



**UNIVERSIDADE DE BRASÍLIA (UnB)
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E
CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO (FACE)
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO (CID)
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (PPGCinf)**

**Análise do conteúdo de um sistema de
informação destinado à microempresa
brasileira por meio de aplicação da Descoberta
de Conhecimento em Textos**

HÉLIA DE SOUSA CHAVES RAMOS

**BRASÍLIA - DF
2008**

Análise do conteúdo de um sistema de informação destinado à microempresa brasileira por meio de aplicação da Descoberta de Conhecimento em Textos

HÉLIA DE SOUSA CHAVES RAMOS

Dissertação apresentada à banca examinadora como requisito parcial à obtenção do Título de Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Área de concentração: Arquitetura da Informação.

Orientadora: Dra. Marisa Bräscher Basílio Medeiros

**BRASÍLIA - DF
2008**

R175a Ramos, Hélia de Sousa Chaves

Análise do conteúdo de um sistema de informação destinado à microempresa brasileira por meio de aplicação da Descoberta de Conhecimento em Textos / Hélia de Sousa Chaves Ramos. – 2008.

124 f. : il.;

Dissertação (Mestrado) – Universidade de Brasília, Departamento de Ciência da Informação, 2008.

Orientadora: Profa. Dra. Marisa Bräscher Basílio Medeiros

1. Descoberta de Conhecimento em Texto (DCT). 2. Mineração de textos. 3. Indicadores de CT&I. 4. Serviços de informação. 5. Micro e pequenas empresas (MPÉs). 6. Empreendedores. 7. Inclusão social. I. Brascher, M. II. Título

CDU:002.1 : 004.62



FOLHA DE APROVAÇÃO

Título: “Análise do conteúdo de um sistema de informação destinado à microempresa brasileira por meio de aplicação da Descoberta de Conhecimento em Textos”

Autor: Hélia de Sousa Chaves Ramos

Área de concentração: Transferência da Informação

Linha de pesquisa: Arquitetura da Informação

Dissertação submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília como requisito parcial para obtenção do título de **Mestre** em Ciência da Informação.

Dissertação aprovada em: 28 de fevereiro de 2008.

Aprovado por:

Prof.ª. Dr.ª. Marisa Bräscher Basílio Medeiros
Presidente – Orientador– (UnB/PPGCInf)

Prof.ª. Dr.ª. Suzana Pinheiro Machado Mueller
Membro Interno – (UnB/PPGCInf)

Luc Quoniam
Professeur des Universités

Prof. Dr. Luc Quoniam
Membro Externo – (Universidade de Toulon, França)

Prof. Dr. Murilo Bastos da Cunha
Suplente – (UnB/CID)

A minha mãe, Maria, meu elo com a vida.

Agradecimentos

"O ouvido do homem é o caminho para aprender." Aristóteles

Agradeço ao Criador pelo milagre da vida, um "mistério a ser celebrado".

Minha imensa gratidão a meu pai querido, *in memoriam*, que fez história quando "arribou" do interior do Goiás com os 9 filhos e a esposa, em um caminhão de mudança, em busca da oportunidade de estudo para os filhos na Capital Federal, que acabara de nascer. Pai-amigo, que, com sua visão de futuro, plantou em nós a vontade de lutar pela vida e nos ensinou o valor das coisas conquistadas com méritos próprios. A minha mãe, que, com seu infinito amor e sabedoria, foi companheira e complemento na construção de uma grande família, pautada na união entre os irmãos. A Cléo Chaves, companheiro de muitos anos, que sempre foi uma força estimuladora dos meus estudos. À mana Valéria Varella, que se mostrou companheira incansável nos meus momentos de fragilidade emocional. A meus irmãos, sobrinhos e cunhados, que "seguraram a barra" nas minhas necessárias ausências. A meu sobrinho Nando Di Chaves, pela ajuda no trato com imagens.

À professora Marisa Bräscher, mais que uma orientadora, uma verdadeira amiga, sintonizada com minhas inquietações acadêmicas, que soube dosar o tom de sua interação comigo nos meus momentos de altos e baixos. Foi terna e compreensiva, quando me senti frágil diante dos momentos de dificuldade. Foi firme e direta, quando eu vacilei nos rumos da minha pesquisa. Foi forte e precisa na escolha dos caminhos a trilhar para chegarmos a um resultado concreto.

Ao Ibict – minha segunda casa, onde aprendi o ofício de profissional da informação e me apaixonei pela lida com conteúdos e tecnologias – na pessoa de seu atual Diretor, professor Dr. José Emir Suaiden, um importante estimulador da capacitação dos servidores. Ainda, na pessoa de Cecília Oliveira Leite, pela força, compreensão e disposição no apoio institucional; ao amigo Pablo Nogueira, pela extração dos dados do SBRT, insumo básico desta pesquisa; a Priscilla Araújo e Walter Júnior, pelas preciosidades garimpadas no acervo do Ibict; a Sonia Lourenço, pela valiosa ajuda em todos os momentos em que precisei; a Celina Lamb, pelas amigáveis conversas de força e incentivo; a Milton Shintaku, pelo apoio nos

processamentos utilizando o BR/Search; a Denis Novaes, pela ajuda com figuras; a Cassandra Viana, pela competente ajuda na elaboração da ficha catalográfica; ao amigo Arthur Costa, pelo estímulo em me fazer acreditar que eu poderia iniciar este trabalho e pelas preciosas contribuições para que eu pudesse concluí-lo.

No tocante às minhas relações acadêmicas, ao professor Murilo Cunha, pela confiança e orientação do primeiro trabalho publicado, ainda no papel de aluna especial, aspirante a candidata a uma vaga no Mestrado. À professora Lillian Alvarez, pelas preciosas “dicas”, fruto de conversas sempre alegres e positivas. A Jucilene e Martha, nossa força na Secretaria de Pós-Graduação do CID/UnB, pela presteza, disposição e suavidade no trato, especialmente nos meus momentos de estresse. A Marcelo Schiessl, que utilizou sua perícia no uso do SAS para me apoiar na geração dos arquivos que possibilitaram as análises desta pesquisa.

Ao SAS Institute Brasil Ltda., pela liberação das ferramentas que possibilitaram a realização do estudo, na pessoa de Patricia Montenegro, minha simpática interlocutora.

À Policentro, na pessoa de seu Superintendente Comercial, Paulo Henrique do Espírito Santo, pela liberação do BR/Search.

À Coordenação da rede SBRT, pela liberação dos dados para a realização desta pesquisa.

Ainda, no que tange a minha família, a Nathan (5), Hannah (2) e Noah (1 mês), sobrinhos amados, pela convivência inocente que me deram força extra e me fizeram sentir que a vida é feita de aprendizados constantes.

*"A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original."*

Albert Einsten

Resumo

Esta pesquisa aborda a aplicação da técnica de Descoberta de Conhecimento em Texto (DCT) em bases de dados textuais (de conteúdos não-estruturados), repositórios de informações não evidentes, as quais podem se revelar importantes fontes de informação para aplicações diversas, envolvendo processos de tomada de decisão. O objetivo central da pesquisa é verificar a eficácia da DCT na descoberta de informações que possam apoiar a construção de indicadores úteis à tomada de decisão estratégica, assim como a definição de políticas públicas para a microempresa. O estudo de caso foi o conteúdo textual do Serviço Brasileiro de Respostas Técnicas (SBRT), um sistema de informação tecnológica na *Web* destinado ao setor produtivo, notadamente empreendedores, micro e pequenas empresas, fruto de um esforço compartilhado entre governo, instituições de pesquisa, universidades e iniciativa privada. A metodologia adotada contempla a aplicação da DCT em 6.041 documentos extraídos do sistema de informação SBRT, para a qual foi utilizado o pacote de *software* SAS Data Mining Solution. A técnica utilizada foi a geração de agrupamentos de documentos a partir dos termos minerados na base de dados. Foram realizadas análises comparativas entre agrupamentos semelhantes e foi selecionado um dos agrupamentos para análise mais aprofundada. Os resultados dessas análises demonstram a eficácia do uso da DCT para extrair informações ocultas em documentos textuais, as quais não poderiam ser visualizadas a partir de recursos tradicionais de recuperação da informação. Uma importante descoberta foi a de que a preocupação com o meio ambiente é um forte componente nas demandas feitas pelos usuários do serviço SBRT. Observou-se a possibilidade de se extraírem informações úteis para apoio à construção de indicadores e à orientação de políticas internas à rede SBRT, assim como para o setor de pequenas e médias empresas. Evidenciou-se, ainda, o potencial da DCT para subsidiar a tomada de decisão, podendo, inclusive, ser utilizada para fins de inteligência competitiva nas organizações.

Palavras-chave: Descoberta de Conhecimento em Texto (DCT), mineração de textos, indicadores de CT&I, serviços de informação, micro e pequenas empresas (MPEs), empreendedores, inclusão social.

Abstract

This research addresses the application of Knowledge Discovery in Texts (KDT) in textual databases (of non-structural contents), repositories of non-evident information that can reveal to be important sources of information for several purposes involving decision-making processes. The main objective of the research is to verify the effectiveness of KDT for discovering information that may support the construction of ST&I indicators useful for the strategic decision-making process, as well as for the definition of public policies destined to microenterprises. The case study of the research was the textual content of the Brazilian Service for Technical Answers (Serviço Brasileiro de Respostas Técnicas – SBRT), a technological information database, available in the Web, geared to the Brazilian production sector, specially micro and small enterprises or entrepreneurs. SBRT is a shared effort accomplished by government, research institutions, universities and the private sector. The methodology adopted encompasses the application of KDT in 6.041 documents extracted from SBRT database by using the SAS Data Mining Solution software package. The technique adopted was document clustering from terms mined in the database. A comparative analysis of similar clusters was carried out and one of the clusters was selected to be subject of more profound investigation. The results of these analyses demonstrate the efficacy of using KDT to extract hidden information – that could not be found by using the traditional information retrieval – from textual documents. An important discovery was that environmental concerns are strongly present in the demands posted by SBRT’s users. It was observed the possibility to extract useful information to construct ST&I indicators and to orient policies for SBRT network and for the microenterprise sector as a whole. It was also evidenced the potential of KDT to support decision-making processes in organizations, and, in addition, to be used for competitive intelligence purposes.

Keywords: Knowledge Discovery in Texts (KDT), text mining, ST&I indicators, information services, micro-enterprises, entrepreneurs, social inclusion.

Lista de Figuras

FIGURA 1 – Escopo de atuação do SBRT

FIGURA 2 – Fluxo do Serviço SBRT

FIGURA 3 – Configuração do sistema de informação SBRT

FIGURA 4 – Organização da Informação Tecnológica na visão do SEBRAE

FIGURA 5 – Pacote de informação personalizado compilado por especialista

FIGURA 6 – Escopo e etapas da pesquisa

FIGURA 7 – Visualização do conteúdo da base de trabalho no SAS

FIGURA 8 – Lematização de Termos

FIGURA 9 – Os agrupamentos e suas interações

FIGURA 10 – Os alcances da pesquisa

Lista de Quadros e Gráficos

QUADRO 1 – Conteúdos do Sistema de informação SBRT

QUADRO 2 – Lista de palavras não significativas removidas do texto

QUADRO 3 – Média de termos por documento

QUADRO 4 – Lista de agrupamentos e termos – primeiro processamento

QUADRO 5 – Lista de agrupamentos e termos – segundo processamento

QUADRO 6 – Lista de erros de lematização

QUADRO 7 – Lista de agrupamentos e termos – terceiro processamento

QUADRO 8 – Lista de termos comuns aos agrupamentos

QUADRO 9 – Dados específicos dos agrupamentos 1, 3 e 12

QUADRO 10 – Análise comparativa: agrupamentos 1, 3 e 12

GRÁFICO 1 – Agrupamento de documentos da base de trabalho

GRÁFICO 2 – Distribuição dos assuntos no agrupamento 10

Lista de Siglas e Abreviaturas

BNDES	Banco Nacional de Desenvolvimento Econômico e Social
CI	Ciência da Informação
CNAE	Classificação Nacional de Atividades Econômicas
C&T	Ciência e Tecnologia
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CT&I	Ciência, Tecnologia e Inovação
DCD	Descoberta de Conhecimento em Dados
DCT	Descoberta de Conhecimento em Textos
DT	Dossiê Técnico
FINEP	Financiadora de Estudos e Projetos
FNDCT	Fundo Nacional de Desenvolvimento Científico e Tecnológico
IBGE	Instituto Brasileiro de Geografia e Estatística
Ibict	Instituto Brasileiro de Informação em Ciência e Tecnologia
KDD	<i>Knowledge Discovery in Databases</i>
KDT	<i>Knowledge Discovery in Texts</i>
MCT	Ministério da Ciência e Tecnologia
MPEs	Micro e Pequenas Empresas
PADCT	Programa de Apoio ao Desenvolvimento Científico e Tecnológico
RI	Recuperação da Informação
RT	Resposta Técnica
SAS	Programa SAS
SBRT	Serviço Brasileiro de Respostas Técnicas
SEBRAE	Serviço Brasileiro de Apoio às Micro e Pequenas Empresas
SENAI	Serviço Nacional de Aprendizagem Industrial
STI/MIC	Secretaria de Tecnologia Industrial do Ministério de Indústria e Comércio

SUMÁRIO

RESUMO	
ABSTRACT	
LISTA DE FIGURAS	
LISTA DE QUADROS, GRÁFICOS E TABELAS	
LISTA DE SIGLAS E ABREVIATURAS	

CAPÍTULO 1

INTRODUÇÃO	16
1.1 O contexto	18
1.2 Sobre o SBRT	20
1.2.1 O foco de atuação do SBRT	22
1.2.2 Configuração do Sistema de Informação SBRT	23
1.2.3 A dimensão social do SBRT	25
1.3 Motivação e relevância da pesquisa	26

CAPÍTULO 2

A PESQUISA	28
2.1 Apresentação do Problema	28
2.1.1 Objeto de estudo	28
2.1.2 A questão	29
2.1.3 A proposta	30
2.2 Objetivos	30
2.2.1 Objetivo geral	30
2.2.2 Objetivos específicos	30
2.3 Justificativa e relevância do tema	30

CAPÍTULO 3

FUNDAMENTAÇÃO TEÓRICA	35
3.1 Gestão de C&T	35
3.1.1 A construção de indicadores	37
3.1.2 A micro e pequena empresa	40
3.2 Gestão da Informação	43
3.2.1 A informação tecnológica	43
3.2.2 Serviços de Informação especializados	46
3.2.3 A Descoberta de Conhecimento em Texto (DCT)	48
3.2.3.1 Mineração de dados e Mineração de textos	50
3.2.3.2 Mineração de textos e Recuperação da Informação	52

CAPÍTULO 4

PROCEDIMENTOS METODOLÓGICOS	54
4.1 Caracterização da pesquisa	54
4.2 Delimitação da pesquisa	55

4.3 Etapas da pesquisa	57
4.4 Seleção das ferramentas de trabalho e preparação do ambiente de pesquisa	58
4.5 Extração dos dados para a pesquisa	59
4.6 Pré-processamento utilizando o BR/Search	60
4.7 Sobre o SAS	61
4.8 Aplicação da DCT utilizando o SAS	61
4.8.1 A criação da base de trabalho	62
4.8.2 Preparação dos dados (pré-processamento, limpeza e transformação)	62
4.8.2.1 Remoção de palavras não significativas	62
4.8.2.2 Lematização	64
4.8.3 A mineração do texto	65
4.8.3.1 Os primeiros resultados da mineração	65
4.8.3.2 Agrupamento dos documentos	66

CAPÍTULO 5

ANÁLISE DOS RESULTADOS	67
5.1 Os primeiros resultados da mineração - erros detectados	67
5.1.1 Primeiro processamento	68
5.1.2 Segundo processamento	69
5.2 Terceiro processamento - objeto da análise final	72
5.2.1 Breve análise dos termos determinantes dos agrupamentos	74
5.2.2 Análise comparativa entre agrupamentos	75
5.2.2.1 Os agrupamentos 1, 3 e 12 e o termo comum entre eles: “químico”	76
5.3 Agrupamento 10 - em busca de informações para indicadores	82
5.3.1 Classificação dos documentos	82
5.3.2 Análise dos termos minerados	85
5.3.3 Informações para construção de indicadores	86
5.3.4 Informações específicas sobre a base de Respostas Técnicas	87
5.4 Dificuldades enfrentadas	88

CAPÍTULO 6

CONSIDERAÇÕES FINAIS	91
6.1 Conclusões	91
6.2 Recomendações de estudos futuros	95

REFERÊNCIAS	97
--------------------	-----------

ANEXOS E APÊNDICES	102
Anexo A – Exemplo de Resposta Técnica (RT)	103
Anexo B – Amostra do dicionário de termos em língua portuguesa	109
Apêndice A – Amostra da lista de palavras extraídas pelo BR/Search (Ordem de frequência e ocorrência na base)	112
Apêndice B – Amostra da lista alfabética de palavras extraídas pelo BR/Search	116
Apêndice C – Amostra da Base de Trabalho do SBRT	119
Apêndice D – Palavras não significativas extraídas da base de trabalho pelo SAS	123

CAPÍTULO 1

INTRODUÇÃO

“Pior que não terminar uma viagem é nunca partir.” Amir Klink

O objetivo central desta pesquisa foi testar a eficácia da DCT na descoberta de informações que possam apoiar a construção de indicadores úteis à tomada de decisão estratégica, assim como a definição de políticas públicas para o setor produtivo de pequeno porte. O estudo de caso foi o conteúdo textual de um sistema de informação criado para prover soluções a questões de natureza tecnológica apresentadas por empreendedores e microempresários brasileiros – o Serviço Brasileiro de Respostas Técnicas (SBRT).

O conteúdo textual analisado contém soluções elaboradas por especialistas em atendimento a questões de natureza tecnológica levantadas por micro-empresários de todo o país. Constitui-se, assim, em rica fonte de conhecimento tecnológico que pode se revelar importante origem de indicadores sobre as necessidades dos pequenos negócios e nortear investimentos para solucioná-las.

A pesquisa aplicou a técnica de Descoberta de Conhecimento em Texto (DCT) nos documentos técnicos que compõem o conteúdo textual do sistema de informação SBRT. A aplicação da DCT vem se alargando no contexto científico, dada a sua adequada aplicabilidade no manuseio e exploração da grande massa de conteúdos textuais eletrônicos que se avolumam nas organizações e à compreensão da importância de se explorarem esses conteúdos na busca dos conhecimentos lá depositados e invisíveis aos olhos humanos.

Buscou-se com este estudo contribuir para a construção de indicadores em Ciência e Tecnologia – tarefa complexa a cargo do Ministério da Ciência e Tecnologia, órgão responsável pela iniciativa de criação do SBRT e um de seus principais financiadores –, motivação maior para a descoberta de associações entre os documentos e identificação de tendências que apoiem a tomada de decisão governamental com relação ao setor produtivo de pequeno porte.

A pesquisa foi estruturada em seis capítulos, iniciando por este introdutório, o Capítulo 1, onde é oferecido um panorama geral da pesquisa em termos de

contexto, motivação e relevância da pesquisa, onde é feita uma apresentação geral do estudo de caso, o SBRT.

O Capítulo 2 é dedicado à pesquisa, iniciando pela apresentação do problema – que abrange o objeto de estudo, a questão e a proposta – passa pelos objetivos e conclui com a justificativa e relevância do tema, enquanto que o Capítulo 3 trata da fundamentação teórica, a qual foi organizada de forma a contemplar os seguintes assuntos relacionados à pesquisa: gestão de C&T – construção de indicadores, micro e pequena empresa), e gestão da informação (informação tecnológica, serviços de informação especializados, DCT, mineração de dados *versus* mineração de textos, e mineração de textos *versus* recuperação da informação).

O Capítulo 4 detalha os procedimentos metodológicos realizados durante a pesquisa, desde a sua caracterização, a delimitação e etapas da pesquisa até a descrição dos primeiros resultados, passando pela seleção e uso das ferramentas de trabalho, extração e tratamento dos dados.

O Capítulo 5 abrange a análise dos resultados da pesquisa, onde são explorados os conteúdos gerados a partir da mineração dos dados. São abordados os erros que causaram a necessidade de se realizarem novos processamentos, assim como o detalhamento das soluções a esses erros. São analisados os agrupamentos gerados com a mineração dos textos, de forma independente e comparativa, apontando similaridades e diferenças entre eles. É realizada uma minuciosa exploração de um dos agrupamentos, em busca de informações que possam contribuir para a construção de indicadores. A seguir, são listadas as dificuldades encontradas e os resultados colhidos ao longo da pesquisa.

As conclusões e recomendações de estudos futuros são tratadas no Capítulo 6, onde são feitas reflexões sobre os alcances da pesquisa como um todo e se apresentam algumas sugestões de pesquisas a com aplicação da DCT, tanto no âmbito da rede SBRT quanto em nível mais amplo, como a melhoria de recursos tecnológicos para facilitar o tratamento de conteúdos em língua portuguesa.

Os Anexos e Apêndices abrigam informações complementares sobre o sistema de informação SBRT e sobre a pesquisa. O Anexo A apresenta um exemplo de Resposta Técnica (RT) elaborada por uma das instituições membros da rede SBRT; e o segundo, o Anexo B, consiste de uma amostra do dicionário de termos

em língua portuguesa utilizado na simplificação do vocabulário a ser analisado, pela redução dos termos a seus radicais (lematização), uma das etapas de limpeza e preparação dos dados para a aplicação da DCT.

Os Apêndices A e B oferecem informações sobre os tratamentos preliminares do conteúdo, realizados por meio da extração de listas de termos da base de RTs, em ordem de frequência e ocorrência na base, e em ordem alfabética. O Apêndice C expõe uma amostra da base de trabalho criada para a aplicação da DCT, e o Apêndice D traz a lista completa das palavras não significativas extraídas da base de trabalho criada para o estudo.

1.1 O contexto

A produção de conteúdos informacionais em formato digital vem se avolumando a cada dia, à medida que a constante evolução das tecnologias de informação e comunicação tem facilitado o registro, tratamento, armazenamento, recuperação e uso da informação. A velocidade e a amplitude com que o conhecimento gerado passou a ser compartilhado geraram uma dinâmica de reaproveitamento e produção de novos conhecimentos, provocando o surgimento de novas necessidades ao trato com a informação.

Para suprir essas novas necessidades, ferramentas e técnicas para tratamento de grandes massas de dados foram criadas e aperfeiçoadas, tratamentos estatísticos aplicados no processamento e análise de dados e informações, em busca de retratar o que não seria possível com a limitada capacidade humana de leitura e registro.

Da mesma forma, as informações estruturadas em bases de dados – organizadas, indexadas e dotadas de ferramentas cada vez mais sofisticadas e velozes para busca e recuperação da informação – têm sido objeto de estudo com a finalidade de extrair conhecimento para apoio à tomada de decisão.

A única técnica utilizada até recentemente para lidar com o grande volume de informação a que as pessoas estão expostas todos os dias era *data mining*, até surgir o *text mining*, que trabalha com a informação armazenada em documentos textuais não-estruturados, por meio da descoberta de relações entre as informações

armazenadas. (SEMIO Corporation). Weiss *et al.* (2005) afirmam que as técnicas de mineração de dados já estão bastante desenvolvidas e adquirindo certa maturidade em determinados tipos de análises.

Hearst (1999) comenta o uso errôneo do termo *data mining* e a escassez de pesquisas sobre *text mining*. Fala da importância de se diferenciar *text data mining* de acesso à informação (mais conhecido como recuperação da informação). Para ele, a meta do *data mining* é descobrir ou derivar novas informações a partir dos dados analisados, encontrar padrões em conjuntos de dados e/ou separar o signo do ruído (HEARST, 1999). O autor define *text mining* como a descoberta, pelo computador, de novas informações, previamente desconhecidas, pela extração automática de informações de diferentes recursos escritos.

Nos últimos anos, os olhos da pesquisa têm se voltado para aqueles conteúdos armazenados sem a preocupação com o rigor da estruturação com vistas a futuros estudos, os documentos textuais, comumente chamados de “informação não estruturada”.

Esses conteúdos se revelaram portadores de informações valiosas, camufladas em grandes volumes textuais, que passaram a ser explorados em busca de padrões de conhecimento até então desconhecidos, para tomada de decisão e geração de novos conhecimentos.

[...] o ser humano coleta e armazena uma quantidade muito grande de dados textuais, que necessitam ser vasculhados, estudados, conhecidos e organizados de forma a fornecerem informações que lhe dêem o conhecimento para a execução de uma tarefa que exija a tomada de uma decisão. (WIVES, 2004)

Segundo Tan (1999), mais de 80% das informações de uma organização estão contidos em documentos textuais, que são, na sua opinião, a forma mais natural de armazenamento de informações.

Não há dúvida de que esse tipo de documento adquire uma importância fundamental para a descoberta do conhecimento gerado dentro das organizações e se torna ainda mais rico quando está associado a uma rede de instituições e especialistas dos mais diversos campos do conhecimento humano.

Como a pesquisa visa utilizar essas técnicas de tratamento de conteúdos textuais em documentos de natureza tecnológica voltados para os pequenos empreendimentos, cabe ressaltar, conforme idéias expostas por Takahashi (2005), que as MPEs são, no Brasil, um dos pilares de sustentação da economia, em razão de seu número, abrangência, capilaridade e capacidade de geração de emprego.

O serviço SBRT, objeto de estudo desta dissertação, constitui uma iniciativa governamental encampada por instituições de ensino e pesquisa, públicas e privadas, atuantes na prestação de serviços de informação tecnológica. O item a seguir traz informações detalhadas sobre SBRT e seu funcionamento.

1.2 Sobre o SBRT

O suporte à criação da Rede de Serviços de Informação Tecnológica (rede SBRT), composta por governo, universidades e iniciativa privada, figura entre as iniciativas governamentais voltadas ao apoio à micro e pequena empresa brasileira.

O primeiro produto dessa rede é o Serviço Brasileiro de Respostas Técnicas (SBRT), concebido pelo Ministério da Ciência e Tecnologia (MCT) como um serviço "que pudesse responder à demanda dos pequenos e médios empresários por informações de fácil acesso e que contribuísse para a melhoria de seus produtos ou processos, por meio da articulação das competências instaladas no País." (BRASIL, 2006). Responderam a esse anseio do MCT sete renomadas instituições brasileiras¹, atuantes nos campos de ensino, pesquisa e prestação de serviços tecnológicos, que se reuniram para a criação da Rede SBRT com o apoio do governo federal.

Assim, em novembro de 2004, foi lançado, em nível nacional, um serviço de informação tecnológica na *Web* voltado para o registro e atendimento gratuito de demandas a problemas tecnológicos de baixa complexidade postados por empreendedores e micro e pequenos empresários de todo o País.

¹ São elas: o Centro de Desenvolvimento Científico e Tecnológico da Universidade de Brasília (**CDT/UnB**), o Disque-Tecnologia da USP (**CECAE/USP**), a Fundação Centro de Tecnologia de Minas Gerais (**CETEC/MG**), a Rede de Tecnologia do Rio de Janeiro (**REDETEC/RJ**), o Instituto de Tecnologia do Paraná (**TECPAR/PR**), o Instituto Euvaldo Lodi – Núcleo Regional da Bahia (**IEL/BA**), e o Serviço Nacional de Aprendizagem Industrial – Departamento Regional do Rio Grande do Sul (**SENAI/RS**). Somam-se a estas, na qualidade de parceiras, o Instituto Brasileiro de Informação em Ciência e Tecnologia (**IBICT**) e o Serviço de Apoio às Micro e Pequenas Empresas (**SEBRAE Nacional**).

A criação do Serviço SBRT teve como objetivos facilitar o rápido acesso a soluções tecnológicas de baixa complexidade e em áreas específicas, difundir e potencializar conhecimentos acumulados nas instituições de C,T & I e contribuir para o processo de transferência de tecnologia (conexão entre as demandas e as competências), especialmente para as empresas de menor porte. Propicia, ainda, às instituições membros a rica experiência de atuarem em rede, unindo suas competências no fornecimento de soluções tecnológicas elaboradas sob medida. (SBRT, 2005)

Os conteúdos gerados pela rede de instituições que compõem o SBRT estão armazenados em um sistema de informação *online*, de acesso público e gratuito. Parte fundamental desse conteúdo – as respostas técnicas às questões postadas pelos empreendedores – encontra-se em formato textual.

O Serviço SBRT é materializado por meio de um sistema de informação *online*, que possibilita o cadastramento dos interessados em obter soluções personalizadas a seus problemas tecnológicos em forma de documentos elaborados com apoio de especialistas das diversas instituições que compõem a rede SBRT.

Os interessados em ter um problema de natureza tecnológica solucionado pelo SBRT podem encaminhar suas perguntas por telefone, correspondência, pessoalmente, nos postos espalhados pelo País – as próprias instituições membros da rede SBRT e balcões SEBRAE – ou pela Internet (<http://sbrt.ibict.br>), por meio do cadastramento gratuito como cliente da rede SBRT e do posterior preenchimento de formulário eletrônico, registrando sua demanda.

As solicitações são encaminhadas, automaticamente, às instituições membros da rede SBRT e respondidas com o auxílio de especialistas, que elaboram respostas personalizadas na forma de documentos técnicos, as chamadas Respostas Técnicas (RTs). As RTs são validadas pela instituição responsável e enviadas aos respectivos clientes. Em seguida, são publicadas na base de dados do sistema de informação SBRT – sem a identificação dos solicitantes – para ampla divulgação e livre utilização pela população interessada, novos potenciais clientes do serviço.

As RTs têm por finalidade proporcionar aos microempresários e empreendedores condições de aplicar com facilidade as soluções apresentadas, de forma a propiciar a melhoria do processo ou produto e contribuir para a melhoria da

sua competitividade ou para a formalização de novos empreendimentos e geração de emprego e renda.

O Anexo A apresenta um exemplo de Resposta Técnica (RT), uma solução elaborada por uma das instituições membros da rede SBRT, em resposta a uma questão colocada por um cliente.

1.2.1 O foco de atuação do SBRT

As respostas técnicas publicadas pelo SBRT são geradas para responder a demandas consideradas de baixa complexidade – a partir da pesquisa em fontes de informação confiáveis – e de média complexidade, que envolvem análises mais específicas incluindo opiniões de especialistas.

As questões consideradas complexas, que exigem conhecimento altamente especializado e pesquisa mais aprofundada recebem outro tipo de tratamento, podendo gerar consultorias específicas, dependendo do interesse do cliente, com custos previamente negociados. Não fazem parte, portanto, do escopo de Respostas Técnicas gratuitas.

A Figura 1 ilustra o escopo de atuação do SBRT, considerando os níveis de complexidade das questões tecnológicas colocadas pelos clientes da rede.

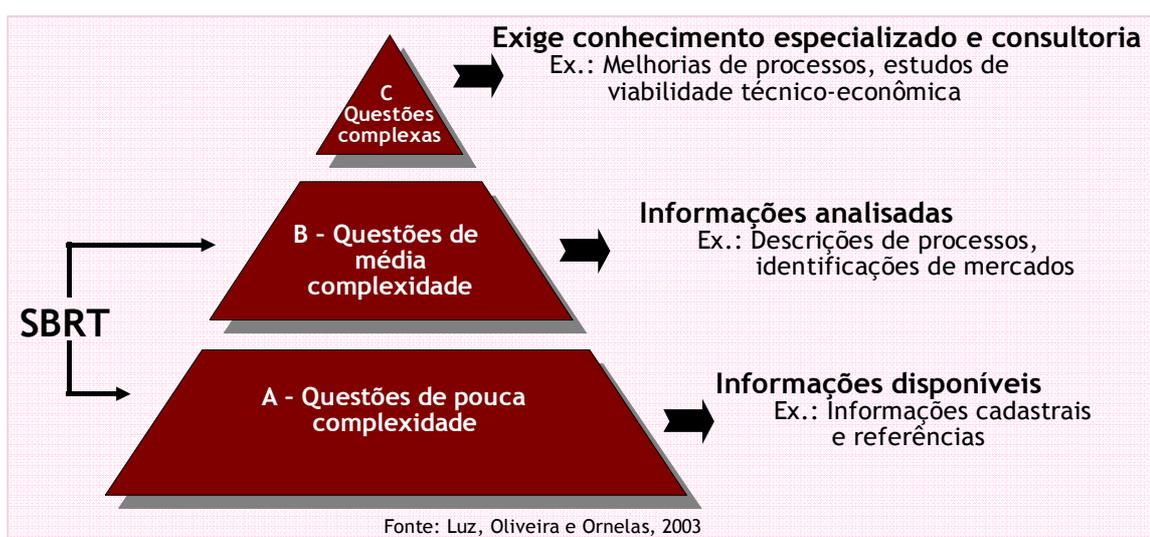


FIGURA 1 – Escopo de atuação do SBRT

Fonte: SBRT

1.2.2 Configuração do Sistema de Informação SBRT

O sistema de informação do SBRT é composto por três ambientes distintos: 1) ambiente de acesso aberto, via Internet (*site*); 2) ambiente de acesso restrito a Clientes SBRT; e 3) ambiente de acesso restrito aos administradores membros do SBRT.

O primeiro é composto por informações gerais sobre o Serviço SBRT, as instituições que compõem a Rede SBRT, descrição dos produtos oferecidos (respostas técnicas e dossiês técnicos), e uma busca por assunto e palavras-chave.

O segundo ambiente é destinado ao Cliente SBRT, após seu cadastramento, onde ele tem acesso ao acompanhamento da sua solicitação, a formulários para registro de novas solicitações, atualização de cadastro, alteração de senha e registro de sugestões/reclamações.

O ambiente dedicado à administração do sistema permite a visualização dos cadastros dos clientes, às solicitações feitas por eles e a formulários específicos para permitir o acompanhamento de todo o fluxo da pergunta postada pelo cliente até o momento de envio da solução ao cliente e sua publicação no *site*.

O esquema apresentado na Figura 2 ilustra o funcionamento do sistema de informação SBRT, desde o instante em que o usuário acessa o *site* até o momento em que ele recebe sua resposta técnica (RT) personalizada, conforme descrito a seguir:

- a) o usuário faz o seu primeiro acesso ao sistema, no endereço <http://www.sbrt.ibict.br>;
- b) consulta o banco de RTs e verifica se já existe alguma que atenda integralmente a sua necessidade;
- c) se houver, preenche o “Cadastro” para se tornar Cliente do SBRT (pessoa física ou pessoa jurídica);
- d) faz *download*, gratuitamente, da(s) RT(s) de seu interesse;
- d) se não houver, formula a sua pergunta ao SBRT por meio do “Formulário de Solicitação”;

- f) sua Solicitação é encaminhada, automaticamente, a uma das 7 instituições membros da Rede SBRT;
- g) especialistas analisam a solicitação e apresentam a solução em forma de um documento técnico padrão (RT);
- h) a RT gerada é enviada diretamente ao Cliente interessado (e-mail, fax, Correios) e publicada no banco de RTs, sem a identificação do Cliente.

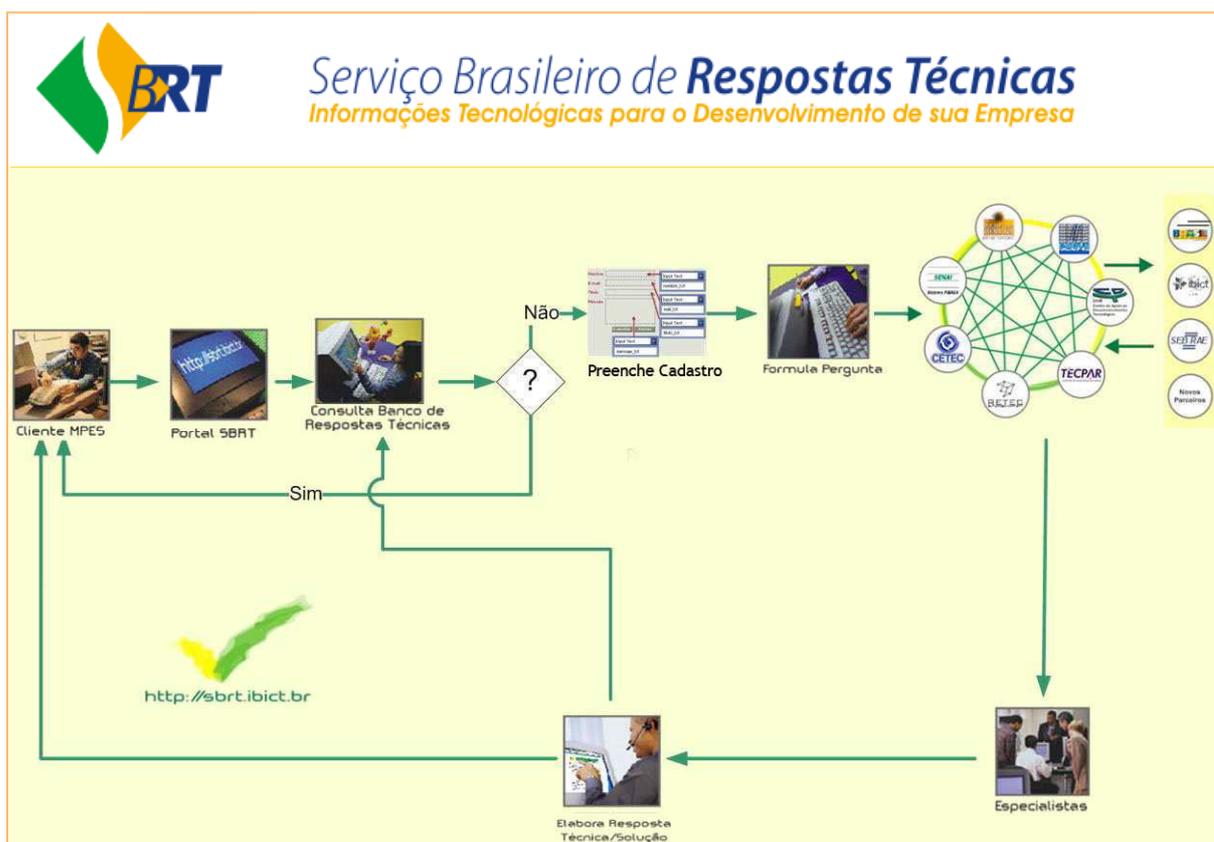


FIGURA 2 – Fluxo do Serviço SBRT

Fonte: Adaptado de SBRT

O esquema representado na Figura 3 ilustra a configuração do Sistema de Informação SBRT (V.2.0), implantado em outubro de 2007 no centro de processamento de dados do Ibict e nas instituições membros da Rede SBRT. Trata-se de um sistema dinâmico, com armazenamento de dados local, isto é, em cada instituição geradora de Resposta Técnica, e também central, por meio da tecnologia *Web Services*. A disponibilização dos dados para recuperação pelo usuário final, na

Web, é feita principalmente pelo subsistema central, ou portal do SBRT, localizado nos servidores do Ibict.

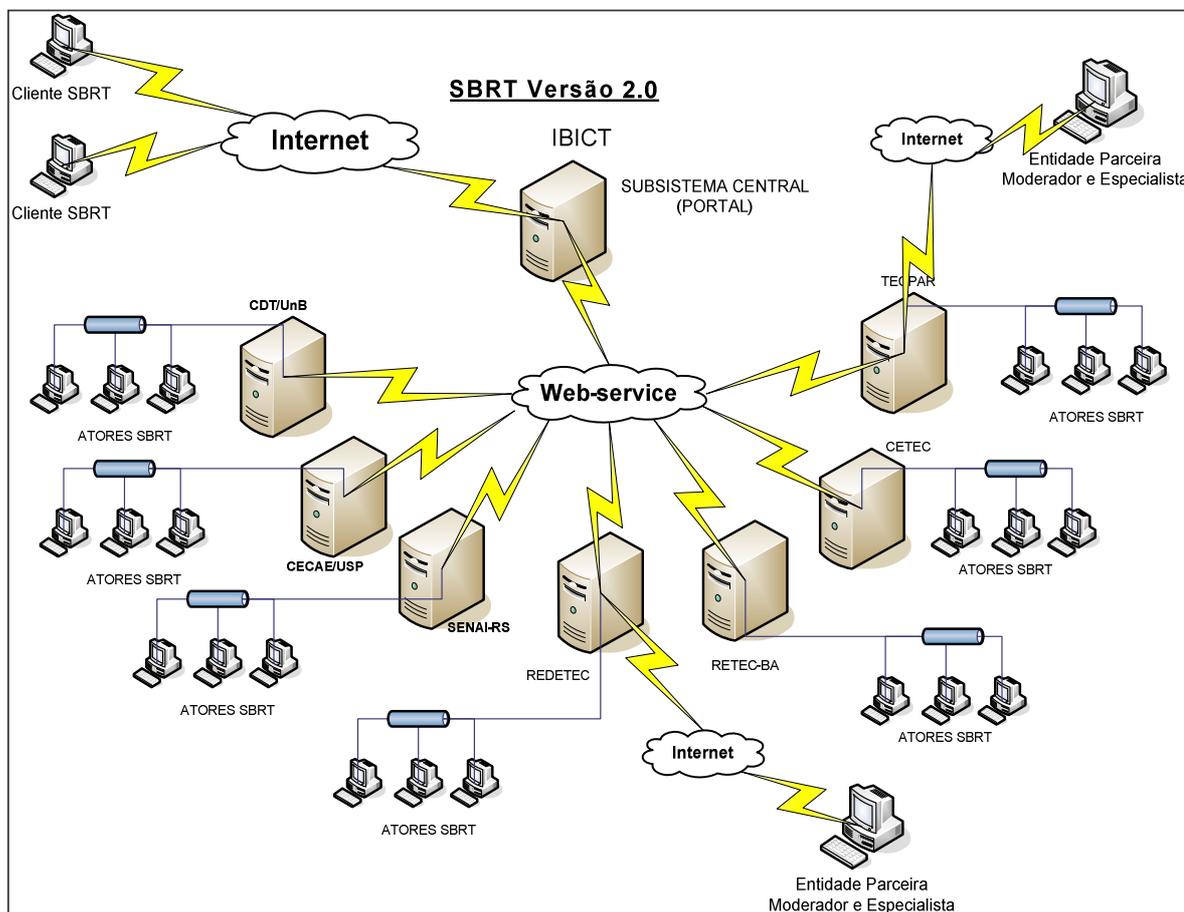


FIGURA 3 – Configuração do sistema de informação SBRT

Fonte: SBRT

1.2.3 A dimensão social do SBRT

Em pesquisa realizada em dezembro de 2005 sobre o uso do SBRT, chegou-se à conclusão de que 92% dos usuários do serviço são pessoas físicas, empreendedores em “busca de novos negócios”. (CHAVES, CORDEIRO e BASTOS, 2006). Mais recentemente, em março de 2007, verificou-se, em estudos internos à rede SBRT, que essa tendência se manteve, quando se observou que apenas 9% das solicitações provêm de empresas formalmente constituídas.

Há que se destacar, portanto, o papel social que o SBRT exerce na sociedade, na medida em que fomenta o empreendedorismo e, como consequência,

apóia a formalização de pequenos negócios e contribui para a geração de emprego e renda. Além disso, contribui para a capacitação do empreendedor, que se utiliza das técnicas sugeridas para solução de seus problemas, e para o combate à conhecida alta mortalidade da microempresa brasileira².

Essa dimensão social do SBRT é reconhecida pelo MCT, que declara em seu Relatório de Gestão 2003-2006 que a “informação tecnológica, por permitir a consolidação de empreendimentos de pequeno porte, se presta ao combate à informalidade e, conseqüentemente, estimula a inclusão social.” (BRASIL, 2006)

Acredita-se, portanto, ser o SBRT uma relevante fonte de informação para a aplicação da DCT, em busca das relações entre seus conteúdos, para apoio à construção de indicadores – importantes insumos para o direcionamento de ações voltadas ao desenvolvimento das micro e pequenas empresas.

1.3 Motivação e relevância da pesquisa

A realização do estudo foi motivada pela identificação da importância da ação inovadora que é o SBRT e seu potencial para contribuir na construção de indicadores de CT&I. O SBRT já é reconhecido como detentor de um rico *corpus* de conteúdo sobre a necessidade de informação tecnológica por parte dos micro e pequenos empresários brasileiros. Esse conteúdo possui informações preciosas, ocultas dentro dos textos das Respostas Técnicas, que, reveladas pela mineração, podem se tornar fontes de novos conhecimentos.

Sabe-se que as micro, pequenas e médias empresas são responsáveis pela maior parte dos empregos gerados no Brasil, fato estimulador de se estudar o conteúdo do sistema de informação de um serviço desenhado especialmente para apresentar soluções aos problemas enfrentados por empreendedores.

É fato comprovado, mediante depoimentos escritos à Rede SBRT, que soluções fornecidas pelo serviço de respostas técnicas vêm contribuindo para que empreendedores melhorem seus negócios, saiam do regime de informalidade e gerem empregos.

² Em pesquisa sobre a mortalidade das empresas, o SEBRAE estima que, dentre as empresas abertas entre os anos 2000, 2001 e 2002, 772.679 mil foram fechadas nos três anos, representando a perda de 2,4 milhões de empregos. Para os empresários, pressupõe-se desperdício de investimentos da ordem de R\$ 19,8 bilhões. (SEBRAE, 2004)

Outro fator considerado importante para a escolha do tema foi a compreensão da relevância de se propiciar a aproximação da pesquisa realizada nos centros geradores de conhecimento dos setores da sociedade que dele necessitam.

Nada mais motivador, portanto, acreditar que se está realizando uma pesquisa cujos resultados poderão trazer reais contribuições à sociedade. As palavras de Léa Velho levam a crer que estamos no caminho certo:

“No contexto atual, a ciência deixou de ser valorizada simplesmente por avançar o conhecimento e passou a ter sentido por seus resultados em termos de impacto na sociedade e na produção.” (VELHO, 2001)

Busca-se com esta pesquisa apresentar resultados concretos dos investimentos públicos empregados e servir de apoio à geração de novas e mais aprofundadas pesquisas no setor.

CAPÍTULO 2

A PESQUISA

2.1 Apresentação do Problema

2.1.1 Objeto de estudo

O objeto deste estudo é o Serviço Brasileiro de Respostas Técnicas (SBRT), sistema de informação tecnológica na *Web* destinado ao setor produtivo brasileiro, notadamente empreendedores, micro e pequenas empresas, fruto de um esforço compartilhado entre governo, instituições de pesquisa, universidades e iniciativa privada.

O serviço SBRT contém um conjunto de informações que expressam a demanda do micro e pequeno empresário brasileiro, assim como a resposta a essa demanda, na forma de conhecimento produzido por instituições de ensino e pesquisa atuantes no setor de informação tecnológica.

Armazenado em forma textual, esse conteúdo não está codificado para exploração por máquina e sua recuperação por meio de buscas com palavras-chave não reflete sua riqueza, dada a sua dispersão.

A base de dados é composta por documentos técnicos (RTs), cuja indexação é feita com o uso de uma tabela de assuntos adaptada da CNAE (Classificação Nacional de Atividades Econômicas)³ e por palavras-chave extraídas do texto das RTs. Essa indexação é fundamental para a busca e recuperação de documentos específicos, para atendimento a demandas pontuais. É insuficiente, no entanto, para realizar estudos do conteúdo da base de dados para fins de extração de informações de interesse estratégico e gerencial, e para identificação de tendências e relações

³ A CNAE é o instrumento de padronização nacional dos códigos de atividade econômica e dos critérios de enquadramento utilizados pelos diversos órgãos da Administração Tributária do país. Aplica-se a todos os agentes econômicos que estão engajados na produção de bens e serviços, podendo compreender **estabelecimentos** de empresas privadas ou públicas, estabelecimentos agrícolas, organismos públicos e privados, instituições sem fins lucrativos e agentes autônomos (pessoa física). Para maiores informações, consultar o site da Secretaria da Receita Federal, órgão coordenador da elaboração da CNAE: <http://www.receita.fazenda.gov.br/PessoaJuridica/CNAEFiscal/txtcnae.htm>

entre os documentos técnicos produzidos pela rede de instituições que compõem o SBRT.

O iminente crescimento da base de dados de respostas técnicas, dada a popularização do serviço SBRT mediante sua ampla divulgação nos meios de comunicação de massa, representará maior produção e disponibilização de conteúdos e a impossibilidade de realizar seu tratamento exclusivamente pelo homem.

Observa-se também que uma mesma RT pode conter informações úteis para várias empresas, podendo servir de base orientadora para novas solicitações, mais pontuais ou complementares. Além disso, um conjunto de RTs sobre o mesmo assunto pode servir de base para o desenvolvimento de dossiês técnicos⁴ para serem disponibilizados às empresas.

Assim, torna-se necessária a aplicação de ferramentas e técnicas que possibilitem a exploração e sistematização desses conteúdos, de forma a melhor utilizá-los, seja na associação dos conteúdos para geração de novos conhecimentos, seja na extração de informações estratégicas para gestão da rede SBRT e definição de políticas públicas voltadas ao setor produtivo brasileiro.

2.1.2 A questão

A pesquisa visou responder à seguinte indagação: a Descoberta de Conhecimento em Texto, aplicada no conteúdo da base de dados do SBRT, pode propiciar a extração de informação para apoio à construção de indicadores e à tomada de decisão?

⁴ Documento disponibilizado no site do SBRT com o objetivo de disseminar informações que possam promover o incremento de melhorias junto as MPEs ou com a função de alerta/antecipação de suas necessidades. Aborda diversos aspectos de natureza tecnológica de um determinado tema de interesse das MPEs. O conteúdo do Dossiê Técnico pode envolver aspectos relativos à tecnologia de produção/processo, matérias-primas, infraestrutura de produção, equipamentos, planta industrial, legislação, regulamentação, normas técnicas dentre outros. (<http://sbrt.ibict.br/oquedossie.php>)

2.1.3 A proposta

Propôs-se com este estudo investigar os conteúdos veiculados pelo sistema de informação do SBRT, com vistas a extrair informações para apoio à tomada de decisão, seja em nível governamental, seja em nível institucional, ou ainda, extrapolando os limites organizacionais, contemplar a gestão da própria rede SBRT e a construção de indicadores de CT&I.

2.2 Objetivos

2.2.1 Objetivo geral

Testar a eficácia da Descoberta de Conhecimento em Texto para exploração do conteúdo textual da base de dados do SBRT, com a finalidade de apoiar a construção de indicadores para subsidiar a tomada de decisão e a definição de políticas para o setor de micro e pequena empresa.

2.2.2 Objetivos específicos

- a) Caracterizar as demandas por informação tecnológica de empreendedores e micro e pequenas empresas brasileiras que utilizaram o SBRT.
- b) Extrair das Respostas Técnicas do SBRT padrões e relacionamentos até então desconhecidos, para apoiar a gestão e/ou a geração de novos conhecimentos.
- c) Extrair informações específicas sobre a base de dados em estudo, visando identificar necessidades de melhoria.

2.3 Justificativa e relevância do tema

Muito se tem investido na organização e disponibilização de conteúdos na *Web* com a finalidade de ofertar o conhecimento produzido ao maior número possível de usuários interessados em utilizá-lo, no menor espaço de tempo desde a sua geração. Fatores como a rápida evolução das tecnologias de informação e as mudanças de paradigma no que diz respeito ao compartilhamento de informações permitiram que se somassem aos tradicionais bancos de dados referenciais um

número cada vez maior de conteúdos textuais, com acesso direto à fonte primária do conhecimento, ou seja, ao próprio documento produzido.

À medida que se avolumam esses conteúdos, mais se desenvolvem técnicas de tratamento automático da informação para possibilitar o manuseio de grandes massas de dados visando seu melhor aproveitamento, seja para geração de novos conhecimentos, seja para extração de informações estratégicas para gestão. Isso ocorre graças à tomada de consciência sobre a importância de se explorar o conhecimento que se encontra oculto nos conteúdos acumulados ao longo do tempo, principalmente naqueles armazenados de forma natural, não planejada.

Sabe-se das necessidades de acesso à informação pelas micro e pequenas empresas, para se manterem no mercado cada vez mais competitivo. Assim, serviços que disponibilizem informações e conhecimentos para a capacitação tecnológica de empresários e, por consequência, melhoria de suas empresas, se tornam indispensáveis. O SBRT não só atende a essa necessidade do ponto de vista de disponibilizar a resposta certa, personalizada, para o problema colocado pelo microempresário, mas também promove a aproximação dessa clientela do setor produtivo com as unidades de pesquisa que produzem o conhecimento técnico necessário para o atendimento a essa necessidade.

Vale ressaltar os aspectos inovadores do SBRT, um serviço dinâmico que leva a informação tecnológica ao micro e pequeno empreendedor onde quer que ele esteja, promove a cooperação entre instituições de ensino e tecnologia do país para socializar o conhecimento por elas produzido, elabora as RTs em ambiente de cooperação interinstitucional *on line*, estimula a demanda por informação tecnológica e seu uso, suprimindo uma carência no setor em nível nacional.

O banco de dados gerado pelo serviço é de acesso público gratuito e constitui um dinâmico repositório de conhecimento gerado a partir de demandas reais postadas por empreendedores de todo o território nacional.

O atendimento às necessidades de informação dos clientes do serviço é potencializado pelo fato de a informação poder ser acessada no momento imediatamente após a sua geração e publicação. Some-se a isso a possibilidade de uma solução poder ser utilizada, na íntegra, por vários outros empreendedores, o que caracteriza um crescente potencial de auto-atendimento. Observa-se, ainda,

que uma mesma RT pode servir de base orientadora para novas solicitações, mais pontuais ou complementares, o que configura a tendência de atualização do conhecimento no que tange às necessidades mais presentes no mercado dos pequenos negócios.

Visualiza-se, assim, o SBRT como indutor da utilização da informação tecnológica como estratégia de desenvolvimento do setor produtivo de pequeno porte.

Os conteúdos gerados pelo SBRT estão armazenados em bases de dados e compõem-se de informações cadastrais e textuais. Entende-se a importância do estudo desses conteúdos na extração de informações com vistas à construção de indicadores para apoio à tomada de decisão, seja no campo da gestão interna da rede SBRT, seja no âmbito da política governamental de apoio à promoção da competitividade da micro e pequena empresa brasileira, buscando, assim, dar maior retorno à sociedade brasileira dos recursos públicos investidos na pesquisa tecnológica.

Entende-se que a pesquisa atende aos requisitos da linha de pesquisa Arquitetura da Informação na forma como é proposto pelo Programa de Pós-Graduação da Universidade de Brasília (PPGCInf): “Estudos teóricos e práticos sobre a análise da informação, indexação, estruturas informacionais, representação do conhecimento e recuperação da informação”. Além disso, identifica-se também a pesquisa com aspectos da linha de pesquisa Gestão da Informação e do Conhecimento, dada sua aplicação dos seus resultados na tomada de decisão. “Estudos teóricos, metodológicos e práticos sobre gestão da informação e do conhecimento em sistemas de informação...” (UnB/CID)

A relevância da pesquisa para a Ciência da Informação reside na abordagem de alguns temas ainda pouco explorados na literatura da área, notadamente a informação tecnológica e a questão do setor produtivo como usuário do conhecimento armazenado em sistemas de informação para a melhoria de seus produtos e processos.

Considerando que o objeto da Ciência da Informação, conforme definição clássica de Griffith (1980), citado por Capurro (2003), abarca *produção, seleção, organização, interpretação, armazenamento, recuperação, disseminação,*

transformação e uso da informação, entende-se que a abordagem da presente pesquisa está inserida com propriedade no escopo dessa Ciência.

Entende-se, também, que o estudo trará contribuições para a área no que diz respeito à disseminação, em língua portuguesa, do uso da DCT para extrair conhecimentos ocultos em grandes massas de dados armazenados em repositórios *online*, assunto ainda pouco explorado no Brasil.

Essa compreensão encontra suporte nas idéias de Soares (2003), que defende a expansão da Ciência da Informação no que diz respeito à investigação de outros ambientes, nos quais a informação também se destaca como um dos principais elementos: o ambiente das empresas de base tecnológica. Segundo o autor, essas empresas constituem um primeiro extrato de um novo modelo, tendo como principal insumo o conhecimento. Define informação tecnológica como a “aplicação de conhecimento no desenvolvimento de um processo/produto/serviço.”

A importância da construção de indicadores de C,T&I para apoio ao crescimento de um país é proporcional às dificuldades de gerá-los. Em um país com as dimensões territoriais do Brasil e a multiplicidade de ações no setor científico e tecnológico, essa tarefa se torna ainda mais complexa, por fatores que vão desde a pulverização de atividades tanto no setor público quanto no privado até as dificuldades de se computarem os resultados das iniciativas. Essa questão é levantada pelo MCT e encontra-se explicitada com mais detalhes em item específico no Capítulo 4.

Acredita-se, portanto, que o estudo do conteúdo de um sistema de informação voltado para o desenvolvimento do setor produtivo de pequeno porte, que atua em nível nacional e conta com a participação de instituições públicas e privadas comprometidas com a aplicação do conhecimento para o bem-estar da sociedade, possa levantar questões até então desconhecidas e de real valor para a tarefa de se computarem indicadores nesse setor.

Levando-se em consideração a conhecida alta mortalidade de micro e pequenas empresas no Brasil, é de se supor que a construção de indicadores que apoiem a gestão de ações e definição de políticas voltadas para o setor possa contribuir para minimizar esse problema. Assim, os resultados pretendidos nesta pesquisa podem somar-se a outros esforços no sentido de ampliar a utilização da

informação tecnológica como subsídio ao desenvolvimento e fortalecimento das MPEs.

Acredita-se, ainda, que este estudo possa trazer benefícios ao SBRT, como a sistematização do conteúdo, de forma a contribuir para a sua rápida recuperação, a associação de respostas técnicas que se complementem, assim como a construção de indicadores para apoio à definição de políticas internas da rede SBRT.

CAPÍTULO 3

FUNDAMENTAÇÃO TEÓRICA

3.1 Gestão de C&T

O Ministério da Ciência e Tecnologia (MCT), responsável pela formulação e implementação da Política Nacional de Ciência e Tecnologia, foi criado em 15 de março de 1985, pelo Decreto nº 91.146, como órgão central do sistema federal de Ciência e Tecnologia. Suas ações estão pautadas no Capítulo IV da Constituição Federal de 1988 e as seguintes responsabilidades estão inseridas na sua área de competência: o patrimônio científico e tecnológico e seu desenvolvimento; a política de cooperação e intercâmbio concernente a esse patrimônio; a definição da Política Nacional de Ciência e Tecnologia; a coordenação de políticas setoriais; a política nacional de pesquisa, desenvolvimento, produção e aplicação de novos materiais e serviços de alta tecnologia. (MCT)

Os parágrafos a seguir descrevem alguns marcos históricos da evolução do setor de C&T no Brasil que antecederam a criação do MCT, sem a pretensão de esgotar o assunto⁵.

O processo de “institucionalização de políticas e o desenvolvimento de um sistema articulado de C&T” teve início no pós-guerra, com a criação do Conselho Nacional de Pesquisas (CNPq), hoje Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com a finalidade estratégica de transformar o Brasil numa “potência emergente” razoavelmente independente dos pólos de poder existentes em nível mundial. (LONGO, 2000)

O autor enumera instituições e fundos de investimentos que tiveram presença marcante, como o BNDES, que se configurou em primeiro instrumento financeiro de apoio ao desenvolvimento de ciência e tecnologia, por meio do Fundo de Desenvolvimento Tecnológico (FUNTEC), criado em 1964, exerceu um papel relevante nos anos 60, mas foi extinto em 1975. A FINEP, empresa do setor público criada em 1967, assume os direitos e obrigações do Fundo e passa a assumir

⁵ Para maior detalhamento, sugere-se consultar os escritos de Longo (2000), reunidos no documento “O desenvolvimento científico e tecnológico do Brasil e suas perspectivas frente aos desafios do mundo moderno”, onde o autor faz um minucioso levantamento histórico sobre o tema.

diversas funções preponderantes, como passar a ser a Secretaria Executiva do FNDCT (Fundo Nacional de Desenvolvimento Científico e Tecnológico), criado em 1969, tornando-se, assim, a “mais importante fonte de recursos para atividades de pesquisa e desenvolvimento do País”. No início da década de 80, foi criado, junto ao Banco Mundial, o Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT), concebido e operacionalizado de maneira cooperativa pelas principais agências de fomento científico e tecnológico: FINEP, CNPq, CAPES e a Secretaria de Tecnologia Industrial do Ministério de Indústria e Comércio (STI/MIC). Em 1985, foi criado o Ministério de Ciência e Tecnologia - MCT, o qual foi extinto em 1989, devido à crise econômica em que o Brasil havia mergulhado no início da década. A extinção do órgão foi considerada pelo autor como um “retrocesso na evolução da área devido à perda de sua importância estratégica relativa e a aguda escassez de recursos que passou a ocorrer”. O MCT foi recriado em 1992 e restabeleceu-se a continuidade nas políticas e estratégias de C&T. (LONGO, 2002)

No que diz respeito ao papel e prioridades da política de Ciência e Tecnologia (C&T) no Brasil, primeiras discussões ocorreram nas décadas de 1960 e 1970, no âmbito das discussões acerca da economia da inovação, e permeavam dois campos considerados opostos. O primeiro deles defendia os investimentos em P&D, principalmente na pesquisa básica, com a finalidade de promover os avanços científicos e tecnológicos, considerados os “principais alavancadores do progresso técnico (teorias classificadas como *science & technology-push*)”. O segundo enfatizava as “forças do mercado e da demanda como o determinante primordial do progresso técnico (teorias classificadas como *demand-pull*)”. Nessa última visão, acreditava-se não serem os esforços de P&D os responsáveis pela maior parte das inovações, mas a atuação de outras partes da empresa (como as áreas de engenharia, produção e controle de qualidade), de outros elementos da cadeia produtiva (produtores de equipamentos, insumos e prestadores de serviços) ou dos próprios consumidores. (LASTRES, 1995)

Observa-se, ao longo do tempo de atuação do MCT, uma crescente preocupação com a inserção do setor produtivo nas prioridades de investimento dos recursos destinados à Ciência e Tecnologia e nas iniciativas de estímulo ao crescimento desse setor como propulsor da economia nacional. A esse respeito, Helena Lastres cita o relatório de atividades do MCT relativo ao período de 1992-

1994, cujos objetivos principais contemplavam a consolidação da base científica e tecnológica “de forma a permitir um desenvolvimento endógeno capaz de oferecer soluções criativas e duradouras aos principais problemas nacionais e a mobilização da maior participação do setor produtivo neste esforço (uma vez que se reconhece que a participação das empresas nos gastos nacionais não passa de 10%).” (LASTRES, 1995)

Nessa linha, o MCT lidera, hoje, ações de estímulo e financiamento do setor produtivo, no âmbito de programas específicos, que abrangem: atividades de pesquisa, desenvolvimento e inovação em empresas e projetos de cooperação entre instituições científicas e tecnológicas (ICTs) e empresas; articulações e iniciativas para a inovação nos setores estratégicos; desenvolvimento de “áreas portadoras de futuros”⁶; estímulos a investimentos em pesquisa, desenvolvimento e inovação, promoção da competitividade da indústria brasileira e incentivo a parceria de instituições acadêmicas e de P&D com empresas; apoio à inovação e competitividade para melhoria da eficiência do setor produtivo; consolidação de redes e serviços tecnológicos de suporte a empresas brasileiras de forma a permitir seu melhor posicionamento no mercado nacional e maior acesso ao mercado de outros países. (BRASIL, 2006).

3.1.1 A construção de indicadores

Atualmente, há uma preocupação em se adaptar a construção de indicadores – tradicionalmente voltados para a eficiência econômica – ao bem-estar social. Nesse sentido, acredita-se que o SBRT esteja inserido nessa nova concepção de conteúdos para apoio à construção de indicadores, vez que se trata de um estímulo à aplicação do conhecimento tecnológico gerado para melhoria da competitividade da microempresa e a conseqüente contribuição, tanto para a economia brasileira quanto para o bem-estar social.

Romão, Pacheco e Niederauer (2000) definem indicadores de C&T como “parâmetros numéricos capazes de resumir informações generalizadas sobre investimentos, produção e tendências no campo da C&T.”

⁶ São consideradas “portadoras de futuro” as áreas da pesquisa, a exemplo da biotecnologia e da nanotecnologia, nas quais se depositam as maiores expectativas do desenvolvimento futuro da ciência e da inovação (células-tronco, bioprodutos, biologia molecular, nanotecnologia, energia de hidrogênio e energia de biomassa, entre outras). (BRASIL, 2006 p.64)

Na visão do MCT⁷ – órgão responsável pela formulação e implementação da Política Nacional de Ciência e Tecnologia – há uma série de fatores que tornam “extremamente complexa a seleção e construção de indicadores”, que vão desde a heterogeneidade e amplitude de abrangência das atividades que envolvem o setor, passando pelo envolvimento de uma multiplicidade de agentes e instituições públicas e privadas, até a questão do tempo transcorrido entre as iniciativas e os resultados das ações. Outra questão levantada é o fato de os resultados produzidos não serem “facilmente computáveis, como é o caso dos ativos intangíveis”. Para vencer esses desafios e fazer face à responsabilidade de organizar e divulgar as informações de C&T no País, o MCT conta com a colaboração de instituições públicas, no âmbito federal e estadual, e de organizações privadas que produzem informações de interesse para a construção de indicadores de C&T e para o desenvolvimento de estudos sobre o tema. (BRASIL, 2004)

Ainda, o MCT afirma que “o conjunto de indicadores de C&T hoje disponível para o Brasil será continuamente enriquecido, na medida em que as dificuldades metodológicas e de acesso aos dados forem sendo superadas e novos indicadores produzidos.” Inicialmente, os indicadores limitavam-se ao dimensionamento dos recursos financeiros e humanos investidos em ciência e tecnologia – os chamados “indicadores de insumo”. Em seguida, foram criados os “indicadores de resultados”, contendo o registro da produção científica, a produção de patentes e a transferência de tecnologia entre países. Há, mais recentemente, a preocupação em se mensurar “como determinado resultado científico ou tecnológico afeta as várias dimensões das condições de existência dos indivíduos, seja no próprio campo científico e tecnológico, seja na dimensão econômica, seja na dimensão social”. Esses são os chamados “indicadores de impacto”, os quais têm se demonstrado de difícil elaboração. Os mais desenvolvidos atualmente são os relativos à dimensão científica e tecnológica, especialmente aqueles construídos no campo da bibliometria. (BRASIL, 2004)

É nessa natureza de indicadores que se insere esta pesquisa, onde os resultados dos investimentos públicos em um sistema de informação voltado a

⁷ A partir de 1999, o MCT assumiu a responsabilidade pela organização e a divulgação das informações de C&T do país, de forma centralizada. (BRASIL, 2004)

empreendedores e microempresários possam reverter em informações úteis à tomada de decisão nesse campo.

É desejável que iniciativas envolvendo a aplicação de recursos públicos voltados para o desenvolvimento sejam acompanhadas e mensuradas sistematicamente como fontes de retroalimentação, para o norteamento de ações futuras visando o melhor aproveitamento dos recursos investidos. Entende-se que a construção de indicadores deveria ser uma responsabilidade para as instituições que geram conteúdos no âmbito de programas governamentais.

Nesse sentido, o MCT chama atenção para as “reconhecidas e marcantes especificidades nacionais” relativas à base técnico-científica, as quais evidenciam a necessidade de “associar à produção de informação quantitativa o desenvolvimento de estudos mais aprofundados para validar ou redefinir os pressupostos sobre os quais se apoiam os indicadores.” (BRASIL, 2004)

Dentro da ótica de adaptação dos indicadores de C&T propostos pela Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) para melhor atender às necessidades dos países em desenvolvimento, Kondo (1998) sugeriu expandir o foco da construção de indicadores de C&T, tradicionalmente voltados para a eficiência econômica, para abranger indicadores “vinculados ao bem-estar social”.

Essa temática é abordada por Velho (2001), quando levanta questões relativas ao estabelecimento de um sistema de indicadores de C&T “útil e relevante para a tomada de decisão” e chama atenção para a importância do uso do conhecimento científico na produção, com a finalidade de propiciar melhoria da qualidade de vida da sociedade. Em sua opinião, os indicadores tradicionais passaram a ser questionados para se considerar a mudança técnica, o conceito de sistema nacional de inovação. De acordo com a autora, a inovação tem uma dimensão local e contingente.

Dessa forma, acredita-se que serviço SBRT corresponde a essa nova concepção de conteúdos para apoio à construção de indicadores, já que se trata de um estímulo à aplicação do conhecimento tecnológico gerado para melhoria da competitividade da microempresa brasileira e a conseqüente contribuição tanto para a economia quanto para o bem-estar social citado por Kondo (1998).

De acordo com o SEBRAE, há mais de 15 milhões de empreendimentos informais no país, três vezes mais que o número de micro e pequenas empresas formalmente constituídas: 4,6 milhões. “As micro, pequenas e médias empresas oferecem a absoluta maior parte dos empregos no país. De sua modernização e expansão dependem a curto prazo o surgimento de mais trabalho e emprego para a população”. (TAKAHASHI, 2005).

No Relatório de Gestão do MCT para o período 2003-2006, o SBRT figura entre as iniciativas que compõem o chamado Eixo Política Industrial, Tecnológica e de Comércio Exterior, como uma das ações de apoio à inovação e à competitividade que “tem tido ótimos resultados”.

“A importância estratégica do projeto SBRT para o aumento da competitividade nacional é, ainda, reforçada por sua contribuição para o estabelecimento de uma cultura de geração e difusão da informação tecnológica e para o desenvolvimento de negócios no setor produtivo”. (BRASIL, 2006)

Acredita-se, portanto, que a análise do conteúdo de um serviço de informação com as características do SBRT possa fornecer subsídios para estudos mais aprofundados relativos a indicadores de CT&I.

3.1.2 A micro e pequena empresa

De acordo com o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE), atualmente, há duas leis federais básicas que definem microempresa e empresa de pequeno porte. A primeira, o Estatuto da Microempresa e da Empresa de Pequeno Porte (Lei nº 9.841/99) – que estabelece incentivo através da simplificação, eliminação ou redução de suas obrigações administrativas, previdenciárias e creditícias – assim as define: "Microempresa é a pessoa jurídica com receita bruta anual igual ou inferior a R\$ 433.755,14"; "Empresa de pequeno porte, de R\$ 433.755,15 a R\$ 2.133.222,00. BRASIL, 2000). (BRASIL, 1999)

A segunda, a “Lei do Simples Federal” (Lei nº 9.317/96), que concede benefícios do ponto de vista tributário e fiscal, as define da seguinte forma: "Microempresa, pessoa jurídica com faturamento anual de até R\$ 240.000,00 (duzentos e quarenta mil reais)"; "Empresa de pequeno porte, pessoa jurídica receita

bruta superior a R\$ 240.000,00 (duzentos e quarenta mil reais) e igual ou inferior a R\$2.400.000,00 (dois milhões e quatrocentos mil reais)". (BRASIL, 1996).

O SEBRAE adota o critério número de empregados para a classificação das empresas brasileiras e expõe o seguinte conceito, adotado como referência aos tipos de empresa abordadas neste trabalho:

Considera-se como microempresa aquela com até 19 empregados na indústria e até 9 no comércio e no setor de serviços; as pequenas empresas são as que possuem, na indústria, de 20 a 99 empregados e, no comércio e serviços, de 10 a 49 empregados; as médias empresas de 100 a 499 empregados na indústria e de 50 a 99 no comércio e serviços. (SEBRAE)

De acordo com Araújo (2005), há poucas pesquisas acadêmicas sobre as microempresas, sendo essas mais freqüentes sob o ponto de vista econômico, onde são abordados temas relacionados a geração de emprego e crescimento nacional. A autora cita alguns estudos que abordaram a informação como estratégia para tomada de decisões no âmbito das micro, pequenas e médias empresas: Campos (1977), Borges (2002), Costa (2003) e Dias e Belluzzo (2003). Afirma que, ao longo das últimas três décadas, o governo Federal tem sido o maior investidor em serviços de informação voltados para as micro, pequenas e médias empresas.

Em suas reflexões sobre o setor empresarial na América Latina, Cubillo (1997) afirma que a pequena e média empresa é considerada, hoje, como um dos atores-chave do desenvolvimento, e isso se confirma quando se examinam os indicadores econômico-sociais, assim como na sua consolidação como tema relevante nas agendas do desenvolvimento.

As MPEs são, no Brasil, um dos pilares de sustentação da economia, em razão de seu número, abrangência, capilaridade e capacidade de geração de emprego. De acordo com o SEBRAE, há mais de 15 milhões de empreendimentos informais no país, três vezes mais que o número de micro e pequenas empresas formalmente constituídas: 4,6 milhões. "As micro, pequenas e médias empresas oferecem a absoluta maior parte dos empregos no país. De sua modernização e expansão dependem a curto prazo o surgimento de mais trabalho e emprego para a população". (TAKAHASHI, 2005).

É conhecido também o alto índice de mortalidade da microempresa brasileira. Em pesquisa realizada em 2004, o SEBRAE noticia que foram extintas 59,9% das empresas com até 4 anos de existência, 56,4% daquelas que tinham até 3 anos, e 49,4% das que tinham apenas 2 anos de existência. Dentre as causas apontadas pelos entrevistados da pesquisa realizada em todo o território nacional, destaca-se, em primeiro lugar, as falhas gerenciais na condução dos negócios, seguida de causas econômicas conjunturais e tributação. Deduz-se, no relatório, que as falhas gerenciais podem estar relacionadas à “falta de planejamento na abertura do negócio”. (SEBRAE, 2004).

Não é difícil associar essa carência à falta de preparo do microempresário, não só do ponto de vista de acesso a informações gerais de extrema relevância para a abertura do seu negócio e a manutenção do sucesso de seu empreendimento, ou de conhecimentos específicos sobre seu ramo de negócio.

Nesse sentido, Blaise Cronin, citado por Araújo, Freire e Mendes (1997), sustenta que “na sociedade atual, o sucesso na produção tem sido caracterizado pela busca de informação, pela comunicação com fontes de conhecimento relevantes, pela capacidade de absorção de tecnologias nas unidades produtivas e, especialmente, pela capacidade para produção e avaliação de informações”. As autoras abordam a demanda da indústria por informação que represente acesso a “conhecimento para ação” e afirmam que “cada vez mais a informação (e sua efetiva comunicação) é um dos recursos mais importantes para a produção de bens e serviços.” (ARAÚJO, FREIRE e MENDES, 1997)

Ferreira e Tarapanoff (2005) chamam atenção para o papel crucial do conhecimento nos processos econômicos, sustentam que a inovação está no cerne da mudança econômica e que, em prática gerencial, a inovação pode estar ligada à utilização de tecnologia. Fundamentam suas afirmações no pensamento de Schumpeter (1934) de que inovações radicais provocam mudanças no mundo, enquanto inovações “incrementais” preenchem continuamente o processo de mudança.

Entende-se que o serviço SBRT é um forte aliado à capacitação do microempresário na aplicação das soluções geradas pelas instituições membros da rede SBRT para melhoria de seus produtos e processos e, conseqüentemente, da competitividade do setor produtivo brasileiro de pequeno porte.

3.2 Gestão da Informação

3.2.1 A Informação Tecnológica (IT)

As buscas por literatura sobre “informação tecnológica” revelaram que há pouca pesquisa no Brasil sobre o tema. Januzzi e Montalli (1999) abordam questões terminológicas na área de informação referente à indústria/empresa, que inclui a informação tecnológica e para negócios, chamando atenção para a importância da qualidade no uso da informação como insumo para a competitividade brasileira e na consolidação das redes de informação, fundamentadas na especialização: “Ciência, tecnologia e negócios são as palavras de ordem no mundo atual, que formam o tripé da competitividade global.” As autoras apresentam uma diversidade de termos encontrados na literatura, que indicam a inexistência de harmonização conceitual: informação em ciência e tecnologia, informação científica e tecnológica, informação industrial e tecnológica, informação industrial, informação para indústria, informação tecnológica e informação para negócios.

Para efeito deste trabalho, foi adotado o termo “informação tecnológica”, comumente utilizado no âmbito da comunidade de Ciência e Tecnologia.

São encontradas na literatura definições diversas para o termo “informação tecnológica”, com alguns enfoques diferenciados. A antiga Federação Internacional de Documentação (FID) associou esse tipo de informação à busca por aperfeiçoamento e inovação e o definiu como:

Todo conhecimento de natureza técnica, econômica, mercadológica, gerencial, social entre outros, que por sua aplicação, favoreça o progresso na forma de aperfeiçoamento e inovação. (FID *apud* SEBRAE)

Outro enfoque, desta vez voltado para a melhoria da produtividade e desenvolvimento de pesquisa tecnológica, é apresentado no Glossário de Informação Tecnológica, publicado pelo Senai em 2001:

Informação tecnológica é aquela relacionada com o modo de fazer um produto ou prestar um serviço para utilizá-lo no mercado, servindo para difundir tecnologia de domínio público para possibilitar a melhoria da qualidade e da produtividade de empreendimentos existentes e construir insumo para o desenvolvimento de pesquisa tecnológica. (RODRIGUES, ABE, DIB, 2001)

O SEBRAE chama atenção para a importância do uso adequado e eficaz da informação, afirma que isso que pode representar uma vantagem competitiva para as empresas, e propõe a organização da informação tecnológica em quatro segmentos (informação tecnológica referencial, monitoramento tecnológico, informação cadastral e resposta técnica). Apresenta uma definição para o termo informação tecnológica atrelado à tomada de decisão:

Toda informação ou conhecimento, de caráter tecnológico ou não, que contribui para a mudança e o aperfeiçoamento de serviço, processo ou produto industrial, em todo o sistema produtivo, agregando conhecimentos necessários à tomada de decisão. (SEBRAE)

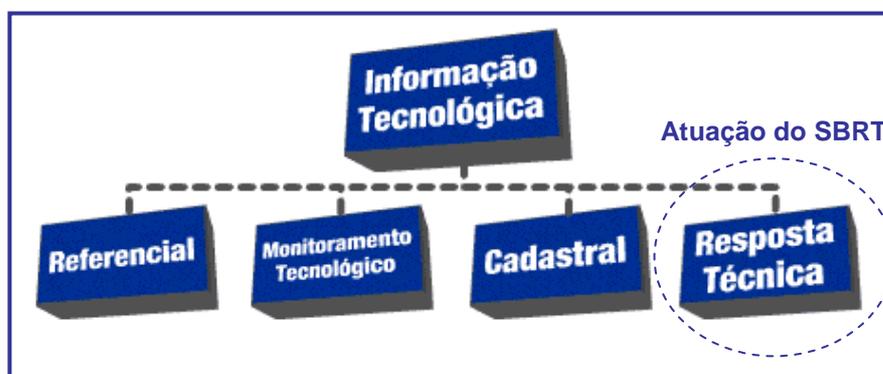


FIGURA 4 – Organização da Informação Tecnológica na visão do SEBRAE

Fonte: Adaptado de SEBRAE (http://www.SEBRAE-sc.com.br/SEBRAEtib/Hp/conceitos/info_tec/)

Informação Tecnológica Referencial: informações que mostram o caminho e a fonte de informações tecnológicas, como: referência bibliográfica patentes de invenção, incluindo o nome do depositante, o título da patente e a classificação por assunto. Essas informações estão organizadas em listas ou bases de dados e possibilitam recuperar informações publicadas, tais como: leis e regulamentos; normas técnicas e padrões; patentes; características de materiais e de matéria-prima; técnicas e processos de fabricação; novas tecnologias.

Monitoramento Tecnológico: serviço de Informação que objetiva acompanhar o desenvolvimento tecnológico, os concorrentes, a clientela, os fornecedores de mercado e a ambiência social política, econômica e cultural, visando orientar estrategicamente uma empresa ou organização, em suas decisões

de negócios. O monitoramento tecnológico mostra a evolução de uma determinada tecnologia (produto ou processo) ao longo do tempo ou em uma determinada região. Utiliza-se de documentos de patentes, publicações em artigos técnicos, revistas especializadas, boletins informativos, normas técnicas, discussões e realização de eventos.

Informação Tecnológica Cadastral: informações cadastradas que listam fornecedores de produtos e serviços tecnológicos. As páginas amarelas do catálogo telefônico é um exemplo claro desse tipo de informação. Classificam as empresas pelos produtos e serviços ofertados e permitem identificar qual delas possui a informação de que se necessita. São também exemplos os cadastros que identificam produtos, equipamentos e serviços especializados: fabricantes de peças e componentes de máquinas e equipamentos; fornecedores de insumos e de matéria-prima; fabricantes de equipamentos especializados; técnicos e especialistas consultores e prestadores de serviços técnicos e tecnológicos; feiras e congressos especializados.

Resposta Técnica (Tecnológica): serviço personalizado, sendo mais abrangente que a Informação Referencial ou que a Informação Cadastral. Raramente está pronta, exigindo consulta a especialistas ou a técnicos, envolvendo certo grau de incerteza, pois normalmente apresenta opções de solução do problema. (SEBRAE)

Na Figura 5, o SEBRAE apresenta um paralelo entre a informação científica e a informação tecnológica, apontando suas principais fontes e o uso dessas duas categorias de informação para a composição dos chamados “pacotes de informação personalizada”, onde estão situadas as soluções tecnológicas denominadas Respostas Técnicas.

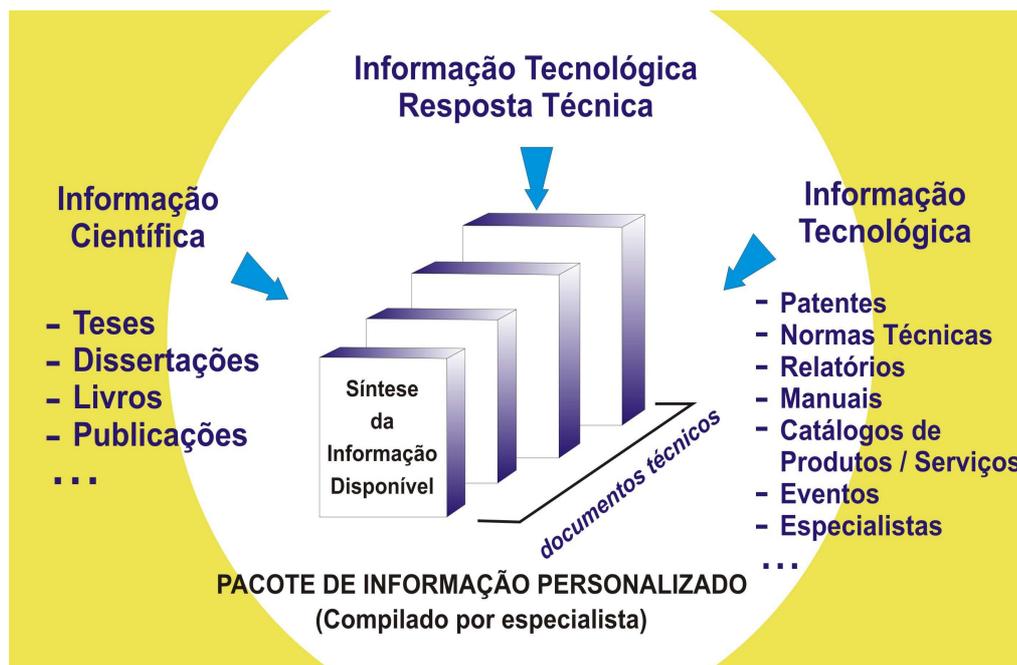


FIGURA 5 – Pacote de informação personalizado compilado por especialista

Fonte: SEBRAE (http://www.SEBRAE-sc.com.br/SEBRAEtib/Hp/conceitos/info_tec/resposta.htm)

Independentemente do foco, esteja ele na tomada de decisão, na melhoria da competitividade das empresas ou na difusão de tecnologias de domínio público, as definições de informação tecnológica encontradas na literatura trazem à tona a preocupação central com a aplicação do conhecimento científico e tecnológico gerado na melhoria de bens, processos e serviços do setor produtivo.

3.2.2 Serviços de informação especializados

Santos (2002) afirma que “a descoberta (semi-)automática de regras que capturem inter-relacionamentos entre itens de informação em um texto pode ser bastante importante para a tomada de decisões estratégicas.” Segundo Vieira (1993), na evolução do processo de desenvolvimento de sistemas de informação para apoiar organizações, a inteligência humana e tecnológica têm se associado, trazendo avanços e novos desafios de pesquisa.

A importância e o impacto de serviços especializados de informação na transferência e exploração tecnológica são abordados por Cysne (1996), que afirma que a informação está afetando a natureza da competição e do sucesso das empresas e tem se tornado um ativo recurso econômico e fator no processo de inovação tecnológica. Segundo a autora,

As redes e sistemas de informação, utilizando as novas tecnologias de informação e de comunicação, podem realmente ser consideradas um novo paradigma da inovação tecnológica e, por isto mesmo, devem se constituir em temas de estudo em qualquer análise sobre desenvolvimento tecnológico e social. (CYSNE, 1996)

Silva e Marinho (2002) abordam a questão da interação universidade-empresa como ação propulsora da competitividade do setor produtivo e do país e citam os serviços de informação como facilitadores para suprir suas necessidades de informação científica e tecnológica.

Vale um mergulho no ponto de vista de Borges (2006) sobre a interação universidade-empresa, sob a ótica da construção de serviços de informação especializados no atendimento a necessidades do setor produtivo. Segundo a autora, a interação universidade-empresa sempre foi “indireta, difícil, pouco freqüente”, dado que, a despeito do “reconhecimento mútuo, são dois modos de produção distintos que, apesar de fazerem parte de um mesmo processo de desenvolvimento têm regras, ritos, mecanismos de validação e dinâmicas próprias”. Sustenta que a TI é um “instrumento de aproximação das duas realidades”, e que no estabelecimento dessa relação, é preciso criar “mecanismos organizacionais e operacionais, como uma estrutura sistêmica de informação, capaz de ser um instrumento efetivo de resposta ao atendimento das necessidades específicas de cada um, com capacidade de atender aos questionamentos”. Afirma que, para organizar a demanda, é preciso identificar as necessidades do setor produtivo, verificar a possibilidade de atendimento, encaminhar ao setor responsável, administrar a resposta e fazer a comunicação dos serviços e a difusão do conhecimento. A autora apresenta como alternativa para que a interação ocorra:

[...] criar estruturas formais, de uma organização ou de um conglomerado de organizações, centros ou núcleos de interação, que teriam a missão de se constituírem em canais de informação e comunicação facilitadores do acesso e da interação com os setores prestadores de serviços, com

descentralização da ação, e a criação de mecanismos ágeis e de fácil acesso ao próprio setor realizador do serviço". (BORGES, 2006, p. 117)

Entende-se, portanto, que os serviços de informação especializados têm um papel preponderante na aplicação direta, pelo setor produtivo, do conhecimento gerado nos centros de pesquisa. Vale ressaltar a importância do uso de recursos tecnológicos para tornar isso uma realidade, não somente do ponto de vista de organização da demanda e seu atendimento, mas também para extrair de seus conteúdos informações que possam proporcionar o surgimento de novos conhecimentos.

Dado o seu caráter especializado, o estudo do conteúdo do sistema de informação SBRT pode representar uma forte contribuição para o norteamento da tomada de decisão voltada ao setor do microempresariado no tocante à transferência de conhecimento.

3.2.3 A Descoberta de Conhecimento em Texto (DCT)

A partir da literatura analisada, observa-se a diversidade de termos utilizados para se definirem as várias técnicas utilizadas na extração de informações relevantes de grandes massas de dados. Esta pesquisa adota a DCT e explora as suas potencialidades para apoiar a construção de indicadores, a partir do tratamento automático e da análise de conteúdos armazenados em forma textual.

Grandes repositórios textuais contêm informações adormecidas, camufladas até que o minerador as encontre e as torne informações preciosas para a organização.

Descobrir conhecimento significa identificar, receber informações relevantes e poder computá-las e agregá-las ao seu conhecimento prévio, mudando o estado de conhecimento atual, a fim de que determinada situação ou problema possa ser resolvido. (WIVES, 2004)

Para melhor compreensão dos processos que envolvem a DCT, vale a pena abordar alguns conceitos básicos apresentados por Trybula (1999) em sua revisão de literatura sobre a mineração de textos. Ele define base de dados como uma "coleção organizada de dados armazenados", os quais normalmente se referem a dados ativos e não a compilações de dados históricos e podem compor sistemas

armazenados em locais distintos. As bases textuais, por sua vez, são repositórios de informações textuais, compilações históricas de informações eletrônicas que podem não ter sido previamente organizadas. (TRYBULA, 1999).

Essa definição pode ser ampliada para abranger repositórios atuais de informação textual, onde se observam grandes volumes de informação não estruturada inserida em sistemas de informação previamente organizada e estruturada. Exemplo disso é o sistema de informação SBRT, que possui uma estrutura de campos padronizados de informações associadas a textos elaborados na forma de documentos técnicos.

São muitas as discussões em torno das definições das técnicas de extração automática de conhecimento em grandes volumes de informação, onde conceitos e termos se misturam, por vezes sendo utilizados como sinônimos ou com pequenas nuances na diferenciação. De forma abrangente, utiliza-se o termo “descoberta de conhecimento”, passando-se a qualificá-lo a partir do conteúdo a ser analisado: se este foi previamente organizado e estruturado ou se se encontra disperso em documentos textuais dos mais diversos formatos e tamanhos.

Essa observação encontra sustentação nos estudos de Wives (2004), para quem a descoberta de conhecimento é um “processo de aquisição de novos conhecimentos auxiliado por computador”, que pode ser classificado em Descoberta de Conhecimento em Dados Estruturados (DCD, ou *KDD* em inglês) e Descoberta de Conhecimento em Textos (DCT, ou *KDT* em inglês).

Weiss *et al.* (2005) abordam o tema de forma simplificada e direta. Para eles, as técnicas para tratamento de grande quantidade de dados para análise futura podem ser resumidas em duas: a mineração de dados (*data mining*) e mineração de textos (*text mining*).

Schiessl (2007) considera a DCD como uma “combinação interdisciplinar de técnicas e métodos” concebida para resolver problemas principalmente em conteúdos numéricos e estruturados armazenados em bancos de dados. O autor acrescenta a necessidade de adaptação da DCD “para que a linguagem natural seja passível de processamento automático visando à extração de conhecimento”.

Na visão de Trybula (1999), a descoberta do conhecimento (sigla em inglês *KD* para o termo *knowledge discovery*), é o “processo de transformação de dados

em relações previamente desconhecidas e insuspeitas, que podem ser empregadas como previsores de futuras ações".

De certa forma, os conceitos expressos pelos autores se assemelham, embora utilizem termos distintos para representá-los.

Araújo Júnior (2005) enumera alguns termos, como prospecção, descoberta de conhecimento em banco de dados, mineração de dados, descoberta de conhecimento em textos, mineração de textos, que correspondem às siglas em inglês: *KDD (Knowledge Discovery in Databases)* e *KDT (Knowledge Discovery in Texts)*. O autor afirma que os termos "mineração de textos", "descoberta de conhecimento em textos" e "mineração de dados" têm sido utilizados como sinônimos na literatura.

Observa-se, entretanto, a preocupação de vários autores em esclarecer a diferenciação entre mineração de dados (*data mining*) e mineração de textos (*text mining*) como um elemento crucial para a compreensão das potencialidades dessas técnicas de tratamento de grandes massas de dados.

3.2.3.1 Mineração de dados e Mineração de textos

Embora se encontre na literatura diferentes abordagens de mineração de dados e mineração de textos, onde fica clara a diferenciação dessas duas técnicas, há alguns autores que consideram a mineração de textos como uma variação da mineração de dados.

Ambas as abordagens buscam encontrar padrões valiosos em grandes volumes de dados armazenados e têm por base amostras de exemplos do passado. A primeira distinção entre elas reside no fato de, na mineração de dados, esses serem extraídos em formato numérico, enquanto que na mineração de textos ocorre a análise de coleções de documentos onde seus conteúdos são legíveis e seu significado é óbvio. Embora seja clara a distinção número *versus* texto, isso não significa que ambas as abordagens representem conceitos distintos, visto que, para a conclusão do processo de mineração de textos, este será processado e transformado em representação numérica. (WEISS, 2005).

Tan (1999) defende a idéia de que a mineração de textos, também conhecida como mineração de dados ou descoberta de conhecimento em bases de dados textuais, refere-se geralmente ao processo de extração de padrões interessantes e não triviais de documentos textuais. Segundo o autor, a técnica pode ser vista também como uma extensão da mineração de dados ou da descoberta de conhecimento em bases de dados estruturadas. Chama atenção, ainda, para a complexidade da mineração de textos em relação à mineração de dados, dada a sua característica inerentemente não estruturada e difusa.

Quoniam et al. (2005) definem *Data Mining* como "todas as técnicas que permitem extrair conhecimento de uma massa de dados que, de outra maneira permaneceria escondido nas grandes bases". Afirmam que sua aplicação "torna possível comprovar o pressuposto da transformação de dados em informação e posteriormente em conhecimento" e, por esta razão, ela se configura em uma técnica imprescindível para o processo de tomada de decisão.

A mineração de dados é definida por Trybula (1999) como o processo básico empregado para analisar padrões em dados e extrair informações – processo esse que inclui a "limpeza" e a validação dos dados – e tem por objetivo gerar, além de verificar, uma hipótese sobre o dado selecionado. Segundo ele, a mineração de dados costuma ser utilizada em grandes bases de dados organizacionais contendo informações sobre clientes para obtenção de informações sobre seu comportamento diante de incentivos do mercado.

Já a mineração de textos, na visão do autor, é o processo básico para analisar padrões em textos e apresentar informações, processo esse que inclui também a "limpeza" e validação dos dados com maior grau de dificuldade, dada a necessidade de isolamento de raízes de palavras e identificação de suas categorias gramaticais (substantivo, verbo, advérbio etc.). (TRYBULA, 1999)

Hearst (2003) comenta o uso errôneo do termo *data mining* e a escassez de pesquisas sobre *text mining*. Considera a mineração de textos como uma ferramenta de suporte ao e valorização do conhecimento gerado, já que proporciona sua exploração e reutilização, e a define como:

Descoberta, por computador, de novas informações, previamente desconhecidas, pela extração automática de informações de diferentes recursos escritos, onde o elemento chave é a interligação das informações

extraídas para formar novos fatos e novas hipóteses a serem posteriormente exploradas pelos meios de experimentação mais convencionais. (HEARST, 2003)

De acordo com Hearst (2003), a mineração de textos, por facilitar a transferência de informação em conhecimento, propicia não somente o manuseio, mas a possibilidade de se manter atualizado no controle a vasta quantidade de informações relevantes para a organização.

Marcelo Scheissl (2007) realizou minucioso estudo sobre a literatura acerca da Descoberta de Conhecimento em Dados e os diversos processos que a compõem, onde registra as particularidades apontadas por diversos autores. Segundo ele,

O processo de descoberta de conhecimento em dados compreende a seleção de dados, o pré-processamento que envolve sua adequação aos algoritmos, a efetiva mineração de dados, isto é, o uso de técnicas de mineração, a validação dos resultados e, finalmente, a análise e interpretação dos resultados para a aquisição do conhecimento. (SCHIESSL, 2007)

O autor afirma, ainda, que a DCD vem sendo consolidada como um poderoso ferramental para auxiliar o homem na exploração da grande quantidade de informação disponível em formato eletrônico, dadas as limitações humanas no manuseio e interpretação dessa informação. (SCHIESSL, 2007)

3.2.3.2 Mineração de textos e Recuperação da Informação

Há autores que abordam a questão da diferenciação entre a mineração de textos e a recuperação da informação.

Hearst (2003) afirma que a mineração de textos, ao contrário do que muitas pessoas pensam, não é uma forma de facilitar a busca de informação na *Web* e não deve, portanto, ser confundida com o processo de Recuperação da Informação (RI). Segundo o autor, na RI ocorre a descoberta de informação já conhecida, inserida em um documento pelo autor, uma forma de o usuário selecionar, em uma coleção de documentos, aqueles que lhe interessam e desprezar os demais. Enquanto que a mineração de textos tem como meta descobrir informação desconhecida, que não está escrita em nenhum documento individualmente, ou derivar novas informações a

partir dos dados analisados, encontrar padrões em conjuntos de dados e/ou separar o signo do ruído. (HEARST, 2003)

Da mesma forma, Lucas (2007) afirma que mineração não é RI; na mineração de dados textuais (*Text Data Mining – TDM*), os relacionamentos entre os documentos podem gerar novos fatos, não previamente conhecidos.

Esse diferencial em relação à RI é o que a torna especialmente adequada ao manuseio do crescente volume de informação contido em repositórios diversos de documentos textuais para extração de informações aparentemente inexistentes e a identificação de padrões e o relacionamento de conceitos nessas coleções.

Outra definição que reforça essa diferenciação entre mineração de textos e RI é apresentada por Aires (2005), que aborda o tema do ponto de vista da necessidade do usuário. A autora enfatiza que a mineração de textos objetiva “analisar coleções de documentos como um todo para extrair informações que possam ser úteis para um determinado propósito”. Segundo ela, tais informações “podem ser esperadas ou não e podem também mostrar relacionamentos totalmente desconhecidos”.

Mineração de Textos não é recuperação de informação; a recuperação de informação atende às necessidades de um usuário que foram expressas através de uma consulta retornando documentos, sendo que a Mineração de Textos explora relacionamentos entre documentos de forma independente das necessidades de um usuário. (AIRES, 2005).

Rogério Araújo Júnior (2005) afirma que a mineração de textos só terá sentido se for aplicada a uma situação concreta e sugere que seja utilizada, por exemplo, para enriquecer os instrumentos de apoio ao processo de indexação em um sistema de RI.

Embora a conceituação das técnicas de tratamento automático de grandes volumes de dados ainda se encontra de certa forma difusa, é possível identificar pontos comuns fundamentais para a compreensão do seu funcionamento. Fica clara a forte evolução dessas técnicas com a possibilidade de tratamento de conteúdos em linguagem natural, que representa a maciça maioria dos conteúdos gerados pela organização de modo geral (80%, conforme dados levantados por TAN (2003) e já citados no capítulo introdutório desta pesquisa).

CAPÍTULO 4

PROCEDIMENTOS METODOLÓGICOS

A proposta desta pesquisa foi o estudo do conteúdo textual da base de dados de soluções tecnológicas personalizadas fornecidas aos micro e pequenos empresários usuários do Serviço Brasileiro de Respostas Técnicas (SBRT).

Há na literatura diferentes termos para se definirem as técnicas utilizadas na extração de informações relevantes de grandes massas de dados, e alguns deles até se confundem: prospecção, descoberta de conhecimento em bancos de dados, mineração de dados, mineração de textos, descoberta de conhecimento em textos. Esta pesquisa adota esta última, isto é, a DCT, e explora as suas potencialidades para construção de indicadores, a partir do tratamento automático e da análise de conteúdos armazenados em formato textual.

4.1 Caracterização da pesquisa

Esta pesquisa pode ser caracterizada como estudo de caso de natureza descritiva, dada a intenção de se conhecer uma realidade, descrevê-la e interpretar os fatos observados. Reforça essa idéia sobre a natureza da pesquisa o fato registrado na literatura de que na pesquisa descritiva não ocorre a interferência do pesquisador no fenômeno estudado e o de que “pode haver procura de relação causal, associativa, ou co-relação entre variáveis, mas será por inferência de dados sobre fenômenos já ocorridos”. (MUELLER, 2006)

A pesquisa descritiva é definida por Marconi e Lakatos (2002) como sendo aquela que “delineia o que é”, em contraposição à pesquisa histórica, que “descreve o que era” e à pesquisa experimental, que “descreve o que será”. Segundo as autoras, esse tipo de pesquisa abrange quatro aspectos: descrição, registro, análise e interpretação de fenômenos atuais, objetivando o seu funcionamento no presente.

4.2 Delimitação da pesquisa

O sistema de informação SBRT é composto por um banco de dados contendo várias bases de dados. Algumas delas são *cadastrais*, estão estruturadas em metadados e reúnem informações sobre os vários atores da rede SBRT, quais sejam: os usuários da Internet que já baixaram alguma Resposta Técnica (RT)⁸; os usuários chamados de clientes, por já terem feito alguma solicitação de RT, as instituições membros da rede SBRT, as instituições parceiras, os usuários internos do sistema: gestores, técnicos e especialistas.

Há também duas bases de dados *textuais*: i) *Respostas Técnicas (RTs)*, contendo as solicitações dos clientes – da forma como apresentaram suas perguntas à rede SBRT – e as RTs, soluções elaboradas para responder a essas perguntas, e ii) *Dossiês Técnicos (DTs)*⁹.

O universo de estudo desta pesquisa é o conteúdo textual da base de dados de RTs, abrangendo, portanto, a pergunta feita pelo cliente. É importante frisar que o segundo conjunto de informações textuais do sistema, contendo os Dossiês Técnicos, não faz parte do escopo desta pesquisa.

O conteúdo a ser estudado foi extraído do sistema de Informação SBRT no dia 8 de agosto de 2007, mediante autorização do Comitê Gestor da Rede SBRT. Os dados representavam, naquela data, a totalidade das RTs enviadas aos clientes e publicadas no *site*: 6.041 documentos.

As informações constantes do corpo do texto da RT são: título da RT, resumo, data de publicação, palavras-chave, assunto, demanda (a pergunta feita pelo cliente) e instituição respondente (responsável pela elaboração da RT).

Decidiu-se pela extração de algumas informações da base de dados cadastral que pudessem vir a ser importantes também para as análises: i) Cliente Pessoa

⁸ A partir da versão 2.0 do Sistema de Informação SBRT, implantado em setembro de 2007, passou a ser exigido dos usuários interessados na visualização do texto integral e *download* de Respostas Técnicas o preenchimento de um cadastro simplificado.

⁹ Dossiê Técnico (DT) – documento elaborado pelas instituições membros do SBRT, que aborda diversos aspectos de natureza tecnológica de um determinado tema de interesse das MPEs. Pode envolver aspectos relativos a: tecnologia de produção/processo, matérias-primas, infra-estrutura de produção, equipamentos, planta industrial, legislação, regulamentação, normas técnicas, dentre outros. Dissemina informações que possam promover o incremento de melhorias junto às MPEs ou com a função de alerta/antecipação de suas necessidades. Fonte: SBRT. (<http://sbrt.ibict.br/dossies.do?&idMenu=13>)

Física (gênero, cidade, estado (UF), escolaridade); ii) Cliente Pessoa Jurídica (nome da empresa, ou razão social, cidade, estado, natureza do vínculo do contato).

O Quadro 1 apresenta detalhamento do conteúdo do sistema de informação SBRT e destaca a parte desse conteúdo que constitui o objeto do presente estudo: as perguntas dos clientes e as respostas a essas perguntas, as RTs.

QUADRO 1 – Conteúdos do Sistema de informação SBRT

Bases de Dados	Conteúdos	Descrição	Tipo de Acesso		
			Restrito		Público
Cadastrais	Dados sobre Instituições	Cadastro das instituições membros da rede SBRT e de seus parceiros.	Rede SBRT	-	-
	Dados sobre Mediadores e Especialistas	Cadastro do pessoal interno (especialista ou mediador) e externo (especialista) que acompanha as demandas e elabora RTs.	Rede SBRT	-	-
	Dados sobre Usuários Internet	Cadastro de usuários da Internet que pesquisaram e visualizaram RTs, mas não preencheram o cadastro de Cliente para fazer perguntas.	Rede SBRT	Usuário Internet	-
	Dados sobre Clientes	Cadastro de empreendedores, micro e pequenos empresários que já fizeram pelo menos uma solicitação ao SBRT.	Rede SBRT	Cliente	-
Textuais	Solicitações dos Clientes	Pergunta feita ao SBRT em busca de solução tecnológica.	Rede SBRT	Cliente	-
	Respostas Técnicas (RTs)	Soluções personalizadas às perguntas colocadas pelos clientes *	-	-	Internet
	Dossiês Técnicos (DTs)	Documentos que abordam diversos aspectos de natureza tecnológica de um tema específico	-	-	Internet

* Cada Resposta Técnica é acompanhada também da pergunta que motivou sua elaboração. Portanto, as solicitações dos clientes estão presentes também na base de dados de RTs e fazem parte do corpo desta pesquisa.

Um exemplo completo do tipo de documento analisado na pesquisa é apresentado no Anexo A – Exemplo de Resposta Técnica (RT).

4.3 Etapas da pesquisa

A mineração de textos é um campo multidisciplinar que envolve recuperação da informação, análise de textos, extração de informações, *clustering* (agrupamento), categorização, visualização, tecnologia de base de dados, aprendizagem de máquina e mineração de dados. (TAN, 1999)

Para a realização da pesquisa, foram realizadas as seguintes etapas:

- i) seleção do conteúdo dentre as bases de dados do sistema SBRT;
- ii) extração dos dados;
- iii) conversão dos dados para o formato legível pela ferramenta de mineração;
- iv) remoção de ruídos (limpeza e padronização);
- v) construção da base de trabalho;
- vi) mineração do texto e agrupamento de documentos;
- vii) análise dos dados para extração de informações visando apoiar a construção de indicadores.

A Figura 6 ilustra, de forma esquematizada e resumida, o escopo da pesquisa, sua abrangência, as etapas percorridas durante todo o processo de desenvolvimento e os resultados pretendidos.

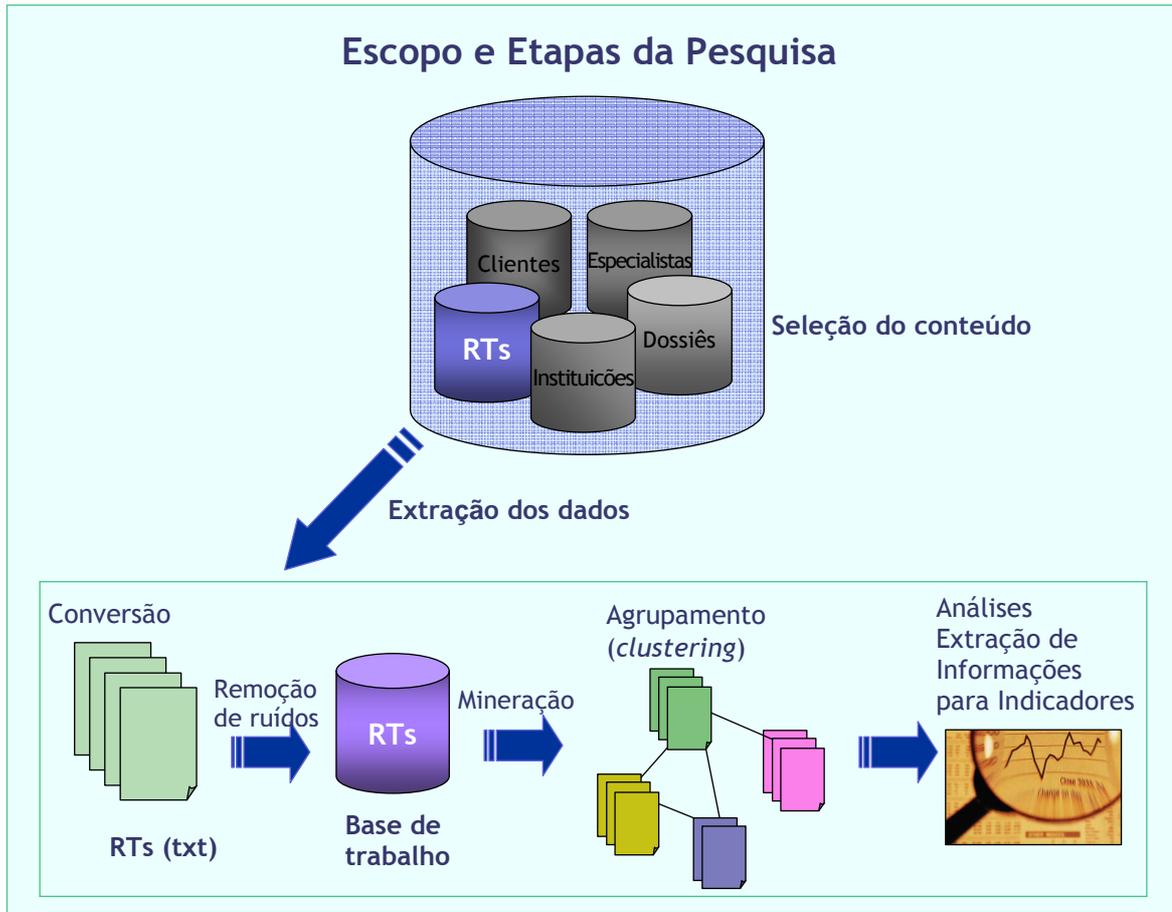


FIGURA 6 – Escopo e etapas da pesquisa

4.4 Seleção das ferramentas de trabalho e preparação do ambiente de pesquisa

Há várias ferramentas disponíveis no mercado para a realização de estudos por meio da descoberta de conhecimento em conteúdos armazenados em formato textual. Optou-se, preliminarmente, pelo uso de um indexador textual de documentos, o BR/Search, para o tratamento inicial dos dados e sua preparação para a pesquisa.

Diante da apresentação da proposta de pesquisa acadêmica, a empresa representante dessa ferramenta no Brasil, a Policentro, cedeu uma versão genérica do *software*, chamada BRS/Fácil.

Assim, foi providenciado, no Ibict, um equipamento servidor (Athlon XP 1800Mhz, 512 MB RAM HD 40 GB), com sistema operacional Linux, onde foi instalado o BRS/Fácil, para a manipulação e preparação dos dados para a pesquisa.

Ao longo da pesquisa, o acesso ao servidor foi realizado por meio de navegadores na *Web*, com controle de senha.

4.5 Extração dos dados para a pesquisa

Como o *software* selecionado no início do trabalho, o BR/Search, opera com unidades de parágrafos, decidiu-se pela adoção de um formato de extração de dados no qual cada documento a ser incluído na base de trabalho fosse composto de três parágrafos: i) campo de identificação do documento; ii) parágrafo contendo os metadados; iii) parágrafo de texto integral.

Originalmente, os dados extraídos do SBRT faziam parte de uma base de dados MySQL (metadados) e parte em arquivos PDF (textos completos) armazenados no servidor. Os procedimentos para sua extração foram realizados por meio de um backup da base MySQL, com o apoio da ferramenta phpMyAdmin, e posterior restauração desses dados em um PC contendo um ambiente de programação.

Em seguida, foi desenvolvida uma página PHP para “ler” os registros de Respostas Técnicas válidas da base MySQL e gerar um Shell Script, responsável pela conversão dos arquivos em PDF para o formato TXT, que é universal e se adapta perfeitamente ao exigido pelas ferramentas de mineração de dados. Finalmente, os dados foram mesclados em um único arquivo (BANCO_SBRT.txt).

Essas operações foram realizadas em servidor com sistema operacional Linux Ubuntu 7.04; servidor Web Apache versão 2.0, linguagem de Script PHP versão 5 e Shell Script, Servidor de Banco de Dados MySQL versão 5.0 e o conversor PDFTOTEXT.

Os dados em formato TXT foram gerados na estrutura de três "parágrafos", conforme definido anteriormente, sendo que o primeiro parágrafo contém o identificador do documento, o segundo os metadados relativos ao documento e o terceiro o texto do documento, conforme ilustrado a seguir:

..DOCN (Identificador do registro)

..METADATA (conjunto de metadados relativos às Respostas Técnicas)

..TXT (conteúdo da resposta técnica).

4.6 Pré-processamento utilizando o BR/Search

As técnicas de pré-processamento são utilizadas para minimizar as diferenças de vocabulário existentes nos documentos, a fim de maximizar os resultados corretos de uma busca de documentos. (WIVES, 2004)

Os documentos extraídos da base de dados de RTs do SBRT e devidamente convertidos para o formato TXT foi importado para o BR/Search, onde foi criada uma estrutura de parágrafos. Realizou-se a limpeza inicial para eliminação dos caracteres de controle que foram inseridos automaticamente quando de sua importação para o formato TXT.

Iniciou-se a primeira exploração do conteúdo textual, por meio da visualização dos documentos e de testes de busca e recuperação, com a finalidade de se identificarem relações entre conteúdos. Extraiu-se uma lista de todas as palavras contidas no texto da base, com indicação de sua ocorrência em toda a base e na quantidade de documentos em que aparecem. O Apêndice A apresenta uma amostra da lista ordenada pelas palavras de maior frequência na base, o que facilitou a visualização das palavras não significativas presentes no texto geral da base. Utilizou-se essa lista para análises preliminares. Essa mesma lista foi organizada em ordem alfabética (Apêndice B) para checagem de ocorrências de termos semelhantes e verificação de incorreções na digitação.

Após uma série de testes, verificou-se que o trabalho de mineração do texto com o BR/Search seria possível utilizando-se de recursos de busca com termos pré-definidos, caso a pesquisa estivesse voltada para o campo da indexação.

Nesse sentido, com base na natureza definida para a pesquisa, optou-se pela busca por uma ferramenta de mineração de textos que fosse dotada de algoritmos internos para a mineração em todo o conteúdo da base em busca de agrupamento dos textos com base nas semelhanças entre eles (*clustering*, em inglês). Encontrou-se na literatura uma diversidade de traduções para esse termo, como, por exemplo, conglomerados (WIVES, 2004) e agrupamentos (SCHIESSL, 2007).

Com base em experiência bem sucedida de pesquisa anterior realizada no CID/UnB, decidiu-se pelo programa SAS Text Miner. Procedeu-se, então, à negociação com o produtor desse programa a liberação de uma versão, em caráter temporário, exclusivamente para a realização da pesquisa.

4.7 Sobre o SAS

A aplicação da DCT foi realizada utilizando-se o pacote SAS *Data Mining Solutions*, composto de dois aplicativos – o SAS *Enterprise Miner* e o SAS *Text Miner for Portuguese* – desenvolvidos para revelar padrões e relações ocultos em dados objetivando contribuir com empresários e pesquisadores no entendimento de tendências históricas e na previsão de oportunidades futuras.

O SAS incorpora tecnologias de integração de dados para limpeza e organização de dados brutos e sistemas de inteligência que convertem em conhecimento as informações ocultas em textos não estruturados, proporcionando uma tomada de decisão pró-ativa.

Dessa forma, pode-se afirmar que o SAS é dotado de ampla aplicação, constituindo-se em uma poderosa ferramenta para analistas de dados e de mercado, investigadores de fraudes, gerentes de negócios, analistas de riscos, técnicos e cientistas, que exercem papéis estratégicos na identificação e solução de situações críticas, tanto na área de negócios quanto na de pesquisa.

4.8 Aplicação da DCT utilizando o SAS

Conforme amplamente discutido na literatura, para que se seja possível realizar qualquer tratamento automático de uma coleção de documentos escritos em linguagem natural em busca do conhecimento nela embutido, torna-se necessária a limpeza e padronização do texto. Sob essa ótica, Tan (1999) considera dois componentes estruturais da técnica de mineração de textos: o *refinamento do texto*, que transforma os documentos com textos não estruturados para o que ele chamou de “formato intermediário”, e a *destilação do conhecimento*, que deduz padrões ou conhecimento a partir desse formato intermediário.

A DCT é aplicada por meio de técnicas diversas, sendo as mais conhecidas: análise de conglomerados (*clustering*), classificação, extração de informações, sumarização, análise qualitativa e quantitativa e identificação de regras de associação. (WIVES, 2004)

Esta pesquisa concentrou-se na técnica de análise desses agrupamentos de documentos, as Respostas Técnicas do Sistema de Informação SBRT.

Sabe-se que o processo de DCT é “interativo e iterativo, correspondendo à aplicação repetida de métodos de mineração e interpretação dos resultados pelo usuário.” (WIVES, 2004) A experiência demonstra que se gasta muito tempo na remoção de ruídos com o intuito de padronizar os dados, de forma a possibilitar maior precisão e acurácia no processo de mineração.

A seguir, estão descritos algumas etapas, desde a criação da base de trabalho até a completa remoção de ruídos, necessárias à preparação dos dados para a aplicação da DCT.

4.8.1 A criação da base de trabalho

Os documentos em formato TXT foram migrados para o SAS, procedendo-se, assim, à criação do que se denominou “base de trabalho”. A partir de então, este passou a ser o objeto de manipulação e análise no âmbito da pesquisa. Os dados foram organizados em formato de tabela, de forma que cada Resposta Técnica pudesse ser identificada por um número seqüencial, conforme pode ser observado no exemplo exposto no Apêndice C. A Figura 7 apresenta uma tela contendo a visualização de três documentos da base de trabalho.

4.8.2 Preparação dos dados (pré-processamento, limpeza e transformação)

A seguir, são apresentadas as etapas da pesquisa relativas à preparação do conteúdo textual da base de trabalho, com a finalidade de possibilitar a aplicação da DCT.

4.8.2.1 Remoção de palavras não significativas

As palavras não significativas – as chamadas *stopwords*, em inglês – são palavras comuns, encontradas em grande quantidade em um arquivo textual e não carregam significados em si próprias. Em geral, as palavras não significativas pertencem às seguintes classes gramaticais: artigos, conjunções, preposições, pronomes e advérbios. Wives (2004) apresenta três outras possibilidades de

denominação do termo em língua portuguesa: “palavras negativas”, “palavras-ferramenta” ou “palavras-vazias”.

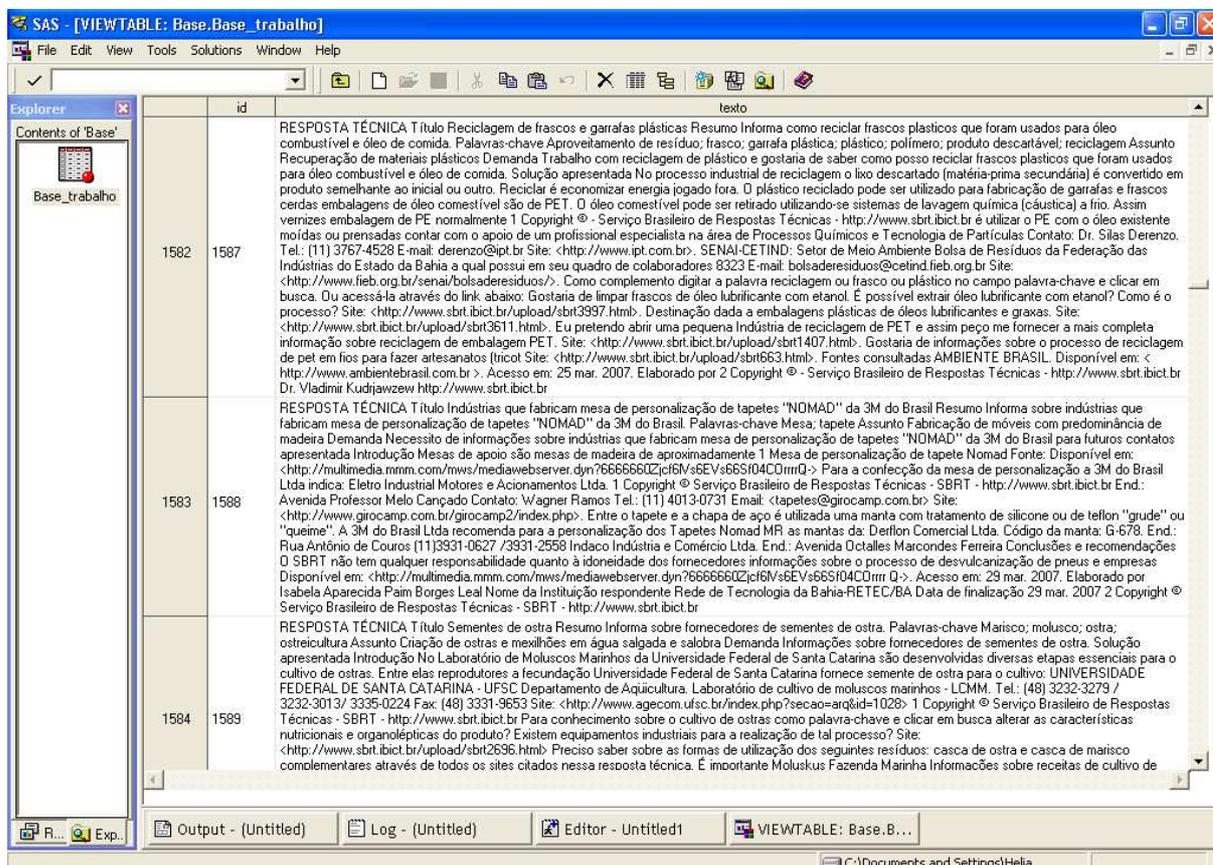


FIGURA 7 – Visualização do conteúdo da base de trabalho no SAS

Para uma primeira remoção das palavras não significativas da base de trabalho, foi aplicada uma lista básica com palavras dessas classes gramaticais e consoantes soltas localizadas ao longo do texto.

Em contraponto à lista de palavras não significativas, o programa SAS oferece uma lista de termos que caracterizam o domínio do assunto a ser pesquisado, os quais denomina *startwords*. Estão contidos nesta lista todos os termos que serão considerados na mineração. (SCHIESSL, 2007).

O Quadro 2 apresenta alguns exemplos das palavras não significativas encontradas nos documentos da base de trabalho. A lista completa de palavras não significativas extraídas da base de trabalho, composta por 416 termos, está disponível no Apêndice D.

QUADRO 2 – Lista de palavras não significativas removidas do texto

CLASSE GRAMATICAL	TERMO
Artigos	a, as, o, os, um, uma, umas, uns
Conjunções	e, mas, ou, que, quando, porque, ou, nem, não, logo, pois, como, quanto, embora, apesar, ainda, segundo, tal, tão, tanto, quando, depois, antes.
Consoantes	c, n, s, r
Preposições	a, à, ao, apos, ate, até, com, contra, da, de, do, dos, na, no, para, por, sem, sob, sobre, como, entre, per
Pronomes	a, algo, algumas, alguns, aquela, aquelas, aquele, aqueles, aquilo, as, cada, ela, elas, ele, eles, essa, essas, esse, esses, esta, estas, este, estes, eu, isso, isto, me, meu, meus, nada, nenhuma, nos, nossa, nossas, nosso, nossos, o, onde, os, outra, outras, outro, outros, próprio, quanto, que, quem, se, seu, seus, sua, tal, te, teu, teus, todo, tu, tua, tuas, tudo, um, uma, umas, uns, você, vocês, vos, vossa, vossas, vosso, vossos.
Vogais	a, e, o

4.8.2.2 Lematização¹⁰

Após a remoção das palavras não significativas, foi realizada outra etapa de preparação de dados para a mineração. Trata-se da lematização, ou extração de inflexões de termos, reduzindo-os a seus radicais, na intenção de se criarem padrões e proporcionar maior confiabilidade na mineração. Obviamente, há que se cuidar para que os radicais expressem um conceito comum. Essa tarefa é possível por meio da criação de um dicionário de termos com suas respectivas categorias gramaticais. O Anexo B traz uma amostra do dicionário de termos utilizado.

O exemplo esquematizado por Schiessl (2007) utilizando o termo *informação* e suas possíveis variações ilustra bem as dificuldades que podem estar presentes na lematização: *informações*, *informar*, *informado* e *informando* pode ser reduzido a uma única forma padrão que seria “inform”.

Segundo Schiessl (2007), a lematização é utilizada também para reduzir a quantidade de termos com a finalidade de facilitar a análise e reduzir o custo computacional, visto que restringe a quantidade de termos que serão processados.

¹⁰ Frequentemente escrito em Inglês – *Stemming*. Ressalta-se que o termo Lematização foi encontrado somente em dicionários de português europeu, como por exemplo: http://www.priberam.pt/dlpo/definir_resultados.aspx. (SCHISSL, 2007)

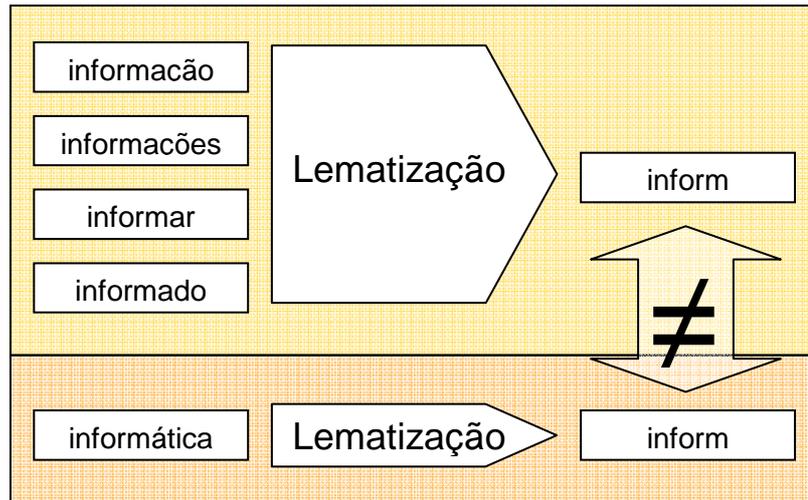


FIGURA 8 – Lematização de Termos

Fonte: Schiessl, 2007

Após as etapas de limpeza e padronização, os dados textuais estarão prontos para a aplicação da técnica de mineração.

4.8.3 A mineração do texto

Após as idas e vindas do processo de limpeza e padronização das palavras significativas do texto, foi iniciada a mineração propriamente dita. Utilizou-se, então o recurso do SAS para agrupamento dos textos com base nas semelhanças entre eles. Esse agrupamento é referenciado com frequência na literatura a partir do seu termo em inglês, *clustering*, ou “geração de *clusters*”.

4.8.3.1 Os primeiros resultados da mineração

A base de trabalho extraída do BR/Search em formato TXT, com separadores, foi exportada para o SAS, que gerou uma tabela com duas colunas: identificador (número seqüencial) e texto completo.

O Quadro 3 apresenta a variação da quantidade de termos nos 6041 documentos da base de trabalho.

QUADRO 3 – Média de termo por documento

Total de Documentos (RTs)	Média de Termos	Desvio Padrão	Mínimo de termos	Máximo de termos
6041	510,5	337,1	1,0*	3.005,0

* Documentos com menos de 10 termos: 302, 619, 954, 1150, 1416, 1808, 1832, 1912, 2576, 2688, 2774, 3048, 3244, 3407, 3535, 4398.

O desvio padrão de 337,1 aponta para uma diversidade muito acentuada no tamanho dos documentos. A média de termos em cada documento é de 510,5. Entretanto, há documentos com mais de 3.000 termos. Isso caracteriza a diversidade dos conteúdos das RTs, que apresentam soluções tecnológicas das mais simplificadas às mais elaboradas.

A partir do primeiro processamento realizado no SAS, em que foram extraídos os termos brutos constantes do *corpus* de texto, pôde-se verificar, de pronto, algumas inconsistências na base de dados, como a presença de 20 documentos com menos de 10 termos, sendo que um deles possui 6 termos e 19 um só termo. Mesmo antes de se proceder a análise desses documentos, optou-se por eliminá-los do sistema. A base de trabalho, portanto, foi reduzida em 20 documentos, permanecendo com 6021 documentos dos 6041 iniciais. Os documentos eliminados da base serão objeto de análise posterior.

4.8.3.2 Agrupamento dos documentos

O agrupamento de documentos constitui o grande diferencial da técnica de mineração de textos, visto que identifica associações entre documentos aparentemente sem nenhuma relação. Ou seja, são apresentadas possibilidades de extração de conhecimentos totalmente novos e imprevistos.

O agrupamento permite que novas classes sejam descobertas, já que consegue agrupar documentos mesmo que estes não pertençam a assuntos conhecidos. Isso porque não há essa necessidade de conhecimento prévio sobre os assuntos (ou os possíveis assuntos dos documentos). Os assuntos ou as classes dos objetos (documentos) sendo agrupados são descobertos após o agrupamento, em um processo de análise dos grupos obtidos. (WIVES, 1999)

CAPÍTULO 5

ANÁLISE DOS RESULTADOS

As etapas descritas no Capítulo 4 – Procedimentos Metodológicos foram realizadas com a utilização das ferramentas discriminadas e os resultados das análises estão detalhados neste Capítulo 5.

As ferramentas SAS aplicadas na mineração do texto da Base de Trabalho geraram as saídas compostas de três arquivos contendo o texto completo da RT, seu identificador numérico na base de trabalho e o agrupamento ao qual pertence. O primeiro desses arquivos, denominado “texto_com_cluster”, contém 2.408 documentos, que representam 40% da base. Este foi o arquivo utilizado para a realização da análise.

O segundo arquivo, “texto_com_cluster-validação”, composto de 1.806 documentos, representando 30% da base de trabalho, poderá ser utilizado para validação. Finalmente, o terceiro arquivo, o “texto_com_cluster-teste”, possui 1.807 documentos, 30% da base, e tem por finalidade propiciar testes de validação, se necessário. Juntos, esses três arquivos perfazem o total de 6.021 RTs válidas da base de trabalho. No decorrer da pesquisa, não se percebeu a necessidade de se realizarem testes de validação, tendo a pesquisa se concentrado na exploração do primeiro arquivo citado acima, contendo os 2.408 documentos.

5.1 Os primeiros resultados da mineração – erros detectados

Sabe-se que no processo de mineração textual, a etapa de preparação dos dados (limpeza e padronização) corresponde a 60% de todo o processo (Quoniam *et al.*, 2005). Segundo os autores, “a preparação é crucial para a qualidade final dos resultados”. Nesta pesquisa isso ficou bastante evidenciado nos diferentes processamentos que se fizeram necessários até que se chegasse aos textos considerados prontos para análise.

Após a limpeza inicial dos arquivos gerados no SAS, pela extração de palavras não significativas e lematização descritos nos procedimentos metodológicos do capítulo anterior, iniciou-se o processo de agrupamento dos

documentos a partir da frequência de termos na base de trabalho de Respostas Técnicas (RTs).

Com a finalidade de fixar melhor o aprendizado e registrar importantes passos na mineração para que se obtenha o melhor resultado, estão detalhados a seguir alguns dos problemas detectados e os procedimentos adotados.

5.1.1 Primeiro processamento

No primeiro processamento, quando já se tinha idéia de que os dados estariam prontos, foram identificados 14 agrupamentos. O Quadro 4 especifica os termos que o SAS utilizou como elementos de união entre esses agrupamentos, a representação percentual de sua presença na base, a quantidade de documentos em que esses termos aparecem e a sua variabilidade no agrupamento. Sobre este último item, vale lembrar que quanto menor for o coeficiente, maior será a precisão de sua representatividade. Os sinais de + que aparecem antes de determinados termos significam que esses termos foram lematizados e na verdade “representam” vários termos de mesmo radical.

Já na primeira observação do quadro, percebeu-se que o agrupamento 3 apresentou uma particularidade: os 46 documentos foram agrupados por caracteres, em vez de palavras (p, t, s, ca, r). Não foi difícil identificar de onde teriam sido extraídos tais caracteres. Ao analisar o conteúdo textual desse agrupamento, identificou-se que em todos os documentos, o termo “Resposta Técnica” que aparece no início de todos os documentos textuais da base de trabalho estava grafado com espaços entre algumas letras: “R E S P O S T A T É C N I C A”. Assim, o *software* identificou algumas dessas letras como palavras independentes e agrupou tais documentos.

Está claro que esse tipo de erro comprometeria sensivelmente a qualidade da análise do conteúdo, visto que os 46 documentos que compõem o agrupamento são documentos válidos que foram agrupados sem coerência e, portanto, ficaram de fora do agrupamento com termos reais.

Erros dessa natureza justificam a necessidade de reprocessamento até que se produzam agrupamentos de fato significativos, daí a iteratividade típica da operação de mineração de textos.

QUADRO 4 – Lista de agrupamentos e termos – primeiro processamento

Agrupamento	Presença na Base	Quantidade Documentos	Variabilidade dos termos no agrupamento	Termos de Agrupamento
1	4%	96	0,1009710982	+ óleo, + usar, + podar, + processar, + apresentar
2	7%	178	0,1121993772	+ material, + usar, + utilizar, + processar, + podar
3	2%	46	0,1040250544	p, t, s, ca, r
4	14%	328	0,1191981235	+ alimentar, + processar, + produto, + dever, + podar
5	5%	117	0,0973484234	+ químico, + produto, + usar, + palavra-chave, + técnico
6	7%	176	0,0954315898	+ solar, + cultivar, + plantar, + dever, + podar
7	9%	216	0,1098847201	+ utilizar, + técnico, + parir, + datar, + palavra-chave
8	6%	133	0,099463168	responsável, + dever, + informação, + produto, + datar
9	4%	94	0,0926823476	+ elétrico, + equipamento, + fonte, + dever, + utilizar
10	7%	169	0,0976685313	+ produto, + alimentar, + sanitário, + dever, + técnico
11	9%	226	0,0978484272	+ fornecedor, + indústria, + equipamento, + técnico, + máquina
12	5%	125	0,0984701312	+ alimentar, + animal, + dever, + apresentar, + podar
13	14%	329	0,109850216	+ material, + processar, + utilizar, + usar, + podar
14	7%	175	0,1101032352	+ químico, + usar, + utilizar, + dever, + podar
Total de documentos:		2.408		

5.1.2 Segundo Processamento

Ao se realizar o segundo processamento, eliminando-se o agrupamento 3, observou-se que não só a quantidade de agrupamentos havia mudado, mas também alguns dos termos captados como importantes nos agrupamentos anteriores. Assim, ao se reprocessarem os dados pela segunda vez, foram gerados 10 novos agrupamentos, conforme detalhado no Quadro 5.

QUADRO 5 – Lista de agrupamentos e termos – segundo processamento

Agrupamento	Peso (Presença na Base)	Frequência (Quantidade Documentos)	Variabilidade dos termos no agrupamento	Termos de Agrupamento
1	19 %	451	0,11489	+ produto, + alimentar, + dever, + processar, + podar
2	8%	195	0,11424	+ resíduo, + processar, + fonte, + utilizar, + usar
3	16%	390	0,11159	+ utilizar, + podar, + usar, + apresentar, + parir
4	5%	127	0,09193	end, bahia, + fornecedor, site, + informação
5	10%	249	0,11720	+ produto, + químico, + usar, + técnico, + utilizar
6	13%	301	0,11492	+ cultivar, + plantar, + podar, + dever, + usar
7	10%	242	0,11053	+ utilizar, + técnico, + usar, + parir, + datar
8	5%	127	0,09852	responsável, + dever, + informação, + produto, + demandar
9	5%	120	0,09712	+ alimentar, + animal, + dever, + apresentar, + podar
10	8%	199	0,10388	+ equipamento, + fornecedor, + técnico, + datar, + palavra-chave
Total de documentos:		2.401		

Observe-se que a quantidade de documentos diminuiu de 2.408 para 2.401, o que significa a eliminação de 7 documentos que estavam presentes no primeiro processamento. Realizou-se, assim, uma análise comparativa dos conteúdos em busca dos documentos eliminados. Após a identificação desses documentos, verificou-se que os seus conteúdos textuais eram compostos de caracteres estranhos, não sendo possível recuperar nenhuma informação que pudesse auxiliar na busca desses documentos na base *online*. Assim, considerou-se pertinente a sua eliminação pelo SAS e o *corpus* da análise ficou fixado em 2.401 documentos.

Passou-se à análise detalhada dos termos, para se verificar a qualidade dos novos agrupamentos gerados. Observou-se, assim, a forte presença de alguns termos e passou-se a investigar uma planilha, gerada pelo SAS, contendo todos os termos da base – 43.271 – e suas respectivas lematizações. Confirmou-se, assim, que alguns termos haviam sido lematizados de forma equivocada, onde alguns termos representavam palavras de diferentes categorias. Além disso, alguns dos termos utilizados no agrupamento estariam invalidados, dada a sua frequência em praticamente todos os documentos.

O Quadro 6 apresenta um detalhamento desses erros de lematização, assim como os termos inadequados, com a respectiva explanação sobre a necessidade de eliminá-los e realizar novo processamento.

QUADRO 6 – Lista de erros de lematização

Agrupamentos	Termos representantes	Termos representados	Por que há necessidade de reprocessamento
1, 3, 6, 9	+ podar	<p> poda podar podadas podas podado pode podados pode-se podá-la podem podam podem-se podar podemos podam </p>	O infinitivo “podar” não diz respeito somente ao verbo podar, mas “poder” = corrigir essa relação.
2	+ fonte	fontes consultadas fonte:	Termos comuns a todos os documentos textuais da base = extrair Manter as demais formas do termo
3, 7	+ parir	para	Palavra não significativa = extrair
4	end	end end.: end:	Termos comuns na maioria dos documentos textuais da base = extrair
	bahia	Aparece em nomes próprios, exemplo: Rede de Tecnologia da Bahia - RETEC/BA	Nome de uma das instituições do SBRT, constante em todas as RTs respondidas por ela = extrair
	site	site sites	Termos comuns a todos os documentos textuais da base = extrair
7, 10	+ datar	data data de finalização	Termos comuns a todos os documentos textuais da base = extrair
		date	Termo presente em endereço de <i>site</i> referenciado = extrair
	+ palavra-chave	palavra-chave palavras-chave	Termos comuns a todos os documentos textuais da base = extrair

Cabe aqui fazer uma observação sobre a importância da lematização na língua portuguesa. Como observado em pesquisa de Bräscher (1999), com a lematização obtêm-se todos os contextos onde o lema foi empregado no *corpus*, independentemente de sua forma, o que enriquece as análises lingüísticas que são objeto da pesquisa e evita a dispersão das freqüências. No entanto, a mesma autora alerta para o fato de a homografia provocar erros de lematização, uma vez que o sistema não tem como determinar, *a priori*, o lema correto para uma forma homógrafa, segundo seu emprego no *corpus*. Os exemplos do Quadro 6 ilustram esse problema e ressaltam a necessidade da análise humana para solucioná-lo.

O *software* SAS utilizado já possui uma adaptação para a língua portuguesa, o que facilitou o processo de lematização. Ainda assim, erros como os descritos não puderam ser evitados. Melhora sensível nos resultados dos agrupamentos poderá ser obtida no futuro se o *software* incorporar técnicas de processamento automático da linguagem natural para o tratamento de homografias.

Após as correções dos erros detectados, realizou-se o terceiro processamento, que deu origem a 12 agrupamentos, os quais foram considerados representativos da realidade da base de trabalho e adequados para a realização das análises subseqüentes. Esses agrupamentos estão descritos com detalhes no item 5.2 a seguir.

5.2 Terceiro processamento – objeto de análise final

A partir deste ponto, passa-se a considerar apenas os agrupamentos gerados pelo terceiro processamento como objeto das análises. O Quadro 7 apresenta o detalhamento desses agrupamentos, isto é, os termos que provocaram a união dos documentos que os compõem, assim como a representatividade percentual de cada um deles, a quantidade de documentos em que os termos aparecem e a variabilidade desses termos no agrupamento.

Vale observar que quanto menor for o índice de variabilidade dos termos no agrupamento, maior é a sua precisão na representatividade e coesão de conteúdo. Identificou-se, portanto, o agrupamento de número 10 como sendo o mais coeso de todos, por ter apresentado o menor coeficiente (0,0876160473). Some-se a este fato o de que os termos que formam este agrupamento (material, + reciclar, + resíduo, + plástico, + processar) sugerem que o seu tema central está voltado para a preocupação com o meio ambiente, estímulo interessante para o objetivo geral desta pesquisa. Assim, o agrupamento 10 foi considerado o ideal para aprofundamento das análises em busca de comprovação da questão central da pesquisa: a possibilidade de uso da mineração de textos para extração de informações para apoiar a construção de indicadores de Ciência e Tecnologia. O agrupamento 10 será objeto de análise detalhada no item 5.3.

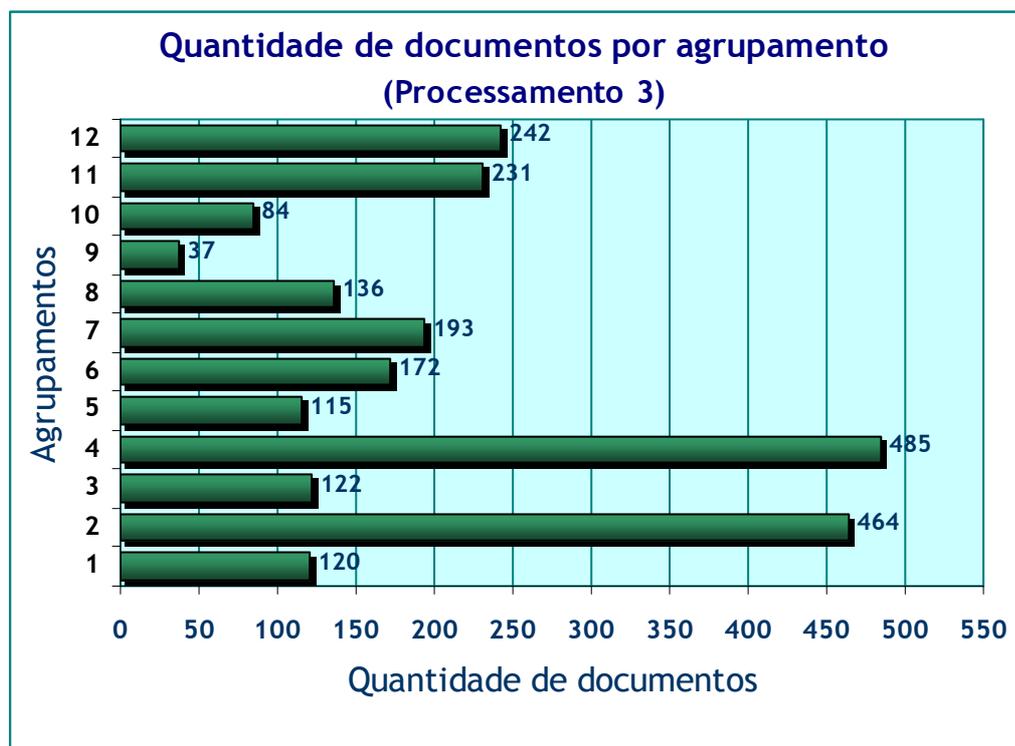
QUADRO 7 – Lista de agrupamentos e termos – terceiro processamento

Agrupamento	Peso (Presença na Base)	Frequência (Quantidade Documentos)	Variabilidade dos termos no agrupamento	Termos de Agrupamento
1	5%	120	0,1032064562	+ óleo, + químico, + usar, + processar, + utilizar
2	19%	464	0,1139042878	+ produto, + alimentar, + processar, + dever, + comer
3	5%	122	0,1075986813	+ químico, + produto, + produzir, + técnico, + resumir
4	20%	485	0,1144369151	+ material, + utilizar, + fonte, + usar, + processar
5	5%	115	0,1017080841	+ animal, + alimentar, + agricultura, + dever, + produção
6	7%	172	0,0941117077	+ cultivar, + solar, + plantar, + dever, + apresentar
7	8%	193	0,1034777388	+ máquina, + fornecedor, + utilizar, + elaborar, + usar
8	6%	136	0,099722373	+ identificação, responsável, + necessário, + dever, + informação
9	2%	37	0,093348348	+ espécie, + criação, + animal, + alimentar, + formar
10	3%	84	0,0876160473	+ material, + reciclar, + resíduo, + plástico, + processar
11	10%	231	0,0976318382	fax, + fornecedor, + equipamento, + máquina, + indústria
12	10%	242	0,1193337586	+ químico, + utilizar, + usar, + resumir, + processar
Total de documentos:		2.401		

Antes disso, far-se-á, no item 5.2.1, uma breve análise dos termos determinantes dos 12 agrupamentos gerados, e, no item 5.2.2, ainda à luz desses termos, uma análise comparativa entre alguns agrupamentos, que trazem termos comuns entre si. Busca-se, com isso, demonstrar as especificidades encontradas na aplicação da DCT no *corpus* textual que não seriam encontradas a “olho nu”.

O Gráfico 1 ilustra a visualização da distribuição dos documentos nos agrupamentos. Observa-se que não há equilíbrio com relação à quantidade de documentos dos agrupamentos, estando a concentração maior entre os agrupamentos 2 e 4 (464 e 485, respectivamente), enquanto que o agrupamento 9 possui apenas 37 documentos. Seria necessário fazer uma análise aprofundada dos conteúdos de todos os agrupamentos, para a identificação precisa das tendências da base como um todo. Entretanto, dada à indisponibilidade de tempo para tal, serão feitas as análises parciais citadas no parágrafo anterior.

GRÁFICO 1 – Agrupamento de documentos da base de trabalho



5.2.1 Breve análise dos termos determinantes dos agrupamentos

Logo na primeira leitura dos termos apresentados no Quadro 7, pôde-se observar a presença da natureza tecnológica do conteúdo da base de dados SBRT, a partir dos verbos característicos de orientações para se realizar a aplicação do conhecimento contido na solução tecnológica fornecida, como por exemplo: alimentar, comer, cultivar, dever, elaborar, formar, plantar, processar, produzir, reciclar, resumir, usar, utilizar.

Da mesma forma, é possível inferir os principais temas de que trata a base, por meio da leitura dos substantivos mais freqüentes na lista de termos representativos dos agrupamentos: agricultura, animal, criação, equipamento, espécie, máquina, óleo, plástico, produção, produto, químico, resíduo, técnico.

É possível perceber, também, uma correlação entre temas e ações, quando se analisa a presença de termos comuns em diferentes agrupamentos (Quadro 8).

QUADRO 8 – Lista de termos comuns aos agrupamentos

Termos	Agrupamentos		Termos	Agrupamentos
agricultura	5		material	4, 10
alimentar	2, 5, 9		necessário,	8
animal	5, 9		óleo	1
apresentar	6		plantar	6
comer	2		plástico	10
criação	9		processar	1, 2, 4, 10, 12
cultivar	6		produção	5
dever	2, 5, 6, 8		produto	2, 3
elaborar	7		produzir	3,
equipamento	11		químico	1, 3, 12
espécie	9		reciclar	10
fax	11		resíduo	10
fonte	4		responsável	8
formar	9		resumir	3, 12
fornecedor	7, 11		solar	6
identificação	8		técnico	3
indústria	11		usar	1, 4, 7, 12
informação	8		utilizar	1, 4, 7, 12
máquina	7, 11			

Há forte presença de alguns termos em vários agrupamentos, como ocorre com os termos “processar” (agrupamentos 1, 2, 4, 10, 12), “dever” (2, 5, 6, 8), “alimentar” (2, 5, 9) e “químico” (1, 3, 12). Os termos “utilizar” e “usar” estão presentes exatamente nos mesmos agrupamentos (1, 4, 7, 12), fica claro, portanto, que estão sendo utilizados como sinônimos. Isso pode ser caracterizado como uma impropriedade, que poderia ter sido evitada com a criação de um dicionário de sinônimos¹¹.

Essas repetições de termos nos agrupamentos provocam a necessidade de se olhar mais de perto cada um desses agrupamentos, em busca de sinais que comprovem a pertinência dos documentos estarem juntos.

5.2.2 Análise comparativa entre agrupamentos

A seguir, são apresentadas algumas reflexões em torno dos agrupamentos 1, 3 e 12, que se destacaram devido à presença de um termo comum: “químico”. Objetiva-se com isso investigar as peculiaridades de cada agrupamento que teriam

¹¹ Para maior detalhamento sobre este recurso da DCT, ver Schiessl (2007)

provocado o “distanciamento” entre os documentos, a despeito da similaridade entre seus termos.

5.2.2.1 Os agrupamentos 1, 3 e 12 e o termo comum entre eles: “químico”

QUADRO 9 – Dados específicos dos agrupamentos 1, 3 e 12

Agrupamento	Freqüência (Quantidade Documentos)	Termos de Agrupamento
1	120	+ óleo, + químico, + usar, + processar, + utilizar
3	122	+ químico, + produto, + produzir, + técnico, + resumir
12	242	+ químico, + utilizar, + usar, + resumir, + processar

Observa-se, à primeira vista, uma forte intersecção entre os agrupamentos 1 e 12, a julgar pela presença de 4 termos comuns dos 5 que os caracterizam (químico, usar, processar, utilizar). Isso provocou a necessidade de um aprofundamento na análise desses, adicionando-se a esta o agrupamento 3, que apresentou a presença marcante de um termo comum a esses dois: “químico”.

A Figura 9 busca representar, em forma de esquema, as interações existentes entre os três agrupamentos, a partir dos termos comuns. A seguinte convenção foi adotada, para maior facilidade de visualização:

- i) os termos que trazem a mesma cor do nome do agrupamento aos quais estão conectados são aqueles que ocorrem exclusivamente naquele agrupamento;
- ii) os termos em preto são aqueles comuns a dois agrupamentos;
- iii) destaque maior é dado ao termo “químico”, por ocorrer nos três agrupamentos.

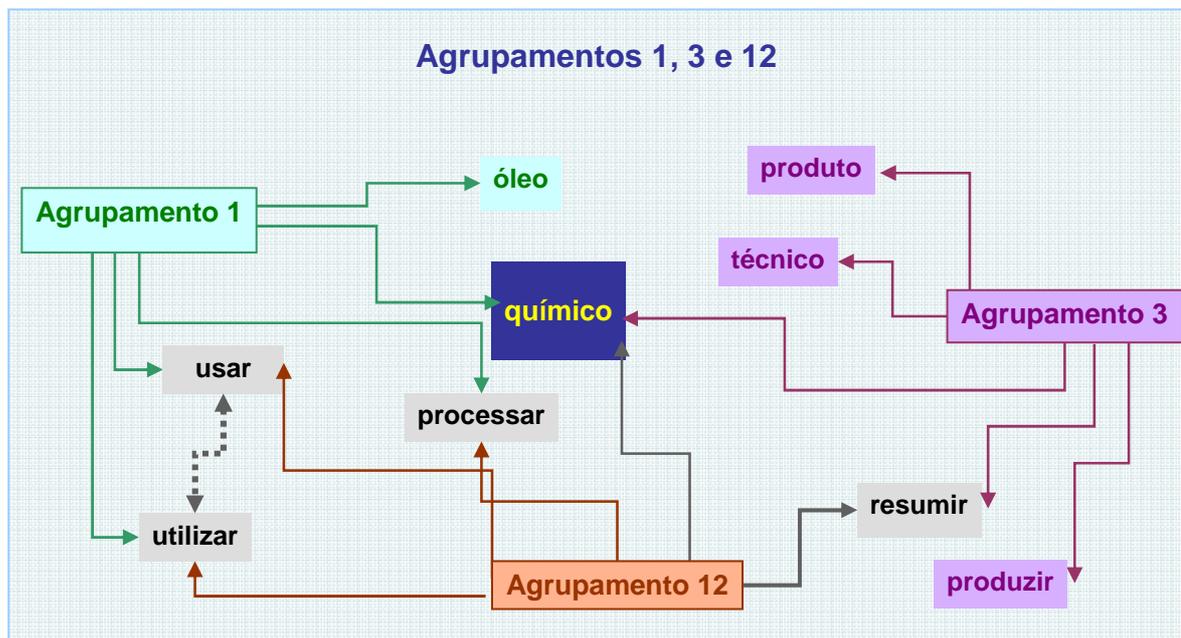


FIGURA 9 – Os agrupamentos e suas interações

Para que se tivesse uma visão mais aprofundada da mineração que pudesse propiciar o esclarecimento do fenômeno presente na Figura 9, foi criada a Tabela 10 com os termos relevantes extraídos dos três agrupamentos.

QUADRO 10 – Análise comparativa: agrupamentos 1, 3 e 12

	Agrupamento 1	Agrupamento 3	Agrupamento 12
Termos	óleo, químico, usar, processar, utilizar	químico, produto, produzir, técnico, resumir	químico, utilizar, usar, resumir, processar
Assuntos – Genéricos	<ul style="list-style-type: none"> • Produtos químicos 22 • Agricultura e Pecuária 12 • Serviços Industriais 9 • Alimentos e bebidas 4 • Meio ambiente 4 	<ul style="list-style-type: none"> • Produtos químicos 51 • Alimentos e bebidas 7 • Agricultura e Pecuária 6 • Meio ambiente 5 • Serviços industriais 3 	<ul style="list-style-type: none"> • Produtos químicos 29 • Meio ambiente 17 • Agricultura e Pecuária 13 • Couro e calçados 13 • Serviços industriais 10 • Metal 9 (Metalurgia 2) • Serviços de usinagem 6 • Alimentos e bebidas 4 • Pecuária 1 • Plásticos e Borrachas 1 • Mobiliário 1
Assuntos – Específicos	<ul style="list-style-type: none"> • Fabricação de óleos vegetais em bruto • Fabricação de aditivos de uso industrial • Geração de energia elétrica • Fabricação de catalisadores • Fabricação de óleos vegetais refinados • Fabricação de cosméticos • Fabricação de óleo essencial de limão • Fabricação de óleos essenciais vegetais • Fabricação de óleos lubrificantes básicos • Cultivo de mudas em viveiros florestais • Fabricação de conservas de frutas • Fabricação de intermediários para detergentes e tensoativos • Cultivo de oleaginosas de lavoura temporária • Produção de semente de linho • Criação de peixes • Fabricação de munição para armas de fogo • Fabricação de outros produtos derivados do petróleo 	<ul style="list-style-type: none"> • Fabricação de cosméticos • Fabricação de cosméticos, produtos de perfumaria e de higiene pessoal • Fabricação de produtos de limpeza e polimento • Equipamentos de medida • Fabricação de sabões e detergentes sintéticos • Fabricação de Produtos químicos orgânicos não especificados anteriormente • Fabricação produtos químicos orgânicos • Fabricação produtos químicos inorgânicos • Fabricação de tintas vernizes esmaltes e lacas • Fabricação de medicamentos para uso veterinário • Fabricação de sabonetes nas formas líquida ou em barras • Fabricação de ácido benzóico • Produtos químicos • Fabricação de aparelhos elétricos para espantar insetos • Resinas e elastômeros • Leite de vaca • Minerais não-metálicos • Material eletrônico e aparelhos e equipamento de comunicação • Brinquedos e jogos recreativos • Equipamentos de medida 	<ul style="list-style-type: none"> • Fabricação de tintas • Captação • Fabricação de produtos diversos • Usinas de compostagem • Edição • Impressão de material para outros usos • Imunização e controle de pragas urbanas • Reparação de artigos do mobiliário • Eletricidade • Fabricação de cal e gesso • Fabricação de águas envasadas • Processamento de gás natural • Produção de gás • Abate e fabricação de produtos de carne • Preparação de couros • Beneficiamento de peixe • Beneficiamento da castanha de caju • Reciclagem de cartucho de toner • Recobrimento de Superfície • Serviços de lavagem e polimento de veículos automotores • Fabricação de papel • Fabricação de resinas termoplásticas • Beneficiamento de café não associado ao cultivo • Ensaio e inspeções de materiais e de produtos para fins de cumprimento de normas legais • Serviço de anodização • Fabricação de partes para calçados • Frigorífico • Descontaminação e outros serviços de gestão de resíduos • Fabricação de pigmentos e corantes • Lavanderias

			<ul style="list-style-type: none"> • Fabricação de artigos de vidro • Couros e peles de animais silvestres e domésticos • Fabricação de adubos e fertilizantes • Fabricação de álcool • Tratamento e disposição de resíduos perigosos • Fabricação de produtos de madeira • Usinas de compostagem • Testes e análises técnicas • Medição da poluição • Obtenção de gás a partir da decomposição biológica de matéria orgânica • Fabricação de canetas • Produção de esterco de bovinos • Fabricação de defensivos agrícolas
Produtos/ Aplicações	<p>Perfumes – essências, fragrância, fixadores, componentes,</p> <p>Fabricação de óleos essenciais vegetais em bruto ou refinados (abacate, amendoim, aroeira, artemísia, babaçu, buriti, castanha-do-pará, cidreira, citronela, copaíba, cravo-da-índia, eucalipto, mamona, maracujá, manjerição, melissa, palmeira, pau-rosa, pequi, pinhão, semente de uva, soja).</p> <p>Fabricação de óleos – essenciais, lubrificantes, óleos aromáticos, corantes, óleos e essências de uso culinário, glicerina vegetal, repelentes.</p> <p>Higiene e limpeza – sabonetes (abacate, glicerina, artesanais, aromaterapêuticos), óleos minerais, sabão, produtos de limpeza.</p> <p>Geração de energia elétrica alternativa (babaçu, gesso, biomassa, lenha).</p> <p>Produção de biodiesel a partir de oleaginosas.</p>	<p>Cosméticos e higiene pessoal – shampoos, sabonetes, creme dental, hidratantes, ceras e linhas depilatórias, gel dental, creme capilar, alisante para cabelo, talco para os pés,</p> <p>Artigos de limpeza e aromatizantes – sabão, detergente, limpa-alumínio, aromatizantes, shampoos automotivos, sachês para carros, silicones para limpeza de automotivos, desengraxante, solução desinfetante fungicida, detergente, fumaça artificial, limpa-pedras.</p> <p>Produtos veterinários – shampoos, sabonetes antipulga, ração animal.</p>	<p>Vidro, alumínio, peles, gás natural, plásticos, borrachas ossos bovinos.</p> <p>Impermeabilizantes, tintas, resinas, húmus, cal, produtos de metal, gesso. metais não-ferrosos, vidro cristal, alumínio, zinco.</p> <p>Água – gaseificação, alcalinização, contaminação e análise, captação, coleta, desinfecção, dessalinização, tratamento, filtragem, reutilização, eutrofização, osmose reversa, uso água residuária, reuso)</p> <p>Couro – tratamento, curtimento, beneficiamento, avaliação, oxidação, salga, secagem, rasgamento progressivo, curtume, raspa hidrofugada.) couro de tilápia, búfalo, equino, rã.</p> <p>Tintas – para cabelo, látex, carimbos)</p> <p>Testes e análises técnicas; Impressão de material; Imunização e controle de pragas urbanas</p> <p>Tratamento – água, metais, madeira, couro, osso, resíduos de galvanoplastia, efluentes, resíduos de serviço de saúde, resíduos de beneficiamento de café, químico, biológico, metais, resíduos perigosos e não perigosos, dejetos industriais.</p>

Com a finalidade de facilitar a leitura do quadro 9, os assuntos de indexação foram organizados pelo termo mais genérico, colocando-se ao lado da classificação a quantidade de ocorrências, em ordem decrescente, no intuito de se identificarem os temas em destaque.

Em seguida, estão relacionados os demais termos também utilizados para a indexação de assunto, mais específicos, sem a preocupação com a quantificação, mas também organizados em ordem decrescente de ocorrência na base. Não se incluem aqui todos os assuntos específicos abordados, dada a pulverização de temas. Buscou-se registrar aqueles com maior ocorrência nos documentos do agrupamento. Há uma forte presença de classificações únicas, ou seja, com apenas uma ocorrência, entretanto, algumas delas apresentam marcante semelhança com o conteúdo geral do agrupamento.

Ao se analisarem atentamente os termos dos três agrupamentos em discussão, foi possível observar que, embora todos estejam relatando ações que empregam processos químicos, há claros indícios de elementos que os diferenciam.

Ao contrário da primeira impressão citada anteriormente, quando se inferiu haver forte relação entre os agrupamentos 1 e 12, dada a intersecção de 4 termos em 5, a conclusão que se chega após olhar mais de perto os demais termos extraídos é exatamente oposta. A semelhança maior se mostrou entre os agrupamentos 1 e 3, embora houvesse apenas um único termo comum entre eles, principalmente do ponto de vista das finalidades do uso de processos químicos.

Ambos os agrupamentos 1 e 3 detêm seu maior conteúdo voltado para a fabricação de produtos cosméticos, material de limpeza e aromatizantes em geral, e aditivos de uso industrial. Entretanto, nota-se a presença marcante, no agrupamento 1, de produtos derivados de óleos, extraídos de produtos naturais, de perfumes a lubrificantes, abarcando também o cultivo de plantas oleaginosas e produção de sementes, óleos e essências de uso culinário. Ainda, destaca-se na produção de energia e combustíveis alternativos. Esses são seus principais diferenciais em relação ao agrupamento 3. Já este último apresenta uma diversidade de produtos cosméticos e de higiene pessoal, entretanto, destaca-se na linha de produtos de limpeza e aromatizantes, como desengraxantes, solução

desinfetante fungicida, detergente, fumaça artificial, limpa-pedras, entre outros, além de produtos de uso veterinário.

Já o agrupamento 12 apresenta aplicações químicas de naturezas diferentes dos outros dois agrupamentos, com conteúdo mais voltado para a indústria de metalurgia, usinagem, couro e calçados, trazendo abordagens específicas, tais como tratamentos de água, pinturas, fundição, edição, gravação, impressão, estamperia, impermeabilização, defensivos agrícolas, imunização e controle de pragas.

Observa-se, nos três agrupamentos analisados, a marcante presença do termo “meio ambiente”, seja como foco central, como assunto indexado, seja em segundo plano, como uma preocupação com os impactos causados pela indústria de forma geral. Verificou-se, ainda, que esse termo está presente em 14% de todo o conteúdo da base de Respostas Técnicas e que é tema central de um dos agrupamentos, o de número 10, que será analisado no item 5.3.

É importante ressaltar que essa porcentagem diz respeito apenas ao termo “meio ambiente”, não tendo sido feito aprofundamento para a detecção de outros termos específicos que estão presentes na base de RTs e que tratam de questões ambientais muitas vezes sem carregar consigo a classificação “meio ambiente”. Citam-se alguns exemplos dessa diversidade de palavras-chave relacionadas ao tema: impacto ambiental, preservação ambiental, tratamento ambiental, gestão ambiental, controle ambiental, licenciamento ambiental, licença ambiental, produção mais limpa, poluição, seqüestro de carbono, alimento orgânico, agrotóxico, agricultura orgânica, agroecologia, reciclagem, aproveitamento de resíduos, recuperação do solo, tratamento de efluentes, produto biodegradável, reflorestamento; efeito estufa, resíduo sólido, aproveitamento de resíduo, transporte de resíduo, ecologia, resíduo hospitalar, lixo hospitalar, autoclave, coleta seletiva, descontaminação, tratamento de resíduo, biodigestor, biogás, efluente orgânico, lixo cultivado protegido, tinta ecológica.

Obviamente, a análise seria mais precisa e valiosa se se pudesse recorrer a profissionais técnicos do SBRT para sua validação e discussões mais apuradas. Entretanto, acredita-se que as discussões levantadas serviram para comprovar a eficácia da DCT para a geração de agrupamentos de documentos textuais.

5.3 Agrupamento 10 – em busca de informações para indicadores

O agrupamento 10 foi selecionado para o aprofundamento das análises em busca de informações para apoiar a construção de indicadores, conforme explanado no item 5.2. Esse agrupamento foi analisado sob dois aspectos, conforme detalhamento dos itens 5.3.1 e 5.3.2 a seguir.

5.3.1 Classificação dos documentos

A primeira análise que se fez no Agrupamento 10 foi a de assunto, campo da base que apresenta a classificação da Resposta Técnica (RT), de acordo com uma tabela de assuntos adaptada da Tabela CNAE.

Os 84 documentos do agrupamento trazem uma diversidade de assuntos característica da base de RTs. Mesmo estando agrupados, dada a sua inter-relação em torno de um tema central, esses documentos obtiveram 44 diferentes classificações, 34 das quais ocorrendo apenas uma vez, conforme detalhado no Quadro 11.

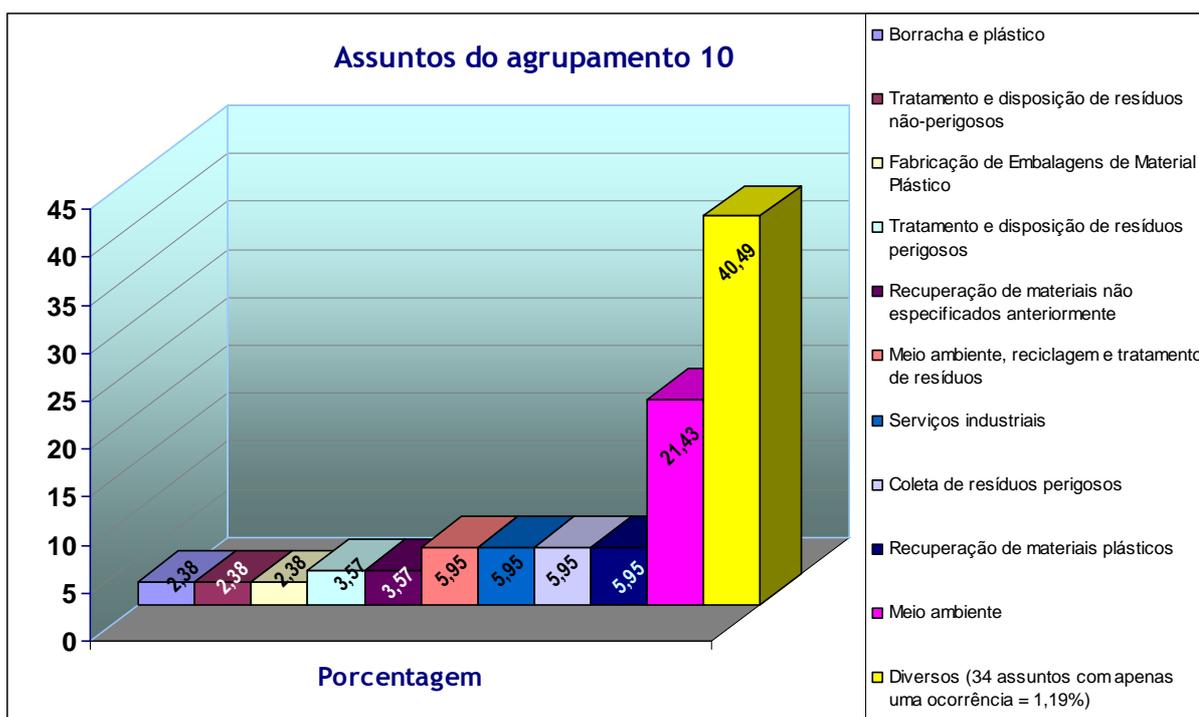
À primeira leitura dos assuntos, já se pode observar a coerência entre os documentos, que, de maneira geral, sugerem tratar dos assuntos mais diversos, contudo, com uma visão central: o de tratamento e/ou reaproveitamento de materiais de diferentes naturezas para fins diversos.

QUADRO 11 – Lista geral de assuntos do agrupamento 10

No.	Assunto	Ocorrências	Porcentagem %
1	Alimentos e bebidas	1	1,19
2	Borracha e plástico	2	2,38
3	Brinquedos e jogos recreativos	1	1,19
4	Cerâmica	1	1,19
5	Coleta de entulhos e refugos de obras e demolições	1	1,19
6	Coleta de resíduos não-perigosos	1	1,19
7	Coleta de resíduos perigosos (2)	5	5,95
8	Descontaminação e outros serviços de gestão de resíduos	1	1,19
9	Extração de minério de níquel	1	1,19
10	Fabrição de brinquedos e jogos recreativos	1	1,19
11	Fabricação de aditivos de uso industrial	1	1,19
12	Fabricação de artefatos de material plástico para outros usos não especificados anteriormente	1	1,19
13	Fabricação de bolsas de plástico	1	1,19
14	Fabricação de Embalagens de Material Plástico	2	2,38
15	Fabricação de móveis de outros materiais	1	1,19
16	Fabricação de peças e acessórios para o sistema motor de veículos automotores	1	1,19
17	Fabricação de produtos de material plástico	1	1,19
18	Fabricação de produtos químicos orgânicos não especificados anteriormente	1	1,19
19	Fabricação de resinas de poliuretano	1	1,19
20	Fabricação de resinas termoplásticas	1	1,19
21	Fabricação de sacolas de material plástico	1	1,19
22	Fabricação de tecidos especiais	1	1,19
23	Fabricação de vassouras	1	1,19
24	Medição da poluição	1	1,19
25	Meio ambiente	18	21,43
26	Meio ambiente, reciclagem e tratamento de resíduos	5	5,95
27	Minerais não metálicos	1	1,19
28	Mobiliário	1	1,19
29	Reciclagem	1	1,19
30	Reciclagem de sucatas não-metálicas	1	1,19
31	Recuperação de materiais metálicos produto reciclado.	1	1,19
32	Recuperação de materiais não especificados anteriormente	3	3,57
33	Recuperação de materiais plásticos	5	5,95
34	Recuperação de resíduos contendo produtos químicos	1	1,19
35	Recuperação de sucatas de alumínio	1	1,19
36	Reforma de pneumáticos usados	1	1,19
37	Reparação e manutenção de computadores e de equipamentos periféricos	1	1,19
38	Seleção latas de alumínio usadas	1	1,19
39	Serviço de coleta acondicionamento e transporte de lixo hospitalar	1	1,19
40	Serviços industriais	5	5,95
41	Tecelagem de fios de fibras artificiais e sintéticas	1	1,19
42	Tratamento e disposição de resíduos	1	1,19
43	Tratamento e disposição de resíduos não-perigosos	2	2,38
44	Tratamento e disposição de resíduos perigosos	3	3,57
Total		84	99,97%

O Gráfico 2 sintetiza essa visão, onde fica evidenciada a presença dominante do meio ambiente como assunto de classificação (Meio ambiente = 21,43% e Meio ambiente, reciclagem e tratamento de resíduos = 5,95%, perfazendo um total de 27,38%).

GRÁFICO 2 – Distribuição dos assuntos no agrupamento 10



Para a realização da análise dos assuntos, algumas dificuldades foram encontradas e estão relatadas a seguir.

Dentre os 84 documentos do Agrupamento 10 analisado, 9 não possuíam os dados referentes à classificação de Assunto. Assim, foi necessário pesquisar essas 9 RTs na base de dados *online* (www.sbrt.ibict.br), em busca dessa classificação. Observou-se que uma delas não possuía de fato essa informação dentro da RT; entretanto, a classificação pôde ser recuperada nos metadados presentes no resultado da pesquisa. Observou-se, também, que, ao contrário, alguns metadados de determinados documentos traziam no campo de assunto a expressão: “Subclasse inexistente na migração”, embora as RTs trouxessem essa classificação internamente.

Foi necessário, ainda, pesquisar na base *online* uma RT cuja indexação apresentava texto diferente do padrão das demais RTs do mesmo assunto: “Recuperação de materiais plásticos descartados”, em vez de “Recuperação de

materiais plásticos”. A indexação adotada para a elaboração da estatística de assunto, portanto, foi esta última, encontrada nos metadados referentes à RT e coerente com a classificação dos demais documentos.

A despeito desses problemas todas as classificações referentes aos 9 documentos foram recuperadas, possibilitando, assim, a análise dos assuntos em 100% dos documentos que compunham o agrupamento em estudo.

5.3.2 Análise dos termos minerados

A partir da extração dos termos, foi possível aprofundar um pouco mais nas análises, onde ficou evidenciado o diferencial de se utilizar a DCT para exploração de informações ocultas em documentos textuais. A seguir, estão descritas algumas particularidades da análise que fundamentam esta afirmativa.

Tome-se como base, por exemplo, o termo “resíduo”, que está presente em 70% dos documentos do agrupamento 10, sob vários aspectos – coleta, estocagem, embalagem, acondicionamento, tratamento, disposição, transporte, descarte, incineração, aproveitamento. Este termo está relacionado a vários tipos de materiais de diversas origens (alimentos, plásticos, vidros, nylon, gesso, madeira, borracha, sucatas metálicas, couro, dejetos humanos, óleos, materiais de construção, entulhos, embalagens de remédios, lixo hospitalar, lixo biológico, aterro sanitário, materiais eletroeletrônicos, resinas) e de diferentes naturezas: química, industrial, biológica, urbana, farmacêutica.

Pode-se observar que todos os documentos do agrupamento trazem consigo, implícita ou explicitamente, uma preocupação, e até mesmo certo comprometimento, com o meio ambiente. Isso fica claro quando se analisam os termos resultantes da mineração nesses documentos, embora eles não tenham sido classificados sob a categoria “Meio ambiente”.

Em suma, a aplicação da DCT proporcionou a visualização das fortes relações existentes entre documentos que aparentemente não teriam interação entre si, caso fossem consideradas apenas as classificações e palavras-chave que lhes foram atribuídas.

Outro exemplo interessante de ser citado é o fato de algumas RTs do agrupamento 10 tratarem de assuntos tão diversificados e aparentemente distanciados do tema central identificado, ou seja, o meio ambiente. Exemplos disso são: fabricação de brinquedos (jogos recreativos, bolinhas de gude); fabricação de vassouras, produção de energia elétrica; fabricação de sofás.

No entanto, um olhar mais aproximado revelou que essas RTs se enquadram com perfeição neste tema, dada a natureza das matérias-primas utilizadas. A fabricação de vassouras e sofás utiliza como matéria-prima garrafas PET recicladas, assim como a produção de energia elétrica, fruto do calor emanado pela queima desse produto. Os jogos recreativos são feitos de papel reciclado e as bolinhas de gude são produzidas a partir de sucatas de vidro.

Detectou-se, ainda, uma outra natureza de envolvimento com o meio ambiente em RTs do agrupamento, pela presença de expressões como: “padrões de emissão de efluentes”, “normas técnicas”, “legalização”, “legislação”, “procedimentos legais”, “licenciamento ambiental”, as quais sugerem uma preocupação dos microempresários em exercer suas atividades de forma regulamentar, dentro de padrões que preservem o meio ambiente.

5.3.3 Informações para construção de indicadores

Em todo o agrupamento, visualiza-se a presença marcante de termos tais como: “aproveitamento”, “reaproveitamento”, “reciclagem”, “resíduos”, “recuperação”, “sucatas”, “descontaminação”, “produção limpa”, os quais evidenciam processos condizentes com questões ambientais.

Esse pode ser visto como um dos temas de destaque do SBRT que merece ser explorado para a extração de informações úteis à construção de indicadores, vez que se trata de um assunto de extrema importância para a sociedade. Em sendo o meio ambiente uma preocupação nacional, a análise aprofundada dessas soluções tecnológicas poderia facilmente nortear ações governamentais de incentivo a iniciativas que de fato contribuíssem para o bem-estar social.

Uma forma de melhor explorar esse conteúdo seria o cruzamento dos dados expostos na análise feita com a DCT com os demais dados disponíveis no sistema de informação, que poderia permitir a identificação, por exemplo, de regiões onde

estariam sendo empreendidos determinados esforços, que tipo de empresa ou o perfil do empreendedor que estaria buscando essas orientações, que tipos de iniciativas poderiam ser associadas umas às outras em busca de melhor aproveitamento dos esforços de cada uma.

Enfim, acredita-se que a partir de dados como esses, que expõem a preocupação do microempresário e do empreendedor brasileiro em buscar orientações para atuar no setor produtivo em consonância com os anseios do bem-estar público, poder-se-á chegar a níveis de decisões estratégicas que facilitem essas ações e produzam melhores resultados a menores custos.

5.3.4 Informações específicas sobre a base de Respostas Técnicas

A título de exemplificação das inúmeras potencialidades da DCT para extração de informação oculta em conteúdos textuais, apresentam-se a seguir algumas informações específicas identificadas durante os estudos realizados nos conteúdos da base de RTs. Trata-se de particularidades do sistema de informação em si, que vieram à tona com a aplicação da DCT, as quais – acredita-se – podem servir de subsídio a decisões internas à rede SBRT.

A primeira particularidade detectada foi a diversidade de níveis de indexação, que comprometem a acurácia dos dados quando do processamento automático. É possível vislumbrar a possibilidade de esses problemas detectados servirem de base para a realização de estudos visando a readaptação da tabela para abrigar as especificidades das demandas e soluções da base de RTs do SBRT. Há que se cuidar para que as análises sejam validadas por profissionais de indexação atuantes no SBRT, detentores de conhecimentos específicos voltados à área de informação tecnológica.

Ainda, a partir do agrupamento de documentos, foram identificadas algumas RTs repetidas, mesmo tendo sido produzidas com grande espaço de tempo entre elas. Dada à exigüidade do tempo, não foi possível realizar uma análise mais apurada nesses documentos, mas é possível inferir que essa descoberta pode apoiar a localização, desde possíveis erros em numeração de RTs, até a possibilidade de reincidências de pesquisas no tempo ou, ainda, a complementaridade entre pesquisas realizadas em diferentes épocas.

5.4 Dificuldades enfrentadas

Tanto nas análises comparativas entre agrupamentos quanto na exploração mais aprofundada do agrupamento 10, algumas dificuldades encontradas tornaram o processo lento e trabalhoso.

A principal delas diz respeito à indexação dos termos da base. Como a indexação não faz parte do objeto deste trabalho, não foram feitos aprofundamentos na busca de explicações para esses fenômenos, não podendo, portanto, ser feitas afirmações quanto a sua origem. Entretanto, buscou-se registrar as dificuldades para nortear ações futuras relativas ao tema e apoiar a identificação de indicadores específicos que possam servir de apoio à gestão do sistema de informação SBRT.

Ao se analisarem os assuntos utilizados para classificar as RTs do agrupamento 10, os quais estão dispostos no Quadro 11, verificou-se que alguns deles aparentemente poderiam ter sido agrupados em uma mesma categoria, como por exemplo: “Meio ambiente” e “Meio ambiente, reciclagem e tratamento de resíduos”. Algumas classificações sugerem um nível genérico, como: “Alimentos e bebidas”, “Meio ambiente”, “Serviços industriais”, “Borracha e plástico”, enquanto outras retratam um nível de especificidade maior que a indexação de primeiro nível: “Fabricação de produtos de material plástico”, “Fabricação de sacolas de material plástico”, “Fabricação de embalagem de material plástico”, “Fabricação de bolsas de plástico”, “Fabricação de artefatos de material plástico para outros usos não especificados anteriormente”. Essa diferenciação nos níveis de classificação dificulta as análises e produz resultados pouco precisos, já que algumas poderiam estar indexadas sob uma mesma categoria. Pareceria coerente, portanto, que as cinco classificações específicas citadas acima estivessem sob a categoria genérica “Borracha e plástico”. Percebeu-se, portanto, uma pulverização em termos de indexação, que de certa forma retrata a especificidade e diversidade dos temas abordados pela base de Respostas Técnicas.

Ainda, dentre os 34 documentos que apresentaram apenas uma ocorrência, foram encontradas algumas classificações que possivelmente poderiam estar juntas. Exemplos disso são: “Recuperação de sucatas de alumínio” e “Recuperação de materiais metálicos”; “Fabricação de móveis de outros materiais” e “Mobiliário”.

Outro exemplo de dificuldade, ainda com relação à indexação, diz respeito à padronização dentro de um mesmo tema, presente no agrupamento 3, um dos três objetos da análise comparativa. Uma das Respostas Técnicas, tratando de “shampoo e sabonetes veterinários”, foi indexada sob o assunto “Fabricação de medicamentos para uso veterinário”, enquanto que a outra, sobre “xampu para cães (antipulga)”, foi indexada como “Produtos Químicos”.

Da mesma forma, deparou-se com esse mesmo tipo de variação na indexação no agrupamento 1, onde Respostas Técnicas tratando de “perfumes” apresentaram cinco categorias diferentes de assunto: “Fabricação de cosméticos”, “Fabricação de aditivos de uso industrial”, “Serviços industriais”, “Produtos químicos”, “Cosméticos”. Há que se fazer uma análise mais aprofundada, para verificar se esse fato caracterizaria problemas na indexação ou se seria um indicativo da especificidade da base de RTs, cujos documentos estariam enfocando aspectos diferenciados da fabricação de um mesmo produto, o perfume.

Essas diferenças de indexação dificultam o processo de tratamento automático, comprometendo a precisão dos resultados. Daí a importância da análise visual dos documentos agrupados. Obviamente, se pudesse ser realizada com a presença de especialistas em indexação, certamente esta análise seria mais apurada e a sua validação incorporaria maior confiabilidade aos resultados. Entretanto, para a realização dessa validação junto aos técnicos do SBRT, responsáveis pela indexação da base de RTs, seria necessário um tempo de que não se dispôs até a conclusão desta pesquisa.

Outra dificuldade, desta vez relacionada à autora da dissertação, que não tem formação na área de exatas, diz respeito à falta de domínio da ferramenta de mineração de textos, dada a complexidade dos recursos – os quais exigem conhecimentos específicos de informática e estatística – e à exigüidade do tempo para a realização da pesquisa, que não propiciou treinamento específico. Essa falta de conhecimento provocou limitações de processamento automático para realização de refinamentos na pesquisa, tendo sido realizados alguns procedimentos manualmente. Isso provocou algum desperdício de tempo, aumentando ainda mais a carga destinada à preparação dos dados para a pesquisa.

A despeito das dificuldades relatadas, o trabalho de análise obteve resultados que comprovaram a eficácia de utilização da DCT na busca de informações ocultas em documentos.

As análises realizadas permitiram a identificação de aspectos importantes para apoio à construção de indicadores. Visualiza-se como exemplo disso a presença marcante da preocupação de empreendedores de diversos setores da produção com o reaproveitamento de resíduos, a reciclagem. Esse tipo de resultado pode contribuir para a produção de indicadores nacionais voltados, por exemplo, a investimentos em ações locais que potencializem as iniciativas de pequenos atores do setor produtivo. Obviamente, o mapeamento dessas iniciativas é factível a partir do cruzamento dos dados minerados com informações outras constantes da base de Respostas Técnicas, como: tipo de empresa, área de atuação, região do país onde ela se encontra estabelecida, especialistas envolvidos no tema, entre outras.

À luz dos resultados apontados, é possível inferir que a aplicação da DCT em conteúdos textuais de diversas naturezas poderá propiciar a realização de estudos de inteligência competitiva na organização que os detém.

CAPÍTULO 6

CONSIDERAÇÕES FINAIS

A aplicação da DCT em grandes massas de dados ainda é uma atividade muito pouco explorada no Brasil, conforme foi possível detectar pela escassez de literatura em língua portuguesa sobre o tema. Muitos dos documentos localizados estão inseridos em áreas como Informática e Estatística, onde se pesquisam temas relacionados a descrição de metodologias, funcionalidades de ferramentas de mineração, ou tratamentos estatísticos, lingüísticos, indexação automática e bases de dados.

As palavras de Fayyad *et al.*, registradas há mais de uma década são representantes da visão:

Há uma necessidade urgente de uma nova geração de teorias e ferramentas computacionais para apoiar o ser humano na extração de informação útil (conhecimento) do rápido crescente volume de dados digitais. (FAYYAD *et al.*, 1996). (Tradução nossa)¹²

Não foram localizados muitos documentos em língua portuguesa tratando de análise de conteúdos utilizando a DCT. Entende-se, portanto, que esta pesquisa apresentou um grau de ineditismo, por ser pioneira no estudo desta técnica aplicada a conteúdos de informação tecnológica, voltada à aplicação dos conhecimentos gerados em prol do setor produtivo de pequeno porte e, por consequência, ao bem-estar social.

6.1 Conclusões

Em recente estudo de caso sobre a inovação e incubadoras de empresas, o Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (Inmetro) registra a intensificação do movimento de incubação de empresas, em nível internacional, como resultado de três movimentos: de condomínios de empresas, de programas voltados ao fomento de empreendedores, e de transferência de

¹² "There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data." FAYYAD *et al.*, 1996)

tecnologia da universidade para a empresa (ARANHA, 2005 *apud* SOUZA *et al.*, 2006).

O estudo do Inmetro afirma que esse movimento também se configura em uma tendência no Brasil, com adaptação às necessidades locais, e cita alguns desafios e oportunidades que se colocam para o futuro do movimento das incubadoras e parques tecnológicos do Brasil: 1) diminuir mortalidade das PME; 2) promover cultura empreendedora; 3) garantir competitividade justa; 4) gerar emprego e renda; 5) exportar e agregar valor; e 6) atrair investimentos. Aborda a importância do adensamento de conteúdo tecnológico como uma das soluções para que incubadoras e parques tecnológicos possam atuar cada vez mais fortemente. (FIATES, 2005 *apud* SOUZA *et all*, 2006).

O SEBRAE aborda a importância da inovação tecnológica para o crescimento econômico e o desenvolvimento sustentado do País:

É crescentemente aceito pela sociedade brasileira que, para acelerar o crescimento econômico com equidade e promover o desenvolvimento sustentado, o País precisa valorizar as suas microempresas e empresas de pequeno porte (MPE) e estimular a inovação – especialmente a inovação tecnológica – em todos os setores da atividade privada e nos serviços públicos. (SEBRAE/NA, 2007)

À luz dessas idéias, entende-se que os resultados da análise do conteúdo do SBRT – um sistema de informação tecnológica que registra as demandas advindas de empreendedores e microempresas e as soluções geradas em universidades e instituições de pesquisa para atender a essas demandas – contém informações relevantes, que podem ser aplicadas na direção das políticas citadas pelas duas instituições, Inmetro e SEBRAE.

A aplicação da DCT nos conteúdos textuais do Serviço Brasileiro de Respostas Técnicas trouxe à tona uma diversidade de informações agrupadas que não poderiam ter sido visualizadas sem o uso dessa técnica, cuja capacidade de extrair informações ocultas em acervos textuais gera possibilidades novas de se produzirem mais conhecimentos.

Lembrando os demais tipos de documentos gerados pelo SBRT e veiculados no sistema de informação descritos no item 2.1.1, os Dossiês Técnicos (DTs), vale observar que análises como as realizadas nesta pesquisa poderão servir de

alavanca para a geração desse tipo de documento. Sabe-se que os DTs oferecem as mais diversas informações sobre determinado setor ou produto em evidência pelo mercado, com a intenção de antecipar as necessidades das microempresas. Assim, a localização de RTs de interesse para a elaboração de DTs pode ser facilitada com a mineração dos textos, enriquecendo as buscas anteriormente realizadas com os recursos tradicionais de recuperação da informação.

Dessa forma, a aplicação da DCT traz à tona informações até então desconhecidas, que, seguramente, quando reveladas, podem se transformar em preciosas fontes de novos conhecimentos, e, portanto, constituir verdadeiros tesouros a serviço de novos DTs, ainda mais completos. A identificação de RTs semelhantes vai muito além da utilização da busca e recuperação de termos indexados de forma semelhante. O estudo demonstrou que a DCT pode aproximar RTs de temas aparentemente díspares e, assim, proporcionar um mergulho diferenciado no conteúdo existente para o aproveitamento dos esforços empreendidos e seu aperfeiçoamento em função da geração de novos DTs.

A aplicação da DCT na base de Respostas Técnicas revelou alguns sérios problemas de indexação e classificação que, se bem explorados, poderão contribuir para um estudo aprofundado da tabela de assuntos utilizada pelo SBRT, seguindo os caminhos apontados pela natureza das demandas postas pelos clientes e respondidas pelos especialistas da Rede de instituições membros.

Considera-se, portanto, que os objetivos propostos pela pesquisa foram alcançados, conforme ilustrado no esquema da Figura 10, onde estão detalhados resultados relativos a cada um deles.

Extrapolando os limites do estudo de caso da pesquisa, entende-se que os resultados alcançados podem ser vistos como teste de utilização da DCT, uma prática aplicável a outros conteúdos informacionais com características semelhantes ao analisado, ou seja, base de textos não estruturados que potencialmente contenham informações ocultas passíveis de serem reveladas por essa técnica.

Acredita-se, ainda, na viabilidade de aplicação prática da DCT no ambiente governamental, notadamente nos conteúdos do Ibict¹³, órgão parceiro da rede de instituições que compõem o Serviço Brasileiro de Respostas Técnicas (SBRT),

¹³ Para maiores detalhes sobre esses conteúdos, consultar a URL do Ibict: <http://www.ibict.br>.

responsável pela disponibilização *online* do serviço na Internet. Essa crença se fundamenta na compreensão de que se trata de um instituto de pesquisa detentor de importantes conteúdos informacionais de interesse nacional na *Web* e tem demonstrado, ao longo de sua história, ser um órgão estimulador de iniciativas que visem explorar esses conteúdos para promover sua melhor utilização pela sociedade como um todo.

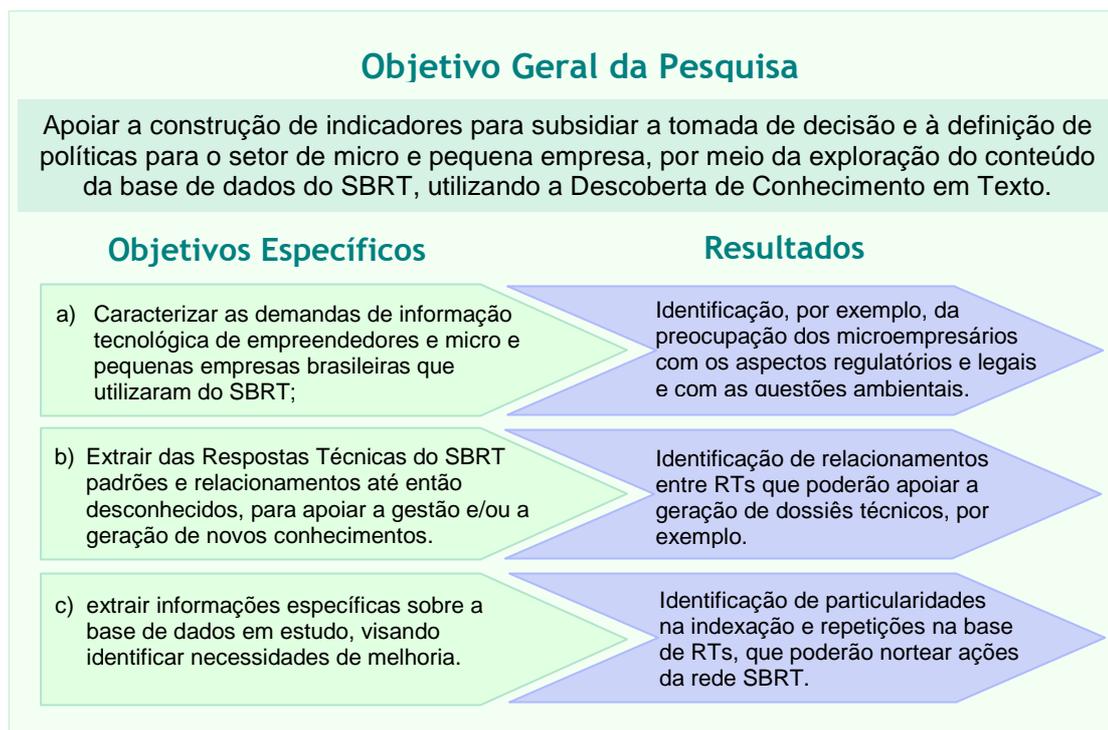


FIGURA 10 – Os alcances da pesquisa

Entretanto, é prudente lembrar que essa aplicação só será factível com o forte envolvimento institucional, dada a necessidade de investimentos financeiros para a aquisição de ferramentas específicas para aplicação da DCT e de destinação de equipe multidisciplinar para a realização de todas as etapas necessárias. Há que se considerar também treinamento específico para a otimização do uso dos recursos de que dispõem as ferramentas e, dessa forma, obter resultados concretos precisos. Faz-se necessário também o relacionamento dessa equipe com profissionais ligados ao conteúdo, para validação desses resultados e compartilhamento de idéias em torno dos indicadores que se pretender gerar.

No tocante às contribuições da pesquisa realizada para a Ciência da Informação vislumbradas no início deste trabalho, entende-se que os resultados aqui

alcançados revelam a pertinência e a importância de se aplicar a DCT em grandes repositórios de informação digital. Daí a compreensão de que este estudo possa contribuir para investigações futuras que indiquem novos rumos para a CI, onde os serviços de informação possam ser analisados e identificados como fontes para geração de novos e diferenciados serviços, à luz das demandas cada vez mais específicas e desafiadoras.

Além disso, vislumbra-se também a larga utilização, pela CI, de novas técnicas que possam representar um avanço na busca e identificação de indicadores e contribuam para ampliar a contribuição dessa Ciência para a gestão de CT&I, assim como para outros setores da sociedade, cumprindo o papel que lhe cabe como Ciência que pode fazer a diferença no uso dos avanços do conhecimento em prol do bem-estar social.

6.2 Recomendações de estudos futuros

A partir dos achados desta pesquisa e das limitações detectadas durante a sua realização, foi possível levantar algumas reflexões sobre estudos futuros que possam complementar as idéias aqui tratadas e extrair os melhores resultados da aplicação da DCT no estudo de conteúdos textuais.

Identificou-se uma lacuna no que se refere à lematização em língua portuguesa de forma automatizada. Há que se considerar o aprofundamento de pesquisas nesse campo, que permitam adaptações em ferramentas de mineração de textos, para atender a necessidades específicas do nosso idioma.

Do ponto de vista da aplicação interna ao SBRT, entende-se que a DCT pode ser largamente utilizada com vistas a propiciar análises voltadas para a indexação da base de Respostas Técnicas e, assim, promover as melhorias que se fizerem necessárias, de acordo com as especificidades presentes nos anseios dos microempresários e nas respostas elaboradas pelos especialistas da rede SBRT.

Sugere-se, ainda, a realização de estudos a partir do cruzamento dos dados extraídos na mineração de conteúdos textuais (Respostas Técnicas e Dossiês Técnicos) com os metadados disponíveis na base de dados de Respostas Técnicas. Com isso, é possível extrair os mais diversos tipos de indicadores que possam nortear ações futuras, como, por exemplo, as regiões do país em que determinados

temas estão sendo mais explorados; o tipo de cliente que lida com determinado assunto e com que finalidades; ou que tipo de técnicas estão sendo utilizadas para a produção determinados produtos; ou, ao contrário, que tipos de produtos podem ser fabricados utilizando-se uma determinada técnica inovadora que represente economia de energia, por exemplo, e quais os especialistas detêm essa técnica. Enfim, são inúmeras as possibilidades de exploração do que há armazenado no sistema de informação SBRT para produzir novos conhecimentos que possam nortear decisões em todos os níveis.

Considerando a já citada preocupação do empresariado cliente do SBRT com as questões relativas ao meio ambiente, sugere-se considerar a incorporação, no corpo da RT, de informação adicional relativa aos impactos da solução apresentada no meio ambiente. Acredita-se que seria um diferencial importante, uma antecipação que representaria uma agregação de valor à resposta do que foi solicitado.

Sugere-se, ainda, agora em nível mais amplo, a criação, e incorporação no *site* do SBRT, de um novo serviço que poderá representar um diferencial no atendimento a microempresários, podendo fomentar o empreendedorismo e até mesmo as relações de negócios entre pequenos produtores. Trata-se da identificação de fornecedores de serviços ou de matérias-primas. Essa foi uma demanda que se verificou constante nas leituras das RTs analisadas durante a pesquisa.

Extrapolando os limites institucionais, empresariais e acadêmicos, espera-se que a presente pesquisa possa estimular a realização de estudos futuros ou o empreendimento de ações que contribuam com o enriquecimento da aplicação do conhecimento gerado no meio acadêmico para a solução de problemas da sociedade como um todo.

REFERÊNCIAS

AIRES, Rachel Virgínia Xavier. **Uso de marcadores estilísticos para a busca na Web em português**. Orientadora: Profa. Dra. Sandra Maria Aluísio, Co-orientadora: Dra. Diana Santos. 2005, 202 p. Tese (Doutorado em Ciências de Computação e Matemática Computacional)-USP-São Carlos- Instituto de Ciências Matemáticas e de Computação - ICMC-USP. (Ver página 40)

ARAÚJO, Nelma Camêlo. **Análise do uso efetivo da informação por empresários de microempresas alimentícias do Estado de Minas Gerais**. Dissertação (Mestrado em Ciência da Informação), Universidade Federal de Santa Catarina, Florianópolis, 2005. 120f.

ARAÚJO, Vania M. R. Hermes de, FREIRE, Isa Maria e MENDES, Teresa Cristina M. Demanda de informação pelo setor industrial: dois estudos no intervalo de 25 anos. **Ciência da Informação**, Brasília, v.26, n.3, p. 283-289, set/dez. 1997. (Cap Inf. Tec.)

ARAÚJO JUNIOR, Rogerio Henrique de. **Precisão no processo de busca e recuperação da informação**. Orientador: Tarapanoff, Kira Maria Antônia. 2005. xiv, 223 p. Tese(Doutorado em Ciência da Informação)-Universidade de Brasília. Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação. Departamento de Ciência da Informação e Documentação. Inclui bibliografia e anexos.

BORGES, Maria Alice Guimarães. **A tríplice hélice e o desenvolvimento do setor de tecnologia da informação no Distrito Federal**. Orientador: Cunha, Murilo Bastos da. 2006. 317 p. Tese(Doutorado em Ciência da Informação)-Universidade de Brasília. Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação. Departamento de Ciência da Informação e Documentação.

BRÄSCHER, M. **Tratamento automático de ambigüidades na recuperação da informação**. 290 f. Tese (Doutorado em Ciência da Informação) - Curso de Pós-graduação em Ciência da Informação, Universidade de Brasília, Brasília, 1999.

BRASIL. **Estatuto da Microempresa e da Empresa de Pequeno Porte (1999)**. Estatuto da microempresa e da empresa de pequeno porte : lei n. 9.841, de 5 de outubro de 1999 e Decreto n. 3.474, de 19 de maio de 2000. – 2. ed. – Brasília : Câmara dos Deputados, Coordenação de Publicações, 2003. 33 p. – (Série fontes de referência. Legislação; n. 51). Disponível em: <<http://www.camara.gov.br/internet/infdoc/Publicacoes/html/pdf/microempresa1.pdf>>. Acesso em: 26 Jul. 2007.

BRASIL. **Lei das Microempresas e das Empresas de Pequeno Porte "SIMPLES"**. LEI nº 9.317, de 05 de dezembro de 1996. DOU de 06/12/96, pág. 25.973/7.

BRASIL. Ministério da Ciência e Tecnologia. **Indicadores de Ciência & Tecnologia – 2002**. Brasília: MCT, 2004, 140 p. ISSN 1413-3148.

BRASIL. Ministério da Ciência e Tecnologia. **Relatório de Gestão** Janeiro de 2003 a Dezembro de 2006. Brasília, MCT, 140 p.

CAPURRO, R. Epistemologia e Ciência da Informação. In: **V Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB**, Belo Horizonte, 11-2003.

CHAVES, H., CORDEIRO, F. e CUNHA, M. B. de. Avaliação do uso do Serviço Brasileiro de Respostas Técnicas: um serviço de informação destinada à microempresa brasileira. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 255-269, set./dez. 2006.

CUBILLO, Julio. La inteligencia empresarial en las pequeñas y medianas empresas competitivas de América Latina : algunas reflexiones. **Ciência da Informação**, Brasília, v.26, n.3, p. 260-267, set/dez 1997.

CYSNE, F. P. Transferência de tecnologia e desenvolvimento. **Ciência da Informação**, Brasília, v.25, n.1, p. 26-35, jan/abril 1996.

FERREIRA, José Rincon e TARAPANOFF, Kira. Micro e pequenas empresas – competitividade e inclusão digital. **Inclusão Social**, v. 1, n. 1, p. 43-46, out./mar., 2005.

FAYYAD *et al.* From data mining to knowledge discovery in databases. **AI Magazine**. American Association for Artificial Intelligence, 0738-4602, 1996. Disponível em: <<http://www.aaai.org/aitopics/assets/PDF/AIMag17-03-2-article.pdf>>. Acesso em 11 fev. 2008.

HEARST, Marti. **Untangling Text Data Mining**. In: *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20-26, 1999 (invited paper)*. Disponível em: <<http://www.sims.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html>>. Acesso em 31 jul. 2007.

HEARST, Marti. **What is Text Mining?** SIMS, UC Berkeley. October 17, 2003. Disponível em: <<http://www.ischool.berkeley.edu/~hearst/text-mining.html>>. Acesso em: 31 jul. 2007.

JANUZZI, Celeste Aída Sirotheau Corrêa e MONTALLI, Katia Maria Lemos. **Informação tecnológica e para negócios no Brasil: introdução a uma discussão conceitual**. Ciência da Informação, IBICT: Brasília, v. 28, n. 1, p. 28-36 1999.

KONDO, Edson Kenji. Desenvolvendo indicadores estratégicos em ciência e tecnologia: as principais questões. **Ciência da Informação**, 1998, vol.27, no.2, p. 128-133.

LASTRES, H. M. M. Dilemas da política de desenvolvimento científico e tecnológico. **Ciência da Informação**, v. 24, n. 2, p. 189-193, 1995.

LONGO, W.P., “O desenvolvimento científico e tecnológico do Brasil e suas perspectivas frente aos desafios do mundo moderno”, coleção **Brasil: 500 anos**, vol. II, Editora da Universidade da Amazônia, Belém, 2000, 99 p.. Disponível em: <<http://www.waldimir.longo.nom.br/artigos/97.doc>>. Acesso em 12 de março de 2008.

LUCAS, Marty. Mining in textual mountains. **Mapa Mundi Magazine**, disponível em <http://mappa.mundi.net/trip-m/hearst/>, acessado em 31 jul. 2007.

MARCONI, Marina de Andrade e LAKATOS, Eva Maria. **Técnicas de Pesquisa: planejamento e execução de pesquisas, elaboração, análise e interpretação de dados**. 5 ed. São Paulo : Atlas, 2002, 282 p.

MUELLER, Suzana Pinheiro Machado. Anotações de aulas durante o curso **Metodologia da Pesquisa em Ciência da Informação**. Brasília, UnB/CID, 1º sem. 2006.

QUONIAM, Luc, TARAPANOFF, Kira, ARAÚJO JÚNIOR, Rogério Henrique, ALVARES, Lillian. Inteligência obtida pela aplicação de *data mining* em base de teses francesas sobre o Brasil. **Ciência da Informação**, Brasília, v.30, n.2, p.20-28, maio/ago. 2001. Disponível em: <http://www.scielo.br/pdf/ci/v30n2/6208.pdf>, acesso em 15 ago. 2007.

RODRIGUES, Margarete de Luna; ABE, Naguiça; DIB, Simone Faury (Org.) **Glossário de informação tecnológica – GLIT**. Brasília, SENAI/DN, 2001. 51 p.

ROMÃO, W. ; PACHECO, R. C. S. ; NIEDERAUER, C. A. P. . Planejamento em C&T: uma abordagem para descoberta de conhecimento relevante em banco de dados de grupos de pesquisa. **Revista Tecnológica**, Maringá, v. 9, p. 139-152, 2000. Disponível em: <<http://www.din.uem.br/wesley/Planejamento.pdf>>. Acesso em: 15 mar. 2008.

SANTOS, Maria Ângela M.R. dos. **Extraindo regras de associação a partir de textos**. Orientador: FREITAS, Alex Alves. Curitiba, 2002, 51p. Dissertação (Mestrado em Informática Aplicada)-Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada.

SBRT. **Plano de negócios do Serviço Brasileiro de Respostas Técnicas**. Rio de Janeiro, 2005. 31 p.

SCHIESSL, José Marcelo. **Descoberta de Conhecimento em Texto aplicada a um sistema de atendimento ao consumidor**. Orientador: Profa. Dra. Marisa Bräscher, 2007. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília.

SEBRAE. **Serviço Brasileiro de Apoio às Micro e Pequenas Empresas**. Site: <http://www.SEBRAE.com.br>.

SEBRAE. **Fatores condicionantes e taxa de mortalidade de empresas no Brasil** : relatório de pesquisa. Brasília, Agosto/2004. 56 p. Disponível em:

<[http://www.biblioteca.SEBRAE.com.br/bte/bte.nsf/9A2916A2D7D88C4D03256EEE00489AB1/\\$File/NT0008E4CA.pdf](http://www.biblioteca.SEBRAE.com.br/bte/bte.nsf/9A2916A2D7D88C4D03256EEE00489AB1/$File/NT0008E4CA.pdf)>. Acesso em: 21 dez. 2007.

SEBRAE/NA. **Atuação do sistema SEBRAE em acesso à inovação e tecnologia.** Brasília, 2008, p. Disponível em: <<http://www.SEBRAE.com.br/customizado/inovacao-e-tecnologia/atuacao-do-sistema>>. Acesso em: 10 fev. 2008.

SEMIO Corporation. **Text Mining and the Knowledge Management Space** : A Semio Corporation white paper. DM Review. 1999.09.17. Disponível em: http://www.dmreview.com/whitepaper/paper_sub.cfm?whitepaperId=10080>. Acesso em: 15 nov. 2006.

SILVA, Eudes Garcez de Souza e MARINHO, Raimunda Ramos. **Serviço de informação: um instrumento para a integração universidade/empresa.** In: XII SNBB 2002.

SOARES, Bruno Jorge. **Comportamento de gestores de empresas de base tecnológica na busca e uso de informações.** Orientador: Costa, Sely Maria de Souza. Brasília, 2003. Dissertação(Mestrado em Ciência da Informação)-Universidade de Brasília. Departamento de Ciência da Informação e Documentação. 163 p.b

SOUZA, Taynah Lopes De. et al. Valorização do Inmetro como agente de inovação: o caso da incubadora de empresas. 06 julho 2006. Disponível em <http://www.inmetro.gov.br/producao/intelectual/obras_intelectuais/122_obraIntelectual.pdf>. Acesso em: 18 fev. 2008.

TAN, A.-H. Text mining: The state of the art and the challenges. In: **Proceedings of the Pacific Asia Conference on Knowledge Discovery and Data Mining – PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases**, Beijing, p. 65–70. Disponível em: <http://www.ntu.edu.sg/home/asahtan/Papers/tm_pakdd99.pdf>. Acesso em: 19 nov. 2007 .

TAKAHASHI, Tadao. Inclusão social e TICs. **Inclusão Social**, v. 1, n. 1, p. 56-59, out./mar., 2005.

TRYBULA, W. J. Text mining. **Annual Review of Information Science and Technology**, vol. 34, 1999, p. 385-419.

UnB/CID. **Programa de Pós-Graduação em Ciência da Informação (PPGCInf)** : Linhas de Pesquisa. Arquitetura da Informação. Disponível em: http://www.cid.unb.br/M001/M0011000.ASP?txtID_PRINCIPAL=146. Acesso em: 09 dez. 2007.

VELHO, L. Estratégias para um sistema de indicadores de C&T no Brasil. **Parcerias estratégicas**, Brasília, Brasil, v. 13, n. -, p. 109-121, 2001.

VIEIRA, Anna da Soledade. Conhecimento como recurso estratégico empresarial. **Ciência da Informação**, Brasília, v.22, n.2, p.99-101, maio/ago. 1993.

WEISS, Sholom, INDURKHYA, Nitin, ZHANG, Tong e DAMERAU, Fred. **Text Mining: Predictive Methods for Analyzing Unstructured Information**. Springer, New York, 2005. 237 p.

WIVES, Leandro Krug. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (*clustering*) de documentos**. Orientador: Oliveira, José Palazzo Moreira de, 2004. 126f : il. Tese(Doutorado)-Universidade Federal do Rio Grande do Sul. Porto Alegre: Programa de Pós-Graduação em Computação. (<http://www.leandro.wives.nom.br/>)

ANEXOS E APÊNDICES

ANEXO A

Exemplo de Resposta Técnica (RT)

	RESPOSTA TÉCNICA	
-----------------------------------------------------------------------------------	-------------------------	-------------------------------------------------------------------------------------

Título

Gerenciamento de resíduos ambientais em indústrias de calçados

Resumo

Apresenta informações sobre produção de calçados e formas de diminuir os impactos ambientais causados pela atividade.

Palavras-chave

Calçado; meio-ambiente; resíduos; couro

Assunto

TRATAMENTO E DISPOSIÇÃO DE RESÍDUOS NÃO-PERIGOSOS

Demanda

Como as indústrias de calçados podem contribuir para a conservação do meio-ambiente durante o processo produtivo, com ações como tratamento de efluentes/resíduos sólidos, análise do ciclo de vida de produto, modificações na fabricação, etc.?

Solução apresentada

1 Introdução

A preocupação ambiental com a produção de calçados deve, primeiro, atentar para os processos e materiais utilizados. O conceito chamado *ecodesign* preocupa-se justamente com esses aspectos, e apregoa a substituição de materiais, a adoção de processos mais eficientes, recuperação e reaproveitamento de resíduos, entre outras medidas.

2 Substituição de matérias-primas

Um calçado é composto de diversos componentes, que precisam ser feitos de material diferentes.

A fim de facilitar a reciclagem futura, os componentes devem ser feitos de materiais simples, não compostos. O calçado deve, também, possibilitar o reaproveitamento de suas partes – por exemplo, a troca do solado, ou mesmo a reutilização de componentes em novos produtos.

O método de produção dos componentes pode ser alterado. Por exemplo, o contraforte, um componente que fica entre o forro avesso (traseiro) e o cabedal, e que serve para “armar” e enrijecer a parte traseira do calçado, pode ser feito de material injetado (espuma de poliuretano, por exemplo) no formato desejado. Com esta substituição, deixam de ser geradas as rebarbas que sobram do corte e da chanfração das peças. A fim de ilustração, para uma produção de 1.000 pares de calçados por dia, tem-se um consumo diário de 27m² de material para confecção dos contrafortes, o que num mês representa cerca de 540m²,

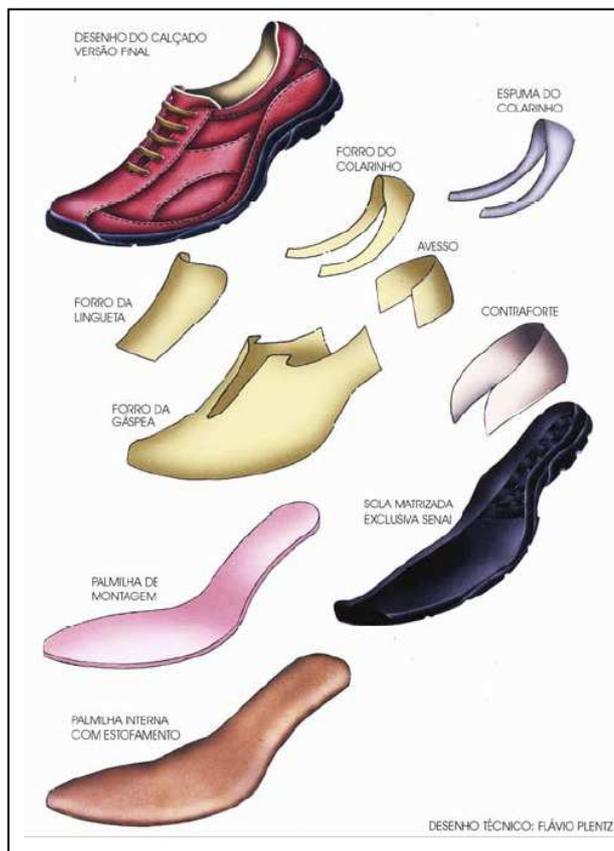


FIGURA 1 - Componentes mais comuns de um calçado.

Fonte: Aplicação do conceito de Eco Design em sapato. Disponível em: <http://sbirt.ibict.br/upload/dossies/sbirt-dossie14.pdf>. Acesso em: 30 jun. 2007.

dos quais 28% (151m²) são desperdiçados em rebarbas. O desperdício em um ano seria suficiente para fabricar contrafortes para 63.100 pares de sapatos, sem levar em conta os gastos com disposição desse material (FIG.1).

Além disso, os adesivos utilizados no calçado devem ser substituídos por produtos à base de água, que é um recurso mais abundante e não tóxico. Atualmente, existem diversos produtos que substituem solventes orgânicos, como meio condutor, por adesivos termoplásticos à base de resina poliéster e aditivos especiais, dispersão aquosa à base de poliuretano alifático ou dispersão aquosa à base de borracha natural e aditivos especiais.

3 Aproveitamento de Resíduos

A indústria de calçados de couro tem como principal resíduo a aparas de couro curtido que sobram do processo de produção. Estes resíduos de couro curtido, que à primeira vista aparentam ser inofensivos, podem representar uma séria ameaça para o meio ambiente e para a saúde humana. O perigo decorre do processo tradicionalmente utilizado para realizar o curtimento do couro, no qual este é tratado com compostos contendo cromo para evitar o seu apodrecimento. Os resíduos desse processo são classificados como de Classe 1, os mais perigosos.

A Associação Brasileira de Normas Técnicas (ABNT), através da NBR 12235, fornece informações sobre os métodos de disposição e armazenamento desses resíduos. Todavia, o gerenciamento e manutenção de depósitos para esse tipo de material é dispendioso.

Além da disposição em aterros controlados, as aparas e outros resíduos de couro podem servir à fabricação de couro reconstituído. São resíduos como serragem (também chamada de farelo de serradeira), pó de rebaixadeira, etc. Sua constituição química é muito discutível, pois varia de acordo com o tipo de curtimento efetuado.

A utilização da serragem para fabricação de couro regenerado não pressupõe o descurtimento prévio do resíduo e, basicamente, seu processo é simples. Consiste na mistura dos resíduos de rebaixe, de aparas de couro curtido e de retalhos provenientes das indústrias de calçados e artefatos de couro em geral, previamente moídos e misturados com aglutinantes – tais como o látex natural ou com resinas sintéticas. A massa obtida é prensada (em prensa manual ou hidráulica) e seca. O couro regenerado é então processado da mesma forma com que é trabalhado o couro natural.

Existe ainda a opção de se destruir os resíduos e reaproveitar os sais de cromo, através da queima controlada. Alguns reatores dispensam o fornecimento contínuo de energia para o processo, sendo auto-sustentáveis. O reator é alimentado com o material a ser incinerado, que pode ou não precisar ser previamente triturado, dependendo do equipamento. Abaixo um incinerador e um lavador de gases (FIG.2).



FIGURA 2 - Incinerador (à esquerda) e lavador de gases (à direita).

Fonte: Luftech. Disponível em: <<http://www.luftech.com.br/arquivos/incinerador.htm>>. Acesso em: 30 jun. 2007.

A câmara de combustão retém a quase totalidade dos particulados originados na queima do resíduo. Na seqüência, as substâncias particuladas se precipitam para um cinzeiro situado na parte inferior do equipamento. Na queima de materiais com alto poder calorífico, a temperatura da câmara de combustão pode atingir até 1.600 °C. O gás quente originado no ciclone pode ser utilizado para gerar energia térmica ou elétrica. Após o resfriamento, o ar passa pelo lavador de gases para eliminação de eventuais materiais particulados ou outros compostos. Ao final da operação as cinzas são coletadas, resfriadas e removidas do incinerador. O gás é totalmente queimado em razão da sua boa mistura e do tempo de residência no ciclone.

Os incineradores liberam uma grande quantidade de calor sem necessitar de combustível auxiliar; dependendo da quantidade de resíduos incinerados, é possível utilizar o calor gerado como fonte de energia elétrica, ou ainda direcioná-lo para um sistema central de refrigeração (sistema *chilling*), que use vapor de água como acionador do compressor.

4 Análise do Ciclo de Vida de Produtos

A Análise de Ciclo de Vida (ACV) tem se mostrado uma importante ferramenta de avaliação ambiental, devido ao seu enfoque “sobre função”, que permite considerar o desempenho ambiental de produtos, procedimento cada vez mais requerido por diversos atores sociais. Nas diferentes unidades desta cadeia, são levantados dados quantitativos sobre aspectos ambientais importantes, tais como emissões, consumo de recursos e geração de resíduos. A ACV enfoca o ciclo de vida de um produto, desde a extração de matérias-primas, passando pelas etapas de produção, distribuição e utilização, até sua destinação final, ou seja, uma abordagem do berço ao túmulo (FIG.3).

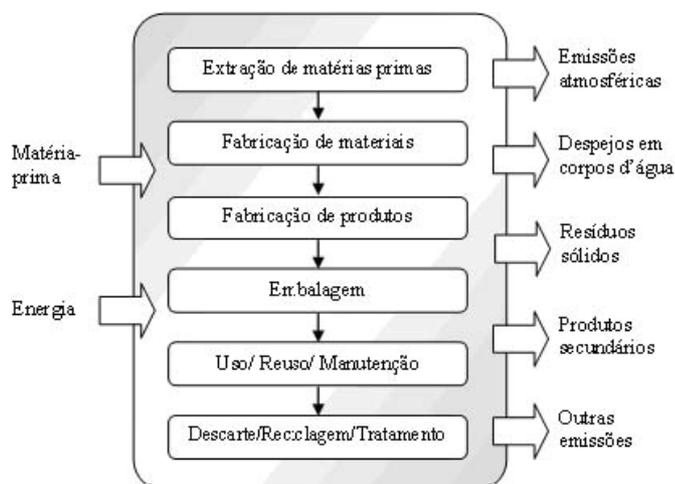


FIGURA 3 - Esquema geral de ACV.

Fonte: disponível em <<http://sbri.ibict.br/upload/dossies/sbri-dossie44.pdf>>.

Acesso em 30 jun. 2007.

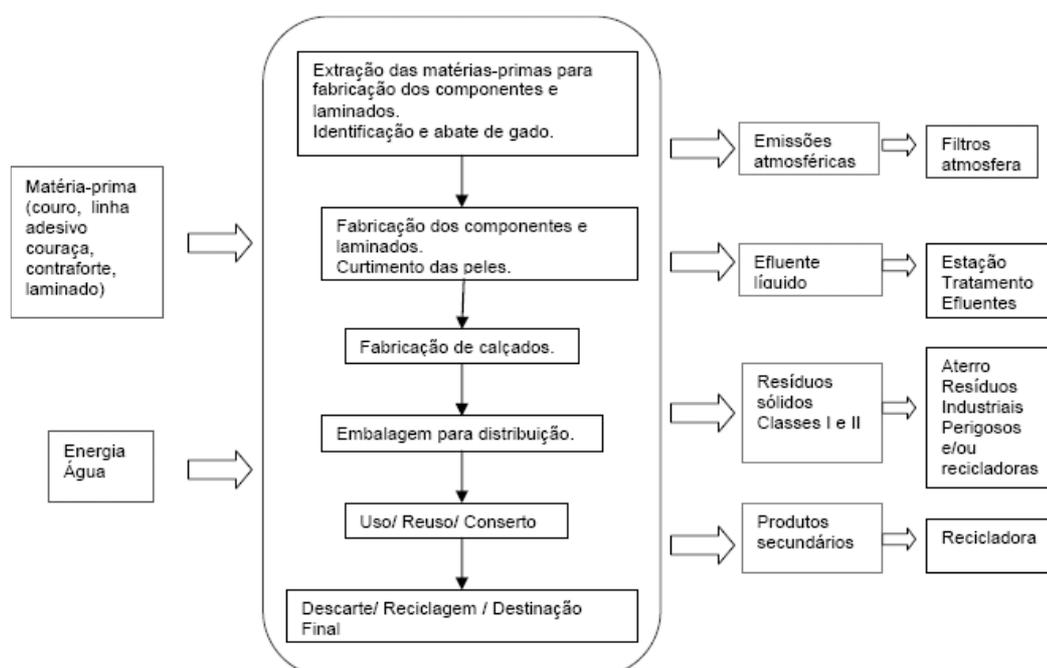


FIGURA 4 - Análise de ciclo de vida do calçado.

Fonte: Disponível em <<http://sbri.ibict.br/upload/dossies/sbri-dossie44.pdf>>. Acesso em 30 jun. 2007.

Imaginando-se um dos possíveis ciclos, numa fábrica que produza sapatos de couro, boa parte dos resíduos é gerada nas etapas de curtimento e de fabricação de componentes (FIG.4).

5 Outras Medidas

A iluminação é considerada a maior fonte de desperdícios de energia elétrica na maior parte das instalações elétricas em residências, comércio e indústria. A arquitetura definida normalmente não otimiza a iluminação natural. Por isso, cada local da fábrica deve ter seu nível de luminância necessário determinado, a fim de se implantar um sistema de iluminação adequado a cada tipo de atividade realizada.

Muitas vezes a luz natural é suficiente, mas para ser utilizada, é necessário que o local de trabalho possua telhas translúcidas, que permitam a entrada de luz. Além disso, paredes, estruturas de sustentação e forro devem ser pintados com cores claras, a fim de refletirem a maior quantidade possível de luz.

É recomendável também instalar sistema de sensores de presença em locais de pouca movimentação de pessoas.

Em algumas empresas que possuem jornada noturna, muitas vezes há áreas não utilizadas mantidas iluminadas, devido ao uso de disjuntores para acionamento de lâmpadas. Recomenda-se a instalação de interruptores individualizados, ou para conjuntos menores de lâmpadas.

6 Fornecedores de maquinário

BERNAUER

Av. do Oratório, 2.635 - CEP: 03221-100 - São Paulo/SP

Telefone de Contato: (11) 601-1855

Fax: 11 6918-0493

E-mail: bernauer@bernauer.com.br

Site: <<http://www.bernauer.com.br/Default.aspx>> Acesso em: 30 jun. 2007.

JAG MÁQUINAS

Estrada Municipal, 387 - Jd. Cristina - CEP: 13320-000

Salto/SP

Telefone de Contato: (11) 4021-3735

Fax: 11 2828-5459

Site: <<http://www.jagmaquinas.com.br/index.php?menu=1>>. Acesso em: 30 jun. 2007.

LUFTECH TECNOLOGIA AMBIENTAL LTDA

Incineradores, lavadores de gases, tratamento de efluentes, monitoramento contínuo, analisadores de gases.

Rua Cândido Pinheiro de Barcellos nº 217 - Distrito Industrial Alvorada/Viamão - Alvorada/RS CEP 94820-260

FONE/FAX: +55 (51) 3447 6687 OU (51) 3447 0614

e-mail: <vendas@luftech.com.br>

Site <<http://www.luftech.com.br/>>. Acesso em 30 jun. 2007.

METALTECH SOROCABA EQUIPAMENTOS E SERVIÇOS LTDA.

Fornos e geradores de gases para tratamento térmico e combustão, medidores de vazão, fornos, incineradores e forno crematório.

Rua Artur Gomes, nº. 427 - Centro

Sorocaba/SP - CEP: 18035-490

Tel: 55 (15) 3224-2466

Fax: 55 (15) 3224-2817

Site: <<http://www.metaltech.com.br/>>. Acesso em 30 jun. 2007.

O CETEC não tem qualquer responsabilidade quanto à idoneidade dos fornecedores, cabendo ao empreendedor optar por aquele que melhor atender às suas necessidades, qualidade, preço, prazo de entrega.

Conclusões e recomendações

Recomenda-se a consulta do Dossiê Técnico existente no site do SBRT - Serviço Brasileiro de Respostas Técnicas, no endereço:

SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. **Aplicação do conceito de ecodesign em calçados**. SENAI-RS, 2006. (Dossiê Técnico). Disponível em <<http://sbrt.ibict.br/upload/dossies/sbrt-dossie14.pdf>>. Acesso em 30 jun. 2007.

Sugere-se a consulta das Respostas Técnicas no site do SBRT, nos endereços:

SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. **Levantamento dos aspectos e impactos ambientais das indústrias de calçados**. CETEC-MG, 2007. Disponível em <<http://sbrt.ibict.br/upload/sbrt4336.pdf>>. Acesso em 30 jun. 2007.

SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. **Resíduos de couro**. Agência USP de Inovação/Disque-Tecnologia, 2007. Disponível em: <<http://sbrt.ibict.br/upload/sbrt5006.pdf>>. Acesso em 30 jun. 2007.

Fontes consultadas

CASTRO, Fernando de. **Gaúchos reaproveitam sal de cromo**. Química. Disponível em: <<http://www.quimica.com.br/revista/qd434/atualidades1.htm>>. Acesso em: 30 jun. 2007.

SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. Disponível em: <www.sbrt.ibict.br>. Acesso em 30 jun. 2007.

Elaborado por

Daniel Igor D. Silva

Nome da Instituição respondente

Fundação Centro Tecnológico de Minas Gerais – CETEC/MG

Data de finalização

30 jun. 2007.

ANEXO B

Amostra do dicionário de termos em língua portuguesa

O dicionário completo compõe-se de 881.880 itens. Trata-se de uma coletânea de dicionários da língua portuguesa, organizado por Marcelo Schiessl para verificação de classes gramaticais no apoio à técnica de lematização.

Term	Parent	Category
a	a	Abbr
a	a	Noun
a	a	Prep
a	ele	Pron
a	o	Pron
a	o	n/a
aa	aa	Abbr
aaron	aaron	Noun
aarão	aarão	Noun
ab	ab	Prx
ab-rogamento	ab-rogamento	Noun
ab-rogamentos	ab-rogamento	Noun
ab-rogação	ab-rogação	Noun
ab-rogações	ab-rogação	Noun
ab-rogáveis	ab-rogável	Adj
ab-rogável	ab-rogável	Adj
aba	aba	Noun
aba	abar	Verb
ababada	ababadar	Verb
ababadada	ababadar	Verb
ababadadas	ababadar	Verb
ababadado	ababadar	Verb
ababadados	ababadar	Verb
ababadai	ababadar	Verb
ababadais	ababadar	Verb
ababadam	ababadar	Verb
ababadamos	ababadar	Verb
ababadando	ababadar	Verb
ababadar	ababadar	Verb
ababadara	ababadar	Verb
ababadaram	ababadar	Verb
ababadaras	ababadar	Verb
ababadardes	ababadar	Verb
ababadarei	ababadar	Verb
ababadareis	ababadar	Verb
ababadarem	ababadar	Verb
ababadaremos	ababadar	Verb
ababadares	ababadar	Verb
ababadaria	ababadar	Verb
ababadariam	ababadar	Verb
ababadarias	ababadar	Verb

ababadarmos	ababadar	Verb
ababadará	ababadar	Verb
Term	Parent	Category
ababadarás	ababadar	Verb
ababadarão	ababadar	Verb
ababadaríamos	ababadar	Verb
ababadaríeis	ababadar	Verb
ababadas	ababadar	Verb
ababadasse	ababadar	Verb
ababadassem	ababadar	Verb
ababadasses	ababadar	Verb
ababadaste	ababadar	Verb
ababadastes	ababadar	Verb
ababadava	ababadar	Verb
ababadavam	ababadar	Verb
ababadavas	ababadar	Verb
ababade	ababadar	Verb
ababadei	ababadar	Verb
ababadeis	ababadar	Verb
ababadem	ababadar	Verb
ababademos	ababadar	Verb
ababades	ababadar	Verb
ababado	ababadar	Verb
ababadou	ababadar	Verb
ababadáramos	ababadar	Verb
ababadáreis	ababadar	Verb
ababadásseis	ababadar	Verb
ababadássemos	ababadar	Verb
ababadávamos	ababadar	Verb
ababadáveis	ababadar	Verb
ababalha	ababalhar	Verb
ababalhada	ababalhar	Verb
ababalhadas	ababalhar	Verb
ababalhado	ababalhar	Verb
ababalhados	ababalhar	Verb
ababalhai	ababalhar	Verb
ababalhais	ababalhar	Verb
ababalham	ababalhar	Verb
ababalhamos	ababalhar	Verb
ababalhando	ababalhar	Verb
ababalhar	ababalhar	Verb
ababalhara	ababalhar	Verb

ababalharam	ababalhar	Verb
ababalharas	ababalhar	Verb
ababalhardes	ababalhar	Verb
ababalharei	ababalhar	Verb
ababalhareis	ababalhar	Verb
ababalharem	ababalhar	Verb
ababalharemos	ababalhar	Verb
ababalhares	ababalhar	Verb
ababalharia	ababalhar	Verb
ababalhariam	ababalhar	Verb
ababalharias	ababalhar	Verb
ababalharmos	ababalhar	Verb
ababalhará	ababalhar	Verb
ababalharás	ababalhar	Verb
ababalharão	ababalhar	Verb
ababalharíamos	ababalhar	Verb
ababalharíeis	ababalhar	Verb
ababalhas	ababalhar	Verb
ababalhasse	ababalhar	Verb
ababalhassem	ababalhar	Verb
ababalhasses	ababalhar	Verb
ababalhaste	ababalhar	Verb
ababalhastes	ababalhar	Verb
ababalhava	ababalhar	Verb
ababalhavam	ababalhar	Verb
ababalhavas	ababalhar	Verb
ababalhe	ababalhar	Verb
ababalhei	ababalhar	Verb
ababalheis	ababalhar	Verb
ababalhem	ababalhar	Verb
ababalhemos	ababalhar	Verb
ababalhes	ababalhar	Verb
ababalho	ababalhar	Verb
ababalhou	ababalhar	Verb
ababalháramos	ababalhar	Verb
ababalháreis	ababalhar	Verb
ababalhásseis	ababalhar	Verb
ababalhássemos	ababalhar	Verb
ababalhávamos	ababalhar	Verb
ababalháveis	ababalhar	Verb
abelara	ababelar	Verb
abelarada	ababelar	Verb
abelaradas	ababelar	Verb
abelarado	ababelar	Verb
abelarados	ababelar	Verb
abelarai	ababelar	Verb
abelarais	ababelar	Verb
abelaram	ababelar	Verb
abelaramos	ababelar	Verb
abelarando	ababelar	Verb
abelarar	ababelar	Verb
abelarara	ababelar	Verb
abelararam	ababelar	Verb
abelararas	ababelar	Verb

abelardes	ababelar	Verb
abelarei	ababelar	Verb
abelareis	ababelar	Verb
abelarem	ababelar	Verb
abelaremos	ababelar	Verb
abelares	ababelar	Verb
abelaria	ababelar	Verb
abelariam	ababelar	Verb
abelarias	ababelar	Verb
abelarmos	ababelar	Verb
abelará	ababelar	Verb
abelarás	ababelar	Verb
abelarão	ababelar	Verb
abelaríamos	ababelar	Verb
abelaríeis	ababelar	Verb
abelas	ababelar	Verb
abelasse	ababelar	Verb
abelassem	ababelar	Verb
abelasses	ababelar	Verb
abelaste	ababelar	Verb
abelastes	ababelar	Verb
abelava	ababelar	Verb
abelavam	ababelar	Verb
abelavas	ababelar	Verb
abele	ababelar	Verb
abelei	ababelar	Verb
abeleis	ababelar	Verb
abelem	ababelar	Verb
abelemos	ababelar	Verb
abeles	ababelar	Verb
abelo	ababelar	Verb
abelou	ababelar	Verb
abeláramos	ababelar	Verb
abeláreis	ababelar	Verb
abelásseis	ababelar	Verb
abelássemos	ababelar	Verb
abelávamos	ababelar	Verb
abeláveis	ababelar	Verb
ababosa	ababosar	Verb
ababosada	ababosar	Verb
ababosadas	ababosar	Verb
ababosado	ababosar	Verb
ababosados	ababosar	Verb
ababosai	ababosar	Verb
ababosais	ababosar	Verb
ababosam	ababosar	Verb
ababosamos	ababosar	Verb
ababosando	ababosar	Verb
ababosar	ababosar	Verb
ababosara	ababosar	Verb
ababosaram	ababosar	Verb
ababosaras	ababosar	Verb
ababosardes	ababosar	Verb
ababosarei	ababosar	Verb

APÊNDICE A

Amostra da lista de palavras extraídas pelo BR/Search (Ordem de frequência e ocorrência na base)

Base de dados SBRT

Arquivo importado para o BR/Search

Lista extraída da base em 18/09/2007

Total de registros na base = 6.041 documentos textuais

Palavra	Ocorrências	Ocorrências na base	%	Documentos	Documentos na base	%
DE	369208	544623	0.0	6013	6013	100.0
"E"	178766	278207	0.0	6009	6009	100.0
A	138256	223785	0.0	5992	5998	0.0
O	91259	146344	0.0	5944	5955	0.0
EM	83621	123784	0.0	5938	5957	0.0
PARA	71455	106675	0.0	5949	5977	0.0
DO	70425	107382	0.0	5970	5997	0.0
DA	60562	91201	0.0	6001	6008	0.0
QUE	54686	86459	0.0	5867	5925	0.0
"COM"	39695	65026	0.0	5717	5847	0.0
HTTP	34990	50621	0.0	5818	5883	0.0
AS	29724	48939	0.0	5373	5596	0.0
UM	29187	45627	0.0	5408	5571	0.0
OS	28394	45593	0.0	5247	5484	0.0
UMA	28186	43469	0.0	5354	5545	0.0
"OU"	26440	43482	0.0	5133	5399	0.0
NO	26193	42098	0.0	5316	5568	0.0
COMO	25770	36396	0.0	5293	5482	0.0
NA	24164	38589	0.0	5234	5514	0.0
POR	23988	38040	0.0	5420	5591	0.0
SAO	23524	38204	0.0	5202	5477	0.0
SER	22000	37388	0.0	5001	5310	0.0
SOBRE	19398	24261	0.0	4725	5074	0.0
SE	17722	28126	0.0	4611	5045	0.0
TÉCNICA	16772	22407	0.0	5972	5975	0.0
DOS	16616	27126	0.0	4689	5106	0.0
"NAO"	16421	26130	0.0	4716	5161	0.0
INFORMACOES	16204	18962	0.0	4498	4833	0.0
PROCESSO	16162	21686	0.0	4073	4398	0.0
SERVICO	15932	22168	0.0	4728	4851	0.0
RESPOSTA	15156	20089	0.0	5966	5966	100.0
ACESSO	15131	19678	0.0	4294	4695	0.0
DISPONIVEL	14899	19702	0.0	4262	4702	0.0
SBRT	14519	21845	0.0	5114	5186	0.0
TECNICAS	14433	21255	0.0	4387	4592	0.0
DAS	14083	23330	0.0	4333	4836	0.0
MAIS	13916	22746	0.0	4217	4700	0.0
AO	13285	22105	0.0	4408	4944	0.0
"WWW.SBRT.IBICT.BR"	13178	19039	0.0	5037	5043	0.0
BRASILEIRO	12652	18662	0.0	4195	4343	0.0
PRODUCAO	12394	17418	0.0	3360	3715	0.0
"1"	12200	20087	0.0	4680	5905	0.0
PRODUTO	12108	16914	0.0	3704	4047	0.0
FABRICACAO	11907	14614	0.0	3104	3353	0.0
RESPOSTAS	11661	17509	0.0	3943	4041	0.0
"2"	10771	16367	0.0	5090	5789	0.0
AGUA	10532	17635	0.0	2534	3079	0.0
PRODUTOS	10178	14849	0.0	3207	3696	0.0
"2007"	9815	12287	0.0	2309	2365	0.0
NOME	9128	10208	0.0	5696	5982	0.0
TECNOLOGIA	9020	10171	0.0	4466	4612	0.0
PODE	8149	13597	0.0	3499	4166	0.0
"2006"	8110	9771	0.0	2439	2573	0.0
EQUIPAMENTOS	7780	10630	0.0	2354	2801	0.0
SOLUCAO	7736	8976	0.0	5974	5977	0.0
PELO	7597	12310	0.0	3246	3825	0.0
GOSTARIA	7592	7741	0.0	2726	2754	0.0
FORNECEDORES	7496	9369	0.0	2040	2389	0.0
ASSUNTO	7068	7484	0.0	5984	5984	100.0
MASCULINO	7059	7073	0.0	3520	3524	0.0
"11"	7038	12303	0.0	2017	2582	0.0

SUA	6992	11437	0.0	3262	3903	0.0
"3"	6951	11014	0.0	4042	4619	0.0
DEMANDA	6722	7128	0.0	5980	5982	0.0
TAMBEM	6712	10877	0.0	3356	4024	0.0
SP	6688	10971	0.0	2535	2996	0.0
SITE	6501	10857	0.0	2343	2939	0.0
TIPO	6317	9256	0.0	2678	3291	0.0
QUALIDADE	6280	9379	0.0	2728	3296	0.0
APRESENTADA	6208	6361	0.0	5966	5968	0.0
COPYRIGHT	6180	9382	0.0	2402	2422	0.0
SABER	6108	6428	0.0	2376	2510	0.0
DEVE	6096	10986	0.0	2480	3153	0.0
DATA	6070	6480	0.0	5803	5980	0.0
QUAL	5985	7225	0.0	2302	2713	0.0
TECNICO	5967	7210	0.0	3352	3667	0.0
INICIO	5930	6450	0.0	2821	2968	0.0
INSTITUICAO	5922	6332	0.0	5682	5978	0.0
TEM	5858	9359	0.0	2913	3604	0.0
FINALIZACAO	5850	6022	0.0	5825	5980	0.0
"4"	5789	8561	0.0	3103	3628	0.0
PALAVRAS-CHAVE	5787	5799	0.0	5766	5767	0.0
RESPONDENTE	5690	5979	0.0	5690	5977	0.0
ESTA	5628	8876	0.0	2810	3506	0.0
OUTROS	5584	8207	0.0	2838	3445	0.0
ENTRE	5578	9363	0.0	2698	3388	0.0
CEP	5490	9858	0.0	1888	2389	0.0
CENTRO	5479	6504	0.0	3027	3272	0.0
LTDA	5445	9414	0.0	1789	2216	0.0
TEL	5340	9544	0.0	1566	1955	0.0
MAQUINAS	5294	7489	0.0	1580	1871	0.0
ALIMENTOS	5220	7511	0.0	1504	1683	0.0
MATERIAL	5163	8039	0.0	2022	2508	0.0
FAX	5143	9176	0.0	1875	2384	0.0
PODEM	5136	8699	0.0	2712	3400	0.0
PAULO	5133	8124	0.0	2303	2866	0.0
FORMA	5116	8296	0.0	2600	3263	0.0
NEGOCIO	5000	5539	0.0	2053	2150	0.0
USO	4971	7744	0.0	2285	2907	0.0
PADRAO	4912	7621	0.0	2002	2131	0.0
INTERNET	4852	5062	0.0	2365	2457	0.0
INDUSTRIA	4749	7100	0.0	2075	2574	0.0
PELA	4689	7887	0.0	2562	3262	0.0
ONDE	4686	6824	0.0	2332	2910	0.0
QUAIS	4647	5479	0.0	1831	2127	0.0
QUANDO	4647	8317	0.0	2281	2955	0.0
ATE	4642	7966	0.0	2342	3009	0.0
BRASIL	4540	6888	0.0	2284	2797	0.0
OLEO	4522	6808	0.0	767	1002	0.0
E-MAIL	4503	7762	0.0	1864	2425	0.0
C	4456	7897	0.0	1817	2381	0.0
MUITO	4451	7253	0.0	2329	2913	0.0
FORMULARIO	4374	6829	0.0	1786	1808	0.0
GRADUACAO	4362	4417	0.0	2171	2188	0.0
NOS	4351	7105	0.0	2330	2990	0.0
SEU	4342	7081	0.0	2407	3075	0.0
NORMAS	4326	5354	0.0	1663	1922	0.0
"2005"	4290	5229	0.0	1716	1863	0.0
ESTE	4221	6554	0.0	2273	2922	0.0
RUA	4121	7586	0.0	1623	2140	0.0
EMPRESA	4051	5773	0.0	1751	2180	0.0
JA	4022	6059	0.0	2213	2873	0.0
NAS	4017	6801	0.0	2228	2939	0.0
MERCADO	4003	5889	0.0	1781	2203	0.0
RIO	4002	4989	0.0	1695	2035	0.0
RECOMENDACOES	3996	6197	0.0	3762	5728	0.0
FONTES	3992	5338	0.0	3444	4359	0.0
N	3932	6795	0.0	1620	2104	0.0
TEMPERATURA	3869	7007	0.0	1548	2086	0.0
INDUSTRIAL	3829	5034	0.0	2353	2704	0.0
MADEIRA	3815	5630	0.0	838	1091	0.0
SENDO	3813	6080	0.0	2230	2831	0.0
SEM	3797	5938	0.0	1964	2641	0.0
TITULO	3796	3919	0.0	3680	3703	0.0
ETC	3749	5529	0.0	1997	2507	0.0
RESUMO	3747	3842	0.0	3656	3689	0.0
ATRAVES	3728	6033	0.0	2225	2913	0.0
REDE	3726	4168	0.0	1829	1925	0.0
SISTEMA	3710	6262	0.0	1642	2156	0.0
BEM	3681	5930	0.0	1939	2593	0.0
MAS	3663	5301	0.0	2000	2501	0.0

FONE	3627	6111	0.0	1192	1500	0.0
UPLOAD	3624	5169	0.0	1535	1943	0.0
"5"	3624	5988	0.0	2152	2702	0.0
GRANDE	3609	5532	0.0	2093	2640	0.0
DEVEM	3555	6392	0.0	1680	2296	0.0
"10"	3526	5793	0.0	1706	2294	0.0
FAZER	3516	4661	0.0	1638	2035	0.0
SEJA	3511	5576	0.0	2150	2821	0.0
NACIONAL	3454	4580	0.0	2183	2488	0.0
AREA	3448	5293	0.0	1839	2412	0.0
EMPRESAS	3430	4741	0.0	1970	2414	0.0
MAIOR	3425	6057	0.0	1940	2635	0.0
MATERIAIS	3414	5056	0.0	1484	1893	0.0
CONSULTADAS	3395	4393	0.0	3260	4169	0.0
ELABORADO	3386	3746	0.0	3322	3628	0.0
RS	3380	4004	0.0	1471	1639	0.0
TRATAMENTO	3357	4813	0.0	1149	1469	0.0
APOS	3344	5868	0.0	1756	2404	0.0
CONTATO	3340	5206	0.0	2039	2743	0.0
CASO	3311	5690	0.0	1946	2682	0.0
CADA	3268	5483	0.0	1910	2562	0.0
"16"	3268	4050	0.0	1568	1849	0.0
USP	3249	3467	0.0	1352	1406	0.0
JANEIRO	3230	4000	0.0	1408	1663	0.0
POIS	3163	5143	0.0	1910	2567	0.0
PRODUZIDA	3143	4862	0.0	1365	1480	0.0
RESPONSAVEL	3133	3771	0.0	2580	2878	0.0
ALEM	3122	5163	0.0	2019	2667	0.0
QUANTO	3074	4862	0.0	1945	2692	0.0
INSTITUTO	3054	3706	0.0	1596	1785	0.0
UTILIZADO	3054	4555	0.0	1741	2324	0.0
"MESMO"	2964	4714	0.0	1817	2489	0.0
MEIO	2963	4647	0.0	1610	2090	0.0
TIPOS	2959	4352	0.0	1600	2100	0.0
RESIDUOS	2953	4310	0.0	794	986	0.0
OBTER	2951	3905	0.0	1636	2096	0.0
UTILIZADOS	2936	4263	0.0	1528	2027	0.0
UTILIZACAO	2930	4338	0.0	1531	2034	0.0
TECNOLOGICO	2927	3222	0.0	1579	1677	0.0
TEMPO	2923	4904	0.0	1508	2090	0.0
TECNICOS	2916	3296	0.0	1482	1681	0.0
MELHOR	2912	4365	0.0	1735	2359	0.0
DESENVOLVIMENTO	2869	4242	0.0	1595	1999	0.0
CARACTERISTICAS	2853	4540	0.0	1600	2095	0.0
"20"	2842	4563	0.0	1452	1978	0.0
BASE	2827	4406	0.0	1364	1819	0.0
IDENTIFICACAO	2826	3135	0.0	2509	2575	0.0
PARTIR	2801	4110	0.0	1733	2246	0.0
FONTE	2778	5038	0.0	1535	2131	0.0
SENAI	2776	3109	0.0	1385	1451	0.0
CULTIVO	2752	3755	0.0	610	701	0.0
HA	2748	4307	0.0	1508	1999	0.0
"64"	2737	2800	0.0	1379	1411	0.0
MASSA	2695	4281	0.0	820	1067	0.0
PECAS	2695	3948	0.0	932	1148	0.0
OUTRAS	2690	4290	0.0	1768	2418	0.0
AOS	2685	4479	0.0	1727	2342	0.0
PAPEL	2677	4085	0.0	693	957	0.0
AINDA	2669	4490	0.0	1717	2403	0.0
NECESSARIO	2649	4137	0.0	1638	2271	0.0
SUAS	2649	4098	0.0	1791	2384	0.0
SEBRAE	2643	3427	0.0	1413	1641	0.0
PADROES	2638	2980	0.0	1268	1411	0.0
AMBIENTE	2637	4127	0.0	1403	1837	0.0
GERAIS	2628	3105	0.0	1338	1539	0.0
LEITE	2624	3900	0.0	606	731	0.0
PLASTICO	2609	3630	0.0	942	1176	0.0
POSSIVEL	2608	3825	0.0	1578	2192	0.0
QUANTIDADE	2590	4281	0.0	1419	1952	0.0
CONTROLE	2578	4413	0.0	1150	1634	0.0
CONDICOES	2573	4437	0.0	1615	2204	0.0
TER	2568	3980	0.0	1515	2083	0.0
PARTE	2546	4364	0.0	1520	2090	0.0
ALGUNS	2534	3949	0.0	1647	2265	0.0
FEMININO	2508	2523	0.0	1249	1260	0.0
SOLO	2493	4204	0.0	728	912	0.0
ASSIM	2471	4025	0.0	1657	2251	0.0
BRASILEIRA	2471	3627	0.0	1549	1979	0.0
EMBALAGEM	2451	3696	0.0	816	1077	0.0
"15"	2450	3972	0.0	1332	1858	0.0

DURANTE	2443	4401	0.0	1447	1994	0.0
QUIMICA	2440	3508	0.0	1286	1601	0.0
PLANTAS	2417	3977	0.0	693	854	0.0
IMPORTANTE	2386	3954	0.0	1726	2418	0.0
"30"	2383	3898	0.0	1267	1794	0.0
PRECISO	2374	2693	0.0	1068	1252	0.0
RESISTENCIA	2364	3956	0.0	984	1285	0.0
EMBALAGENS	2358	3553	0.0	759	999	0.0
GRAU	2358	2847	0.0	1251	1498	0.0
INFORMACAO	2338	2733	0.0	1336	1563	0.0
MINAS	2319	2652	0.0	1105	1206	0.0
ACIDO	2317	3796	0.0	745	978	0.0
EXISTE	2283	2936	0.0	1248	1614	0.0
LIMPEZA	2275	3398	0.0	829	1137	0.0
AR	2271	3975	0.0	954	1368	0.0
MEDIO	2262	2686	0.0	1147	1328	0.0
ALGUMAS	2251	3519	0.0	1504	2078	0.0
PROCESSOS	2242	3442	0.0	1202	1594	0.0
CONSTRUCAO	2234	3123	0.0	862	1083	0.0
VIDRO	2233	3340	0.0	484	679	0.0
CONCLUSOES	2230	3646	0.0	2225	3624	0.0
"6"	2227	3705	0.0	1474	1974	0.0
APLICACAO	2217	3533	0.0	1132	1556	0.0
ESTOU	2213	2242	0.0	837	842	0.0
PARANA	2213	2454	0.0	1067	1150	0.0
REGULAMENTOS	2201	2277	0.0	1085	1136	0.0
JUL	2194	2748	0.0	606	643	0.0
CORTE	2189	3542	0.0	818	1095	0.0
MAR	2189	2710	0.0	674	768	0.0
RECOMENDA-SE	2184	3638	0.0	1765	2590	0.0
FINAL	2173	3630	0.0	1308	1833	0.0
MAQUINA	2173	3157	0.0	816	1046	0.0
LEGISLACAO	2171	3124	0.0	1009	1305	0.0
S	2160	3487	0.0	1171	1563	0.0
"12"	2159	3374	0.0	1163	1620	0.0
EXISTEM	2149	3325	0.0	1533	2121	0.0
FOI	2143	3276	0.0	1370	1826	0.0
MAIO	2132	2736	0.0	715	806	0.0
END	2115	4029	0.0	554	678	0.0
ESSE	2113	3259	0.0	1265	1729	0.0
MATERIA-PRIMA	2088	3014	0.0	1073	1344	0.0
UMIDADE	2087	3696	0.0	904	1269	0.0
"19"	2075	3273	0.0	1022	1394	0.0
ISSO	2074	3289	0.0	1307	1816	0.0
ABAIXO	2058	3314	0.0	1615	2261	0.0
CUSTO	2046	3092	0.0	1079	1450	0.0
RELACAO	2046	3171	0.0	1273	1788	0.0
TRABALHO	2046	3007	0.0	1053	1406	0.0
QUALQUER	2040	3276	0.0	1474	2051	0.0
APOIO	2023	2663	0.0	1604	1990	0.0
PO	2018	2997	0.0	671	937	0.0
CRIACAO	2012	2697	0.0	483	571	0.0
CURSO	2008	2229	0.0	986	1081	0.0
ALUMINIO	2005	3022	0.0	506	690	0.0
ACORDO	1991	3340	0.0	1463	2066	0.0
PROJETO	1990	2847	0.0	1089	1445	0.0
DIAS	1987	3432	0.0	920	1240	0.0
PEQUENAS	1985	2790	0.0	1424	1810	0.0
MELHORIA	1982	2141	0.0	1024	1121	0.0
AV	1973	3417	0.0	1148	1588	0.0
REFERENCIAS	1972	2533	0.0	1723	2119	0.0
JUN	1971	2411	0.0	586	622	0.0
BOA	1970	3168	0.0	1224	1612	0.0
QUIMICOS	1965	2777	0.0	1124	1407	0.0
PRODUZIR	1960	2543	0.0	996	1305	0.0
R	1957	3116	0.0	942	1219	0.0
RECICLAGEM	1955	2635	0.0	403	466	0.0
COMERCIO	1940	3194	0.0	1078	1477	0.0
ANIMAIS	1934	3115	0.0	643	814	0.0
MISTURA	1931	3348	0.0	942	1307	0.0
FEITA	1928	3212	0.0	1250	1776	0.0
PRINCIPALMENTE	1925	3047	0.0	1305	1745	0.0
TODOS	1925	3143	0.0	1401	1971	0.0
TENHO	1920	1951	0.0	758	770	0.0
MG	1917	2900	0.0	1200	1465	0.0
ENERGIA	1915	2942	0.0	686	940	0.0
ALTA	1911	3250	0.0	1143	1586	0.0
GERAL	1895	3097	0.0	1303	1853	0.0
"05"	1891	2541	0.0	1023	1253	0.0

APÊNDICE B

Amostra da lista de palavras extraídas pelo BR/Search (Ordem alfabética)

Base de dados SBRT

Arquivo importado para o BR/Search

Lista extraída da base em 18/09/2007

Total de registros na base = 6.041 documentos textuais

Total de termos identificados pelo BR/Search: 149.624 termos

There are 149624 unique terms.

R1	()	1	occs,	1	docs
R2	AA	68	occs,	18	docs
R3	AAA	1	occs,	1	docs
R4	AAAA	1	occs,	1	docs
R5	AAAPIP	1	occs,	1	docs
R6	"AABAV.ASP"	2	occs,	2	docs
R7	AACC	1	occs,	1	docs
R8	AACD	5	occs,	1	docs
R9	AADEMICA	1	occs,	1	docs
R10	AAF	7	occs,	1	docs
R11	AAFÇO	2	occs,	2	docs
R12	AAFLOW	2	occs,	1	docs
R13	AAI	1	occs,	1	docs
R14	AALBORG	2	occs,	2	docs
R15	AAN	1	occs,	1	docs
R16	AANCIA	2	occs,	2	docs
R17	AANCIAS	1	occs,	1	docs
R18	AANDRELLO	2	occs,	2	docs
R19	AANIO	2	occs,	2	docs
R20	AAO	7	occs,	5	docs
R21	AAOCERT	2	occs,	2	docs
R22	"AAOCERT.COM.BR"	2	occs,	2	docs
R23	AARTEDAFIBRA	1	occs,	1	docs
R24	"AARTEDAFIBRA.SITES.UOL.COM.BR"	1	occs,	1	docs
R25	"AARTEDE.HTM"	2	occs,	2	docs
R26	AAS	1	occs,	1	docs
R27	AAT	1	occs,	1	docs
R28	AATCC	1	occs,	1	docs
R29	AAUTICOS	1	occs,	1	docs
R30	AAXTIL	4	occs,	4	docs
R31	AAXTILRECOMENDA	1	occs,	1	docs
R32	AA0CDFB4C6BDF0B083257279004756	3	occs,	1	docs
R33	"AA0006.HTM"	4	occs,	4	docs
R34	"AA0019.HTM"	1	occs,	1	docs
R35	"AA032503A.HTM"	1	occs,	1	docs
R36	"AA040902A.HTM"	1	occs,	1	docs
R37	"AA120797.HTM"	1	occs,	1	docs
R38	AA36	1	occs,	1	docs
R39	AA46	1	occs,	1	docs
R40	AA60	1	occs,	1	docs
R41	"AA.HTM"	1	occs,	1	docs
R42	AB	22	occs,	16	docs
R43	ABA	16	occs,	11	docs
R44	ABAC	2	occs,	2	docs
R45	ABACA	2	occs,	2	docs

R46	ABACAT	1	occs,	1	docs
R47	ABACATE	69	occs,	16	docs
R48	ABACATEIRO	11	occs,	9	docs
R49	ABACATEIROS	2	occs,	2	docs
R50	ABACATES	3	occs,	1	docs
R51	"ABACATE.HTM"	3	occs,	2	docs
R52	"ABACATE.PPT"	1	occs,	1	docs
R53	ABACAXI	180	occs,	55	docs
R54	"ABACAXIGODEMEL.COM.BR"	1	occs,	1	docs
R55	ABACAXIS	4	occs,	4	docs
R56	ABACAXIZAIS	1	occs,	1	docs
R57	ABACAXIZAO	1	occs,	1	docs
R58	ABACAXIZEIRO	10	occs,	3	docs
R59	ABACOS	1	occs,	1	docs
R60	ABAD	2	occs,	2	docs
R61	ABADE	1	occs,	1	docs
R62	ABADIA	1	occs,	1	docs
R63	ABAETE	2	occs,	2	docs
R64	ABAFADAS	1	occs,	1	docs
R65	ABAFADO	3	occs,	3	docs
R66	ABAFADOS	1	occs,	1	docs
R67	ABAFAMENTO	10	occs,	9	docs
R68	ABAFANDO	1	occs,	1	docs
R69	ABAFANDO-0	1	occs,	1	docs
R70	ABAFAR	4	occs,	3	docs
R71	ABAFE	2	occs,	1	docs
R72	ABAGA	2	occs,	2	docs
R73	"ABAGA.COM.BR"	1	occs,	1	docs
R74	ABAI	1	occs,	1	docs
R75	ABAIRA	1	occs,	1	docs
R76	ABAIX	1	occs,	1	docs
R77	ABAIXA	11	occs,	11	docs
R78	ABAIXADA	1	occs,	1	docs
R79	ABAIXADO	2	occs,	2	docs
R80	ABAIXADOS	2	occs,	2	docs
R81	ABAIXAM	1	occs,	1	docs
R82	ABAIXAMENTO	18	occs,	18	docs
R83	ABAIXANDO	1	occs,	1	docs
R84	ABAIXAR	23	occs,	20	docs
R85	ABAIXE	3	occs,	3	docs
R86	ABAIXO	2219	occs,	1719	docs
R87	ABAIXOS	1	occs,	1	docs
R88	ABAIXO-MENCIONADOS	1	occs,	1	docs
R89	"ABAI.ORG.BR"	3	occs,	2	docs
R90	ABAJOUR	1	occs,	1	docs
R91	ABAJUR	45	occs,	5	docs
R92	ABAJURES	5	occs,	4	docs
R93	ABAJURPESQUISAPORRESUMO	4	occs,	1	docs
R94	ABAL	19	occs,	12	docs
R95	ABALA	1	occs,	1	docs
R96	ABALADO	2	occs,	2	docs
R97	ABALAR	1	occs,	1	docs
R98	ABALE	1	occs,	1	docs
R99	ABALO	1	occs,	1	docs
R100	ABALONE	1	occs,	1	docs
R101	ABALOS	1	occs,	1	docs
R102	"ABAL.ORG.BR"	2	occs,	2	docs
R103	ABAM	10	occs,	8	docs
R104	ABAMECTIN	3	occs,	2	docs
R105	ABANANDO	1	occs,	1	docs
R106	ABANDONA	2	occs,	2	docs
R107	ABANDONADA	7	occs,	7	docs
R108	ABANDONADAS	11	occs,	11	docs

R109	ABANDONADO	4	occs,	4	docs
R110	ABANDONADOS	7	occs,	7	docs
R111	ABANDONAM	3	occs,	3	docs
R112	ABANDONANDO	2	occs,	2	docs
R113	ABANDONAR	6	occs,	6	docs
R114	ABANDONARAM	2	occs,	2	docs
R115	ABANDONA-LO	1	occs,	1	docs
R116	ABANDONO	8	occs,	6	docs
R117	ABANORTE	1	occs,	1	docs
R118	ABAP	3	occs,	2	docs
R119	"ABAP.COM.BR"	1	occs,	1	docs
R120	ABARCAM	1	occs,	1	docs
R121	ABARE	1	occs,	1	docs
R122	ABARROTAMENTO	1	occs,	1	docs
R123	ABARROTAR	1	occs,	1	docs
R124	ABAS	26	occs,	15	docs
R125	ABASTECA	1	occs,	1	docs
R126	ABASTECAM	1	occs,	1	docs
R127	ABASTECE	6	occs,	6	docs
R128	ABASTECEDORA	1	occs,	1	docs
R129	ABASTECEM	3	occs,	3	docs
R130	ABASTECENDO	4	occs,	4	docs
R131	"ABASTECENDO.ASP"	1	occs,	1	docs
R132	ABASTECER	20	occs,	19	docs
R133	ABASTECEREM	1	occs,	1	docs
R134	ABASTECE-LO	1	occs,	1	docs
R135	ABASTECIDA	3	occs,	3	docs
R136	ABASTECIDAS	1	occs,	1	docs
R137	ABASTECIDO	14	occs,	11	docs
R138	ABASTECIDOS	4	occs,	4	docs
R139	ABASTECIMENTO	456	occs,	286	docs
R140	ABASTECIMENTOS	1	occs,	1	docs
R141	"ABASTECIMENTO.EMPRESA"	2	occs,	1	docs
R142	ABASTECIMENTO-MAPA	1	occs,	1	docs
R143	ABASTESCIMENTO	1	occs,	1	docs
R144	ABAT	2	occs,	1	docs
R145	ABATE	339	occs,	145	docs
R146	ABATEATE	3	occs,	3	docs
R147	ABATEDOURO	133	occs,	46	docs
R148	ABATEDOUROS	74	occs,	33	docs
R149	ABATEM	3	occs,	2	docs
R150	ABATER	7	occs,	7	docs
R151	ABATEREM	1	occs,	1	docs
R152	ABATES	6	occs,	4	docs
R153	"ABATE.HTM"	1	occs,	1	docs
R154	"ABATE.PDF"	6	occs,	6	docs
R155	ABATE-LAS	1	occs,	1	docs
R156	ABATE-LOS	3	occs,	2	docs
R157	ABATIDA	4	occs,	4	docs
R158	ABATIDAS	7	occs,	7	docs
R159	ABATIDO	24	occs,	19	docs
R160	ABATIDOS	21	occs,	19	docs
R161	ABATIMENTO	11	occs,	11	docs
R162	ABAULADA	3	occs,	3	docs
R163	ABAULADAS	1	occs,	1	docs
R164	ABAULADO	3	occs,	3	docs
R165	ABAULADOS	2	occs,	2	docs
R166	ABAULAMENTO	3	occs,	3	docs
R167	ABAV	2	occs,	2	docs
R168	ABAXIAL	1	occs,	1	docs

APÊNDICE C

Amostra da Base de Trabalho do SBRT

Esta amostra faz parte da base de dados utilizada para a mineração do texto, a chamada Base de Trabalho. Foi extraída do arquivo geral da base, após vários processos: i) extração da base de dados SBRT original; ii) remoção de ruídos; iii) conversão para o formato txt.

ID	TEXTO
1	<p>RESPOSTA TÉCNICA Título Utilização de babaçu na produção de energia elétrica Resumo Utilização de babaçu na produção de energia elétrica Palavras-chave Palmeira; energia termoelétrica; óleo vegetal Assunto Geração de energia elétrica Demanda Utilização de babaçu na produção de energia elétrica Solução apresentada Alternativas para geração de energia elétrica a partir de biomassa De acordo com a reportagem Pacote Ecológico de Crescimento (3): Energia Elétrica da revista Ambiente em Foco formas de se aproveitar a energia da biomassa para a produção de energia elétrica. Uma delas usinas termelétricas e do calor gerado dentro desta para outras aplicações industriais dentro da usina (sistema denominado de co-geração de óleo vegetal e de coqueiros de palmeiras com as cascas de babaçu que já são jogadas fora anualmente. É pouco em termos absolutos outros locais do Brasil necessário plantá-las poderia-se obter uma enorme quantidade de óleos vegetais (que poderiam ser utilizados tanto no plano nacional de biodiesel quanto na geração de energia elétrica e alimentação) e subprodutos. E ainda teria-se as vantagens adicionais de empregar um maior número de pessoas (pois a coleta é manual no caso do babaçu) muito tempo para ocorrer naturalmente conforme o grau de estrago que tenha sido feito. 1 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - http://www.sbrt.ibict.br É importante salientar que a produção de gases do efeito estufa se a usina é movida pela energia de biomassa produzida a partir de plantas cultivadas é a mesma que foi absorvida pela fotossíntese durante o crescimento da planta. Geração de energia elétrica a partir do óleo de babaçu De acordo com Cavaliero C. & Silva E. energia elétrica nos sistemas isolados" (2000) desaconselham seu aproveitamento em nível nacional E. óleo de castanhas podres. O outro utiliza óleo de dendê (Freitas et al. município de Caruaru Amazonas a partir do óleo de andiroba. O projeto está em fase de extração do óleo e faz parte do Programa Trópico Úmido PTU (MCT/CNPq) alternativas motor Elsbett energia elétrica. O aditivo foi desenvolvido pela Companhia Schur da Alemanha e é misturado ao óleo de mamona na proporção de 0 consumo de óleo diesel. De acordo com a matéria da revista Ambientemfoco "Comunidade isolada em Rondônia começa a gerar energia de óleo de Babaçu" interior de Rondônia. Com o apoio de pesquisadores da Unir (Universidade Federal de Rondônia) própria comunidade. A uma distância de cerca de 40 quilômetros da rede pública de energia elétrica mais próxima equipamentos de geração e transferência de energia para um freezer e geladeira comunitários. De acordo com a matéria com a geração de energia uma vez que os alimentos a partir de um motor que gira através da combustão do óleo em energia Brasileiro de Respostas Técnicas - SBRT - http://www.sbrt.ibict.br 2 Obtenção do óleo vegetal de babaçu De acordo com Cavaliero C. & Silva E. sua obtenção ser utilizada como ração para animais e vegetais em motores à diesel causa grandes problemas em função de sua alta viscosidade e baixa volatilidade. A alta viscosidade é responsável pela pobre atomização e combustão incompleta do óleo injetores e nas paredes frias do cilindro (Ózaktas et al (1994) pode-se também utilizar uma mistura de diesel com até 30% de óleos vegetais em motores a diesel convencionais com relativo sucesso firma alemã DMS cilindro e no sistema de injeção do combustível permita se conhecer o real desempenho deste motor com este tipo de combustível. Para mais informações sobre extração de óleo de babaçu recomenda-se a leitura da resposta técnica "extração de óleo de babaçu". Disponível em: < http://sbrt.ibict.br/upload/sbrt3790.pdf >. Acesso em: 26 jul 2007. Benefícios sociais e ambientais De acordo com Cavaliero C. & Silva E. isolados" (2000) desenvolva e se mantenha por si só os salários população nas regiões como recuperação de áreas degradadas e maior taxa de infiltração pluvial ambientais atuais quanto à utilização dos combustíveis fósseis e sua relação com as mudanças climáticas globais das emissões de CO2 para a atmosfera. Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - http://www.sbrt.ibict.br 3 Segundo Freitas et al (1996) citado por Cavaliero C. & Silva E. localidade de Vila Campinas utilizado (30% do consumo energético) por óleo vegetal planta oleaginosa durante o seu crescimento http://sbrt.ibict.br/upload/sbrt3968.pdf >. Acesso em: 26 jul 2007. Conclusões e recomendações A utilização do óleo de babaçu em motores adaptados para funcionar apenas com óleos pode ser uma solução na geração de energia elétrica em pequenas comunidades. A utilização de óleo de babaçu em motores não adaptados requer a sua mistura com óleo diesel sítios ou fazendinhas e a utilização de placas fotovoltaicas para o aproveitamento da energia solar. Recomenda-se a leitura do artigo de Cavaliero C. & Silva E. elétrica nos sistemas isolados" (2000) para mais informações sobre produção de energia elétrica a partir de óleos vegetais. Disponível em: <http://www.proceedings.scielo.br/scielo.php?pid=MSC00000002200000200044&script=sci_arttext>. Acesso em: 26 jul 2007. Fontes consultadas AMBIENTEMFOCO. Pacote Ecológico de Crescimento (3): Energia Elétrica. Disponível em: < http://www.ambienteemfoco.com.br/?p=3475 >. Acesso em: 26 jul 2007. AMBIENTEMFOCO. Comunidade isolada em Rondônia começa a gerar energia de óleo de Babaçu. Disponível em: < http://www.ambienteemfoco.com.br/index.php?s=baba%C3%A7u&submit= >. Acesso em: 26 jul 2007. CAVALIERO C. & SILVA E. sistemas isolados (2000). Disponível em: <http://www.proceedings.scielo.br/scielo.php?pid=MSC00000002200000200044&script=sci_arttext>. Acesso em: 26 jul 2007. SBRT. Disponível em: < http://sbrt.ibict.br/ >. Acesso em: 26 jul 2007. Elaborado por Camila Gomes Victorino Nome da Instituição respondente USP/DT (Agência USP de Inovação / Disque-Tecnologia) Data de finalização 31 jul 2007 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - http://www.sbrt.ibict.br 4</p>
2	<p>RESPOSTA TÉCNICA Título Destinação de resíduos de amianto e pneus Resumo Destinação de resíduos de amianto e pneus Palavras-chave Amianto; aterro sanitário; pneu Assunto Tratamento e disposição de resíduos perigosos Demanda Destinação de resíduos de amianto e pneus Solução apresentada Destinação de resíduos de amianto De acordo com o jornal eletrônico do Ministério Público do Estado do Paraná em aterro sanitário destinado para lixo perigoso. A determinação integra uma resolução do Conama (Conselho Nacional de Meio Ambiente) colocação em aterros especiais. Telhas e demais objetos e materiais que contenham amianto são considerados perigosos saúde oriundos de demolições Saúde (OMS) que também indústria da construção civil sistemas de freio e embreagem. Aterro sanitário para lixo perigoso (classe D) Para receber informações sobre os endereços dos aterros sanitários mais próximos da região da empresa em questão recomenda-se o contato com a Diretoria de Meio Ambiente da FIESP (Federação das Indústrias do Estado de São Paulo) - http://www.sbrt.ibict.br Resíduos de pneus inservíveis Recomenda-se a consulta da resposta técnica "Destinação de pneus inservíveis" do SBRT (Serviço Brasileiro de Respostas Técnicas). Disponível em: < http://sbrt.ibict.br/upload/sbrt6537.pdf >. Acesso em: 27 jul 2007. Conclusões e recomendações O amianto pelo seu potencial cancerígeno e tóxico não pode ser</p>

- reutilizado ou reciclado e deve ser destinado para depósitos específicos para seu grau de toxicidade poderem ser recauchutados. Somente pneus inservíveis devem ter destinação apropriada <http://sbrt.ibict.br/upload/sbrt5882.pdf> >. Acesso em: 27 jul 2007. - Aproveitamento de pneus usados. Disponível em: < <http://sbrt.ibict.br/upload/sbrt5753.pdf> >. Acesso em: 27 jul 2007. - Recauchutagem de pneu. Disponível em: < <http://sbrt.ibict.br/upload/sbrt5135.pdf> >. Acesso em: 27 jul 2007. Recomenda-se: Leitura da reportagem "Conama classifica AMIANTO como lixo perigoso" distribuição e responsabilidade das empresas a respeito do amianto. Disponível em: < <http://www.midiaindependente.org/pt/blue/2007/02/374301.shtml> >. Acesso em: 27 jul 2007. Fontes consultadas FIESP. Disponível em: < <http://www.fiesp.com.br/> >. Acesso em: 27 jul 2007. Midiaindependente. "Conama classifica AMIANTO como lixo perigoso". Disponível em: < <http://www.midiaindependente.org/pt/blue/2007/02/374301.shtml> >. Acesso em: 27 jul 2007. Ministério Público do Estado do Paraná deve ir para aterro especial" <http://www.mma.gov.br> >. Acesso em: 27 jul 2007. SBRT. Disponível em: < <http://sbrt.ibict.br/> >. Acesso em: 27 jul 2007. Elaborado por Camila Gomes Victorino Nome da Instituição respondente USP/DT (Agência USP de Inovação / Disque-Tecnologia) Data de finalização 02 ago 2007 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> 2
- 3 RESPOSTA TÉCNICA Título Resistência da uréia à umidade Resumo Resistência da uréia à umidade. Palavras-chave Uréia; estocagem Assunto Armazenamento Demanda Resistência da uréia à umidade. Solução apresentada URÉIA E UMIDADE A uréia tende a volatilização quando submetida às condições ambientes ideais para a enzima uréase que catalisa a reação de hidrólise da uréia em amônia e dióxido de carbono - sítios de adsorção da amônia. Quando a uréia é alocada superficialmente em solos cobertos por palha catalisa a hidrólise da uréia para dióxido de carbono e amônio enzima em diferentes padrões de umidade (1970) modificado por TABATABAI (1982). IDENTIFICAÇÃO DO PRODUTO: Nome Comercial: Ureia Pecuária ESPECIFICAÇÕES TÉCNICAS: Nome: Fórmula Química: Teor de N: Teor de Biureto: Teor de Amônio: Umidade: Natureza física: Aparência: Solubilidade (em H2O): Pureza: Ureia H2 N-CONH2 46% mínimo 1 INFORMAÇÕES ADICIONAIS: PRINCIPAIS APLICAÇÕES: Pecuária: - Aplicação na pecuária como nutriente na composição de rações animais Ureia Fórmula Química H2N-CONH2 Teor de N 46% mínimo Natureza física: Granulado Aparência Branco Solubilidade(em H2O): 98% mínimo Pureza: 98% mínimo Fonte: Agrária Indústria e Comércio Ltda INFORMAÇÕES ADICIONAIS: Principais aplicações: Agricultura - Aplicação na agricultura como componente de fertilizantes grande variedade de formulações NPK específicas não são cedidas pelas empresas fabricantes do composto químico < http://www.serrana.com.br/f_boletins.asp?Tipo=f&id=12 >. Acesso em: 18 jul 2007. SANTOS Piracicaba Jan./Apr. 1994. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90161994000100019 >. Acesso em: 19 jul 2007. Elaborado por Camila Gomes Victorino Nome da Instituição respondente USP/DT (Agência USP de Inovação / Disque - Tecnologia) Data de finalização 20 jul 2007 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> 2
- 4 RESPOSTA TÉCNICA Título Reciclagem de CD Resumo Informação sobre reciclagem de CD Palavras-chave Aproveitamento de resíduo; CD; compact disk; DVD; digital vídeo disc; reciclagem Assunto Fabricação de produtos de material plástico Demanda O tipo de plástico utilizado na injeção de CD pode ser feito com material reciclado? Solução apresentada De acordo com o Prof. Hélio Wiebeck da Universidade de São Paulo reciclar CDs é um processo difícil e caro. Segundo o Prof. Hélio o policarbonato teria que ser tratado para atingir o grau de transparência necessária e que permita a leitura termos de custos citam-se o custo com o tratamento do policarbonato telefone (11) 3091-2221. Fontes consultadas Wiebeck finalização 07 ago 2007 1 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>
- 5 RESPOSTA TÉCNICA Título Lingüiça blumenau Resumo Fabricação lingüiça blumenau. Palavras-chave Embutido; lingüiça; lingüiça defumada Assunto Embutidos diversos de suínos (porcos); produção de Demanda Gostaria de saber a receita e o processo para a produção de lingüiça blumenau. Solução apresentada Segundo a INSTRUÇÃO NORMATIVA Nº. 4 açougue fresco; produto seco devendo ter o sabor picante característico da pimenta calabresa submetidas ou não ao processo de estufagem ou similar para desidratação e ou cozimento carnes suína Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> Lingüiça Toscana: É o produto cru e curado obtido exclusivamente de carnes suína e adicionado de ingredientes Separada animal). Nota: a CMS utilizada poderá ser substituída pôr carne de diferentes espécies de animais de açougue que o caracterizem Lingüiça Toscana Lingüiça de Carne de Peru Lingüiça de Carne de Frango Lingüiça Mista Lingüiça Tipo Calabresa Lingüiça Tipo Portuguesa Lingüiça Cozida de ... Paio Outros COMPOSIÇÃO de ... Ingredientes Obrigatórios: Carne das diferentes espécies de animais de açougue e sal. Ingredientes Opcionais: Gordura Água Proteína vegetal e/ou animal Açúcares Plasma Aditivos intencionais Aromas proteína agregada. Não sendo permitida a sua adição nas lingüiças toscana Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> Requisitos Características Sensoriais: São definidas de acordo com o processo de obtenção. Textura: Característica Sabor: Característico Odor: Característico Características Físico-Químicas FRESCAIS Umidade (máx) Gordura (máx) Proteína (min) Cálcio (base seca) (máx) Fatores essenciais de qualidade: . É proibido o uso de CMS (carne mecanicamente separada) em Lingüiças Frescais (cruas e dessecadas). . O uso de CMS em Lingüiças Cozidas Envoltórios naturais Envoltórios artificiais Embalagens plásticas ou similares Caixas 70% 30% 12% 0 aprovadas junto ao órgão competente. Higiene Considerações Gerais: - As práticas de higiene para a elaboração do produto recomenda-se estar de acordo com o estabelecido no: "Código Internacional Recomendado de Práticas de Higiene para os Produtos Cármicos Elaborados" (Ref. CAC/RCP 13 - 1976 (rev. 1 Alimentos" (Ref.: CAC/RCP 1 - 1969 (rev. 2 - 1985) - Ref. Codex Alimentarius Industrial e Sanitária de Produtos de Origem Animal" - Decreto nº 30691 pouco ácidos e Alimentos acidificados envasados". 3 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> . Após ter sido inspecionado a carne para Lingüiças manipuladas armazenadas e transportadas em locais próprios de forma que as Lingüiças estejam protegidas da contaminação e deteriorização. - As Lingüiças curadas e dessecadas de gênero não nocivos a saúde humana. - Critérios Macroscópicos/ Microscópicos: O produto não deverá conter substâncias estranhas de qualquer natureza. - Critérios Microbiológicos: O produto deve obedecer à legislação específica em vigor. Rotulagem Aplica-se o Regulamento vigente (Portaria nº 371 com a matéria-prima utilizada seus Ingredientes - Sal e Salmoura - SDA - Ministério da Agricultura e Abastecimento Para fazer dez quilos de embutido gramas de pasta de alho. Modo de preparo: Moer a paleta e toucinho juntos. Adicionar os temperos à carne moída (a pasta de alho deve ser diluída num copo d'água porco ou de boi. Tradicionalmente a lingüiça Blumenau tem o formato de uma ferradura. Cada uma com cerca de 30 centímetros. 4 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> O passo seguinte é amarrar a lingüiça. Aperta-se um pouquinho leve. Frans costuma utilizar eucalipto para a defumação. O Dossiê Técnico - Produção de Lingüiça Frescal produção de lingüiças. Conclusões e recomendações Recomenda-se a leitura do dossiê técnico mencionado acima 2000. Disponível em: <<http://extranet.agricultura.gov.br/sislegisconsulta/consultarLegislacao.do?operacao=visualizar&id=7778>>. Acesso em: 07 ago. 2007. Fontes consultadas Ministério da Agricultura de 31 de Março de 2000. Disponível em: <<http://extranet.agricultura.gov.br/sislegisconsulta/consultarLegislacao.do?operacao=visualizar&id=7778>>. Acesso em: 07 ago. 2007. Globo Rural. Lingüiça Blumenau. Disponível em: <<http://globo.ruraltv.globo.com/GRural/0> 2007 5 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>
- 6 RESPOSTA TÉCNICA Título Seringueira com café Resumo Consorciação café-seringueira Palavras-chave Seringueira; café; cafeicultura Assunto Cultivo de seringueira Demanda Tenho duvidas quanto a viabilidade técnica e financeira de plantio de seringueira com cultura de café irrigado intercalado. Solução apresentada CONSORCIAÇÃO DE SERINGUEIRA COM CULTURAS DE IMPORTÂNCIA ECONÔMICA Em seringais recém-implantados complementar a renda do produtor. A

consorciação com culturas anuais deve ser feita no início do desenvolvimento do seringal possam infestar o seringal e a competição entre as duas espécies não prejudique o desenvolvimento da cultura principal. Muitas são as vantagens da utilização de uma cultura intercalar diminuição dos custos de implantação do seringal até o terceiro ou quarto ano existe luminosidade suficiente nas entrelinhas deve obedecer aos seguintes preceitos: a. Proporção em relação às seringueiras; b. Adaptação às condições edafoclimáticas do local; 1 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> c. Deve suportar certo grau de sombreamento promovido pelo seringal das plantas de seringueira para facilitar os tratos culturais consorciada no seringal mesma densidade de plantio quando comparado ao plantio em fileiras simples. Com isso pois dessa forma o sol percorre a copa das plantas durante todo o dia. h. Quando o excesso de sombra não permitir a exploração de uma cultura intercalar é possível instalar o chamado sistema silvo pastoril pode-se manejar animais como carneiros o tronco da seringueira e possibilitam reduzir os custos de manutenção do seringal seringueira é utilizada na substituição de cafezais em decadência de forma gradual ao café deve ser feito em ruas alternadas colocando as mudas na projeção da copa da planta de café. No segundo esquema linhas e 2 espaçamentos de 1 x 3 m exploração de 4 qualidade; produção de frutos de maior tamanho; microclima mais ameno ao cafeeiro Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> noturna mais alta e diurna mais baixa; redução do ciclo biennial; menor incidência da seca dos ponteiros entanto sobre o plantio de seringueira e a consorciação de outras plantas com a seringueira. As demandas podem ser verificadas em: <<http://www.sbrt.ibict.br/upload/sbrt5702.pdf>>. Acesso em: 07 ago. 2007. <<http://www.sbrt.ibict.br/upload/sbrt3350.pdf>>. Acesso em: 07 ago. 2007. Conclusões e recomendações Recomenda-se a leitura do material selecionado acima e do Boletim Técnico "Conсорciação de Seringueira com Culturas de Importância Econômica". Fontes consultadas O Agrônomo. Boletim Técnico-Informativo do Instituto Agrônomo. CONSORCIAÇÃO DE SERINGUEIRA COM CULTURAS DE IMPORTANCIA ECONOMICA. Disponível em: <<http://www.iac.sp.gov.br/OAgronomico/511/informacoestecnicas/agr5116.htm>>. Acesso em: 07 ago. 2007. Serviço Brasileiro de Respostas Técnicas. Disponível em: <<http://www.sbrt.ibict.br>>. Acesso em: 07 ago. 2007. Elaborado por Adriana Sabino Sotolani Nome da Instituição respondente USP/DT (Agência USP de Inovação / Disque-Tecnologia) Data de finalização 07 ago. 2007 3 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>

- 7 RESPOSTA TÉCNICA Título Empilhadeiras Resumo Uso de cinto de segurança e extintor de incêndio nas empilhadeiras. Palavras-chave Equipamento de proteção; cinto de segurança; empilhadeira; segurança no trabalho Assunto Peças e acessórios para outros veículos; fabricação de Demanda Preciso de confirmação técnica e legal quanto à obrigatoriedade de cinto de segurança e extintor de incêndio nas empilhadeiras. Solução apresentada De acordo com Luiz Henrique G. Camargo da Clark produtos movidos à combustão e que produzem faíscas correm o risco de provocar incêndio motivo pelo o qual a legislação exige que fabricante de empilhadeiras agregue o extintor de incêndios como acessório ao produto. Cinto de segurança Segundo Luiz Henrique a empilhadeira estreita e por ser utilizada normas técnicas relacionadas com empilhadeiras 12/1991 9) NBR 11897 de 12/1991 10) NBR 12152 de 12/1991 11) NBR 153 de 1967 12) NBR 12149 de 12/1991 13) NBR 11896 de 12/1991 14) NBR 12146 de 12/1991 15) NBR 12148 de 12/1991 16) NBR 11893 de 12/1991 1 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> Estas normas só estão disponíveis mediante pagamento aplicadas. As normas podem ser obtidas no formato digital ou impresso (que é bem mais barato). A aquisição pode ser feita diretamente no site da ABNT. O Senai Centro de tecnologia Euvaldo Lodi consultadas ABNT Data de finalização 07 ago. 2007 2 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>
- 8 RESPOSTA TÉCNICA Título Ração de farelo de coco para frango e gado leiteiro Resumo Informações sobre ração de farelo de coco para frango e gado leiteiro. Palavras-chave Casca de coco; coco; farelo; fibra de coco; nutrição animal; ração Assunto Fabricação de alimentos para animais Demanda Procuo informações sobre ração de farelo de coco para frango e gado leiteiro. Solução apresentada O farelo de coco ou torta de coco é um subproduto da extração do óleo de coco energética e protéica na alimentação animal. Torna-se importante uma avaliação deste subproduto e seus efeitos no desempenho e rendimento da carcaça de frangos de corte. Ração é uma mistura de alimentos na forma farelada umidade inferior a 13%. Sua composição deve conter de 19% a 20% de proteína bruta (PB) material mineral e fibroso. A relação entre PB e NDT deve estar próxima de 1/3 falta de alimentos de rações. Desta forma nordestino assumindo posição importante como atividade geradora de emprego e renda polpa do coco ou amêndoa seca e é o produto de mais valor que o coqueiro fornece (Woodroof Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> final de proteína bruta de razoável qualidade e 10% a 12 % de fibra sendo que esta Centro de Ciências Agrárias da Universidade Federal da Paraíba de açúcar e fubá de milho nas primeiras horas. Após esse período foram ofertadas as rações experimentais e água "ad libitum" durante todo o experimento. No 10 dia de idade as aves foram vacinadas cálculos das rações foram feitos com base nas exigências da linhagem e na composição química dos alimentos metabolizável e 20 bicálcico mecânico mineral 5 de carcaça eviscerada aumentou. Já as percentagens de peito superiores em peito e asas. Ração concentrada para gado leiteiro De acordo com RODRIGUES FILHO exigências dos animais. No entanto Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> esquecido ingredientes Tabela 1. Alimentos tradicionais e alternativos para suplementação de gado leiteiro. Alimentos tradicionais Energético Milho Protéico Farelo de soja Tortas de coco e de babaça massa e raspa de mandioca sconcentrada.htm> Cálculos das misturas Existem diversos métodos de cálculo das misturas próxima da ideal. Exemplo de complementação dos nutrientes de uma pastagem A seguir matéria seca (MS) (Batista et al. 1986; Teixeira et al. 2000). O balanço entre os nutrientes fornecidos pela pastagem e os nutrientes necessários para apenas a manutenção da vaca considerada é apresentado na TAB. 2. Tabela 2. Balanço entre os nutrientes fornecidos pela pastagem de quicuío-da-amazônia e os necessários para apenas a manutenção da vaca considerada. 3 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> Itens Pastagem NDT (kg) (10x60 <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Leite/GadoLeiteiroZonaBragantina/paginas/sconcentrada.htm>> A Tabela 2 indica que P. O Senhor Vítor Albino da Embrapa Suíno e Aves - Serviço de Atendimento ao Cidadão [sac@cnpsa.embrapa.br] explica que em relação à solicitação fundamental saber soja ração adequada. Deve-se identificar portanto do coco para a produção de ração para aves. O Farelo de Coco alimentação das aves publicadas no Portal do Serviço Brasileiro de Respostas Técnicas (<<http://sbrt.ibict.br>>) informações sobre o tema solicitado. Basta digitar a palavra casca de coco busca. Informações sobre a fabricação de leite de coco. Disponível em: <<http://sbrt.ibict.br/upload/sbrt2331.pdf>> Informações referentes a fornecedores de equipamentos utilizados no processo de fabricação do coco ralado. Disponível em: <<http://sbrt.ibict.br/upload/sbrt5109.pdf>> Informações detalhadas sobre o coco-da-baía: peso médio coco verde Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> Disponível em: <<http://sbrt.ibict.br/upload/sbrt4015.pdf>> Gostaria de obter informações a respeito do aproveitamento industrial das cascas do coco da Bahia. Disponível em:<<http://sbrt.ibict.br/upload/sbrt1621.pdf>> Desejo informações sobre como fabricar ração de animais e quais equipamentos são utilizados? Disponível em: <<http://www.sbrt.ibict.br/upload/sbrt6305.pdf>> Quais equipamentos são necessários para iniciar uma fábrica de ração para bovinos com capacidade para produzir inicialmente 100 ton/mês? Disponível em: <<http://www.sbrt.ibict.br/upload/sbrt5317.pdf>> Fontes consultadas JÁCOME 2002. Disponível em: <http://www.ppg.uem.br/Docs/ctf/Zootecnia/2002/25_lanflio%20Jacome_Efeitos%20da%20incl_usao_060_02.pdf>. Acesso em: 03 ago. 2007. RODRIGUES FILHO <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Leite/GadoLeiteiroZonaBragantina/paginas/sconcentrada.htm>> . Acesso em: 03 ago. 2007. Elaborado por Isabela Aparecida Paim Borges Leal Nome da Instituição respondente Rede de Tecnologia da Bahia-RETEC/BA Data de finalização 01 ago. 2007 5 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>

- 9 RESPOSTA TÉCNICA Título Ciclones e lavadores de gases Resumo Informações sobre industrialização de ciclones e lavadores de gases. Palavras-chave Ciclone; lavador de gases; resíduo; industrialização Assunto Fabricação de aparelhos para filtrar e depurar gases Demanda Informações sobre a industrialização de ciclones e lavadores de gases causam um menor dano ambiental projeto dos sistemas e também da forma de operação e manutenção dos mesmos pode ser coletado previamente por meio de coletores centrífugos conhecidos como ciclones. Esses equipamentos coletam o particulado mais grosseiro e pode ser lavador sólidos de maior tamanho e elevado peso específico de uma corrente gasosa temperaturas Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> · Funcionamento O Ciclone (FIG. 1) paredes ascendente interna. O gás é descarregado axialmente pela saída inferior do ciclone unidades de pequenos diâmetros com grande aceleração centrífuga associada - Representação de um ciclone Fonte: Galvão Filho. Disponível em: <www.consultoriaambiental.com.br/artigos/21.pdf>. Acesso em: 02 ago. 2007. FIGURA 2 - Representação de um multiciclone Fonte: Galvão Filho. Disponível em: <www.consultoriaambiental.com.br/artigos/21.pdf>. Acesso em: 02 ago. 2007. · Eficiência Vantagens: baixa potência consumida abaixo de 5 um visualizado através da FIG.3: 2 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> FIGURA 3 - Esquema geral da construção de um ciclone Fonte: PERES convencional (FIG.4) FIGURA 4 - Relações geométricas de um ciclone convencional Fonte: MOREIRA Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> · Velocidade e aceleração do gás Quando o gás entra no ciclone · Eficiência de Captação Diversos autores investigaram a duração teórica do movimento das partículas no ciclone e fizeram uma previsão teórica do desempenho do equipamento. Várias expressões teóricas e semi-empíricas têm sido propostas para prever a eficiência de captação de um ciclone (GRAF. 1) MOREIRA considerado. Na prática feitas pelo gás no interior do ciclone (igual a 5); v = Velocidade de entrada do gás no ciclone baseada na área B.H (recomenda-se usar 15 m / s); 4 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> μ = Viscosidade do gás; ρ = Densidade do gás; S = Densidade do sólido. · Dimensionamento Da equação (01) temos: $N = 5 \cdot$ Marcha de Cálculo Depois de estabelecido o percentual da captação eficiência ciclone · Perda de Carga A queda de pressão num ciclone e também a perda de carga são expressas de forma mais conveniente em termos da pressão cinética nas vizinhanças da área de entrada do ciclone. 5 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> As perdas estão associadas: Ao atrito no duto de entrada; À contração/expansão na entrada; Ao atrito nas paredes; Às perdas cinéticas no ciclone; Às perdas na entrada do tubo de saída; Às perdas de pressão estática entre a entrada e a saída. Sheferd das partículas sólidas. $A_e =$ Área da entrada = $B \times H$ Área da saída = $.DS/2$ 4 Lavadores de gases Os lavadores são destinados à limpeza de gases via úmida lavadores disponíveis para as várias necessidades de controle da poluição do ar. Os lavadores promovem o abatimento de vapores ácidos Uma bomba d'água eleva o líquido de lavagem da piscina ao distribuidor no topo do leito de recheio. O líquido de lavagem desce por gravidade através do recheio deste recheio. Como o meio líquido possui mais afinidade com os poluentes do que com os gases Eficiência Por um lado altera e Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> · Materiais Os LGR normalmente são fabricados em polipropileno visualizado através da seguinte Resposta Técnica: SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. Lavador de gases. DisqueTecnologia/CECAE/USP informações teóricas básicas sobre o funcionamento e a construção de um ciclone e lavador de gases convencional. Mas química e de engenharia Quantidade de Movimento. Separação de sólidos e líquidos de gases: Planilha para cálculo do dimensionamento de ciclones. Disponível em: <<http://www.enq.ufsc.br/disci/eqa5313/cyclone.xls>>. Acesso em: 02 ago. 2007. Sugere-se ainda Respostas Técnicas no site do SBRT seguinte dissertação: PERES Ciclones. Disponível em: <www.ecotechsystem.com.br/frame_ciclone.htm> Acesso em: 02 ago. 2007. Galvão Filho IMAPA Indústria Metalúrgica Ambiental do Paraná LTDA. Disponível em: <http://www.imapa.com.br/prod_frames.htm>. Acesso em 02 ago. 2007. MOREIRA Química e Engenharia de Alimentos. Disponível em: <http://www.enq.ufsc.br/disci/eqa5313/Separacao%20Sol_Liq%20de%20Gases.htm>. Acesso em: 02 ago. 2007. PERES Disponível em: <<http://libdigi.unicamp.br/document/?code=vtls000317698>>. Acesso em: 02 ago. 2007. 7 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br> SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. Aproveitamento de resíduos da indústria de beneficiamento de mármore e granito e equipamentos usados no processo de separação. REDETEC 02 ago. 2007. Elaborado por Michele Macedo Moraes Nome da Instituição respondente Fundação Centro Tecnológico de Minas Gerais CETEC Data de finalização 04 ago. 2007 8 Copyright © Serviço Brasileiro de Respostas Técnicas - SBRT - <http://www.sbrt.ibict.br>
- 10 RESPOSTA TÉCNICA Título Processos de impressão Resumo Apresenta os processos de impressão serigráfica e tampográfica e as tintas utilizadas na impressão de objetos em silicone. Palavras-chave Impressão; impressão tampográfica; serigrafia; tampografia Assunto Fabricação de tintas de impressão Demanda Como pode ser feito o processo de impressão gráfica em borracha de silicone? Que tinta deve ser utilizada? Solução apresentada Serigrafia Serigrafia ou silk-screen é um processo de impressão no qual a tinta é vazada pela pressão de um rodo ou puxador através de uma tela preparada. A tela onde a matriz preparada com uma emulsão fotosensível é colocada sobre um fotolito permitindo a passagem da tinta pela trama do tecido tipos de materiais (papel forma mecânica (por pessoas) ou automática (por máquinas). A serigrafia caracteriza-se como um dos processos da gravura retirada de matéria da matriz. O processo se dá no plano estêncil Tampografia A impressão tampográfica é um sistema indireto de impressão que utiliza um clichê em baixo relevo. A imagem é transferida da matriz para o suporte através de uma peça de silicone denominado tampão. O tampão pode ter diferentes formatos o que Aplicações típicas incluem brinquedos ser de silicone ou não uma vez que a peça a ser impressa é de silicone Porto Alegre RS Tel.: 3028-6688 Home Page: <www.tinpel.com.br> Santini Suply Com.e Importação Ltda. R. São Carlos mais informações sobre o tempo de secagem de uma tinta vinílica recomenda-se a leitura da resposta técnica "Tempo de secagem de tinta vinílica" tampografia recomenda-se a leitura da resposta técnica "Tampografia" responsabilidade quanto à idoneidade dos fornecedores. Cabe ao empreendedor entrar em contato com os fabricantes para fornecer as especificações Copyright © - Serviço Brasileiro de Respostas Técnicas - <http://www.sbrt.ibict.br> 2 desejadas e optar por aquele que melhor atender as suas necessidades (qualidade <<http://www.sbrt.ibict.br/upload/sbrt6658.pdf>>. Acesso em: 03 ago. 2007. SERVIÇO BRASILEIRO DE RESPOSTA TÉCNICA. Tampografia. Disponível em: <<http://www.sbrt.ibict.br/upload/sbrt5287.pdf>>. Acesso em: 03 ago. 2007. WIKIPÉDIA. Tampografia. Disponível em: <<http://pt.wikipedia.org/wiki/Tampografia>>. Acesso em: 03 ago. 2007. WIKIPÉDIA. Sistemas de impressão. Disponível em: <http://pt.wikipedia.org/wiki/Sistemas_de_impress%C3%A3o#A_Serigrafia_28silk_screen.29>. Acesso em: 03 ago. 2007. Elaborado por Fabiana Freitas Bibliotecária Márcia Rodriguez Instrutora de Impressão Offset Flávia Santos Twardowski Pinto Nome da Instituição respondente SENAI/RS / CEP SENAI de Artes Gráficas Henrique D'ávila Bertaso Data de finalização 03 ago. 2007 Copyright © - Serviço Brasileiro de Respostas Técnicas - <http://www.sbrt.ibict.br> 3

APÊNDICE D

Lista de palavras não significativas extraídas da base de trabalho pelo SAS

a	contra	embora	há	no
abaixo	custa	entre	isso	noite
abril	cá	era	isto	nome
adeus	da	essa	janeiro	nos
adiante	daquela	essas	julho	nossa
agora	daquelas	esse	junho	nossas
agosto	daquele	esses	junto	nosso
ainda	daqueles	esta	já	nossos
algo	dar	estar	lado	nova
algumas	das	estas	local	novas
alguns	de	estava	logo	nove
ali	debaixo	este	longe	novembro
além	demais	estes	lugar	novo
ano	dentro	estive	lá	novos
anos	depois	estiveram	maio	num
antes	depressa	estiveram	maior	numa
ao	desde	estiveram	maioria	nunca
aos	dessa	estiveram	mais	não
apenas	dessas	estiveram	mal	nível
apoio	desse	estiveram	março	nós
após	desses	estiveram	mas	número
aquela	desta	estiveram	me	o
aquelas	destas	estiveram	meio	obra
aquele	deste	estiveram	menor	obrigada
aqueles	destes	estiveram	menos	obrigado
aqui	devagar	estiveram	meses	oitava
aquilo	deve	estiveram	meu	oitavo
as	deverá	estiveram	meus	oito
assim	dez	estiveram	mil	onde
atras	dezanove	estiveram	minha	ontem
através	dezasseis	estiveram	minhas	onze
atrás	dezassete	estiveram	minhas	os
até	dezembro	estiveram	momento	ou
aí	dezoito	estiveram	muito	outra
baixo	dia	estiveram	muitos	outras
bastante	diante	estiveram	máximo	outro
bem	direito	estiveram	mês	outros
boa	diz	estiveram	na	outubro
boas	dizem	estiveram	nada	para
bom	dizer	estiveram	naquela	para
bons	do	estiveram	naquelas	parece
breve	dois	estiveram	naquele	parte
cada	dos	estiveram	naqueles	partir
catorze	doze	estiveram	nas	paucas
cedo	duas	estiveram	nem	pela
cento	dá	estiveram	nenhuma	pelas
certamente	dão	estiveram	nessa	pelo
certeza	dúvida	estiveram	nessas	pelos
cima	e	estiveram	nesse	perto
cinco	ela	estiveram	nesses	pode
coisa	elas	estiveram	nesta	podem
com	ele	estiveram	nestas	poder
como	eles	estiveram	neste	pois
conselho	em	estiveram	nestes	ponto
		estiveram	ninguem	pontos

por	quarto	seu	terceira	vais
porque	quatro	seus	terceiro	vem
porquê	que	sexta	teu	vens
posição	quem	sexto	teus	ver
possivelmente	quer	sim	teve	vez
posso	quereis	sistema	tive	vezes
possível	querem	sob	tivemos	viagem
pouca	queremas	sobre	tiveram	vindo
pouco	quero	sois	tiveste	vinte
poucos	questão	somos	tivestes	vocês
primeira	quinta	sou	toda	vos
primeiras	quinto	sua	todas	vossa
primeiro	quinto	suas	todo	vossas
primeiros	quinze	são	todos	vosso
própria	quáis	sétima	trabalho	vossos
próprias	quê	sétimo	treze	vários
próprio	relação	tal	três	vão
próprios	sabe	talvez	tu	vêm
próxima	sabem	também	tua	vós
próximas	se	tanta	tuas	zero
próximo	segunda	tantas	tudo	à
próximos	segundo	tanto	tão	às
puderam	sei	tarde	têm	área
pôde	seis	te	um	é
põe	sem	tem	uma	és
põem	sempre	temos	umas	ele
qual	ser	tendes	uns	
quando	seria	tenho	vagarosament	
quanto	sete	tens	e	
quarta	setembro	ter	vai	