

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

IMPLEMENTAÇÃO DA TÉCNICA *TEXT MINING* APLICADO
NO AMBIENTE DO EMPREENDEDOR

ROGER ERIVAN GAULKE

BLUMENAU
2011

2011/2-28

ROGER ERIVAN GAULKE

**IMPLEMENTAÇÃO DA TÉCNICA *TEXT MINING* APLICADO
NO AMBIENTE DO EMPREENDEDOR**

Trabalho de Conclusão de Curso submetido à Universidade Regional de Blumenau para a obtenção dos créditos na disciplina Trabalho de Conclusão de Curso II do curso de Ciência da Computação — Bacharelado.

Prof. Oscar Dalfovo, Dr. – Orientador

**BLUMENAU
2011**

2011/2-28

IMPLEMENTAÇÃO DA TÉCNICA *TEXT MINING* APLICADO NO AMBIENTE DO EMPREENDEDOR

Por

ROGER ERIVAN GAULKE

Trabalho aprovado para obtenção dos créditos na disciplina de Trabalho de Conclusão de Curso II, pela banca examinadora formada por:

Presidente:

Prof. Oscar Dalfovo, Dr. – Orientador, FURB

Membro:

Prof. Everaldo Artur Grahl, Mestre – FURB

Membro:

Prof. Ricardo de Alencar Azambuja, Mestre – FURB

Blumenau, 12 de dezembro de 2011

Dedico este trabalho a todos os meus familiares, amigos e mestres que contribuíram direta ou indiretamente na realização deste.

RESUMO

Neste trabalho é desenvolvido e apresentado um aplicativo on-line para auxiliar os gestores das disciplinas em empreendedorismo da Universidade Regional de Blumenau (FURB), que utilizam o Ambiente do Empreendedor, como método e auxílio de ensino aos seus alunos. O objetivo deste trabalho é apresentar a implementação da técnica *text mining* no ambiente, comparando-se os textos entre os módulos, com isto, verificando as inconsistências do plano de negócio criado no Ambiente do Empreendedor, para que possa, auxiliar o gestor na análise do plano de negócio. A ferramenta foi desenvolvida com foco na interação do usuário, para isso, foi utilizado para o desenvolvimento das interfaces as linguagens de programação PHP e AJAX, com acesso ao banco de dados MYSQL. Como resultado, através dessa nova funcionalidade implementada, o Ambiente do Empreendedor, mostrou-se uma maior agilidade na entrega das avaliações do plano de negocio, pelo administrador da disciplina.

Palavras-chave: *Text mining*. Ambiente do empreendedor. Empreendedorismo.

ABSTRACT

In this work is developed and presented an online application to help managers in entrepreneurship courses at the Universidade Regional de Blumenau (FURB), using the Ambiente do Empreendedor as a method of teaching and assisting their students. This paper present a text mining technique implemented on the environment, comparing the text between modules, checking the inconsistencies of the business plan created in the Ambiente do Empreendedor and helping the manager in the business plan review. The tool was developed with a focus on user interaction, using the AJAX development technique for interfaces, more the PHP programming language, with access to MySQL database. As result, through this new feature implemented, the Ambiente do Empreendedor proved more agile in the delivery of assessments business plan and notes to students, by the administrator of the discipline.

Key-words: Text mining. Ajax. Ambiente do empreendedor. Entrepreneurship.

LISTA DE ILUSTRAÇÕES

Quadro 1 – Nove etapas da metodologia	14
Figura 1 – Fase Stemmer.....	15
Figura 2 – Ambiente do Empreendedor - janela de criação do plano de negócio.....	19
Quadro 2 – Requisitos funcionais.....	21
Quadro 3 – Requisitos não funcionais	21
Figura 3 – Diagrama de casos de uso do administrador	22
Figura 4 – Tela de pesquisa por assunto	26
Figura 5 – Tela de pesquisa por empresa.....	27
Figura 6 – Tela de relatório de assuntos.....	27
Figura 7 – Tela de visualização gráfica.....	28
Quadro 4 – Método para mineração de texto	28
Quadro 5 – Seleção modulo 1	29
Quadro 6 – Seleção modulo 2	29
Quadro 7 – Stopwords modulo 1	30
Quadro 8 – Stopwords modulo 2.....	30
Quadro 9 – Resultado modulo 1	30
Quadro 10 - Resultado modulo 2.....	30
Figura 8 – Resultado mineração de texto.....	31

LISTA DE SIGLAS

AJAX – *Asynchronous Javascript And XML*

CSS – *Cascading Style Sheets*

FURB – Fundação Universidade Regional de Blumenau

HTML – *HyperText Markup Language*

KDD – Knowledge Discovery in Database

PHP – *PHP: Hypertext Pre Processor*

RF – Requisito Funcional

RNF – Requisito Não Funcional

SGBD – Sistema Gerenciador de Banco de Dados

SUMÁRIO

1 INTRODUÇÃO	9
1.1 OBJETIVOS DO TRABALHO.....	10
1.2 ESTRUTURA DO TRABALHO	10
2 FUNDAMENTAÇÃO TEÓRICA	11
2.1 <i>TEXT MINING</i>	11
2.2 PLANO DE NEGÓCIO	16
2.3 AMBIENTE DO EMPREENDEDOR.....	17
2.4 TRABALHOS CORRELATOS	19
3 DESENVOLVIMENTO	21
3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO	21
3.2 ESPECIFICAÇÃO.....	22
3.2.1 Diagramas de Casos de Uso	22
3.2.1.1 Diagrama de Caso de Uso: Administrador	22
3.3 IMPLEMENTAÇÃO	23
3.3.1 Técnicas e ferramentas utilizadas	23
3.3.1.1 PHP.....	23
3.3.1.2 MySQL	24
3.3.1.3 AJAX	24
3.3.1.4 JavaScript e Framework Ext.js.....	25
3.3.1.5 Text Mining.....	25
3.3.2 Operacionalidade da implementação	25
3.3.2.1 Tela de Pesquisa por Assunto	26
3.3.2.2 Tela de Pesquisa por Empresa.....	26
3.3.2.3 Tela de Relatório de Assuntos.....	27
3.3.2.4 Tela de Visualização Gráfica	28
3.4 RESULTADOS E DISCUSSÃO.....	31
4 CONCLUSÕES	33
4.1 EXTENSÕES	33
REFERÊNCIAS BIBLIOGRÁFICAS	35

1 INTRODUÇÃO

O Ambiente Empreendedor da Universidade Regional de Blumenau, utilizado nas disciplinas de empreendedorismo é um ambiente onde se desenvolve o plano de negócio, como um dos recursos do material de apoio. O Plano de Negócio é um documento, que tenta-se apresentar como um arquivo “documento vivo” onde se especifica e planeja o negócio e que deve estar sempre em constante atualização. Para um plano de negócios, utilizam-se informações sobre a empresa, o negócio, o marketing, o financeiro e outros. Deste modo, formalizam-se os estudos a respeito das idéias do empreendedor, transformando-as em negócio (DALFOVO, 2010, p. 1).

Uma das melhorias desenvolvidas na linguagem JAVA/AJAX no Ambiente do Empreendedor foi desenvolvida por Santiago (2008), que teve como método de ensino a criação de uma empresa simulada para criação de um plano de negócio, com intuito de auxiliar os alunos da Universidade Regional de Blumenau (FURB). No trabalho citado, foi sugerido a continuação deste com a implementação de uma rotina para comparação entre os textos digitados nos módulos do plano de negócio. O Ambiente do Empreendedor foi desenvolvido utilizando-se a linguagem de programação JAVA/AJAX para o desenvolvimento de interfaces.

Atualmente, o Ambiente Empreendedor vem sendo utilizado na disciplinas de empreendedorismo nos cursos de Ciências da Computação, Sistemas de Informação, Engenharia de Telecomunicações entre outras da FURB. Porém, um dos problemas existentes é com a verificação e comparação da integridade, consistência do plano de negócio, preenchido no Ambiente do Empreendedor, que são completados manualmente, o que dificulta a atualização dos dados, pelo fato da mesma não estar integrada com o restante do ambiente. Atualmente o gestor (professor) da disciplina tem que conferir manualmente cada item do plano de negócio, digitados pelos alunos, e verificar o que foi descrito no Ambiente do Empreendedor em cada um dos módulos.

Como solução de auxílio do professor na análise do conteúdo do plano de negócio no Ambiente do Empreendedor, foi implementado uma rotina de mineração de texto, que fará a comparação dos textos entre os módulos, verificando as inconsistências entre si. A técnica *Text Mining* ou Mineração de Texto é o processo de obtenção de dados e informações a partir de textos em linguagens naturais, conforme (WIVES, 2004), a qual está sendo usada neste trabalho.

1.1 OBJETIVOS DO TRABALHO

O objetivo geral deste trabalho é implementar a técnica *Text Mining* no Ambiente do Empreendedor.

Os objetivos específicos do trabalho são:

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

1.2 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em quatro capítulos que estão descritos na seguinte forma.

No primeiro capítulo é apresentado a contextualização e justificas para o desenvolvimento do trabalho.

No segundo capítulo é disponibilizada a fundamentação teórica necessária para um razoável conhecimento sobre empreendedorismo, empreendedor e plano de negócio, também são apresentados os trabalhos correlatos.

O terceiro capítulo tem como foco o desenvolvimento do sistema para apoio ao empreendedor, descrevendo os requisitos principais do problema como também a especificação e implementação.

O quarto capítulo apresenta as considerações finais e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados os temas: *Text Mining* e o plano de negócio. É visto também sobre o Ambiente do Empreendedor. Na última seção são descritos os trabalhos correlatos.

2.1 TEXT MINING

Os textos podem ser considerados, em muitos casos, como uma seqüência de dados, similares aos dados coletados por sensores ou outro sistema de informação, que podem ser processados por métodos similares aos utilizados no processo de extração de conhecimento de base de dados. (Wives, 2004, p. 19) A adaptação deste processo a formas textuais dá origem a um processo denominado *text mining*, também conhecido como Descoberta de Conhecimento em Textos (*Knowledge Discovery in Text - KDT*), ou ainda *text data mining*. É usada a nomenclatura *text mining*, que se refere à tecnologia capaz de analisar grandes coleções de documentos não estruturados com o propósito de extrair conhecimento, padrões de interesse previamente desconhecidos e potencialmente úteis.

Nos últimos anos tem aumentado o interesse em aplicar métodos de aprendizado de máquina para uma variedade de problemas em categorização de textos ainda (GADCON, 2006, p. 1) afirma que “80% dos dados de uma organização estão armazenados em forma de dados não-estruturados (texto).” Além disso, existe hoje certa tendência em se combinar mais de um algoritmo para se conseguir melhores resultados, como aprendizado estatístico com aprendizado relacional ou *clustering* com categorização.

Segundo Wives (2004, p. 72), o *text mining* possui cinco passos importantes:

- a) definição de objetivos: compreensão do domínio, identificação do que deve ou do que pode ser descoberto;
- b) recuperação de informações: localização e recuperação dos textos relevantes para o que o usuário necessita descobrir, encontrando palavras-chaves descritas no texto;
- c) extração de informação: identificação de itens (características, palavras) relevantes nos documentos, verificando a coerência entre os textos;
- d) mineração: aplicação de um método de mineração que identifique padrões e

relacionamento entre os dados. Existem vários métodos de descoberta e será estudado o mais adequado;

- e) interpretação: a interpretação e a aplicação dos dados encontrados. Neste passo é o resultado final, mostrando para o usuário o que foi encontrado de similaridade entre os textos.

Text Mining consiste em apresentar como uma possibilidade do proprietário ou usuário dos dados obter novas visões e conhecimento, com isso, procurando padrões nos dados nos quais não seriam reconhecidos usando consultas tradicionais de dados e técnicas relacionadas. Estas técnicas permitem comparações para serem feitas através de dados vindos de muitas fontes de diferentes tipos, extraindo informação que não poderiam ser óbvias ou mesmo visíveis para o usuário, organizando documentos e informação pelos assuntos ou temas (HOESCHL, 2002).

Atualmente, o problema de quem trabalha com informação não é mais a obtenção dos dados; ao contrário, o volume de dados disponível é muitas vezes amplo demais. O problema é encontrar a informação útil entre os dados disponíveis. Como exemplos de fontes de dados úteis à Inteligência Competitiva temos: a Internet oferece a todos os usuários de estações de trabalho a possibilidade de coletar informações no mundo inteiro através do protocolo de comunicação TCP/IP. Outras ferramentas de busca permitem ao usuário final extrair informação dos bancos públicos de empresas, centros de pesquisa e instituições financeiras. A recente acessibilidade aos mais importantes ‘*hosts*’ é o principal avanço para a entrada de usuários *on line* no mundo dos negócios. Essa evolução se caracteriza pela diminuição dos custos de acesso, pelo aumento da velocidade de transmissão de dados, pela simplificação dos processos de conexão e, sobretudo, pelo surgimento de inovadores serviços *on-line* colocados a disposição dos usuários finais.

As pessoas que navegam na *Web* observam aumento na quantidade de documentos, dados e informações que podem ser obtidos via Internet. A principal vantagem é a de que este fluxo de informação e às vezes livre e não validada, ou seja, não se tem certeza que a informação seja correta, completa ou verdadeira; ao contrário pode ser completamente falsa ou não servir aos propósitos de quem consulta.

Empresas podem ter bancos de dados privados que podem ser acessados junto com outras fontes de dados. O banco de dados privado pode conter dados cuja formatação e completamente diferente da dos outros bancos públicos. Por exemplo, define-se a estratégia de pesquisa para identificar melhor os mais importantes conjuntos de dados. Escolhidas as fontes de dados, é necessário definir e implementar uma estratégia de pesquisa documental. A

dificuldade de operação da base de dados acaba determinando as palavras-chaves a serem utilizadas na extração de citações e ‘abstracts’.

Em alguns casos, isto é óbvio. Muitas vezes, porém, trata-se de um processo de tentativa e erro. Um trecho pode trazer informação mais diversificada da que realmente interessa. Assim, alguns descritores para delimitar a pesquisa são necessários. Mesmo assim, tudo que está sendo excluído com o uso de descritores qualificativos, deve ser levado em consideração. Na medida em que os serviços contratados são pagos com base no tempo de conexão e no número de documentos importados, é importante realizar buscas eficientes para evitar gastos exagerados. A busca de documentos é tarefa especializada e difícil; o sucesso ou o fracasso da atividade de análise documental, tanto do ponto de vista da informação, quanto do ponto de vista econômico, está vinculado à habilidade de busca nas bases de dados (HOESCHL, 2003).

Há que se escolher a melhor ferramenta para uma análise automatizada dos conjuntos de dados para chegar ao conhecimento “escondido”. Pode-se afirmar “escondido” porque os dados a serem analisados estão em centenas de milhares de páginas. Este é o motivo pelo qual as estratégias de mineração de dados e textos é ferramenta útil ao processo de análise documental.

Existem vários métodos de análise automática de conjuntos de dados. Uma boa aproximação é o uso desses métodos em etapas sucessivas.

Depois de pesquisas bibliográficas de varios autores, teses e dissertações, encontrou-se o trabalho de Wives (2004), que serviu como a base para o trabalho a ser desenvolvido usando o modelo dos nove experimentos apresentadas na Quadro 1.

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> (reductor 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, 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.

Fonte: Wines (2004, p. 79).

Quadro 1 – Nove etapas da metodologia

Wives (2004) fez um comparativo da utilização do método de “Conceitos como descritores de texto”. Que consiste em fazer primeiramente uma busca normal dos dados sem usar nenhum método pegando todas as palavras do texto a fim de verificar o ganho real da utilização. Depois compara o resultado com o segundo método dele que seria retirar todas as *Stopworks* (são as palavras irrelevantes nos textos onde não agrega nenhuma informação para o usuário como artigos preposições etc.), depois fez uma nova pesquisa só com *Stemmer* (que é palavras com varias terminações, mas significando a mesma coisa, por exemplo: afirmava, afirmam, afirmando e afirmaram), conforme mostra a Figura 1, no seu quarto experimento ele junto os dois métodos anteriores.

MySQL-Front - [Conec - /DB_SIIC_DEV/stemm]

File Edit Tools Im-/Export Window Help

Limit: 0 32.016 OK

Host Database Table Data SQL Query

DB_SIIC_DEV / stemm: 32013 Records (32013 retrieved)

palavra	stemm
<input type="checkbox"/> majuro	majur
<input type="checkbox"/> makron	makron
<input type="checkbox"/> mal	mal
<input type="checkbox"/> mala	mal
<input type="checkbox"/> malabar	malab
<input type="checkbox"/> malabarismo	malabar
<input type="checkbox"/> malabarismos	malabar
<input type="checkbox"/> malaia	mala
<input type="checkbox"/> malaias	mala
<input type="checkbox"/> malaios	malai
<input type="checkbox"/> malan	malan
<input type="checkbox"/> malandragem	malandrag
<input type="checkbox"/> malandras	malandr
<input type="checkbox"/> malária	malár
<input type="checkbox"/> malas	mal
<input type="checkbox"/> malásia	malás
<input checked="" type="checkbox"/> malcasadas	malcas
<input type="checkbox"/> malcolm	malcolm
<input type="checkbox"/> malcom	malcom
<input type="checkbox"/> malconduzida	malconduz
<input type="checkbox"/> malconservados	malconserv
<input type="checkbox"/> malcriação	malcriaçã
<input type="checkbox"/> malcriados	malcri
<input type="checkbox"/> malcuidadas	malcuid
<input type="checkbox"/> maldade	maldad
<input type="checkbox"/> maldaner	maldan
<input type="checkbox"/> maldita	maldit
<input type="checkbox"/> malditas	maldit
<input type="checkbox"/> maldosa	maldos

SQL BLOB-Editor Filter

DB_SIIC_DEV: stemm: 2 field(s)

Figura 1 – Fase Stemmer

Na quinta etapa ele utiliza de conceitos para fazer o processo de agrupamento (que é a definição de uma palavra, como por exemplo, as palavras água e bebida podem ser definidas como sendo um líquido), através do site desenvolvido por Jorge (2004) foi possível capturar a base de conceitos organizada por ele não necessitando criar uma do zero.

A partir dessa pesquisa optou-se por fazer o desenvolvimento dessas três técnicas afim de conseguir o melhor resultado para o problema proposto.

2.2 PLANO DE NEGÓCIO

O plano de negócio é um estudo de projeto para o desenvolvimento de uma empresa onde são feitos questionamentos a respeito do que irá ser criado, como o conceito do negócio, os riscos, os concorrentes, o perfil da clientela, as estratégias de *marketing*, bem como todo o plano financeiro que viabilizará o novo negócio para poder estar sempre atualizado para acompanhar a evolução do mercado e de seus competidores (SEBRAE-SP, 2008).

Através do plano de negócio as futuras empresas vêem a viabilidade de seu negócio, podendo encontrar parceiros ou talvez patrocinadores que viram no negócio, uma potencial empresa de sucesso. De acordo com Dornelas (2001, p. 91), é fundamental para o empreendedor o desenvolvimento de um bom plano de negócio, a fim de planejar suas ações e estratégias para o sucesso da empresa.

Bangs (1999, p. 20-21) propõe um modelo de plano de negócio a ser seguido pelos empreendedores. O modelo proposto pelo autor é dividido em 3 partes. O negócio, dados financeiros e documentação de apoio.

O negócio – é a parte mais importante e mais difícil do plano de negócio, e tem como objetivo obter informações sobre o que é o negócio, quais os produtos ou serviços oferecidos, o segmento do mercado, vantagens sobre os concorrentes e como o dinheiro investido tornará o negócio mais rentável.

Dados financeiros – é necessário um bom controle financeiro a fim de não deixar o negócio tomar conta de você. O objetivo desta parte do plano é documentar as fontes e aplicações dos fundos, a lista de bens de capital, o balanço da empresa, analisar o ponto de equilíbrio financeiro, projeções de fluxo de caixa e análise dos desvios.

Documentos de apoio – nesta parte são anexados os documentos que comprovem as declarações feitas no plano de negócio, tais como currículos, informações sobre crédito, estimativas, cartas de intenção de clientes potenciais, cartas de recomendações entre outros documentos.

A partir do segundo semestre de 1996 foi introduzido, na Universidade Regional de Blumenau, a disciplina de Empreendedor em Informática que tem como objetivo desenvolver o espírito empreendedor dos alunos através do desenvolvimento de um plano de negócio (DALFOVO, et al.; 2005, p. 169). Porém, o principal objetivo da disciplina é discutir com seus alunos como montar uma empresa de sucesso, fazendo de uma pequena idéia um valioso negócio.

2.3 AMBIENTE DO EMPREENDEDOR

De acordo com Dalfovo (2010) o histórico da criação da disciplina empreendedor na FURB, iniciou-se desde o segundo semestre de 1996. A cada semestre observa-se um acontecimento, uma melhoria, que marcaram os pontos fortes das atividades aplicadas na disciplina de Empreendedorismo, que iniciou regularmente no segundo semestre de 1996 com três alunos no período matutino. No primeiro e segundo semestres de 1997, o número de alunos aumentaram para seis, então foi elaborado um cronograma de atividades para o semestre letivo, que disponibilizou aos alunos, além do conteúdo programático, uma coletânea (cópia) de alguns artigos relacionados ao empreendedorismo.

No final do segundo semestre, em julho de 1997, foi realizado, em Brasília, DF, em âmbito nacional, o I encontro dos professores da disciplina Empreendedor em Informática, patrocinado pelo CNPq com o edital chamado Software de Exportação 2000 (SOFTEX2000). No referido encontro, foi criado um exemplo proposto do cronograma de atividades a serem cumpridas no decorrer de cada semestre, bem como algumas estratégias para aumentar e intensificar a disciplina, com intuito de aumentar a demanda por acadêmicos.

As estratégias criadas pelos professores da disciplina Empreendedor em Informática participantes do encontro foram: levar para as aulas depoimentos de ex-alunos que se tornaram empresários; montar minicursos sobre assuntos do momento; realizar, no final do curso, um júri final, julgando para eleger os melhores Planos de Negócios dos alunos; mudar o horário do curso, passando de matutino para noturno, pois o número de acadêmicos no curso de Ciências da Computação é maior no período noturno; e fazer um concentrado de duas semanas antes do início das atividades normais.

Estas alterações iniciaram a trajetória de mudanças da FURB, da disciplina Empreendedor em Informática, começaram a trajetória de mudanças. Como um curso tem suas normas e obrigações a serem cumpridas, sugeriu-se ao Colegiado do curso de Ciências da Computação a troca do curso do período matutino para o noturno e que fosse realizado um concentrado de duas semanas antes do período normal das aulas. O Colegiado do curso deu parecer favorável ao pedido, o qual, a partir do primeiro semestre de 1997, foi posto em prática.

Na seqüência, apresenta-se o cenário escolhido do primeiro semestre de 1998 ao II segundo semestre de 2004, pois em cada semestre, em linhas gerais a evolução e os procedimentos ocorridos na disciplina foram muito similares. Desde o primeiro semestre de

1998, houve incremento de alunos, numa média de 20% a cada semestre. A partir de 1998 vem sendo apresentado um cronograma para as atividades a serem atendidas durante o semestre. Também a disciplina foi em regime concentrado de duas semanas e no período noturno. Em 2000 foi elaborado um folder explicativo com informações básicas sobre a disciplina, e o mesmo foi distribuído aos acadêmicos do quinto ao oitavo semestre do curso de graduação. Ao término do primeiro semestre de 2002, realizou-se o júri final. Uma particularidade interessante, onde foi transmitido ao vivo pela rede de TV FURB da Universidade. Na apresentação dos trabalhos, foram selecionados alguns para serem submetidos ao Concurso Nacional Jovem Empreendedor (COJEM), para identificar e estimular o surgimento de empresas de software de grande potencial. Além das atividades mencionadas, realizou-se em 2003 uma trilha ecológica.

Como trabalho final da disciplina, os acadêmicos Empreendedores apresentaram os planos na I Feira de Informática, na qual tinham a missão de vender seu produto, tanto pelo site da disciplina como fisicamente. Foi montado um *stand* no saguão principal da Universidade e quem tivesse a maior venda teria a maior nota final. A partir de 2003 foi aplicado e testado o ambiente totalmente com acesso via internet para a geração automático do plano de negócio. Vale resaltar que nesta fase esta melhoria foi gerada a partir de uma orientação do trabalho de conclusão de curso. Depois de feitos os testes por meio do ambiente, pôde-se efetuar a entrega dos diversos trabalhos de avaliação desenvolvidos. Com a implementação desta nova rotina o professor passou a ter um maior controle na correção de trabalhos e conseqüentemente uma melhoria no aprendizado. Após o término da disciplina o aluno possui um espaço no site onde ele decide se irá disponibilizar ou não o seu plano final de negócio com objetivo de demonstrar sua idéia.

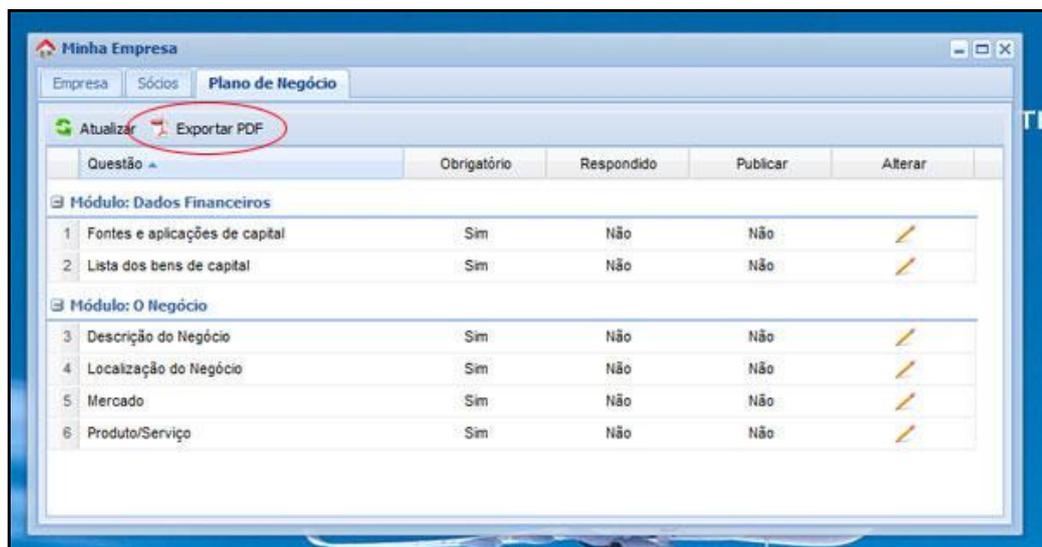
Do primeiro semestre de 2004 até o presente momento o número de alunos permaneceram em torno de noventa. Também em 2004 foram feitos os mesmos conteúdos programáticos dos semestres anteriores. Foi aplicado a feira simulada tanto presencial como no formato de *E-commerce*, identificando-se o ranking das empresas. Em 2004, no segundo semestre, foi apresentado a disciplina e suas atividades na semana de empreendedor do Norte do país (Alagoas – Maceió), como uma visão empreendedora. No segundo semestre de 2004, além de estar sendo utilizado o ambiente nos cursos Sistemas e Computação foi implantado no Programa de Pós-Graduação em nível de Mestrado em Administração. A partir do segundo semestre de 2004, foi liberado o ambiente gratuitamente para os professores de Alagoas e outros de graduação da nossa instituição, bem como de outras IES localmente, e incluído a modalidade de 20% como Ensino à Distância.

A partir de 2005, iniciou-se a transformação e melhorias do ambiente, transformando-se as telas da linguagem PHP com o banco de dados MYSQL, para nova visualização em Java com AJAX, conforme implementações apresentado por Santiago (2008).

De acordo com Borges e Dalfovo (2004, p. 1) “O ambiente surgiu de um Trabalho de Conclusão de Curso, desenvolvido pelo acadêmico Alencar Barbieri.” Seu objetivo é auxiliar os alunos na construção do plano de negócio, permitindo ao professor acompanhar as atividades desenvolvidas pelo aluno, como também um local para os alunos encontrarem informações e materiais da disciplina.

Com este ambiente o processo de criação do plano de negócio ficou mais fácil, tanto para o aluno e para o professor, podendo controlar melhor a disciplina como todas as informações contidas nele, avaliando os alunos conforme seu desempenho. (BARBIERI, 2002, p. 86).

Dalfovo (2005) descreve que o diferencial do Ambiente do Empreendedor em relação a outros ambientes e o Plano de Negócio, onde os acadêmicos desenvolvem o plano a partir de questões previamente cadastradas pelo professor, conforme apresentado na Figura 2.



Fonte: Santiago (2008, p. 54).

Figura 2 – Ambiente do Empreendedor - janela de criação do plano de negócio

2.4 TRABALHOS CORRELATOS

As ferramentas apresentadas a seguir possuem características em comum com a ferramenta proposta neste trabalho.

Wives (2004) utiliza a técnica de *clusterin*. Por ser aplicada em textos, está sujeito a sofrer interferências decorrentes de problemas da própria linguagem e do vocabulário utilizado nos mesmos, tais como erros ortográficos, sinonímia, homonímia, variações morfológicas. Visto este problema, Wines (2004) apresenta uma solução para minimizá-los, que consiste na utilização de “conceitos” (estruturas capazes de representar objetos e idéias presentes nos textos) na modelagem do conteúdo dos documentos.

Com base num software de cadastro de chamados telefônicos, Uber (2004) desenvolveu um programa para minerar os chamados telefônicos que são compostos por data de abertura, software utilizado, versão do software, descrição do problema, situação da ficha de atendimento, dentre outros, sendo que a descrição do problema é um texto livre que foi a variável analisada, permitindo assim com a análise desta variável identificar o motivo do chamado do cliente.

A SPSS Brasil (2007) desenvolveu um software chamado *Text Mining of Clementine* que possibilita a extração de conceitos-chaves, percepções e relações de dados não-estruturados. O diferencial deste software é a quantidade de idiomas que ele diz poder fazer a mineração de texto, sendo que cada idioma existe uma forma diferente de escrever.

3 DESENVOLVIMENTO

O sistema desenvolvido neste trabalho realiza a verificação dos dados inseridos no ambiente do empreendedor, através da criação de ferramentas que possibilitem o entendimento dessas informações.

Este capítulo aborda a realização e análise dos requisitos que definem as características do sistema proposto. A seguir são descritas todas as análises e suas especificações.

3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

Os requisitos, classificados como Requisitos Funcionais (RF) e Requisitos Não Funcionais (RNF) descrevem o que o sistema deve fazer. Os RF demonstram às funcionalidades e o comportamento que o sistema deve possuir em determinadas situações, podendo também explicitar o que o sistema não deve realizar. Os RNF demonstram as restrições que o sistema terá sobre alguns serviços ou funções oferecidas como usabilidade, navegabilidade, portabilidade, segurança, hardware.

No Quadro 2 apresentam-se os RF do sistema.

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.

Quadro 2 – Requisitos funcionais

No Quadro 3 apresentam-se os RNF do sistema.

Requisitos Não Funcionais
RNF01 – O sistema deve utilizar a linguagem de programação PHP 5.0.
RNF02 – O sistema deve utilizar AJAX para a programação da interatividade da ferramenta.
RNF03 – O sistema deve utilizar o SGBD mySQL para armazenamento das informações necessárias.
RNF04 – Aplicar a técnica <i>text mining</i> entre os módulos preenchidos pelo empreendedor, utilizando-se uma técnica de mineração

Quadro 3 – Requisitos não funcionais

3.2 ESPECIFICAÇÃO

Neste item são apresentadas as especificações diagramas de casos de uso.

Para a elaboração do diagrama de caso de uso utilizou-se a ferramenta Enterprise Architect (EA).

3.2.1 Diagramas de Casos de Uso

Os casos de uso têm como função representar as principais funcionalidades que se pode observar em um sistema e dos elementos externos que interagem com o mesmo (BEZERRA, 2002).

A seguir é listado o diagrama de caso de uso modelado na fase de especificação do sistema.

3.2.1.1 Diagrama de Caso de Uso: Administrador

O administrador é o responsável por toda a organização do sistema, incluindo a nova funcionalidade de verificação dos dados inserida pelos alunos.

Na Figura 3 mostram-se as novas funcionalidades disponíveis ao administrador do sistema.

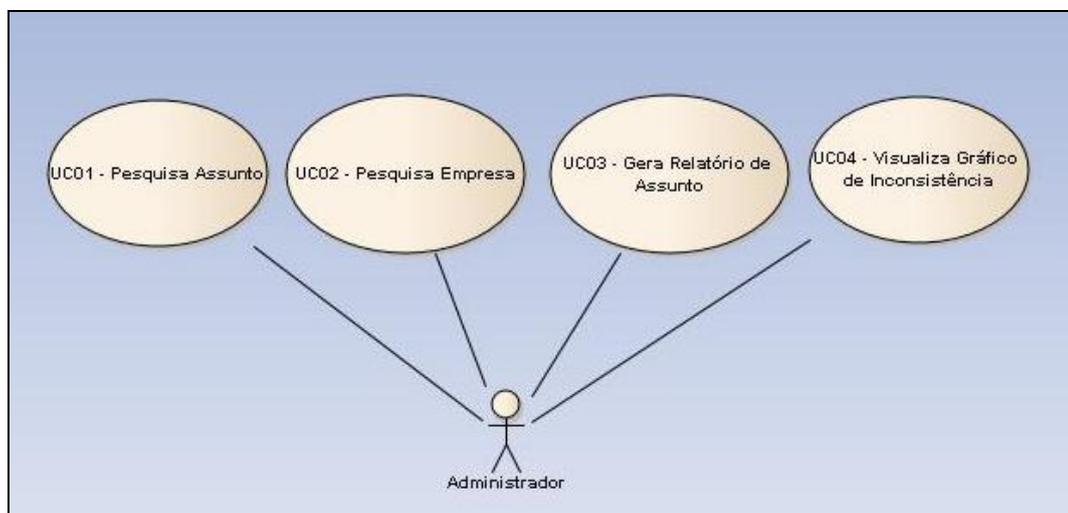


Figura 3 – Diagrama de casos de uso do administrador

A seguir apresenta-se uma breve descrição para os casos de uso do administrador principal:

- a) pesquisa assunto: permite ao administrador pesquisar empresas que já desenvolveram algum plano de negocio por determinado assunto;
- b) pesquisa empresa: permite ao administrador pesquisar empresas que já desenvolveram algum plano de negocio de uma determinada empresa;
- c) visualiza gráfico de inconsistência: permite através de um gráfico verifica se uma empresa esta seguindo de acordo o tema proposto em todos os módulos;
- d) gera relatório assunto: permite ao administrador gerar um relatório com todos os assuntos já desenvolvidos no ambiente do empreendedor.

3.3 IMPLEMENTAÇÃO

Neste item são apresentadas as informações sobre as técnicas e ferramentas utilizadas para o desenvolvimento deste trabalho. Também são apresentadas as principais operacionalidades do sistema juntamente com algumas telas.

3.3.1 Técnicas e ferramentas utilizadas

A seguir são apresentadas as ferramentas utilizadas para o desenvolvimento do sistema proposto, tais como: PHP, MySQL, AJAX, JavaScript, framework Ext.js e a *Text mining*.

3.3.1.1 PHP

PHP é uma linguagem de script embutida do código *HyperText Markup Language* (HTML) criada por Rasmus Lerdof. Sua sintaxe é herdada do C, Java e do Perl com algumas características específicas. Tem como principal característica a facilidade de implementação acarretando em um desenvolvimento rápido aumentando a produtividade do programador (PHP, 2008).

3.3.1.2 MySQL

Para o armazenamento dos dados do sistema foi utilizado o SGBD MySQL. O MySQL tornou-se o banco de dados de código aberto mais popular do mundo pois possui consistência, performance, confiabilidade e é de fácil utilização. Pode ser instalado em mais de 20 plataformas incluindo Linux e Windows (MYSQL BRASIL, 2008).

3.3.1.3 AJAX

O AJAX é um conjunto de tecnologias que está sendo utilizado para o desenvolvimento de uma nova geração de aplicativos totalmente baseados na *web*. Também é uma linguagem de programação baseada em Java, capaz de criar pequenos aplicativos, para serem rodados no *browser*. O AJAX embora seja atual, não é realmente uma nova tecnologia, ele reúne várias tecnologias *web* bem estabelecidas e as usa de maneira nova (CRANE; PASCARELLO; JAMES, 2007, p. 53). A principal característica do AJAX é a função `XMLHttpRequest` do Javascript. Através desta função pode-se fazer com que o navegador faça chamadas assíncronas ao servidor, retornando os dados solicitados, sem a necessidade de atualização da página que está sendo visualizada. Com isso o usuário poderá continuar trabalhando enquanto a requisição está sendo processada.

Ao utilizar-se uma ferramenta *desktop*, podem-se usar recursos como arrastar e soltar, navegação através do teclado, botões que acendem ao passar o *mouse* e assim por diante. Com o AJAX pode-se transportar estes recursos para a plataforma *web*, facilitando a utilização e focando a ferramenta na usabilidade.

As vantagens de utilizar AJAX é a redução de informação que navega na rede, diminuindo a sensibilidade, tornando a navegação mais rápida e sem necessidade de instalar softwares proprietários, é possível fazer aplicações com alta qualidade de interação. Segundo Limeira (2006, p. 13), “Praticamente todos os navegadores, em suas últimas versões, são compatíveis com as técnicas utilizadas no AJAX, viabilizando o seu uso em qualquer aplicação Web.”

3.3.1.4 JavaScript e Framework Ext.js

No desenvolvimento do sistema proposto foi utilizado o JavaScript com objetivo de proporcionar uma interatividade maior do usuário com a ferramenta. Operacionalidades como arrastar e soltar objetos são possíveis através do JavaScript.

O JavaScript em conjunto com o CSS, disponibiliza uma vasta opção de efeitos e recursos que pode-se implementar em sistemas baseados na web. Neste trabalho foi utilizado o *framework* Ext.js, que consiste em uma biblioteca de funções para a construção de sistemas com uma interface rica, conseqüentemente melhorando a usabilidade do mesmo (EXTJS, 2008).

3.3.1.5 Text Mining

Conforme Wives (2003), são chamadas de *Text Mining*, todas as técnicas que permitem localizar e articular documentos, acelerar a análise dos mesmos e extrair conhecimento que, de outra maneira, permaneceria escondido nas grandes bases de dados textuais. Os algoritmos computacionais que tornam isso possível estão baseados na matemática e em novas e inteligentes maneiras de fazer uso do poder de processamento do computador. *Text Mining* é extração de padrões ou conhecimentos interessantes e não-triviais a partir de documentos textuais. A tecnologia de *Text Mining* pode ser aplicada para formalizar e explorar conhecimento. O conhecimento disponível com pessoas pode ser armazenado em textos, os quais serão analisados para se entender seu significado, ou seja, do que tratam os textos. Depois, pode-se explorar o conhecimento extraído dos textos para gerar novos conhecimentos. Também se pode combinar este conhecimento com o conhecimento explícito armazenado em bancos de dados estruturados.

3.3.2 Operacionalidade da implementação

O sistema desenvolvido tem por objetivo exercer o papel de auxiliar na disciplina de empreendedorismo. No ambiente do administrador de turma são encontradas as novas ferramentas desenvolvidas neste trabalho. Entre as novas funções do administrador estão: a

pesquisa por assunto, pesquisa por empresa já desenvolvida, visualização gráfica das porcentagem das comparações efetuadas entre os módulos e a visualização de um relatório com todos os assuntos já desenvolvidos.

A seguir são apresentadas as telas do sistema e detalhadas das principais funcionalidades do mesmo.

3.3.2.1 Tela de Pesquisa por Assunto

Nesta tela de pesquisa por assunto o administrador poderá informar um assunto com isso a ferramenta mostrará todas as empresas que já criaram algum tipo de plano de negocio com esse determinado assunto como apresenta-se na Figura 4.

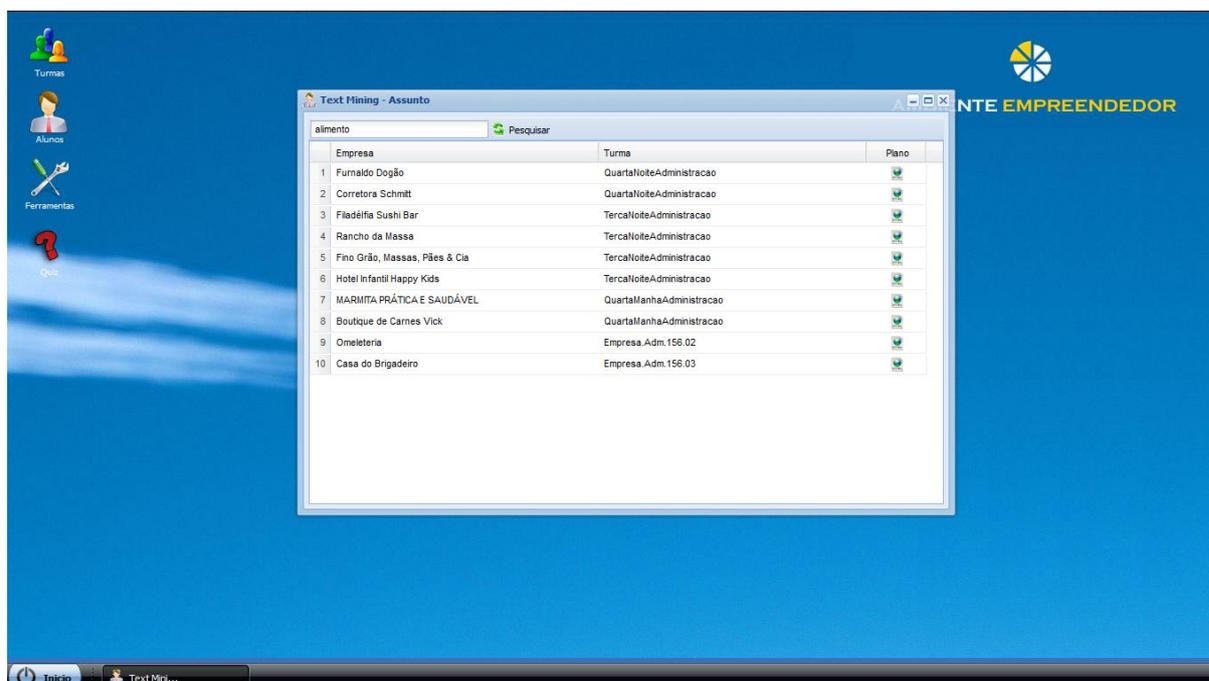


Figura 4 – Tela de pesquisa por assunto

3.3.2.2 Tela de Pesquisa por Empresa

Na tela de pesquisa por empresa o administrador selecionará uma empresa e a ferramenta mostrara os planos de negocio que tem assuntos em comum a fim de verificar uma possível repetições na criação do plano de negócio. Esta ferramenta é apresentada na Figura 5.

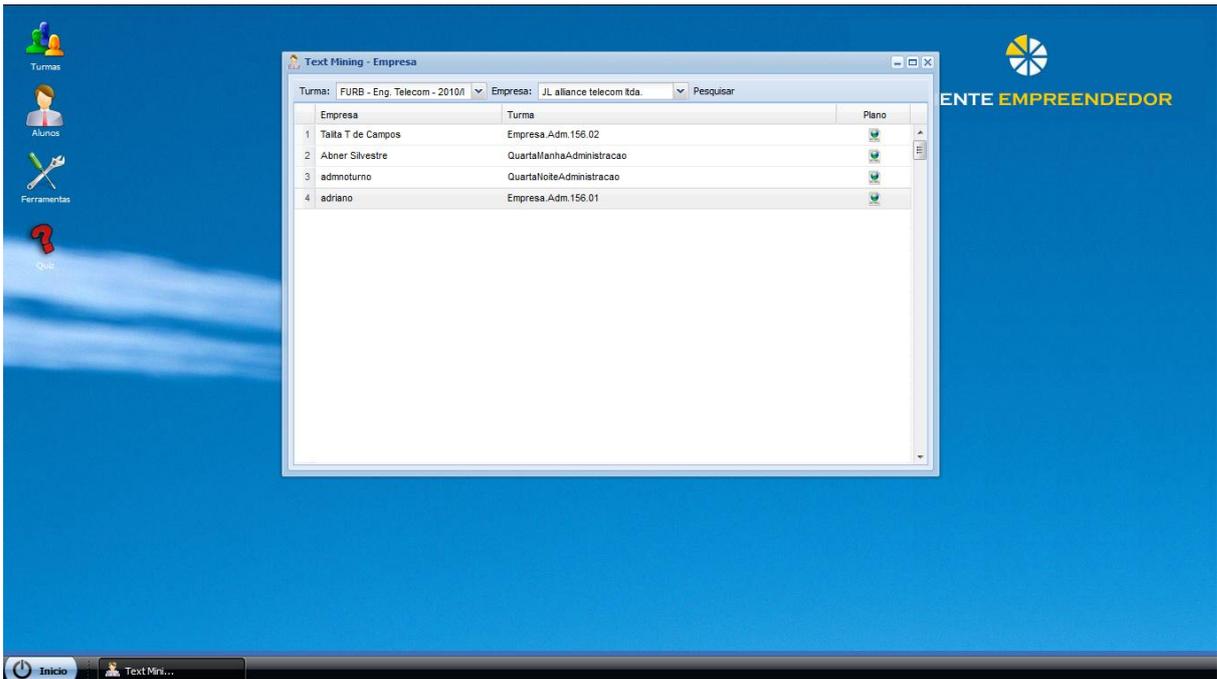


Figura 5 – Tela de pesquisa por empresa

3.3.2.3 Tela de Relatório de Assuntos

Na tela de relatório de assuntos (Figura 6), é apresentada, uma lista com todos os assuntos já abordados na criação de uma empresa no ambiente do empreendedor. Sendo possível verificar os assuntos que mais geraram planos de negócio.

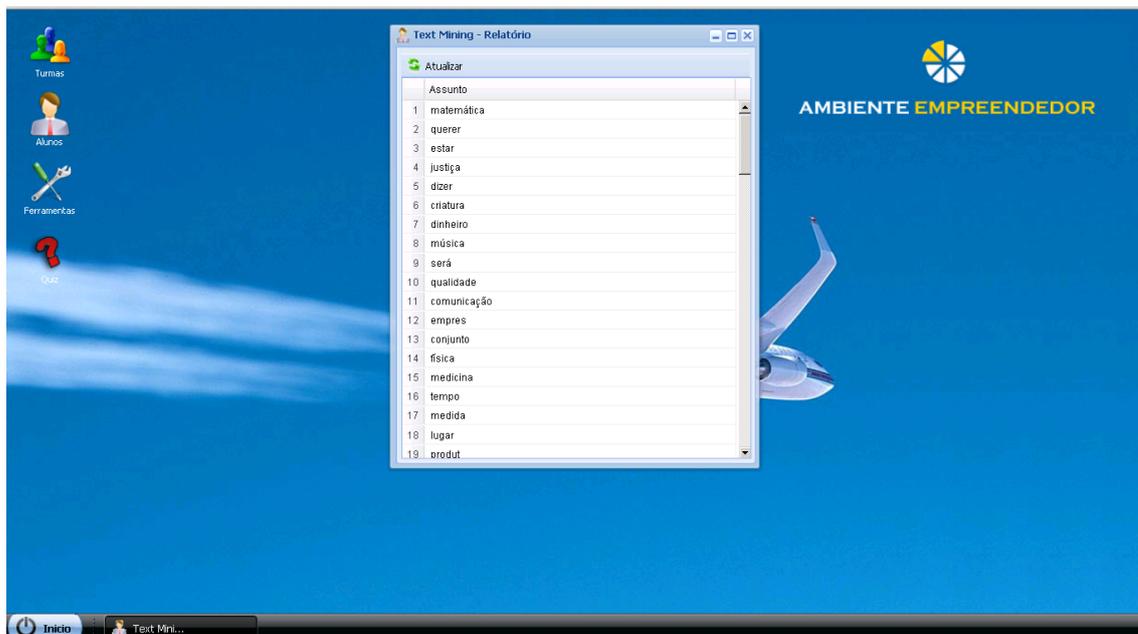


Figura 6 – Tela de relatório de assuntos

3.3.2.4 Tela de Visualização Gráfica

Na tela de visualização gráfica o administrador selecionará uma empresa para fazer a comparação dos módulos utilizando as técnicas de *stopwords*, *stemming* e conceitos para a descoberta das informações contidas a fim de visualizar graficamente se os módulos selecionados estão de acordo com o que o aluno esta desenvolvendo, ver na Figura 7.

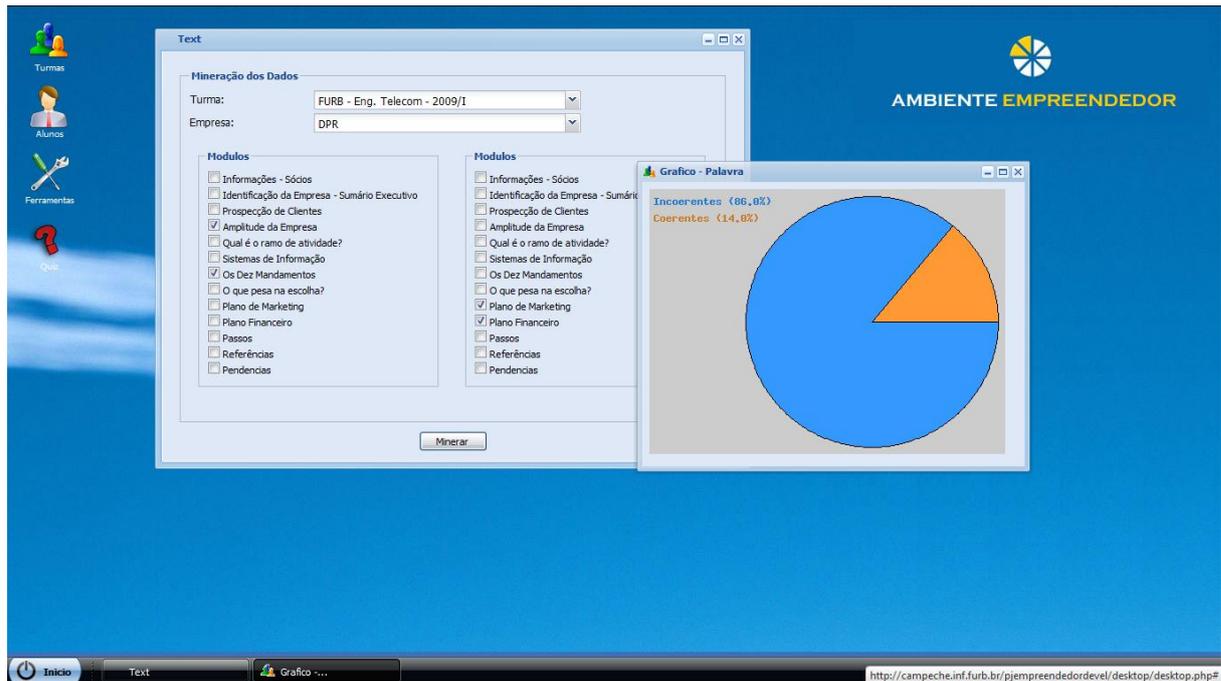


Figura 7 – Tela de visualização gráfica

Com o intuito de avaliar o sistema proposto na tela de Visualização Gráfica, onde é realizado os seguintes passos conforme Quadro 4. Para cada modulo selecionado são realizados os seguintes passos para a limpeza e mineração dos textos conforme será descrito a seguir.

```
//Retira Stop
$SqlResposta =RetiraStop(strtolower($SqlResposta));
$SqlResposta2 =RetiraStop(strtolower($SqlResposta2));

//Retira Stemm
$SqlResposta = RetiraStempart1($SqlResposta);
$SqlResposta2 = RetiraStempart1($SqlResposta2);

//Conceitos
$arrayStem1 = Conceitos($SqlResposta);
$texto1 = ContaPalavras($arrayStem1);

$arrayStem2 = Conceitos($SqlResposta2);
$texto2 = ContaPalavras($arrayStem2);
```

Quadro 4 – Método para mineração de texto

Foi selecionada a empresa Eco Sports onde compara-se os módulos “Identificação da

Empresa” e “Qual é o ramo da atividade” (Quadro 5) com os módulos “Os Dez Mandamentos” e “Plano de Marketing” (Quadro 6) e obtêm-se os seguintes resultados.

O Ecoturismo é uma forma de reencontrar a vida, e a alegria de viver! A Eco Sports oferece meios e serviços para levar os nossos clientes, a lugares magníficos e ter experiências que tornarão cada momento em um momento especial. Conhecer belos recantos de outra perspectiva, em contato com a natureza. Atuando na organização de passeios ecológicos, em sítios turísticos e culturais por toda Santa Catarina e em especial no Vale do Itajaí, atendendo desde o esportista em busca de maiores orientações até o amante da natureza, ou o pai de família que busca um fim de semana especial. Por meio de parcerias com empresas especializadas em esportes radicais a Eco Sports trará mais opções aos seus clientes, de em contato com a natureza ter experiências em novos esportes. Um fator crítico ao negócio é a dependência de um bom clima, para motivar as pessoas a praticar os esportes, ou passeios realizados ao ar livre como propõem a empresa, porém fator este que é atenuado pelo fato da empresa ser itinerante permitindo atuar em todo estado, de acordo com a escolha do cliente ou das condições climáticas. Os gastos iniciais serão com somente a publicidade da empresa e transportes, grande parte dos gastos são diminuídos devido ao fato da empresa trabalhar com empresas parceiras. Espera-se que o retorno deste investimento será proporcional ao investimento em publicidade e na mesma velocidade. Ser reconhecida como uma empresa

Quadro 5 – Seleção módulo 1

Um dos maiores riscos assumidos na constituição da Eco sports, é trabalhar com diversas empresas parceiras, devido a grande variedade de esportes radicais em seus roteiros, dependendo assim de profissionais treinados nestes esportes. A Eco Sports identificou um mercado em expansão, um mercado que se preocupa cada vez mais com o meio ambiente, mas que ao mesmo tempo quer interagir com a natureza de maneira sustentável. -Roteiros de passeios em parques ecológicos, voltados para excursões escolares, famílias que buscam um fim de semana ao ar livre ou simplesmente aos amantes da natureza que queiram conhecer novos locais. -Roteiros de passeios com prática de esportes radicais ao ar livre (trekking, ciclismo, voo livre), acompanhados de instrutores especializados nestes esportes. -Roteiros de excursões para já praticantes dos esportes, que querem ter a comodidade de conhecer novos lugares para a prática de seu esporte sem se preocupar com o transporte até estes locais, ou com a hospedagem. Atuando na organização de passeios ecológicos, em sítios turísticos e culturais por toda Santa Catarina e em especial no Vale do Itajaí, atende desde o esportista em busca de maiores orientações até o amante da natureza, o pai de família que busca um fim de semana especial, ou até excursões escolares. O anseio dos clientes da Eco Sports é o contato com a natureza, através de passeios ecológicos e esportes radicais, de maneira sustentável. CMMI for

Quadro 6 – Seleção módulo 2

Após a captura dos dados é feita a primeira depuração dos dados, retirando os artigos e preposições os chamados *Stopwords* resultando nas Quadros 7 e 8.

Ecoturismo forma reencontram vida alegria viver eco sports oferece serviços levar clientes lugares magníficos ter experiência tornarão momento momento especial conhecer belos recantos outra perspectiva contato natureza atuando organização passeios ecológicos sítios turísticos culturais toda santa catarina especial vale itajai atendendo esportista busca maiores orientações amante natureza pai família busca fim semana especial por parcerias empresas especializadas esportes radicais eco sports trará opções clientes contato natureza ter experiências novos esportes um fator crítico negócio dependência clima motivar pessoas praticar esportes passeios realizados ao livre propõem empresa fator atenuado fato empresa itinerante permitindo atuar todo estado acordo escolha cliente condições climáticas os gastos iniciais serão publicidade empresa transportes parte gastos diminuídos devido fato empresa trabalhar empresas parceiras esperase retorno investimento será proporcional investimento publicidade velocidade reconhecida empresa inspiradora liberdade praticantes simpaticantes esportes ao livre oferecer serviços passeios opções ecoturismo criativos praticantes simpaticantes esportes ao livre gerem sensação desafio emoção liberdade promover média passeios eventos mês amante esportes ao livre geral especial voo livre ciclismo aplica ecologicamente integrado ambiente ecoturismo conjunto atividades esportes praticados

Quadro 7 – Stopwords modulo 1

maiores riscos assumidos constituição eco sports trabalhar diversas empresas parceiras devido variedade esportes radicais roteiros dependendo profissionais treinados nestes esportes eco sports identificou mercado expansão mercado preocupa ambiente tempo interagir natureza sustentável roteiros passeios parques ecológicos voltados excursões escolares famílias buscam fim semana ao livre simplesmente aos amantes natureza queiram conhecer novos locais roteiros passeios pratica esportes radicais ao livre trekking ciclismo voo livre acompanhados instrutores especializados nestes esportes roteiros excursões praticantes esportes que querem ter comodidade conhecer novos lugares prática esporte preocupar transporte locais hospedagem atuando organização passeios ecológicos sítios turísticos culturais toda santa catarina especial vale itajai atende esportista busca maiores orientações amante natureza pai família busca fim semana especial excursões escolares anseio clientes eco sports contato natureza através passeios ecologicos esportes radicais sustentável for services nível pesquisa novos roteiros serem oferecidos feita primeiramente análise secretarias turismo cidades buscando pontos turísticos enquadrem política roteiros elaborados empresa posteriormente

Quadro 8 – Stopwords modulo 2

Em seguida é realizada a redução das palavras a seu radical o chamado *Stemmer* onde as palavras derivadas são transformadas em sua formação primitiva como flor – floricultura e casa – casebre.

Na última etapa é realizada o agrupamentos das palavras em conceitos, onde defini-se o que cada palavra quer expressar no contexto.

qualidade economia meteorologia atmosfera química vento aparência vestuário empresa esporte pratica matemática justiça experiência livre passei passeie lingüística livra física natura tempo comércio filosofia trabalho biologia militar sports arte aos profissão medida trabalha informática lugar oferece música atividade ofereça política pronome cinema excursões finalidade ecologia ave avião inseto opção

Quadro 9 – Resultado modulo 1

matemática qualidade economia música cinema profissão roteiro vestuário empresa esporte justiça química lugar razão finalidade atmosfera vento aparência meteorologia física parente religião transversalmente estar querer dizer criatura filosofia turismo lingüística sports dinheiro natural passei biologia excursões busca começo natura passeie ciência busca comércio será entretenimento normas princípios radicar ecologia

Quadro 10 - Resultado modulo 2

Depois de concluído todas as etapas anteriores para cada seleção e realizada uma comparação dos dados obtidos em cada um dos módulos, resultando em um gráfico com o percentual com as palavras que são coerentes entre as seleções e as incoerentes conforme Figura 8.

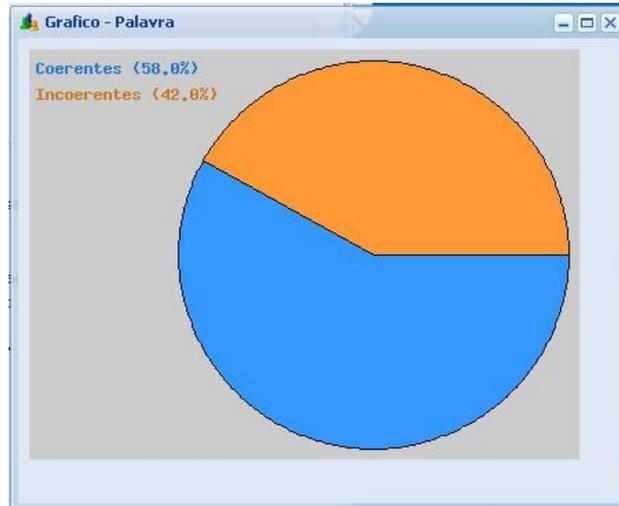


Figura 8 – Resultado mineração de texto

Com o resultado obtido pode-se concluir que os dois módulos selecionados no começo do experimento estão coerentes, que a equipe está desenvolvendo corretamente o plano de negócio.

3.4 RESULTADOS E DISCUSSÃO

Ao final do desenvolvimento da ferramenta, pode-se observar que a mesma atingiu o resultado esperado.

Um dos problemas que não estavam no escopo do projeto, foi inserção de *tags* HTML na lista de *stopwords*, pois o sistema quando da edição das respostas, o aluno pode alterar o formato da letra, sublinhar etc como se fosse um editor de texto porém no banco de dados ele armazena todas essas informações como se fosse uma página HTML.

Através deste trabalho identificou as influências dos assuntos discutidos e dos planos de negócios criados, onde auxiliou o desenvolvimento de conteúdos para plano de negócio, no sentido de criar valor para o relacionamento entre o trabalho acadêmico e o plano de negócio. O empreendedorismo é uma das oportunidades mais avançadas do mundo e os Empreendedores são a mola-mestra deste estágio. Quanto mais preparados com base no Plano

de Negócio, maiores serão as possibilidades de superar as expectativas do mercado na atuação empreendedora.

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, portanto uma boa pesquisa preliminar pode descobrir as vantagens e desvantagens de cada método, podendo-se desenvolver uma grande ferramenta sobre *text mining*.

Com a técnica *Text Mining*, pode-se ver passo a passo dos métodos até ele chegar ao último passo onde nos trouxe às possíveis palavras chaves do módulo. Mas o trabalho contém algumas inconsistências que só através de vários testes poderá ir aperfeiçoando-se a ferramenta.

4 CONCLUSÕES

Pode-se observar que o *Text Mining* ainda é uma grande área a ser abordada, o que foi realizado nesse projeto é uma pequena parte das técnicas de *text mining*, sendo uma área considerada relativamente nova, ainda podendo ser criados outros técnicas até mesmo reaproveitando técnicas existentes do Knowledge Discovery in Database – Mineração de Dados (KDD), somente fazendo algumas modificações para se enquadrarem, usando texto em vez da base de dados.

Os objetivos citados neste trabalho foram alcançados. Apresentou-se graficamente o acompanhamento do percentual de incoerências, inconsistências e número da versão do plano de negócio. Também conseguiu aplicar os passos para o *text mining* nos módulos do plano de negócio descrito no Ambiente do Empreendedor selecionados pelo professor.

Pode-se concluir que a procura de palavras-chave no texto dos módulos, abrange um grande número de temas gerando certa dificuldade tendo em vista a velocidade que novos temas surgem e nem sempre o resultado é preciso, em relação ao tema específico de alguns trabalhos.

A maior dificuldade encontrada foi a montagem de um dicionário de dados eletrônico, dado a escassez de tempo, decidiu-se localizar alguns prontos na internet e utilizou-se um dicionário de sinônimos e de conceitos, no que auxiliou bastante o desenvolvimento e conclusão deste trabalho. Outra dificuldade nesse trabalho foi em relação ao *framework* cuja maioria do material de apoio para o desenvolvimento das suas interfaces foi pesquisada no site oficial da Ext.js.

Por fim esse trabalho foi contribuiu com os gestores do ambiente com uma ferramenta que os permite a avaliar os planos de negócios criados, podendo reduzir o tempo que cada um desperdiça com a análise dos planos de negócios.

4.1 EXTENSÕES

Entre algumas sugestões:

- aplicar a técnica de *Text Mining* em outros ambientes, como nos resumos das monografias, para compará-los com as palavras-chave.

- pesquisar outras técnicas de KDT para ser utilizada no ambiente além da que foi aplicada neste trabalho.
- desenvolver para o ambiente empreendedor, outras rotinas como ferramentas, transformar o ambiente empreendedor em um observatório do Empreendedor, onde além do ambiente plano de negócios, poderá o ambiente ter mais funcionalidade e integração com o Sistemas de Informação Observatório da Educação.

REFERÊNCIAS BIBLIOGRÁFICAS

BANGS, David H. **Planejamento de negócios**: criando um plano para seu negócio ser bem-sucedido. Tradução Rosa Krausz. São Paulo: Nobel, 1999.

BARBIERI, Alencar. **Ambiente de aprendizagem para auxílio na disciplina de empreendedor em informática**. 2002. 89 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau.

BEZERRA, Eduardo. **Princípios de análise e projeto de sistemas com UML**. Rio de Janeiro: Campus, 2002.

BORGES, Jonathan M. DALFOVO, Oscar. Ambiente do empreendedor: ambiente de aprendizagem para auxiliar na disciplina de empreendedor em informática. In: SEMINÁRIO DE COMPUTAÇÃO, 13., 2004, Blumenau. **Anais...** Blumenau: Editora da FURB, 2004. p. 195–196.

CRANE, Dave; PASCARELLO, Eric; JAMES, Darren. **AJAX em ação**. Tradução Edson Furmankiewicz e Carlos Schafranski. São Paulo: Pearson Prentice Hall, 2007.

DALFOVO, Oscar; DALFOVO, Michael S.; FLORES Danusa C. O ambiente do empreendedor como estratégia de ensino e aprendizagem para auxiliar na elaboração do plano de negócio. In: HOELTGEBAUM, Marianne; MACHADO, Denise (Org). **Gestão em empreendedorismo**. 1. ed. Blumenau: FURB, 2005. v.1. p. 169-183.

DALFOVO, Oscar. **Site do empreendedor**. Blumenau, 2010. Disponível em: <<http://www.inf.furb.br/~dalfovo/Empreendedor>>. Acesso em: 05 maio 2010.

DORNELAS, José C. A. **Empreendedorismo**: transformando idéias em negócios. Rio de Janeiro: Campus, 2001.

EXTJS. **Client-side javascript framework**. [S.l], 2008. Disponível em: <<http://www.extjs.com/>>. Acesso em: 26 maio 2008.

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>. Acesso em: 13 ago. 2008

HOESCHL, Hugo C., et al. **AlphaThemis – do texto ao conhecimento**. Florianópolis,[2002]. Disponível em: <<http://www.cs.us.es/ideia/articulos/ideia-9.pdf>>. Acesso em: 12 abr. 2010.

JORGE, Valdir. **Thesaurus da Língua Portuguesa do Brasil**. Montreal, 2004. Disponível em: <<http://alcor.concordia.ca/~vjorge/Thesaurus/indices.html>>. Acesso em: 31 jul. 2008

LIMEIRA, José L. S. **Utilização de AJAX no desenvolvimento de sistemas web**. 2006. 44 f. Trabalho de Conclusão de Curso (Curso de Especialização em Web e Sistema de Informação) - Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre.

MYSQL BRASIL. **Por que Mysql**. São Paulo, [S.l], 2008. Disponível em: <<http://www.mysqlbrasil.com.br/?q=node/2>>. Acesso em: 26 maio 2008.

PHP. **Informações gerais**. [S.l], 2008. Disponível em: <http://br2.php.net/manual/pt_BR/faq.general.php>. Acesso em: 26 maio 2008.

SANTIAGO, Rafael W. **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.

SEBRAE-SP. **O que é plano de negócios?** São Paulo, 2008. Disponível em: <<http://www.sebraesp.com.br/principal/abrindo%20seu%20neg%C3%B3cio/orienta%C3%A7%C3%B5es/cria%C3%A7%C3%A3o%20de%20empresas/planonegocio.aspx>>. Acesso em: 14 ago. 2008

SPSS Brasil. **Text mining for clementine**. São Paulo, 2007. Disponível em: <http://www.spss.com.br/tecnologias/text_mining.htm>. Acesso em: 13 ago. 2008.

UBER, José L. **Descoberta de conhecimento com o uso de text mining aplicada ao SAC**. 2004. 54 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau.

WIVES, Leandro K.; STANLEY Loh; OLIVEIRA, Jose P. M. de. **Text-mining: conceitos, objetivos, técnicas e aplicações**. Porto Alegre, [2003]. Disponível em: <<http://www.inf.ufrgs.br/~palazzo/disciplinas/cmp511/textmining.pdf>>. Acesso em: 28 mar. 2010.

WIVES, Leandro K. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos**. 2004. 136 f. Tese (Curso de Pós-Graduação em Computação) - Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre.