



# IMPLEMENTAÇÃO DA TÉCNICA TEXT MINING APLICADO NO AMBIENTE DO EMPREENDEDOR

**Acadêmico:** Roger Erivan Gaulke

**Prof. Orientador:** Dr. Oscar Dalfovo

# Roteiro da apresentação

- ✓ Introdução
  - Objetivos do Trabalho
  
- ✓ Fundamentação teórica
  - Plano de Negócio
  
- ✓ Text Mining
  - Stopwords
  - Stemmer
  - Conceitos

# Roteiro da apresentação

- ✓ Desenvolvimento do Trabalho
  - Especificação
  - Tecnologias e ferramentas utilizadas
  - Operacionalidade
  - Trabalhos Correlatos
  
- ✓ Considerações Finais
  - Resultados
  - Extensões

# Introdução

## ✓ Contextualização

- 80% dos dados de uma organização estão armazenados em forma de dados não estruturados (texto);  
(Gadcon, 2006)
- mineração de texto é o processo de obtenção de dados e informações a partir de textos em linguagens naturais.  
(Wives, 2004)

# Introdução

✓ Objetivo geral

O objetivo deste trabalho é a aplicação da técnica *Text Mining* nos planos de negocio, desenvolvidos pelos alunos na disciplina de empreendedorismo da Universidade Regional de Blumenau a fim de ajudar os professores a verificar as inconsistências descritas em cada plano.

# Introdução

- ✓ Objetivos específicos
  - identificar informações que possibilitem apresentar inconsistências e incoerências no plano de negócio;
  - apresentar graficamente o acompanhamento do percentual de incoerências, inconsistências;
  - aplicar os passos da técnica *Text Mining* nos módulos do plano de negócio descrito no Ambiente do Empreendedor selecionados pelo professor.

# Fundamentação Teórica

## ✓ Plano de Negócio

- é um documento pelo qual o empreendedor formalizará os estudos a respeito de suas idéias, transformando-as em um negócio;

(Sebrae, 2007)

- deve-se criar o plano de negócio, pois o processo de organizá-lo, obriga o empreendedor a assumir um posicionamento racional, crítico e objetivo em relação ao projeto como um todo;

(Bangs, 1999)

# Fundamentação Teórica

- ✓ Plano de Negócio
  - destina-se aos sócios da empresa, aos funcionários, aos investidores e ao público em geral; (Salin, 2001)
  - pode ser organizado em 3 etapas: o negócio, dados financeiros e documentação de apoio; (Bangs, 1999)



# Text Mining

É a tecnologia capaz de analisar grandes coleções de documentos não estruturados com o propósito de extrair conhecimento, padrões de interesse previamente desconhecidos e potencialmente úteis.

# Text Mining

## Experimento utilizado por Wives

Experimentos e seus objetivos	
Experimento	Objetivo
(1) Normal	Verificar o resultado de um processo de agrupamento tradicional, no qual todas as palavras são utilizadas sem qualquer espécie de pré-processamento, a fim de verificar o ganho real da utilização dos conceitos.
(2) <i>Stopwords</i>	Verificar o resultado do processo de agrupamento com o pré-processamento mais comum e simples que corresponde à exclusão de palavras irrelevantes devido à sua categoria lingüística (artigos, preposições, etc).
(3) <i>Stemmer</i>	Verificar a influência de um <i>stemmer</i> (reduzidor de palavras ao seu radical) no resultado do agrupamento.
(4) <i>Stopwords + stemmer</i>	Avaliar a combinação dos dois métodos anteriores.
(5) Conceitos	Avaliar a influência dos conceitos no processo de agrupamento.
(6) <i>Stopwords + conceitos</i>	Identificar se a remoção de <i>stopwords</i> facilita a identificação de conceitos e se melhora o resultado do agrupamento.
(7) <i>Stopwords + stemmer + conceitos</i>	Identificar se a remoção de <i>stopwords</i> e o <i>stemmer</i> facilitam a identificação de conceitos e se melhoram o resultado do agrupamento.
(8) Refinamento dos conceitos pelo <i>WordWeb</i>	Identificar se a adição manual de sinônimos ( <i>similar</i> , <i>synonyms</i> ) e subtipos ( <i>sub-type</i> ), com base no dicionário <i>WordWeb</i> , melhora a qualidade dos conceitos. OBS: também se exclui <i>stopwords</i> e se realiza <i>stemmer</i> .
(9) Refinamento por subdivisão de conceitos	Verificar se conceitos mais específicos (menos abrangentes) descrevem melhor os documentos e obtêm melhores resultados.

# Text Mining

- ✓ Stopwords

Palavras muito frequentes nos textos, sendo de baixa discriminação e portanto inúteis para representar e distinguir os textos uns dos outros.

Exemplos: “de”, “que”, “e”, “a” e “o”

# Text Mining

## ✓ Stemmer

É a padronização do vocabulário onde as palavras são reduzidas a um único radical, ou seja, busca representar em um único termo ou forma eliminando suas diferenças morfológicas ou léxicas.

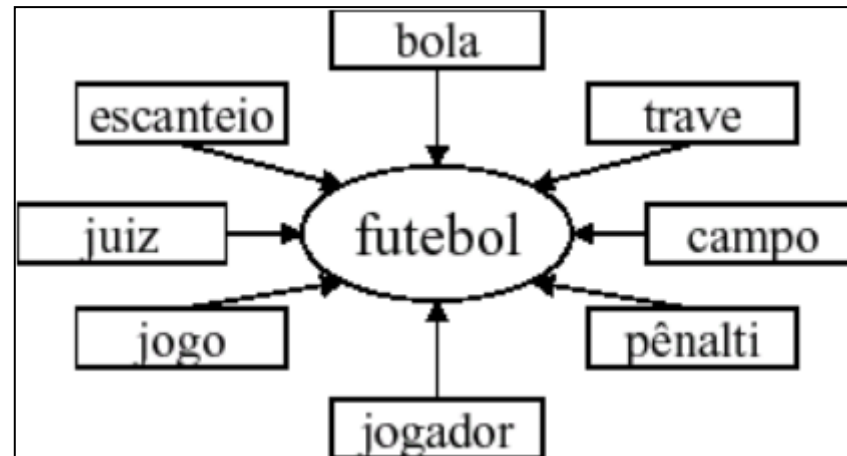
Exemplos: escrever – escrevendo, escreve.

# Text Mining

## ✓ Conceitos

São fragmentos de conhecimento que o ser humano usa para representar ideias, opiniões e pensamentos.

Exemplos:



# Desenvolvimento do trabalho

- ✓ Especificação
  - requisitos funcionais;
  - diagrama de casos de uso.

# Desenvolvimento do trabalho

## ✓ Requisitos Funcionais

RF01: Permitir o acompanhamento das porcentagens das comparações efetuadas entre os módulos.

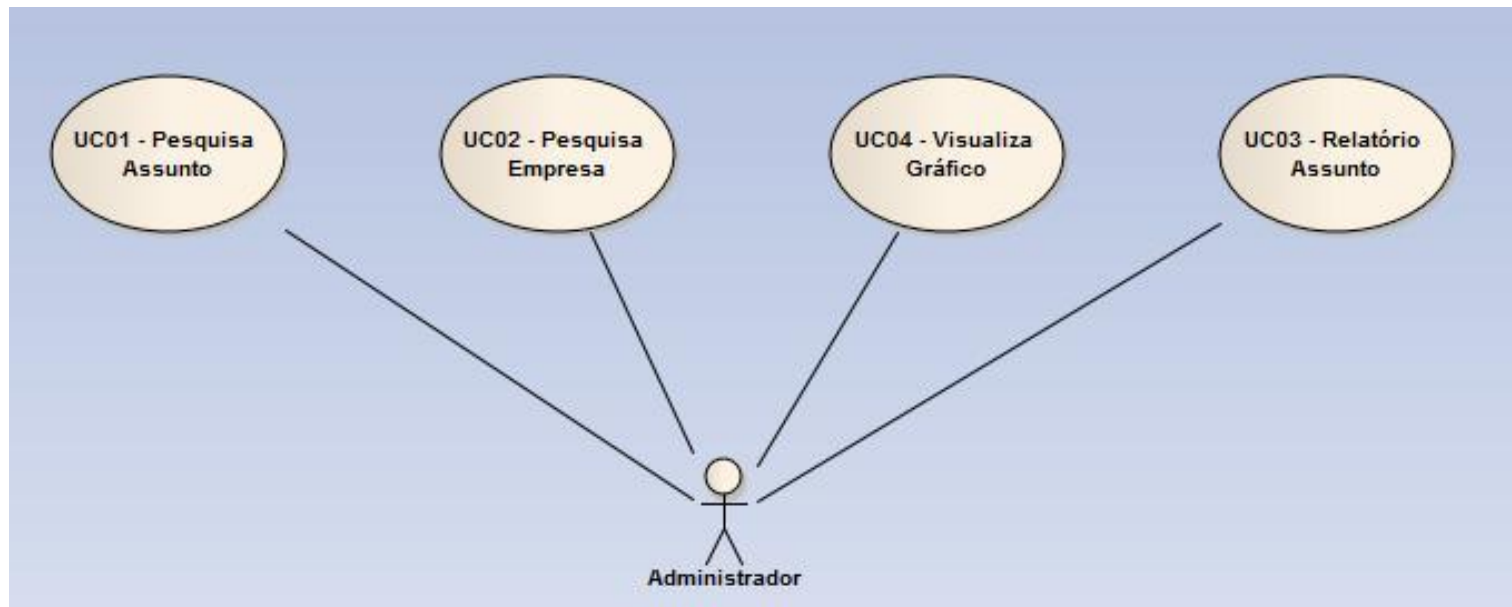
RF02: Permitir a visualização de um relatório com todos os assuntos já desenvolvido plano de negócio.

RF03: Permitir a comparação dos planos de negócios existentes, com os que estão sendo desenvolvidos para verificar a já existência do mesmo assunto.

RF04: Permitir a busca de planos de negócios por assunto.

# Desenvolvimento do trabalho

- ✓ Diagrama de casos de uso do administrador

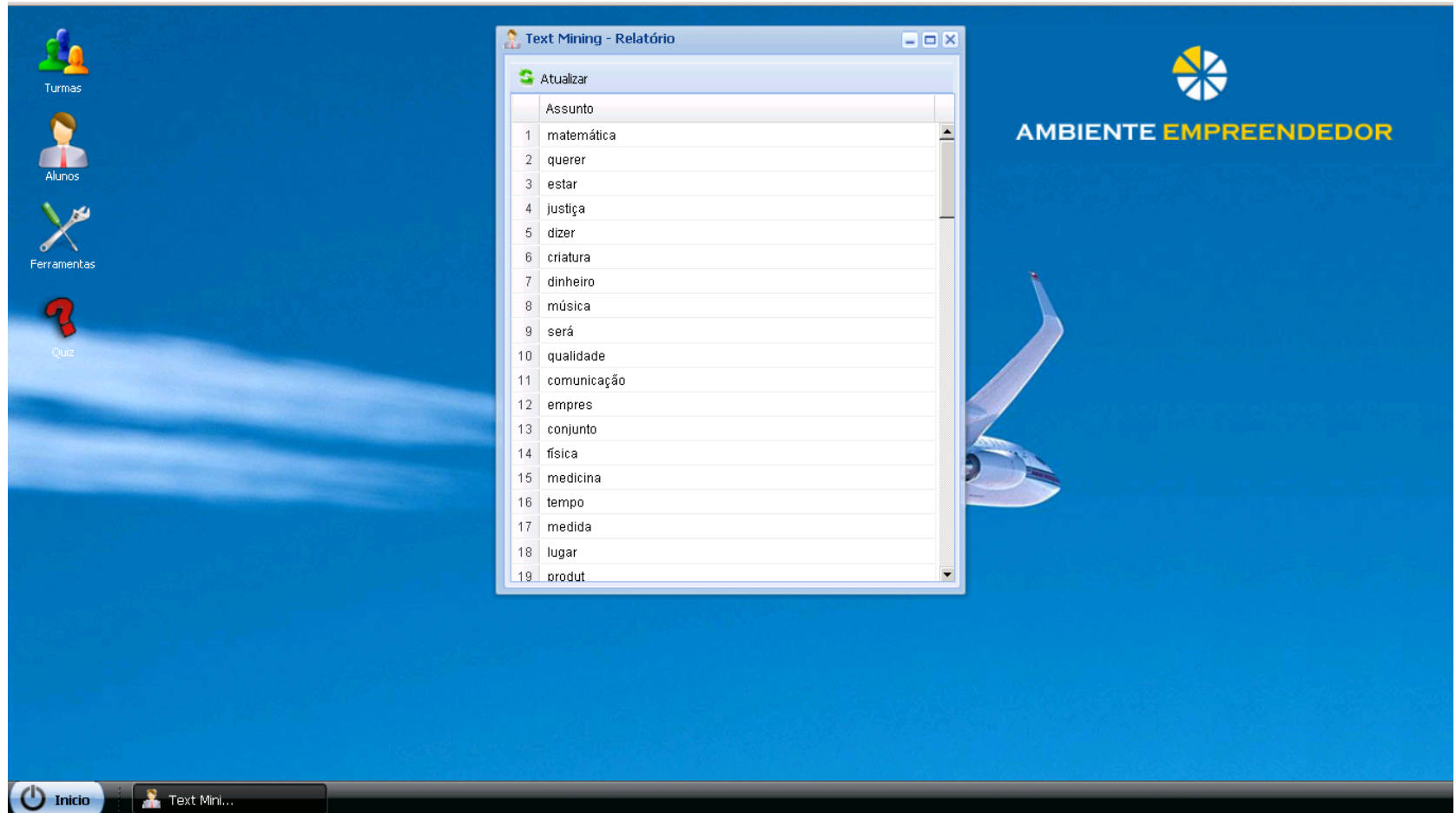




# Desenvolvimento do trabalho

- ✓ Técnicas e ferramentas utilizadas
  - Enterprise Architect - UML:  
criação dos diagramas e casos de uso;
  - PHP e JavaScript:  
programação da ferramenta;
  - MySQL:  
base de dados de informações do sistema;
  - XML e JSON:  
troca de informações entre o php e o javascript;
  - Framework EXT.js:  
programação da interface visual do ambiente.

# Operacionalidade

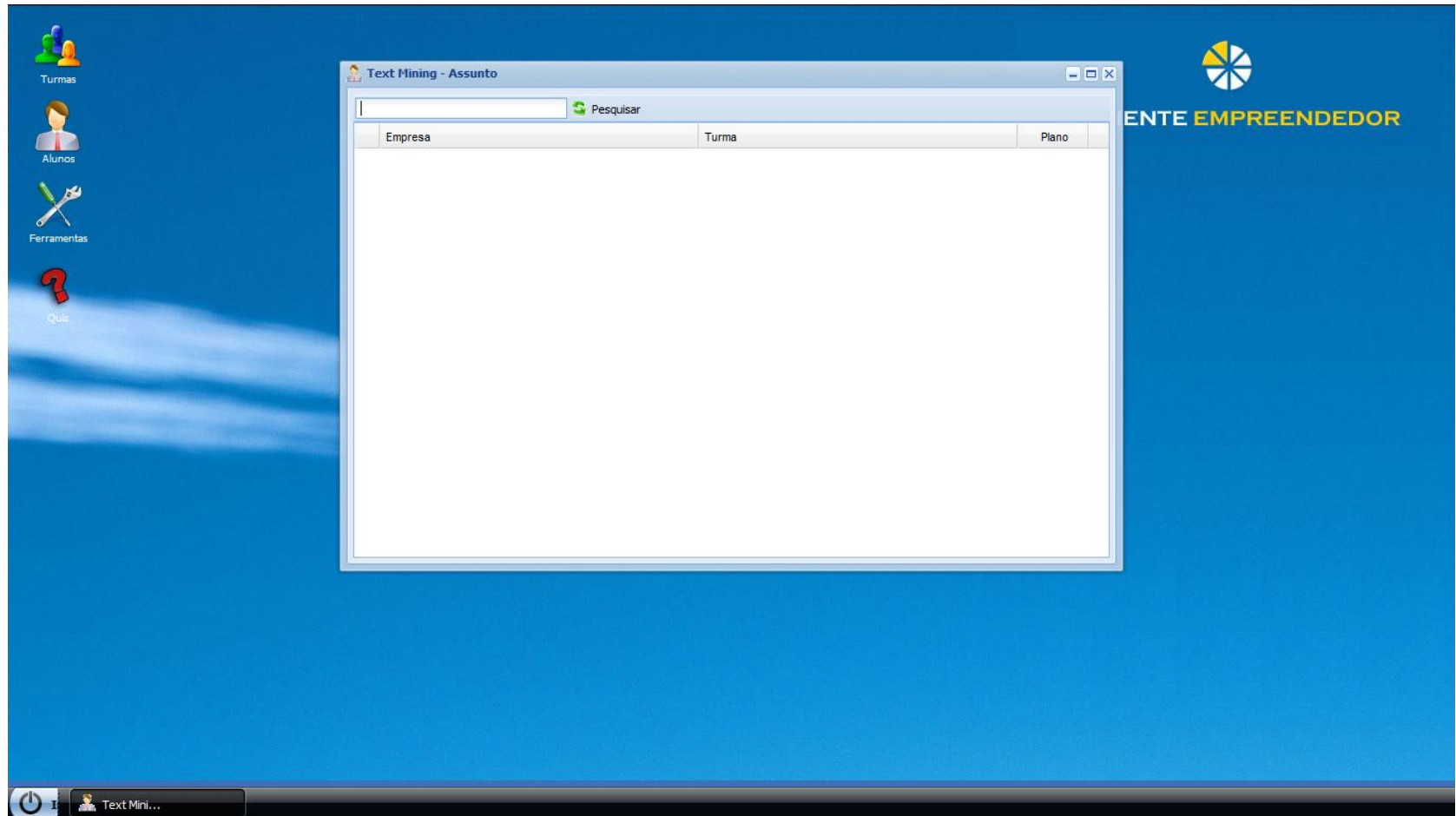


The screenshot displays a software interface with a blue background. On the left, there is a vertical sidebar with four icons: 'Turmas' (classrooms), 'Alunos' (students), 'Ferramentas' (tools), and 'Quiz'. The main area features a window titled 'Text Mining - Relatório' with a table of subjects. The table has a header 'Assunto' and 19 rows of data. To the right of the window, there is a logo for 'AMBIENTE EMPREENDEDOR' and a partial image of an airplane. The taskbar at the bottom shows the 'Início' button and a task for 'Text Mini...'.

	Assunto
1	matemática
2	querer
3	estar
4	justiça
5	dizer
6	criatura
7	dinheiro
8	música
9	será
10	qualidade
11	comunicação
12	empres
13	conjunto
14	física
15	medicina
16	tempo
17	medida
18	lugar
19	produt

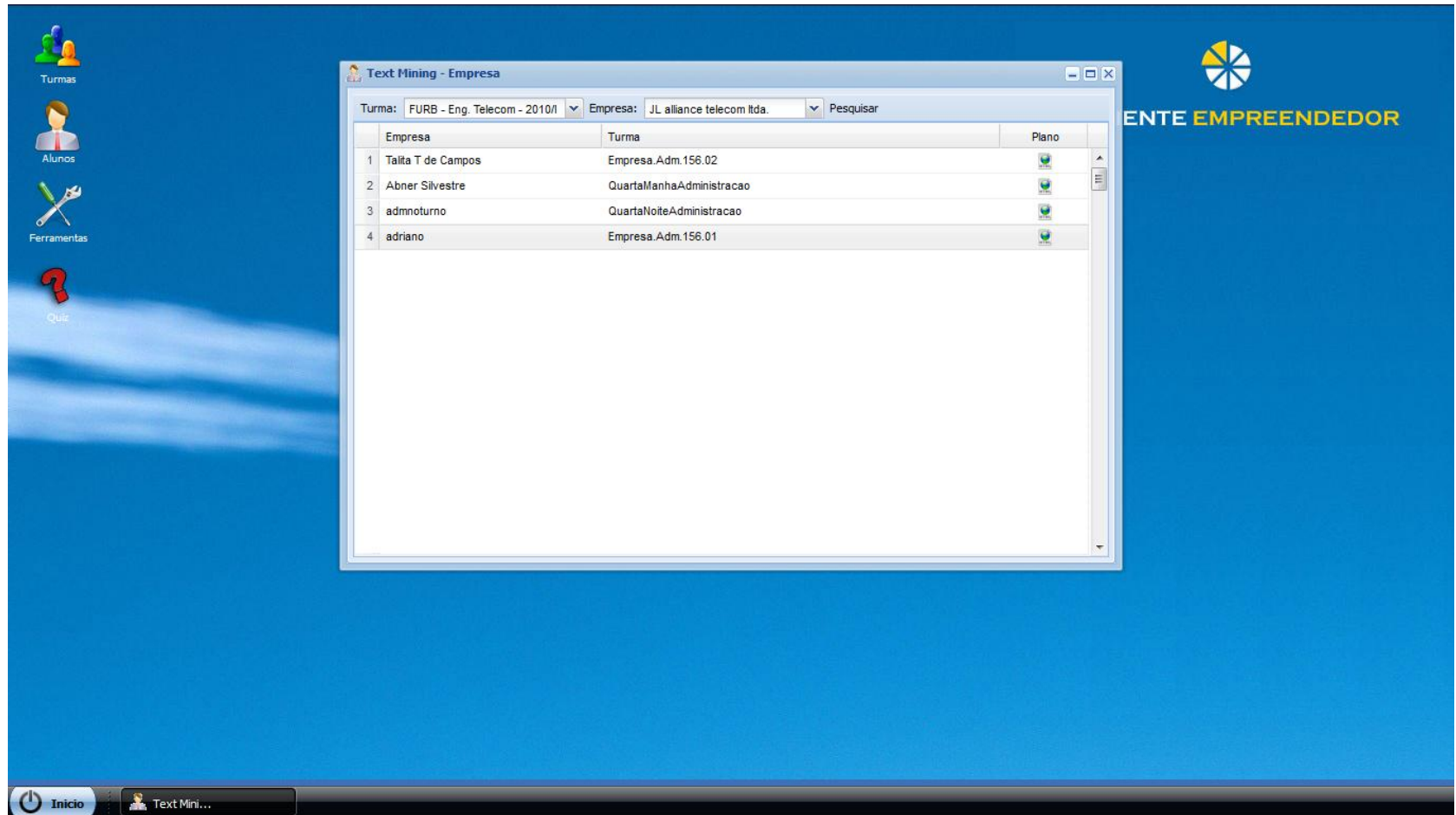
Tela de relatório de assuntos

# Operacionalidade



Tela pesquisa por assunto

# Operacionalidade

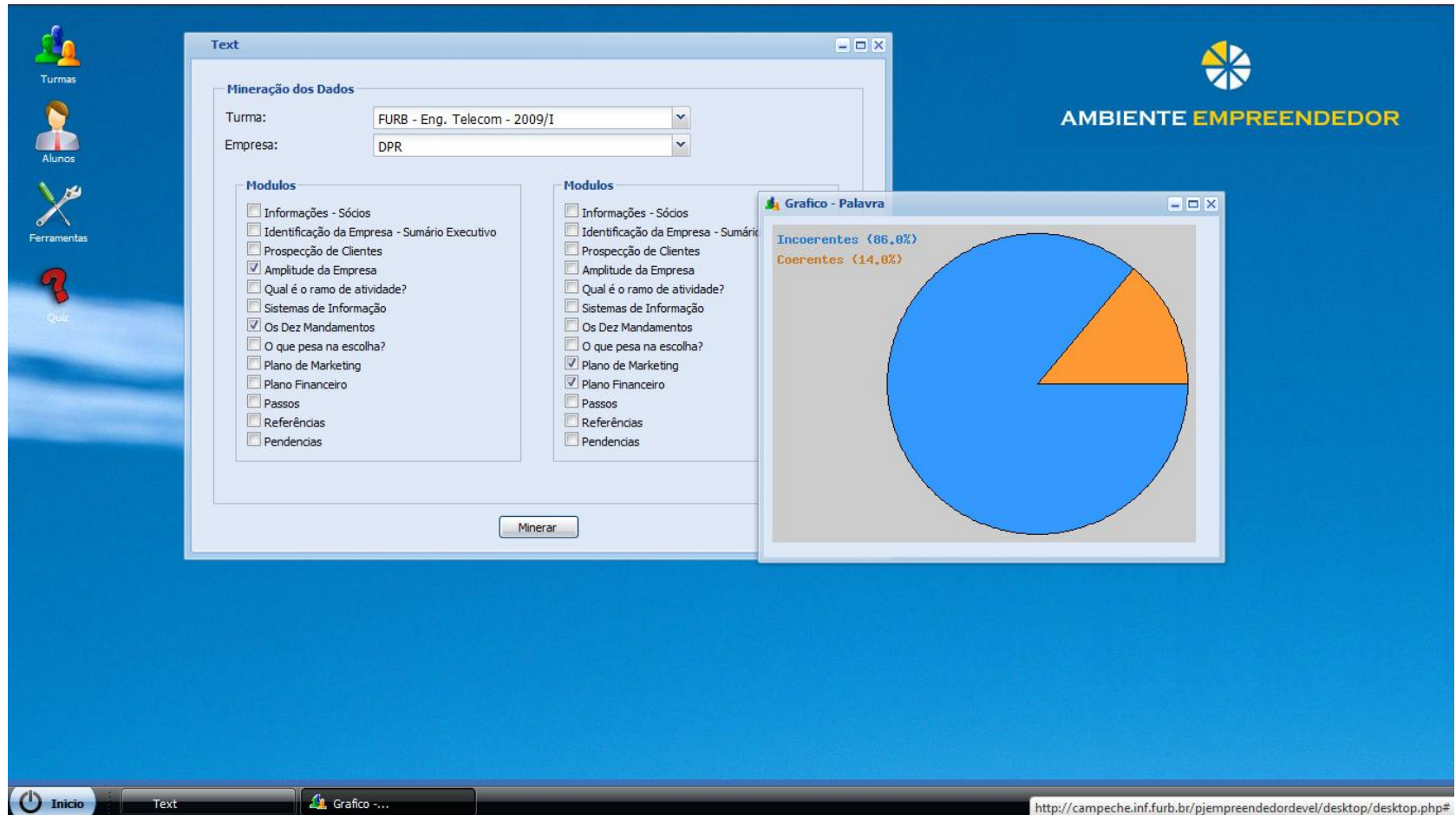


ENTE EMPREENDEDOR

	Empresa	Turma	Plano
1	Talita T de Campos	Empresa.Adm.156.02	
2	Abner Silvestre	QuartaManhaAdministracao	
3	admoturno	QuartaNoiteAdministracao	
4	adriano	Empresa.Adm.156.01	

Tela de pesquisa por empresa

# Operacionalidade



**Mineração dos Dados**

Turma: FURB - Eng. Telecom - 2009/I  
 Empresa: DPR

**Modulos**

- Informações - Sócios
- Identificação da Empresa - Sumário Executivo
- Prospecção de Clientes
- Amplitude da Empresa
- Qual é o ramo de atividade?
- Sistemas de Informação
- Os Dez Mandamentos
- O que pesa na escolha?
- Plano de Marketing
- Plano Financeiro
- Passos
- Referências
- Pendências

**Modulos**

- Informações - Sócios
- Identificação da Empresa - Sumário
- Prospecção de Clientes
- Amplitude da Empresa
- Qual é o ramo de atividade?
- Sistemas de Informação
- Os Dez Mandamentos
- O que pesa na escolha?
- Plano de Marketing
- Plano Financeiro
- Passos
- Referências
- Pendências

**Gráfico - Palavra**

Incoerentes (86,0%)  
 Coerentes (14,0%)

AMBIENTE EMPREENDEDOR

http://campeche.inf.furb.br/pjempreendedorlevel/desktop/desktop.php#

Tela de visualização gráfica

# Desenvolvimento do trabalho

## ✓ Trabalhos correlatos

- programa que analisa os chamados telefônicos, onde é identificado o motivo do problema aberto pelo cliente (Uber, 2004);
- sistema que possibilita a extração de conceitos-chaves, percepções e a relação dos dados não estruturados (SPSS, 2007).

# Considerações Finais

## ✓ Resultados

- Com este trabalho identificou-se as influências motivadas pelos ambientes na capacidade de tomada de decisão no ambiente empreendedor;
- Ele auxiliou o desenvolvimento de conteúdos para plano de negocio, no sentido de criar valor para o relacionamento entre o trabalho acadêmico e o plano de negocio;

# Considerações Finais

## ✓ Resultados

- Com este trabalho pode-se criar um jeito diferente de fazer o trabalho proposto havendo inúmeras possibilidades de serem feitas umas melhores outras nem tanto, esse trabalho vem mostrar que dependendo do modo como é feito pode influenciar no resultado final;
- O trabalho contém algumas inconsistências que só através de vários testes poderá ir aperfeiçoando-se a ferramenta, já que alguns conceitos retirados da técnica *Text Minig* não estão coerentes com o linguajar acadêmico, sendo algumas vezes as palavras-chaves encontradas inadequadas para a ocasião.



# Considerações Finais

## ✓ Conclusões

- a utilização de uma ferramenta de ensino é de suma importância para a disciplina de empreendedor em informática;
- *Text Mining* ainda é uma grande área a ser abordada podendo ser criadas outras técnicas para a verificação dos dados.

# Considerações Finais

## ✓ Extensões

- ferramentas para troca de informações, como salas de bate papo e fórum de discussão;
- ferramenta para controle de versão do plano de negócio;
- pesquisar outras técnicas de KDT para ser utilizada no ambiente além da que foi aplicada neste trabalho.

## Referencias

- ✓ BANGS, D. H. **Planejamento de negócios**: criando um plano para seu negócio ser bem-sucedido. Tradução Rosa Krausz. São Paulo: Nobel, 1999.
- ✓ GADCON. **SPSS lança solução inédita para análise preditiva para português**. São Paulo, 2006. Disponível em:  
<[http://www.gadcom.com.br/ultimas\\_noticias/ultimasnoticias\\_item.asp?cod=1249](http://www.gadcom.com.br/ultimas_noticias/ultimasnoticias_item.asp?cod=1249)>. Acesso em: 13 ago. 2008
- ✓ SALIM, C. S. et al. **Construindo planos de negócios**: todos os passos necessários para planejar e desenvolver negócios de sucesso. Rio de Janeiro: Campus, 2001.
- ✓ SANTIAGO, Rafael Wilson. **Desenvolvimento de um ambiente web para apoio do empreendedor utilizando AJAX**. 2008. 64 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau.
- ✓ WIVES, Leandro Krug. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos**. 2004. 136 f. Tese (Doutorado no Programa de Pós-Graduação em Computação) - Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre.



**Obrigado!**