve estar cadastrado.
<b>Pós-Condições:</b>	Cadastro modificado e novos dados armazenados.

<b>Fluxo Principal:</b>	<ol style="list-style-type: none"> <li>1. O professor efetua pesquisa para localizar os artigos que deseja modificar.</li> <li>2. O professor modifica o(s) dado(s) do(s) artigo(s) que julgar necessário.</li> <li>3. O professor salva as modificações realizadas.</li> </ol>
-------------------------	---

---

<b>Caso de uso:</b>	<b>Cadastrar nova revista como fonte de publicação<sup>1</sup></b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	A nova revista deve conter artigos de interesse da pesquisa.
<b>Pós-Condições:</b>	A nova revista é acrescentada no menu revista do formulário de cadastro de artigo.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"> <li>1. O professor efetua o cadastro da nova revista como fonte de publicação pelo menu principal do sistema.</li> <li>2. O sistema acrescenta a nova revista no menu revista do formulário de cadastro de artigo.</li> </ol>

---

<b>Caso de uso:</b>	<b>Emitir relatório de artigo</b>
<b>Interessados:</b>	Professor coordenador da pesquisa e acadêmicos cadastrados.
<b>Pré-Condições:</b>	Selecionar os artigos que irão compor o relatório.
<b>Pós-Condições:</b>	O sistema gera o relatório solicitado com os atributos dos artigos selecionados.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"> <li>1. O usuário efetua a pesquisa pelos artigos que atendam aos atributos por ele selecionados.</li> <li>2. O usuário seleciona os artigos que irão compor o relatório.</li> <li>3. O sistema gera o relatório solicitado.</li> </ol>

---

<sup>1</sup>Este caso de uso é idêntico para o cadastro de novos valores para sujeito, análise, delineamento e objeto de estudo/categoria. Corresponde ao requisito **Cadastrar Novas Categorias**.

---

<b>Caso de uso:</b>	<b>Emitir relatório de usuário</b>
<b>Interessados:</b>	Professor coordenador da pesquisa.
<b>Pré-Condições:</b>	Selecionar os usuários que irão compor o relatório.
<b>Pós-Condições:</b>	O sistema gera o relatório solicitado.
<b>Fluxo Principal:</b>	<ol style="list-style-type: none"> <li>1. O coordenador efetua a busca pelos usuários que atendam aos atributos por ele selecionados.</li> <li>2. O coordenador seleciona os usuários que irão compor o relatório.</li> <li>3. O sistema gera o relatório solicitado.</li> </ol>

---

## 2.5 Diagrama de caso de uso

A Figura 4.1 representa os relacionamentos existentes entre os atores do sistema e cada um dos casos de uso descritos anteriormente. Os atores são o coordenador e os acadêmicos, representados por bonecos, os casos de uso são aqueles descritos na seção anterior e representados por elipses. Os relacionamentos entre ator e caso de uso são representados por setas.

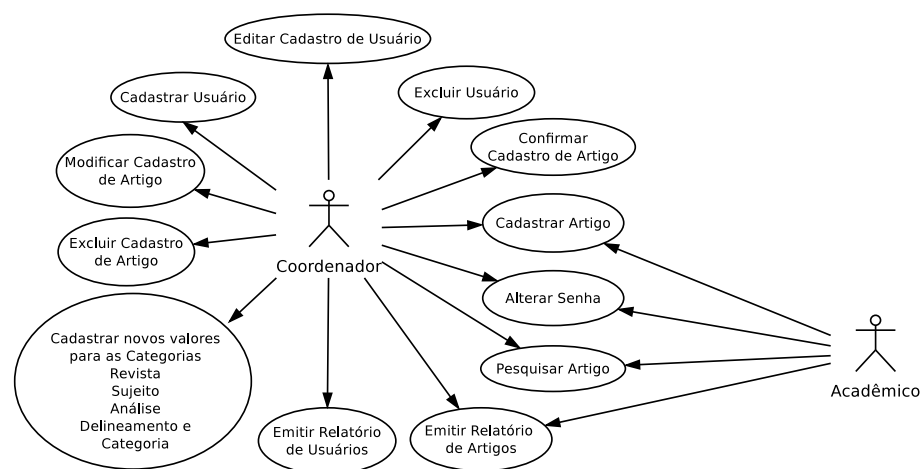


Figura 2.1: Relacionamentos entre os atores e os casos de uso do sistema.

Observe que são comuns ao coordenador da pesquisa e aos acadêmicos as tarefas de pesquisa e cadastro de artigo, alteração de senha e emissão de relatório de artigo. As demais tarefas são exclusivas do coordenador.

Para a elaboração das fases de desenvolvimento do software não foi utilizado um único modelo ou processo de desenvolvimento, mas sim, um conjunto de técnicas abordadas em sala de aula e revistas na referência [10].

O modelo de Prototipação ajudou a compreender o que era necessário estar visível nas interfaces. A cada interface apresentada o coordenador da pesquisa julgava a eficiência da mesma e solicitava as alterações necessárias.

Assim que foram estudadas as técnicas e fases do Processo Unificado elas passaram a nortear - dada a pouca experiência dos desenvolvedores - todo o desenvolvimento do software.

O modelo Incremental foi adotado como modelo principal na elaboração do projeto. Apresentou-se ao coordenador da pesquisa uma primeira versão do aplicativo (*núcleo do produto*) e a partir dela foram sendo elaborados incrementos até que o aplicativo se tornasse não só operacional mas atendesse também a todas as necessidades da pesquisa.

# Capítulo 3

## O Banco de Dados

A referência [9] foi utilizada na elaboração deste capítulo. Com base na referência pode-se identificar os tipos de entidades e atributos, os relacionamentos e o mapeamento dos seus tipos, definir as restrições e todo mapeamento do ER para o Relacional.

Com base nas informações adquiridas durante as reuniões com o professor coordenador da pesquisa e mediante a descrição dos casos de uso foi modelado o banco de dados. Inicialmente identificou-se duas entidades distintas: **artigo** e **usuário** e ambas têm seus atributos definidos a partir dos requisitos funcionais.

A entidade artigo deverá possuir os seguintes atributos: **título**, **autor**, **ano de publicação**, **data de cadastro**, **situação do cadastro** (confirmado ou não), **nome da revista fonte da publicação**, **tipo de sujeito sendo analisado**, **tipo da análise**, **delineamento**, **objeto de estudo/categoria** e **se ele discute ou não a sua aplicação**, **palavras-chave**, **pesquisador responsável pelo cadastro**, **resumo** e **comentário** (quando o pesquisador responsável pelo cadastro julgar necessário.)

A entidade usuário deverá possuir os atributos: **nome**, **login**, **senha**, **identificação** (usuário comum - acadêmico ou super-usuário - coordenador), **data de entrada** e **data de saída** - quando for o caso - **no projeto de pesquisa**, **instituição de ensino à qual pertence**, **e-mail**, **telefone** e **identificação de status na pesquisa** (ativo ou não).

### 3.1 Diagrama Entidade Relacionamento

O diagrama mostrado na Figura 5.1 corresponde ao Modelo Entidade Relacionamento (MER) do sistema proposto.



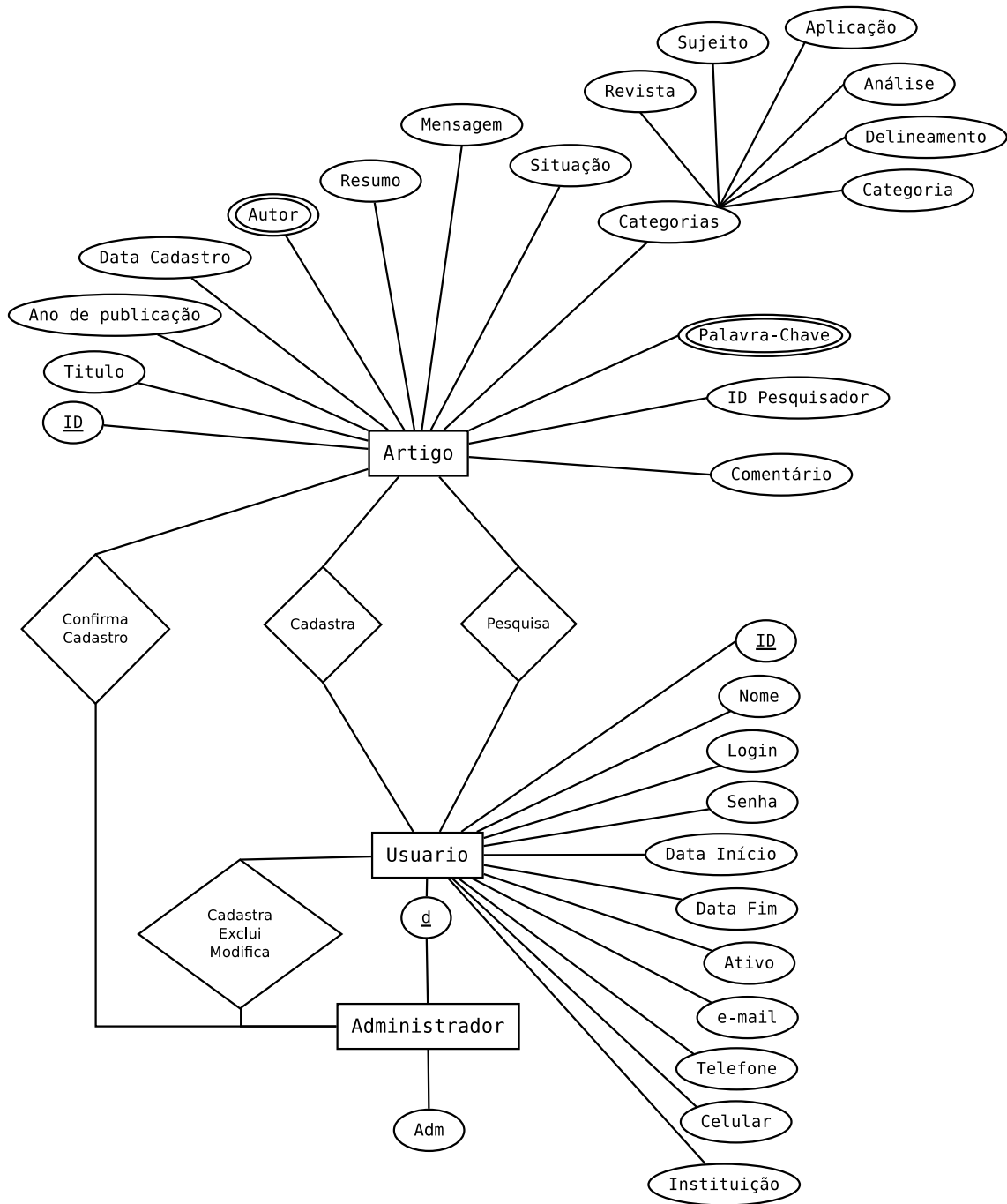


Figura 3.1: Modelo Entidade Relacionamento - Cada entidade é representada por um retângulo contendo o nome da respectiva entidade, cada atributo é representado por uma elipse contendo o nome do atributo a qual é ligada à sua respectiva entidade por uma linha. Os relacionamentos entre as entidades são representados por losângulos contendo o nome do relacionamento em questão. Note que a entidade Administrador é derivada da entidade Usuário e pode cadastrar, excluir ou modificar a entidade Usuário.

Há apenas duas classes distintas: ARTIGO e USUÁRIO. Pela simplicidade das relações e pelas mesmas serem descritas no Modelo Relacional, omitiu-se o diagrama de classes.

## 3.2 O Modelo Relacional

Com base no Modelo Entidade Relacionamento - MER (Figura 5.1) obtém-se o Modelo Relacional - MR, seguindo as regras de mapeamento descritas no Capítulo 7 de [9]. De acordo com tais regras o Modelo Relacional é definido pelas tabelas apresentadas a seguir:

**Tabela ARTIGO**

id	titulo	ano_pb	data_cd	resumo	sit	pesq_id	comentario	categorias

### Descrição dos atributos

- id:** Chave de identificação do artigo.
- titulo:** Título do artigo.
- ano\_pb:** Ano de publicação do artigo.
- data\_cd:** Data da realização do cadastro do artigo.
- resumo:** Resumo do assunto tratado pelo artigo.
- sit:** Sinaliza se o artigo em questão já possui seu cadastro confirmado (1) ou ainda é um pré-cadastro (0).
- pesq\_id:** Pesquisador responsável pelo cadastro do artigo. Chave estrangeira que faz referência à tabela **USUÁRIO** pelo atributo **id**.
- comentario:** Comentário que o pesquisador julgue necessário fazer a respeito do artigo.
- categorias:** Atributos de classificação do artigo. Chave estrangeira que faz referencia à tabela **CATEGORIAS** pelo atributo **id**.

**Tabela USUÁRIO**

id	nome	login	senha	adm	data_i	data_f	inst	telefone	celular	email	ativo

### Descrição dos atributos

- id:** Chave de identificação do usuário.
- nome:** Nome do usuário.
- login:** Identificação do usuário quando logado.
- senha:** Senha de acesso do usuário ao sistema.
- adm:** Identifica o usuário como administrador (super-usuário) (1) ou não (0).
- data\_i:** Data de entrada do usuário no projeto de pesquisa.
- data\_f:** Data de saída - quando for o caso - do usuário no projeto de pesquisa.
- inst:** Nome da instituição de ensino à qual o usuário pertence. Chave estrangeira que faz referencia à tabela **INSTITUIÇÃO** pelo atributo **id**.

telefone: Telefone do usuário.  
celular: Celular do usuário.  
email: E-mail do usuário.  
ativo: Status do usuário em relação ao projeto ativo (1) ou não (0).

### Tabela AUTOR

id	nome

### Descrição dos atributos

id: Chave de identificação do autor.  
nome: Nome do autor.

### Tabela CATEGORIAS

id	revista	sujeitos	analise	delineamento	categoria	discute_aplicacao

### Descrição dos atributos

id: Chave de identificação da categoria.  
revista: Fonte de publicação do artigo.  
sujeitos: Define o tipo de sujeito que está sendo analisado.  
analise: Define se o estudo é sobre um grupo ou um sujeito individual.  
delineamento: Define a classificação prévia do artigo.  
categoria: Define o objeto de estudo, do qual trata o artigo.  
discute\_aplicacao: Define se o conteúdo do artigo discute alguma aplicação do objeto de estudo.

### Tabela PALAVRA\_CHAVE

id	palavra

### Descrição dos atributos

id: Chave de identificação de cada palavra chave.  
palavra: Palavra chave do artigo.

### Tabela INSTITUICAO

id	nome

### Descrição dos atributos

id: Chave de identificação de cada instituição.  
nome: Define o nome da instituição.

#### Tabela MENSAGEM

id	msg

#### Descrição dos atributos

id: Chave de identificação de cada mensagem.  
msg: Define a mensagem a ser registrada.

#### Tabela ARTIGO\_PALAVRA\_CHAVE

ar	chave

#### Descrição dos atributos

ar: Chave estrangeira que faz referência à tabela **ARTIGO** pelo atributo id.  
chave: Chave estrangeira que faz referência à tabela **PALAVRA\_CHAVE** pelo atributo id.

#### Tabela ARTIGO\_AUTOR

ar	autor

#### Descrição dos atributos

ar: Chave estrangeira que faz referência à tabela **ARTIGO** pelo atributo id.  
autor: Chave estrangeira que faz referência à tabela **AUTOR** pelo atributo id.

#### Tabela ARTIGO\_MENSAGEM

ar	msg

#### Descrição dos atributos

ar: Chave estrangeira que faz referência à tabela **ARTIGO** pelo atributo id.  
msg: Chave estrangeira que faz referência à tabela **MENSAGEM** pelo atributo id.

# Capítulo 4

## O Aplicativo Web

A fim de disponibilizar uma interface amigável ao usuário do sistema, um aplicativo web foi desenvolvido. Este aplicativo consiste basicamente de um conjunto de interfaces que permitem ao usuário desenvolver as tarefas descritas pelos casos de uso.

As interfaces e o funcionamento do sistema foram projetados com as seguintes linguagens: HTML, PHP5, JavaScript, CSS e MySQL para o banco de dados. O PHP5 possibilitou o uso de *Design Patterns* - Singleton, na conexão com BD. Para a elaboração e gerenciamento do projeto em questão foram usadas as seguintes ferramentas: Phpmyadmin - para o BD e para o desenvolvimento do código o editor Gedit e a IDE NetBeans 6.9.1.

A escolha pelas linguagens e ferramentas descritas acima foi tomada com base nos conhecimentos adquiridos pelos acadêmicos no decorrer do curso, pela robustez das mesmas, pela capacidade de atender a elaboração de todos requisitos funcionais do sistema e por terem licença pública.

O sistema será hospedado em uma máquina servidora no Departamento de Psicologia da UFMS.

### 4.1 Interfaces do Sistema

O sistema implementado foi denominado SIAP - Sistema de Apoio à Pesquisa e tem por objetivo principal o armazenamento, consulta e gerenciamento de referências bibliográficas relacionadas ao tema Comportamento Verbal, da área de psicologia.

A seguir são apresentadas algumas das interfaces funcionais do sistema, a fim de exemplificar sua estrutura e funcionamento.

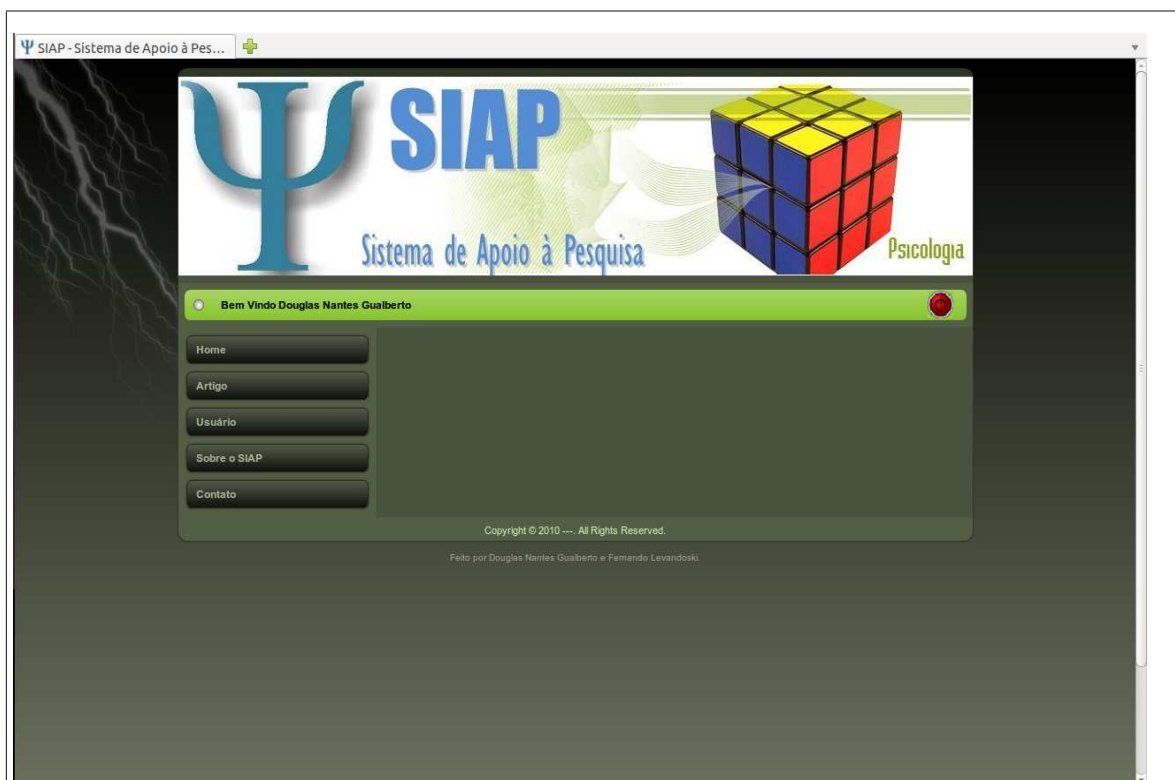


Figura 4.1: Interface referente à tela principal da ferramenta, após efetuado o login. Ela é idêntica para o usuário comum e o administrador. Nessa interface são apresentados o nome do sistema, o nome do usuário logado e o menu de opções na lateral esquerda.



Figura 4.2: Interface referente à tela de cadastro de artigo. Nessa tela todos os atributos do artigo que são interessantes para a pesquisa podem ser inseridos. O botão inserir ao lado dos campos busca no banco de dados e gera uma lista de valores já cadastrados para aquele campo, valores dessa lista podem ser selecionados para preencher o campo em questão. O nome do pesquisador e a data são preenchidos automaticamente.

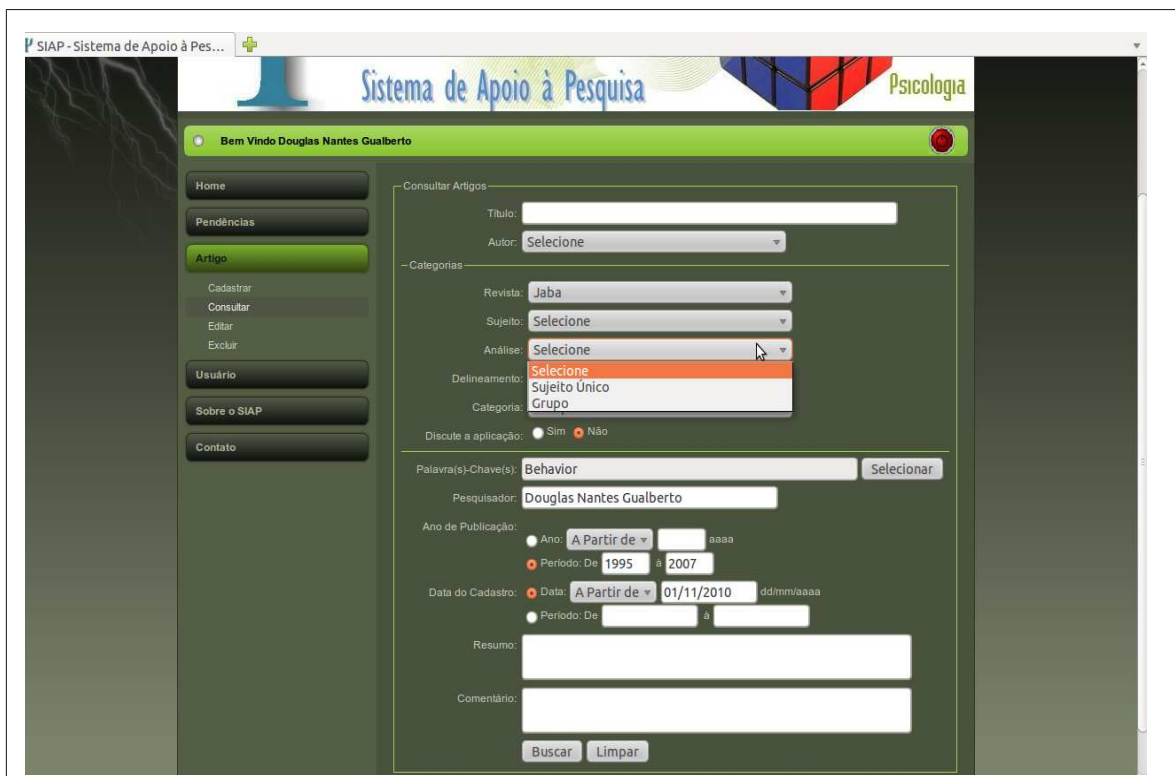


Figura 4.3: Interface referente à tela de consulta à artigos cadastrados. Os artigos cadastrados podem ser buscados a partir de qualquer um dos atributos inseridos na interface anterior. Alguns campos, como aqueles pertencentes à área categorias da interface, exibem uma lista de valores já cadastrados que podem ser selecionados facilitando assim a busca.





Figura 4.4: Exemplo de relatório para apresentação de resultados referentes a uma consulta à artigos. Os links gerados pelo software são referentes aos artigos que atendem à pesquisa solicitada e apresentam tanto o nome do autor como o nome do artigo. Pode ser gerado, através do botão "Gerar PDF" no canto inferior direito um pdf com os links gerados ou então seleciona-se um link e gera-se um pdf com todos os dados do artigo em questão.

# Capítulo 5

## Data Mining

### 5.1 Motivação

O aplicativo web irá agilizar de uma maneira geral todo o processo da pesquisa sobre Comportamento Verbal. Porém, é interessante que o coordenador da pesquisa possa dispor também de uma forma automatizada para realizar a classificação dos artigos dentro de categorias/classes por ele determinadas, visto que isso lhe proporcionará uma agilidade ainda maior uma vez que todo processo de análise dos atributos para saber em qual categoria cada artigo se enquadra passa a ser trabalho do software.

Diferentes ferramentas podem ser aplicadas com o objetivo de classificação de texto, como por exemplo SVN (*Support Vector Machine*) [7], KNN (*K-Nearest Neighbor*) [4] e *Naive Bayes* [1].

O *Naive Bayes* é considerado um dos modelos mais simples [8]. Apesar de sua simplicidade ele vem sendo amplamente utilizado com sucesso em diversas situações de classificação de textos [5] como por exemplo, na identificação de *spams*.

O algoritmo *Naive Bayes* foi escolhido como ferramenta da *Data Mining* para que o processo de classificação dos artigos fosse feito. O restante deste capítulo apresenta um estudo teórico sobre o algoritmo *Naive Bayes*, um exemplo da sua utilização para a classificação de um artigo cuja categoria não é conhecida e por fim a ferramenta Weka, que implementa o algoritmo descrito e é apresentada no contexto da sua utilização neste trabalho.

### 5.2 Classificação Bayesiana

A referência [8] foi utilizada na elaboração do texto que segue.

O algoritmo de Classificação Bayesiana recebe este nome por ser baseado no teorema de probabilidade de *Bayes*. Também é conhecido por classificador de *Naive Bayes* ou somente algoritmo de *Bayes*.

O algoritmo tem como objetivo calcular a probabilidade de que uma amostra desconhecida pertença a cada uma das classes possíveis, ou seja, prever a classe mais provável de uma nova instância. Este tipo de predição é chamada de classificação estatística, pois é completamente baseada em probabilidades.

Por ser um algoritmo de aprendizado supervisionado, ele requer um conjunto de dados já classificados. Baseado neste conjunto de dados classificados, que também é chamado de conjunto de treinamento, o algoritmo recebe como entrada uma nova instância desconhecida, ou seja, que não possui classificação, e retorna como saída a classe mais provável para esta instância de acordo com cálculos probabilísticos.

## 5.3 Processo de Classificação

### 5.3.1 Cálculos das probabilidades das classes

Neste passo, cada classe do conjunto de treinamento possui sua probabilidade calculada. O cálculo é feito dividindo-se o número de instâncias de determinada classe pelo número total de instâncias do conjunto de treinamento.

### 5.3.2 Cálculo da probabilidade de ocorrência de cada atributo da amostra desconhecida

Cada atributo da amostra desconhecida tem sua probabilidade calculada. Dados:

1. os atributos  $t1$ ,  $t2$  e  $t3$  da amostra desconhecida;
2. três classes conhecidas  $x$ ,  $y$ ,  $z$  e;
3. o conjunto de instâncias  $i$  já classificadas.

Para cada um dos atributos da amostra desconhecida verifica-se a ocorrência simultânea - em  $i$  - do valor desse atributo em cada uma das 3 classes já conhecidas. Por exemplo: Se há no conjunto de instâncias conhecidas 6 instâncias pertencentes à classe  $x$  e em 3 situações onde essa classe aparece, também aparece um valor de  $t1$  igual ao valor de  $t1$  da amostra desconhecida então tem-se uma probabilidade - para esse atributo - de 50% dessa amostra ser classificada como pertencente à classe  $x$ .

Este passo é onde o processamento mais pesado do algoritmo ocorre, pois, dependendo do número de atributos, classes e instâncias do conjunto de treinamento, é possível que muitos cálculos sejam necessários para se obter as probabilidades.

### 5.3.3 Cálculo da probabilidade total da amostra desconhecida

Neste passo, as probabilidades calculadas no item 5.3.2 para os valores da amostra desconhecida de uma mesma classe são multiplicadas entre si. Sempre que existir a pro-

bilidade de ocorrência de uma mesma classe e atributos diferentes essas probabilidades devem ser multiplicadas. O mesmo deve ser feito para cada uma das classes restantes. Em seguida, o valor obtido é multiplicado pela probabilidade da respectiva classe calculada no item 7.3.1.

Com as probabilidades de cada classe calculadas, verifica-se qual é a classe que possui maior probabilidade para a amostra desconhecida. Feito isso, o algoritmo termina retornando a classe mais provável para a amostra sendo classificada.

A seção seguinte detalha os passos descritos por esta seção, e as duas seções anteriores, por meio de um exemplo.

## 5.4 Aplicação do Algoritmo de Bayes

Considere um conjunto de treinamento de tamanho 15 como mostra a tabela da Figura 7.1.

Id	Título	Autor	Ano_Publicado	Pesquisador	Palavra-Chave	Revista	Sujeito	Análise	Delimitamento	Aplicação	Categoria
1	Psicologia Moderna	skinner	1908	levandolski	moderna psicologia skinner	jeab	criança	grupo	ab	sim	Independência funcional
2	Psicologia e Sociedade	platão	1500	douglas	plátão sociedade psicologia	jaba	adulto	grupo	aba	sim	Seguimento de regras
3	Comportamento Agressivo	freud	1408	levandolski	aristóteles agressão comportamento	jaba	Animal não humano	sujeito único	abab	não	Correspondência entre dizer e fazer
4	Metodologia de Ensino	Lucas	2010	levandolski	ensino método aula	jaba	criança	grupo	abab	sim	Independência funcional
5	Tendências Suicidas	Freud	1715	douglas	freud suicídio morte	jeab	adulto	sujeito único	abab	sim	Independência funcional
6	Harmonia	skinner	1903	douglas	skinner harmonia paz	jeab	Adulto com diagnóstico	sujeito único	ab	sim	Independência funcional
7	Parapsicologia	Kevedo	2007	douglas	medo parapsicologia fantasmas	jeab	criança com diagnóstico	grupo	ab	não	Independência funcional
8	Doenças mentais	skinner	1915	levandolski	bucura doença mental	jaba	adulto com diagnóstico	sujeito único	aba	não	Independência funcional
9	Alegria e Frustração	Lucas	2006	levandolski	alegria felicidade frustração	jaba	criança	sujeito único	aba	não	Correspondência entre dizer e fazer
10	Limpeza Mental	Lucas	2008	levandolski	mental limpeza paz	jaba	Adulto com diagnóstico	grupo	aba	não	Correspondência entre dizer e fazer
11	Religião e Psicologia	skinner	1512	levandolski	religião psicologia Deus	jaba	adulto	sujeito único	ab	não	Correspondência entre dizer e fazer
12	Idolatria	platão	1600	douglas	religião Deus fanatismo	jeab	adulto com diagnóstico	sujeito único	abab	sim	Correspondência entre dizer e fazer
13	Sexualidade na Escola	freud	1725	levandolski	sexo escola preconceito	jaba	criança com diagnóstico	sujeito único	aba	não	Correspondência entre dizer e fazer
14	Síndrome do Pânico	freud	1717	douglas	medo síndrome pânico	jaba	adulto	grupo	aba	sim	Seguimento de regras
15	Crescer saudável	Lucas	2005	douglas	criança crescer saúde	jeab	criança	grupo	ab	sim	Seguimento de regras

Figura 5.1: Artigos e suas respectivas categorias/objetos de estudo.

Note que são consideradas três classes distintas:

- Independência funcional;
- Seguimento de regras;
- Correspondência entre dizer e fazer.

Seja a instância apresentada na Figura 7.2, cuja categoria é desconhecida.

id	Título	Autor	Ano_Pub	Pesquisador	Palavra-Chave	Revista	Sujeito	Análise	Delimitação	Aplicação	Categoria
23	Psicologia para Leigos	skinner	1970	douglas	skinner leigo psicologia	jaba	adulto	sujeito único	abab	não	???

Figura 5.2: Amostra desconhecida.

Primeiramente calculamos a probabilidade de ocorrência de cada uma das 3 classes para o atributo *categoria*:

Independência Funcional	$6/15 = 0,4$
Correspondência entre dizer e fazer	$6/15 = 0,4$
Seguimento de regras	$3/15 = 0,2$

Agora para cada atributo da amostra desconhecida calculamos a probabilidade de seu valor estar dentro de cada uma das 3 classes do atributo *categoria*:

Atributo *Autor*: *Skinner*

Classe	Probabilidade
Independência Funcional	$3/6 = 0,5$
Correspondência entre dizer e fazer	$1/6 = 0,17$
Seguimento de regras	$0/3 = 0$

Atributo *Pesquisador*: *Douglas*

Classe	Probabilidade
Independência Funcional	$3/6 = 0,5$
Correspondência entre dizer e fazer	$1/6 = 0,17$
Seguimento de regras	$0/3 = 0$

Atributo *Palavra-Chave*: *Skinner, Leigo, Psicologia*

Calcula-se a probabilidade da existência de qualquer uma das 3 palavras-chaves em cada uma das 3 classes do objeto de estudo.

Classe	Probabilidade
Independência Funcional	$2/6 = 0,33$
Correspondência entre dizer e fazer	$1/6 = 0,17$
Seguimento de regras	$1/3 = 0,33$

Atributo Revista: *Jaba*

Classe	Probabilidade
Independência Funcional	$2/6 = 0,33$
Correspondência entre dizer e fazer	$4/6 = 0,67$
Seguimento de regras	$2/3 = 0,67$

Atributo Sujeito: *Adulto*

Classe	Probabilidade
Independência Funcional	$1/6 = 0,17$
Correspondência entre dizer e fazer	$1/6 = 0,17$
Seguimento de regras	$2/3 = 0,67$

Atributo Análise: *Sujeito Único*

Classe	Probabilidade
Independência Funcional	$3/6 = 0,5$
Correspondência entre dizer e fazer	$5/6 = 0,83$
Seguimento de regras	$0/3 = 0$

Atributo Delineamento: *abab*

Classe	Probabilidade
Independência Funcional	$2/6 = 0,33$
Correspondência entre dizer e fazer	$2/6 = 0,33$
Seguimento de regras	$0/3 = 0$

Atributo Aplicação: *não*

Classe	Probabilidade
Independência Funcional	$2/6 = 0,33$
Correspondência entre dizer e fazer	$5/6 = 0,83$
Seguimento de regras	$0/3 = 0$

Conhecendo-se todas as probabilidades dos atributos da amostra desconhecida em relação a cada uma das 3 classes, multiplica-se todas aquelas referentes à mesma classe do objeto de estudo juntamente com a probabilidade da classe em questão. Veja os resultados obtidos para o exemplo utilizado:

**Probabilidade da amostra desconhecida pertencer à alguma das classes já conhecidas:**

Classe	Probabilidades	Probabilidade Final
Independência Funcional	$0,5 * 0,5 * 0,33 * 0,33 * 0,17$ $* 0,5 * 0,33 * 0,33 * 0,4$	0,010
Correspondência entre dizer e fazer	$0,17 * 0,17 * 0,17 * 0,67 * 0,67$ $* 0,17 * 0,83 * 0,33 * 0,83 * 0,4$	0,005
Seguimento de regras	$0 * 1 * 0,33 * 0,67 * 0,67$ $* 0 * 0 * 0 * 0,2$	0

Como a maior probabilidade calculada (0,010) corresponde à classe *Independência Funcional*, o classificador Naive Bayes retorna esta como sendo a classe mais provável para a amostra desconhecida.

## 5.5 Implementação do Algoritmo

Optou-se por não implementar o algoritmo de *Bayes* e sim pela utilização da implementação disponibilizada pela ferramenta Weka - *Waikato Environment for Knowledge Analysis* [6]. A ferramenta Weka forneceu uma solução imediata para o problema de classificação das categorias, porém, se com o uso constante do sistema a ferramenta não se mostrar uma solução eficiente para a pesquisa, será então desenvolvido um classificador específico para a pesquisa.

O aplicativo Weka disponibiliza a implementação de uma coleção de algoritmos de aprendizado de máquina direcionados para a realização de tarefas de mineração de dados. Além de algoritmos de aprendizado, o Weka conta com uma variedade de algoritmos que facilitam a manipulação das principais estruturas de dados utilizadas por algoritmos de aprendizado. Tais facilidades fazem do Weka um ambiente apropriado para o desenvolvimento de novos esquemas de aprendizado de máquina. Weka é software livre e possui licença do general público do GNU.

O Weka foi desenvolvido na plataforma Java, a qual é orientada a objetos. Assim, existe uma coleção de classes que implementam os algoritmos de aprendizado,<sup>1</sup> as estruturas de dados utilizadas por esses algoritmos e outras funcionalidades interessantes em um processo de mineração de dados.

Weka suporta várias tarefas padrões em mineração de dados, mais especificamente, os pré-processamentos, agrupamento, classificação, regressão, visualização e seleção de características. Todas as técnicas de previsão partem do pressuposto de que os dados estão disponíveis como um único arquivo, onde cada dado é descrito por um número fixo de atributos (normalmente, numérica ou nominal). Weka também fornece acesso a bases

<sup>1</sup>Existe uma classe a `weka.classifiers.bayes.NaiveBayes` que implementa o método abstrato `buildClassifier` da classe `Classifier` usando o algoritmo de aprendizado supervisionado Naive Bayes.



de dados SQL usando JDBC e pode processar o resultado retornado por uma consulta à base de dados.

Usando-se dessa característica de conexão com BD do Weka, a cada nova categoria definida pelo coordenador da pesquisa, é consultada a tabela CATEGORIAS (Capítulo 5) e com os dados desta tabela o Weka gera uma nova base de conhecimento para ser usada na classificação dos artigos a partir daquele momento.

### 5.5.1 Resultados obtidos com Weka

Com base no conjunto de treinamento descrito na Figura 7.1 e na amostra desconhecida apresentada na Figura 7.2, a utilização da implementação do *Naive Bayes* disponibilizado pelo weka gerou os resultados apresentados na Figura 7.3.

```

:classifier output
=== Predictions on training set ===
inst#,  actual, predicted, error, probability distribution
1 1:independ 1:independ      *0.984 0.011 0.005
2 2:sequimen 2:sequimen      0.003 *0.995 0.002
3 3:correspo 3:correspo      0.005 0      *0.995
4 1:independ 1:independ      *0.871 0.064 0.065
5 1:independ 1:independ      *0.919 0.061 0.021
6 1:independ 1:independ      *0.987 0.005 0.007
7 1:independ 1:independ      *0.942 0.052 0.006
8 1:independ 3:correspo      + 0.47 0.002 *0.528
9 3:correspo 3:correspo      0.007 0.002 *0.991
10 3:correspo 3:correspo      0.027 0.007 *0.966
11 3:correspo 3:correspo      0.069 0.002 *0.929
12 3:correspo 3:correspo      0.243 0.057 *0.7
13 3:correspo 3:correspo      0.005 0.001 *0.994
14 2:sequimen 2:sequimen      0.01 *0.986 0.003
15 2:sequimen 2:sequimen      0.081 *0.917 0.002
16      ? 1:independ      + *0.56 0.019 0.42

=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances      14      93.3333 %
Incorrectly Classified Instances    1      6.6667 %
kappa statistic                    0.8958
Mean absolute error                 0.0601
Root mean squared error             0.1317
Relative absolute error             13.992 %
Root relative squared error        28.5054 %
Total Number of Instances          15
Ignored Class Unknown Instances    1

=== Confusion Matrix ===
a b c <- classified as
5 0 1 | a = independenciaFuncional
0 3 0 | b = sequimentoDeRegras
0 0 6 | c = correspondenciaDizer e Fazer

```

Figura 5.3: Resultados obtidos com o aplicativo Weka.

Dado o pequeno número de instâncias fornecidas para aprendizagem, ainda assim os resultados foram satisfatórios. Apenas uma instância foi classificada de forma incorreta.

Para a instância da Figura 7.4 a classificação correta para a categoria - segundo a ferramenta Weka - seria *Correspondência entre Dizer e Fazer* e não *Independência Funcional*.

Para a amostra desconhecida (Figura 7.2) inserida no conjunto de treinamento a classificação do Weka foi idêntica à realizada na seção 7.4.

id	Título	Autor	Ano_Pub	Pesquisador	Palavra-Chave	Revista	Sujeito	Análise	Delineamento	Aplicação	Categoria
8	Doenças mentais	skinner	1915	levandoski	loucura doença mental	jaba	adulto com diagnostico	sujeito único	aba	não	<b>independência funcional</b>

Figura 5.4: Amostra com classificação incorreta.

A fim de contornar o problema das instâncias classificadas incorretamente, é permitido ao coordenador da pesquisa (e administrador do BD) alterar uma classificação dada pelo processo de classificação automático.

Cada classificação errada do algoritmo é corrigida pelo coordenador e, através do Weka, pode-se gerar uma nova base de conhecimento para as classificações futuras.

# Capítulo 6

## Conclusão

O desenvolvimento do software proposto neste trabalho envolveu o estudo de áreas relacionadas à Engenharia de Software, Banco de Dados, programação para Web e Data Mining, mais especificamente inteligência artificial, que pode ser visto como um "ramo" de Data Mining. Entretanto o maior desafio enfrentado pelos acadêmicos foi o entendimento do assunto *Comportamento Verbal* - contexto contemplado pelo projeto - devido sua extensa área de abrangência. Tal entendimento se fez necessário para que o projeto do banco de dados, a implementação das interfaces e toda sistemática do software desenvolvido atendessem completamente às necessidades dos pesquisadores envolvidos na pesquisa sobre Comportamento Verbal coordenada pelo professor Dr. Lucas Cordova do Departamento de Psicologia desta Universidade.

Toda a base da documentação do sistema - requisitos, casos de uso, modelos de interface - foi proveniente de constantes reuniões com o coordenador e membros da pesquisa.

Não houve - até o fim do oitavo semestre do curso dos acadêmicos responsáveis pelo projeto - tempo hábil para se mensurar o quanto o coordenador e membros da pesquisa ganharam em eficiência e eficácia em relação à antiga sistemática de cadastro/classificação dos artigos. Tal dado - acreditam os acadêmicos - só será revelado nos próximos meses com o uso constante do aplicativo.

Sendo comprovada a eficácia do sistema proposto, poderá ser estudada a aplicação do mesmo em outras áreas. Com poucas mudanças o SIAP pode atender satisfatoriamente a outras pesquisas desenvolvidas.

# Referências Bibliográficas

- [1] MCCALLUM A. and NIGAM K. *Comparison of event models for Naive Bayes text classification. AAAI98 Workshop on Learning for Text Categorization, Vol. 752.* Citeseer, pp. 41-48. DOI: 10.1.1.46.1529, 1998. 5.1
- [2] MARTINS Ana, NUNES Manuela Barreto, and RODRIGUES Eloy. *Repositórios de informação e ambientes de aprendizagem: Criação de espaços virtuais para a promoção da literacia e da responsabilidade social.* Rede de Bibliotecas Escolares Newsletter 3, 2008. Disponível em: [http://www.rbe.min-edu.pt/newsletter/newsletter3/newsleter\\_n3\\_ficheiros/page00014.htm](http://www.rbe.min-edu.pt/newsletter/newsletter3/newsleter_n3_ficheiros/page00014.htm), acessado em 30 de Julho de 2010. 1
- [3] SKINNER B. F. *O Comportamento Verbal.* São Paulo: Cultrix, 1978. 1
- [4] LIM H. S. *Improving kNN Based Text Classification with Well Estimated Parameters.* Lecture Notes in Computer Science, Neural Information Processing., 2004. Disponível em: [http://dx.doi.org/10.1007/978-3-540-30499-9\\_79](http://dx.doi.org/10.1007/978-3-540-30499-9_79), acessado em 19 de novembro de 2010. 5.1
- [5] ANDROUTSOPOULOS I., KOUTSIAS J., CHANDRINOS K. V., PALIOURAS G., and SPYROPOULOS G. *An Evaluation of Naive Bayesian Anti-Spam Filtering.* 2000. Proceedings of the workshop on Machine Learning in the New Information Age, 11th European Conference on Machine Learning, Barcelona, Spain, pp. 9-17, 2000. 5.1
- [6] WITTEN I. H. and FRANK E. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2ª edição., 2005. 5.5
- [7] YANG Liu Y. *DA Re-Examination of Text Categorization Methods.* ACM Special Interest Group of Information Retrieval (SIGIR), 1999. pp 42-49. 5.1
- [8] PICHILIANI Mauro. *Data Mining na Prática: Classificação Bayesiana.* Rede de Bibliotecas Escolares Newsletter 3, 2010. Disponível em: [http://imasters.com.br/artigo/4926/sql\\_server/data\\_mining\\_na\\_pratica\\_classificacao\\_bayesiana](http://imasters.com.br/artigo/4926/sql_server/data_mining_na_pratica_classificacao_bayesiana), acessado em 10 de novembro de 2010. 5.1, 5.2
- [9] ELMASRI R. and NAVATHE S. R. *Sistemas de Bancos de Dados.* São Paulo: Pearson Addison Wesley, 4 edition, 2005. 3, 3.2

- [10] PRESSMAN R. S. *Engenharia de Software*. São Paulo: McGraw-Hill, 6 edition, 2006. 2, 2.5