

HUGO CESAR HOESCHL, Post Doc. (ORG.)

ILLUMINIS

*Ética Digital  
e  
Tecnologia  
da  
Informação*



H695e

Ética digital e tecnologia da informação / Hugo Cesar Hoeschl (org.) – Florianópolis (SC) : Editora Digital Ijuris, 2004.

ISBN 85-89587-14-2

1. Ética Digital 2. Tecnologia da Informação 3. gestão do conhecimento 4. Sociedade da Informação. I. Autor II. Título.

CDD 025.04

Hugo Cesar Hoeschl, Post Doc. (org.)

# Ética Digital e Tecnologia da Informação

Florianópolis  
Editora Digital Ijuris  
Julho, 2004

Copyright © 2004 IJURIS

Editora: Tânia Cristina D'Agostini Bueno  
Editoração & capa: Marcos Carlson

ISBN: 85-89587-14-2

EDITORA DIGITAL IJURIS

Rua Lauro Linhares, 728 / 212 Trindade  
88036-002 Florianópolis SC BRASIL  
+55 48 3025-6609 / +55 48 234-6801 fax

[editoradigital@ijuris.org](mailto:editoradigital@ijuris.org)  
[www.ijuris.org/editora](http://www.ijuris.org/editora)

Biblioteca Nacional – no. 89587

# SUMÁRIO

<b><i>Prefácio</i></b>	<b>11</b>
------------------------	-----------



IBERAMIA  
VIII Conferencia Iberoamericana de Inteligencia Artificial  
(2002, Sevilla, España)  
[www.lsi.us.es/iberamia2002](http://www.lsi.us.es/iberamia2002)

<b>Alphathemis®</b> <b><i>Do Texto ao Conhecimento</i></b>	<b>17</b>
---	-----------

<b>Aplicando Raciocínio Baseado em Casos para Representar o Conhecimento de Decisões Tributárias em Sistemas de Recuperação de Informação</b>	<b>35</b>
---	-----------

<b>Pesquisa Contextual Estruturada para o Conselho de Segurança da ONU</b>	<b>53</b>
--	-----------

<b>UNL como Elemento Redutor das Barreiras Culturais nos Sistemas CSCW</b>	<b>73</b>
--	-----------



ENEMPRESA  
5º Encontro Nacional de Empreendedorismo  
(2003, Florianópolis)  
[www.ene.ufsc.br/enempre2003](http://www.ene.ufsc.br/enempre2003)

<b>Por Que as Pequenas Empresas Devem Tentar o e-Business?</b>	<b>91</b>
--	-----------



## CIBERÉTICA

II Simpósio Internacional de Propriedade Intelectual,  
Informação e Ética (2003, Florianópolis)

[www.ciberetica.org.br](http://www.ciberetica.org.br)

<b>Centro Nacional de Gestão de Negócios</b> <i>CENABIO</i>	<b>103</b>
<b>A Democracia na Era do Governo Eletrônico</b>	<b>117</b>
<b>Diretriz Jurídica da EAD no Ensino Superior</b>	<b>137</b>
<b>O Governo Eletrônico Respondendo às Propensões da Presença da Administração Pública no Ciberespaço</b>	<b>159</b>
<b>Hackers</b> <i>Atividade e Legislação no Brasil</i>	<b>173</b>
<b>O Impacto da Tecnologia da Informação</b> <i>Oportunidades x Exclusão</i>	<b>193</b>
<b>Inclusão Digital no Direito Brasileiro</b> <i>Direito Difuso</i>	<b>205</b>
<b>NECATI</b> <i>Núcleo Catarinense de Tecnologia da Informação</i>	<b>219</b>
<b>Portal do Consumidor Catarinense</b> <i>Um Modelo de Atendimento ao Cidadão no Âmbito do Governo Eletrônico</i>	<b>231</b>
<b>A Representação do Conhecimento para Criação de uma Central Inteligente para Padronização dos Procedimentos de Licenciamento Ambiental</b> <i>CIPPLA</i>	<b>245</b>

**Sistema Olimpo Web**  
*Tecnologia para Governo Eletrônico e Paz Mundial* **263**

**Usabilidade**  
*A Primeira Fronteira do e-Gov* **281**



SUCESU  
Congresso Nacional de Tecnologia da Informação  
e Comunicação (2004, Florianópolis)  
[www.sucesu2004.org.br](http://www.sucesu2004.org.br)

**Avaliação de Websites do Poder Judiciário Brasileiro** **295**

**Engenharia de Almas**  
*A Sincronicidade entre a Inteligência Artificial e a Gestão do Capital Intelectual, Social e Emocional das Instituições* **309**

**Ferramenta de Desenvolvimento de Software para Representação Visual de Redes de Relacionamento** **327**

**KMAI**  
*Gestão do Conhecimento com Inteligência Artificial* **341**

**Um Modelo de TI para Aplicar a Gestão do Conhecimento em Prefeituras** **355**

**Transparência Eletrônica**  
*A TI e a Informação sobre Negócios Públicos Internacionais* **369**



CATI  
Congresso Anual de Tecnologia de Informação  
(2004, São Paulo)  
[www.fgvsp.br/cati](http://www.fgvsp.br/cati)

## **Fórum Eletrônico**

***Modernização Tecnológica do Poder Judiciário Brasileiro***

**379**



## **DIVULGAÇÃO CIENTÍFICA ONLINE**

Matérias veiculadas no UOL/Conjur -  
Revista Consultor Jurídico - e Ijuris  
(janeiro a julho de 2004,  
São Paulo e Florianópolis)  
[conjur.uol.com.br](http://conjur.uol.com.br) / [www.ijuris.org](http://www.ijuris.org)



e-GOV

**Congresso publicará Sistema Olimpo na Europa**

**395**

85-89587-11-8

**Sai o ISBN para o novo ebook**

**399**

PROPRIEDADE INTELECTUAL

**CESUSC abre as portas para o NECATI**

**405**

INFORMÁTICA Y DERECHO

**Simpósio internacional tem catarinense no Comitê**

**411**

SUCESU 2004

**Ijuris bate recorde brasileiro de aprovações**

**419**

LANÇAMENTO

**Sai livro com tese de pós-doutoramento em e-Gov**

**427**



INCLUSÃO DIGITAL <b>Florianópolis terá Conselho e telecentros nos terminais</b>	<b>435</b>
GOVERNO ELETRÔNICO <b>Cientista brasileiro faz parte de referências no exterior</b>	<b>449</b>
SUCESSO NO SUCESU <b>Ijuris apresenta seis trabalhos em um único Congresso</b>	<b>455</b>
GOVERNO ELETRÔNICO <b>Cientista brasileiro ultrapassa 100 artigos publicados</b>	<b>471</b>
MODERNIZAÇÃO <b>Novo livro catarinense projeta a Justiça do futuro</b>	<b>479</b>



# Prefácio

"I only dream 'cause I am alive"  
**Janick Gers / Steve Harris**

"Forever trusting who we are"  
**James Hetfield / Lars Ulrich**

*Escrevo aqui algumas linhas para você, uma pessoa especial e que decidiu ler um livro em formato exclusivamente digital.*

*Aqui você vai encontrar uma série de trabalhos científicos desenvolvidos por uma equipe totalmente multidisciplinar, que atua nas áreas de Governo Eletrônico, Sociedade da Informação, Gestão do Conhecimento, Inclusão Digital, Ética Hacker, Eco-tecnologia e outros assuntos relacionados. Quando o time de pesquisadores do IJURIS decidiu lançar o seu 14<sup>o</sup> livro, alguns valores foram considerados fundamentais, como a gratuidade e a linguagem direta.*

*Trata-se de um premiado time de pesquisadores, que já foi agraciado com sólidos reconhecimentos nacionais, entre os quais:*

	<b>FENASOFT MAX AWARD</b>	<b>PRÊMIO CONIP</b>		<b>GOVERNO ELETRÔNICO (MOG, ABEP, FGV)</b>	
	2002	2003	2004	2003	2004
<b>Fórum Eletrônico</b>		Vencedor			
<b>i02</b>	Menção de Honra				
<b>KMAI- -SAEI</b>			Finalista		
<b>NECATI</b>			Segundo Lugar		
<b>Plataforma Ômega</b>					Vencedor (categoria B2B)
<b>Portal do Consumidor</b>		Finalista			
<b>Portal OBID</b>				Relevância Social	

*O IJURIS foi considerado Centro de Referência em Gestão do Conhecimento pelo MCT/FINEP, em 2002.*

*Diversas pesquisas do time foram divulgadas em veículos de mídia nacional, entre as quais destacam-se a Revista Veja, o Guia do Estudante, a Revista Istoé Dinheiro, a Gazeta Mercantil, a Agência Estado, a Revista Exame, a Revista Consultor Jurídico (do UOL), o Portal Universia, a Revista ComCiencia, o Portal Radiobras e o Observatório da Imprensa, entre outros.*

*Estou falando de pesquisadores que tiveram trabalhos selecionados para apresentação e publicação em países como França, China, Escócia, Alemanha, Índia, EUA, Noruega, Argentina, Espanha, Inglaterra, Chile, República Theca e Grécia, entre outros, nos seguintes eventos:*

- ICAIL '99 (Oslo), 2001 (St. Louis) e 2003 (Edimburgo);
- DEXA EGOV 2003 (Praga);
- UNL 2001 (Suzou) e 2002 (Goa);
- ICEIS 2003 (Angers) e 2004 (Porto);
- ICCBR 97 (Providence);
- ROBOCUP PanAm 2003 (Carnegie Mellon, Pittsburgh);
- I3E IFIP 2003 (Guarujá, WS6);
- COLLECTER 2003 (Santiago);
- BILETA 2003 (Londres);
- ENVIROINFO 2003 (Alemanha);
- IBERAMIA/IDEIA 2002 (Sevilha);
- IADIS 2004 (Madrid);
- WCC 2004 (Toulouse);
- ATINER 2004 (Athenas/Lefkada);
- JAIIO 2004 (Córdoba);

*Destaca-se, nesse cenário, a aprovação de nove papers selecionados para apresentação e publicação em uma única conferência internacional. Trata-se da ICAIL 2003, realizada na Edinburgh University, e esse número foi obtido somando-se todos os tipos de trabalhos do IJURIS que foram aprovados naquele forum científico. O fato materializa significativa realização internacional, pois a ICAIL é a conferência mais conceituada do mundo em sua área de atuação, sendo referendada pela Association for Computing Machinery - ACM, e todos os trabalhos foram selecionados por um comitê científico*

*altamente consistente, sempre através de blind review. Jamais um único grupo de pesquisa, de qualquer parte do mundo, havia obtido uma marca tão expressiva em número de trabalhos aprovados, e os trabalhos do IJURIS corresponderam a 15% da totalidade das pesquisas apresentadas durante a ICAIL 2003.*

*Depois de produzir tudo isso, é necessário tornar o conhecimento acessível a todos os interessados. Então, foi decidido editar esta obra, com uma coletânea dos principais artigos do grupo que foram publicados no Brasil, em importantes fóruns como o SUCESU, o CIBERETICA, o ENIA, o ENGEMA, o ENEGEP e o CATI, entre outros, e que têm conexão com relevantes trabalhos internacionais.*

*Essa iniciativa foi empreendida também com o objetivo de demonstrar que podemos buscar uma posição de maior reconhecimento no cenário internacional. Quando digo isso, não estou falando somente de produtividade científica. Estou falando de uma nova realidade que podemos construir para o Brasil.*

*Lembre-se, não existe soberania sem ciência e tecnologia. Vamos confiar na nossa energia edificadora, na nossa capacidade produtiva, no nosso poder de transformação, e vamos sonhar um sonho de liberdade plena e de autonomia real. Um sonho que o nosso País não pode deixar de sonhar. O grande sonho da redistribuição da maior riqueza da história da humanidade: o CONHECIMENTO.*

**Hugo Cesar Hoeschl, Post Doc.**  
Organizador





## IBERAMIA

VIII Conferencia Iberoamericana de Inteligencia Artificial  
(2002, Sevilla, España)

[www.lsi.us.es/iberamia2002](http://www.lsi.us.es/iberamia2002)





# **AlphaThemis®**

## ***Do Texto ao Conhecimento***

Hugo Cesar Hoeschl, Dr.  
digesto@digesto.net

Tânia Cristina D'Agostini Bueno, M.Sc.  
bueno@eps.ufsc.br

Andre Bortolon  
bortolon@eps.ufsc.br

Eduardo da Silva Mattos  
emattos@eps.ufsc.br

Marcelo Stopanovski Ribeiro  
marcelos@eps.ufsc.br

Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, SC, Brasil, 88040-000

### RESUMO

Nota-se uma crescente demanda pela extração de conhecimento provenientes de dados não estruturados, haja visto, que tais arquiteturas são as principais fontes de armazenamento de informações para os tomadores de decisões. O mercado está ávido por aplicativos que apresentem as melhores alternativas para dinamizar as suas tarefas diárias e minimizar o tempo na tomada de decisão. Neste artigo apresenta-se um aplicativo de extração de conhecimento de documentos textuais. Os documentos trabalhados são do domínio do direito, notadamente, as súmulas dos tribunais superiores. O aplicativo classifica, norteado por uma métrica de similaridade, as súmulas existentes na base, mediante algumas informações requisitadas pelo usuário. As informações são inseridas em

linguagem natural, ou seja, sem a necessidade de conectores lógicos. O artigo referencia-se pela bibliografia sobre mineração de textos.

## 1. INTRODUÇÃO

O artigo que segue versa sobre mineração de textos. Nele encontra-se, inicialmente, uma abordagem teórica para embasar o enquadramento da aplicação prática à conceituação científica do assunto. Com base na literatura referenciada faz-se a ligação entre a mineração de textos e a extração de informações, utilizando-se técnicas de inteligência artificial, notadamente parte do Raciocínio Baseado em Casos, a métrica de similaridade.

Na continuidade do escrito verifica-se, globalmente, que tipo de vantagens as aplicações de mineração de textos conferem a quem as usa e principalmente as necessidades geradoras das implantações. Neste tópico tratar-se-á de comparar entre si algumas ferramentas disponíveis no mercado, tentando se chegar a conclusões sobre as características principais esperadas de uma solução voltada para a mineração de textos.

Finalmente, faz-se um estudo da aplicação denominada AlphaThemis<sup>®</sup> ressaltando-se que características presentes na ferramenta sustentam sua classificação como pertencente ao universo da mineração de textos. Conceitos e usos de técnicas de classificação e geração de regras são discutidos visando explicar o uso da tecnologia descrita nesse campo.

O objetivo do trabalho é comprovar a hipótese que a tecnologia AlphaThemis<sup>®</sup>, apresenta-se como alternativa sólida para a mineração de textos no ambiente de dados não estruturados, servindo de aplicação neste tema para as mais diversas organizações, extrapolando o ambiente estritamente jurídico. O conhecimento contido no modo textual não estruturado pode ser extraído através das técnicas contidas neste novo sistema.

## 2. EMBASAMENTO TEÓRICO

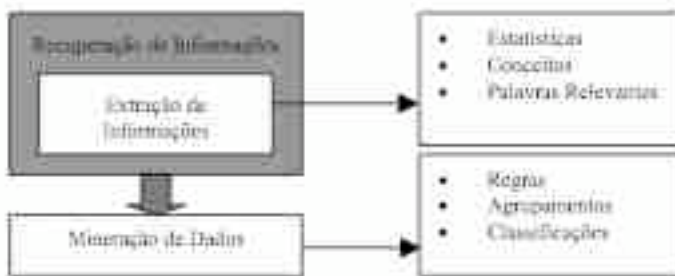
No levantamento de referências bibliográficas verificou-se o extenso material produzido pela empresa IBM, a qual considera a mineração de textos como a "aplicação da idéia de mineração de dados para textos não estruturados ou menos estruturados." Também, entende a mineração de dados como a possibilidade do "proprietário ou usuário dos dados obter novas visões e conhecimento procurando padrões nos dados os quais não seriam reconhecidos usando consultas tradicionais de dados e técnicas relacionadas. Estas técnicas permitem comparações para serem feitas através de dados vindos de muitas fontes de diferentes tipos, extraíndo informação que não poderiam ser óbvias ou mesmo visíveis para o usuário e organizar documentos e informação pelos seus assuntos ou temas." [11] Veja a Tabela 1:

	<b>Mineração de Dados</b>	<b>Mineração de Textos</b>
<b>Objeto de Investigação</b>	Dados categóricos e numéricos	Textos
<b>Estrutura de objeto</b>	Bases de dados relacionais	Textos em formato livre
<b>Objetivo</b>	Prever resultados de situações futuras	Recuperar informações relevantes, purificar o significado, categorizar o resultado
<b>Métodos</b>	Conhecimento de máquina: SKAT, DT, NN, GA, MBR, MBA	Indexação, processamento especial de redes neurais, lingüística, ontologias
<b>Tamanho atual de mercado</b>	100.000 analistas em grandes e médias companhias	100.000.000 de trabalhadores corporativos e usuários individuais
<b>Maturidade</b>	Ampla implementação desde 1994	Ampla implementação iniciando em 2000

**Tabela 1. Comparação entre Mineração de Dados e Mineração de Textos**

Melhor conceituação encontra-se junto ao *Text Mining Research Group* da Universidade de Waikato, desenvolvedores da ferramenta Weka, para quem a "Mineração de textos é a procura por padrões em um texto em linguagem natural e pode ser definido como o processo de análise do texto para extrair informação dele para um propósito em particular."

Diante dessa conceituação verifica-se que a mineração de textos possui duas fases principais e seqüentes: a extração de informações e a própria mineração de dados. A primeira destina-se a extrair conceitos, estatísticas e palavras relevantes de um conjunto textual para estruturá-los minimamente, preparando-os para a aplicação das técnicas de mineração de dados. Neste segundo momento aplicam-se as diretrizes e algoritmos de mineração de dados destinados a gerarem regras, classificações ou agrupamentos (Figura 1).



**Figura 1. Técnica de Mineração de Textos**

O mesmo laboratório da Universidade citada afirma que "A mineração de textos utiliza *recall* e precisão para medir a efetividade das diferentes técnicas de extração de informação, permitindo a realização de comparações quantitativas. Estes conceitos seguem os mesmos princípios da recuperação de informação."

A extração de informações cumpre um importante papel no processo de mineração de textos, visto que ele permite o

pré-processamento do corpo dos documentos textuais de maneira a facilitar o processo de mineração. Geralmente, o processo de extração de informações gera pares de atributos e valores que representam o conteúdo do documento. Normalmente, os atributos são um conjunto previamente definidos (Tabela 2).

<b>Categoria</b>	<b>Palavra-chave 1</b>	<b>Palavra-chave 2</b>	<b>Palavra-chave 3</b>	<b>Maior Valor (R\$)</b>
Aplicações de Intranet	Título do artigo	Aplicações web	Avaliação de Rede	41.000,00
Política de pessoal	Licença Maternidade	Benefícios de saúde	Licença Paternidade	150,00
Informações sobre o Lotus Notes	Replicação do Banco de Dados	e-mail	Aplicações de colaboração	90,00
Informações para viajantes	Linhas de ônibus	Opções de teletrabalho	Estações de trem	2,50
Ergonomia do Escritório	Descanso do pulso	Curvatura da coluna	Reconhecimento de voz	208,00

**Tabela 2. Exemplificação da extração de conceitos em textos usando o Intelligent Miner da IBM**

A mineração de textos deve, também, preocupar-se com os problemas básicos de processamento de linguagem natural, tais como: erros de ortografia e digitação, problemas de normalização de palavras (variações de gênero, número e flexões verbais) e ambigüidades semânticas. Novamente a explicação do *Text Mining Research Group* é elucidativa. "A mineração de textos reconhece que o completo entendimento de linguagem natural é um objetivo que não pode ser atingido imediatamente e foca a extração de uma pequena quantidade de informação. Esta informação extraída pode ser o autor, o título e a data de publicação de um artigo, as siglas definidas em um texto ou os artigos mencionados em uma bibliografia."

Embora a mineração de textos possa trabalhar com qualquer tipo de informação, esta deve seguir alguns critérios quando da visualização pelo usuário:

1. A informação recuperada deve ter valor para o operador do conhecimento. A mineração de textos deve fornecer novas visões sobre os textos para os operadores do conhecimento. Isto permite a adição de valores à base de conhecimento;
2. O conteúdo deve ser baseado em texto. Para dados numéricos as tecnologias existentes estão melhor desenvolvidas;
3. O valor da mineração de textos é diretamente proporcional ao valor do dado que se está minerando. Quanto mais importante for o conhecimento contido na coleção de textos, mais valiosa será a mineração;
4. O conteúdo deve estar explicitamente declarado no texto, tais como textos técnicos e científicos;
5. A mineração de textos é mais valiosa quanto menos estruturado for o texto. Textos desorganizados, contendo acúmulo de informações são os mais indicados.

Complementar à discussão sobre mineração de textos aparece a abordagem semelhante da mineração na *world wide web*. Esta ocupa-se especialmente da procura de indicadores, conceitos e hipertextos contidos em páginas www para a identificação de padrões ou extração de uma análise gráfica do conteúdo dos textos.

## **2.1 RBC como Algoritmo de Recuperação**

A técnica de Raciocínio Baseado em Casos (RBC) encontrada no campo da Inteligência Artificial (IA), notadamente a parte

do estudo que trata da métrica de similaridade é aqui usada para a comparação de textos.

Segundo Tânia Bueno, "o raciocínio baseado em casos constitui uma poderosa ferramenta de utilização das mais variadas experiências humanas. É muito usual, na história da humanidade, buscar-se uma experiência anterior para auxiliar na resolução de um problema atual, fazendo do passado um imenso banco de informações que nos ajuda a pensar sobre o presente e o futuro.(...)" A autora ainda completa o raciocínio ressaltando a utilidade da técnica no universo das ciências jurídicas, escrevendo: "O processo característico do RBC consiste em: identificar a situação atual, buscar a experiência mais semelhante na memória e aplicar o conhecimento desta experiência passada na situação atual. O RBC é uma técnica da IA muito adequada para a modelagem do conhecimento jurídico."

O sistema que será objeto da análise principal deste artigo utiliza parte da teoria de RBC no processo de classificação. A aplicação de uma métrica de similaridade permite que o sistema esteja apto a extrair o conhecimento dos documentos, verificando assim quais que respondem à consulta feita pelo usuário.

A métrica de similaridade foi construída respeitando as características do documento e seguindo o padrão de outras métricas tradicionais utilizadas em sistemas de Recuperação de Informação, principalmente o vizinho mais próximo.

### **3. ESTADO DA ARTE**

#### **3.1 Mercado**

O avanço das tecnologias disponíveis para o trabalho com informações superou a abordagem de dados armazenados para a junção destes em informações e, mais recentemente,

com os processos de mineração de dados ressalta-se a descoberta do conhecimento nessas informações.

Dados da empresa Megaputer Intelligence indicam que 99% das informações analíticas na web existem em forma de textos, falam ainda que a web é o lugar onde rotineiramente os usuários encontram novos textos e finalmente, que 99% do comércio eletrônico atual não aumenta sua vantagem competitiva fornecida por seus websites ricos em conteúdo, porque não utilizam a mineração de textos.

A produção de informações não estruturadas, especialmente textos, supera amplamente a parte armazenada em banco de dados.

Exemplos citados pela IBM indicam a variedade de informações disponíveis nesse formato:

- E-mails de clientes, contendo informações de *feedback* sobre produtos e serviços;
- Documentos na intranet, tais como memorandos e apresentações as quais incorporam a expertise corporativa;
- Relatórios técnicos descrevendo novas tecnologias;
- Novas correntes contendo informações sobre o ambiente de negócios e as atividades dos concorrentes.

Provavelmente o conhecimento de maior valor agregado possuído por uma organização está contido em informações não estruturadas, vide os setores de inteligência empresarial estratégica, por exemplo. Essa constatação sustenta uma previsão de amplo crescimento técnico e comercial das ferramentas que se propõem a fornecer soluções nessa área,



justificando fortes investimentos em pesquisa e desenvolvimento por parte de empresas e instituições tecnológicas.

"A Pesquisa Forrester previu que os dados não estruturados, tais como texto, tornar-se-ão o tipo predominante de dado a ser armazenado. Isto significa uma grande oportunidade: fazer mais efetivo o uso de repositórios das comunicações de negócios e outros dados não-estruturados pelo uso da análise por computador. Mas, o problema é que o texto não foi criado para ser usado por computadores. Diferente da informação tabular tipicamente armazenada em bases de dados, os documentos têm uma estrutura interna limitada, se tiver. Além disso, a informação importante contida no texto não está explícita, mas inserida dentro do texto. Devido a isso, aparece a metáfora de "mineração" - a informação redescoberta pelo computador que estava codificada no texto pelo seu autor." [8]

## 3.2 Ferramentas de Mineração de Textos

A análise das ferramentas presentes no mercado privilegiou três aplicações:

### 3.2.1 TextAnalyst - Megaputer

TextAnalyst é uma ferramenta para análise semântica, navegação e procura em textos não-estruturados. Está disponível como:

- Aplicação *standalone*
- SDK de Componentes COM para fácil integração

Funcionalidade do TextAnalyst:

- Purificação do Resultado (Rede Semântica);
- Navegação;
- Sumarização;
- Explicação em Tópicos;
- Agrupamento;

- Focalização dinâmica;
- Categorização (TextAnalyst COM).

### 3.2.2 Semio

Características da suíte de informação:

- Extração automática dos conceitos chaves de grandes volumes de conteúdos baseados em textos;
- Estruturação das informações em categorias lógicas, traçando conexões entre os conceitos;
- Fornece ao usuário acesso à informação estruturada por meio de interfaces amigáveis.

### 3.2.3 Intelligent Miner for Text - IBM

Segundo a descrição da fabricante o produto oferece aos integradores de sistema, fornecedores de solução e desenvolvedores de aplicação uma ampla variedade de sofisticadas ferramentas de análise de texto, componentes de recuperação *full text* e ferramentas de acesso à web para enriquecer as soluções de inteligência de negócios. Pode-se construir uma grande variedade de aplicações - informações classificadas vindas de fontes de notícias; análises de portfólios de patentes, cartas de admissão de clientes e páginas web de competidores; permitir suporte à decisão; e conduzir buscas direcionadas na intranet.

O Intelligent Miner for Text inclui:

- Ferramentas de análise de texto;
- Um avançado mecanismo de busca;
- Ferramentas de acesso à web.

As ferramentas de análise de texto, em conexão com o mecanismo de busca avançado ou com as ferramentas de acesso à web, fornecem os elementos-chaves para desenvolver soluções de mineração de textos.

Muitas das ferramentas no Intelligent Miner for Text podem ser vistas como metadados sobre os documentos. Metadados são os dados estruturados que podem ser armazenados em uma base de dados e poderia ser usado objetivando a mineração de dados.

## 4. APLICAÇÃO

Pesquisas em inteligência artificial e inteligência jurídica conduzidas pela Web Intelligence Systems S.A. geraram uma aplicação inovadora denominada AlphaThemis<sup>®1</sup>.

Consiste em um sistema inteligente para a extração do conhecimento contido nas súmulas dos tribunais nacionais, configurando-se em uma potente ferramenta de mineração de textos destes documentos norteadores de decisões judiciais. A técnica responsável pela recuperação de súmulas empregada no aplicativo é derivada do desejo de se mesclar uma pesquisa estruturada em texto e atributos que pudessem ensejar a formação de um contexto (próprio para o sucesso da busca).

A modelagem conferida ao software tem como idéia básica tratar o conhecimento contido nas súmulas e seus respectivos enunciados de forma tal que a interface do sistema permita ao usuário formular uma consulta na qual seja possível mesclar atributos fixos com texto livre (Figura 2).

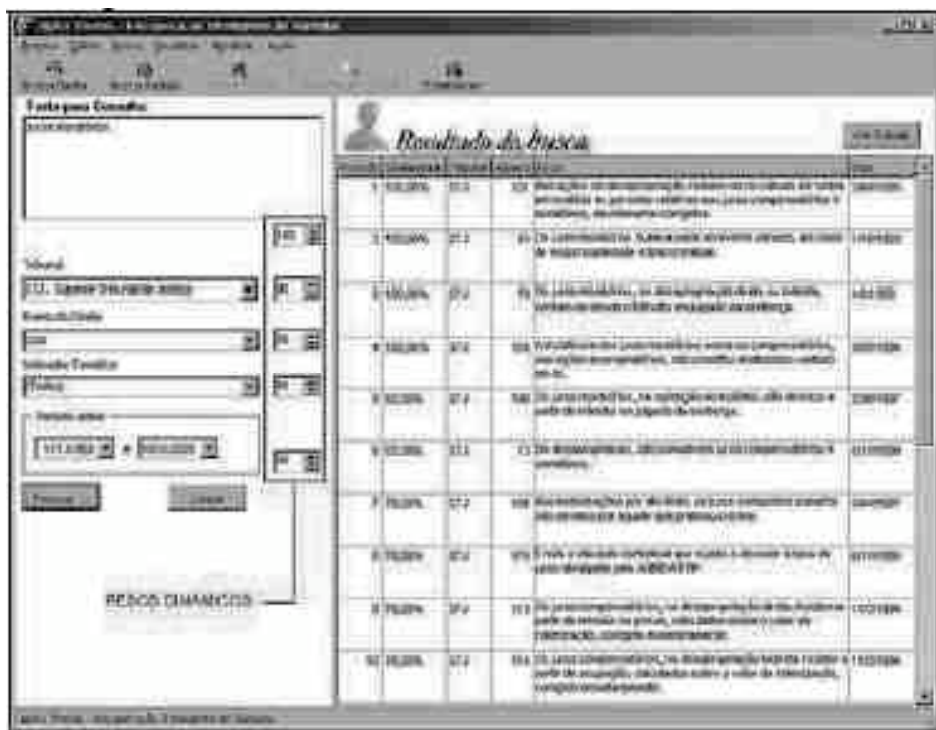


**Figura 2. Identificação de conceitos**

---

<sup>1</sup> Todos os direitos reservados da marca e uso da aplicação reservados para a Web Intelligence Systems S.A. Copyright 2001.

O uso de pesos dinâmicos nos filtros de busca torna o AlphaThemis<sup>®</sup> ainda mais flexível na formulação da pesquisa. Trata-se da possibilidade de alteração dinâmica da relevância dos atributos dentro do contexto pesquisado, calibrando-se, individualmente, os pesos de cada um dos atributos, dentro de uma escala de 0 a 100 (Figura 3).



**Figura 3. Interface AlphaThemis<sup>®</sup>, mostrando a saída dos dados e ressaltando o uso de pesos dinâmicos**

O uso de pesos dinâmicos nos filtros de busca torna o AlphaThemis<sup>®</sup> ainda mais flexível na formulação da pesquisa. Trata-se da possibilidade de alteração dinâmica da relevância dos atributos dentro do contexto pesquisado, calibrando-se, individualmente, os pesos de cada um dos atributos, dentro de uma escala de 0 a 100.

O resultado da pesquisa é apresentado ao usuário em uma interface de saída que mostra os resultados hierarquizados de acordo com o percentual de similaridade constatado entre a consulta e os enunciados das súmulas. O sistema AlphaThemis® apresenta em sua interface de saída, além do texto da súmula recuperada, todos os mecanismos que o usuário acionou na realização da consulta, bem como os valores respectivamente atribuídos (a fim de que o usuário possa checar quais atributos e valores influenciaram na escolha do registro).

O sistema AlphaThemis® conta também com um módulo estatístico e um módulo específico onde podem ser visualizados cada um dos enunciados das súmulas, diretamente.

A forma de representação e extração de conhecimento usada no desenvolvimento do AlphaThemis® foi uma adequação de técnicas de Recuperação de Informação e de Inteligência Artificial (IA) com o conhecimento jurídico do especialista, o que permitiu o desenvolvimento de um eficiente sistema de busca contextual.

## **4.1 AlphaThemis® como Minerador de Textos**

A constatação do aplicativo como uma ferramenta de mineração de textos encontra substrato no próprio processo de mineração de textos. Inicialmente ocorre uma fase de extração de informações do texto da súmula, a qual alimentará a base de conhecimento. Algo semelhante a essa extração ocorre dinamicamente durante o funcionamento do programa, após cada demanda de busca ser solicitada, o sistema faz uma extração para utilizar o resultado na comparação com a base anterior.

### **4.1.1 Características do texto**

Na identificação das características do texto é imprescindível a participação de especialistas no domínio do conhecimento jurídico.

### 4.1.2 Extração das informações

Como técnica de extração de informações identifica-se valores retirados pela localização dos conceitos e palavras no texto, mediante a comparação com o vocabulário previamente construído pelos especialistas do domínio, bem como pela inferência de informações via contexto da súmula.

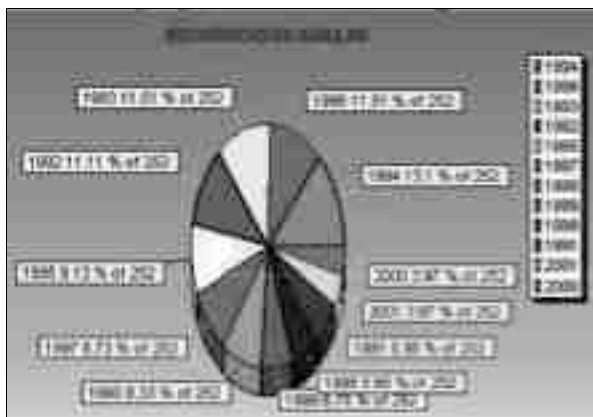
### 4.1.3 Conceitos

Ressalta-se que um dos pontos fortes dessa argumentação é a tarefa executada pelo programa que permite extrair conceitos subjetivos não diretamente explicitados no texto. A junção de várias palavras pode gerar conceitos diferenciados daqueles representados pelas palavras isoladamente.

O principal item no qual as súmulas podem ser agrupadas é o ramo do Direito ao qual elas pertencem. Ramo do Direito é o ramo doutrinário do assunto tratado na súmula, (por exemplo, direito tributário, direito constitucional etc). A totalidade dos ramos no domínio compreende em torno de 20. Infere-se os conceitos tratados na súmula a partir dos termos existentes na mesma. Verifica-se, então, em qual ramo do Direito estes conceitos se encaixam.

Os conceitos podem possuir graus de pertinência a mais de um ramo. A pertinência da súmula a um certo ramo do Direito é determinado pelo somatório dos graus de pertinência dos conceitos aos ramos. O ramo que sair vencedor neste somatório é considerado o ramo ao qual a súmula pertence. Entretanto, os demais ramos não são desprezados.

Após a extração de informações o aplicativo consegue aplicar um tratamento estatístico sobre a base resultante, levantando uma série de informações não visíveis no texto original. Um exemplo pode ser visualizado no Gráfico 1.



**Gráfico 1. Resultado da produção anual de súmulas do Superior Tribunal de Justiça**

## 5. CONCLUSÃO

Identifica-se claramente a mineração de textos como uma fronteira tecnológica, onde as aplicações estão em fase inicial e os conceitos não completamente definidos. No mercado as empresas procuram consolidar excelências no fornecimento de produtos para esse fim e as indicações de crescimento e importância do segmento são claras.

Novas tecnologias possuem a inovação como fundamento e o campo apresenta-se aberto para alternativas criativas que ampliem as possibilidades das ferramentas atuais. A incorporação da abordagem de Inteligência Artificial figura como norte estratégico nessa área, visto as características não estruturadas do conhecimento em foco.

O AlphaThemis<sup>®</sup> aparece como indicação inovadora nessa área, configurando-se como uma ferramenta de busca e mineração de textos própria para o conhecimento jurídico, porém com a tecnologia flexível para aplicação em outros campos.

Inserido no processo definido de mineração de textos o aplicativo descrito inicia seus trabalhos extraindo informações do texto das súmulas, sejam explícitas ou implícitas, por técnicas de localização ou inferência.

Segue essa atividade de forma dinâmica a cada nova entrada na base ou a partir de demandas da consulta. Executa, conseqüentemente, a mineração de dados, extraindo estatísticas e agrupando conceitos.

Como continuidade dos estudos prevê-se a geração de regras de decisão para a extração de conhecimentos ainda não identificados.

Conclui-se que a evolução deste programa acompanhará o andamento tecnológico da mineração de textos, propondo inovações a partir de aplicações reais da pesquisa científica.

## 6. REFERÊNCIAS

- [1] ANANYAN, S. Text mining applications and technologies. Disponível em [www.megaputer.com/down/tm/Text\\_Mining.pps](http://www.megaputer.com/down/tm/Text_Mining.pps). Acesso em 30 jan. 2002.
- [2] BUENO, T. C. D. Recuperação da informação jurídica em Sistema Baseado em Casos. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 1999.
- [3] BUENO, T. C. D.; HOESCHL, H. C. Representação e extração do conhecimento das súmulas do STJ. Florianópolis, 11 p. Trabalho não publicado. Acesso restrito para Web Intelligence Systems S.A.
- [4] IBM. IBM intelligent miner for text. In: IBM Business Intelligence Solutions CD. USA, 1998. CD-ROM.
- [5] NAHM, U. Y.; MOONEY, R. J. Text mining with information extraction. In: AAAI 2002 Spring Symposium on Mining Answers



from Texts and Knowledge Bases. Proceedings... Stanford, American Association for Artificial Intelligence, 2001.

- [6] SEMIO. Semio products and solutions overview. Disponível em [www.semio.com/products/overview.html](http://www.semio.com/products/overview.html). Acesso em 30 jan. 2002.
- [7] SEMIO CORPORATION. Text mining and the knowledge management space. Disponível em [www.dmreview.com/master.cfm?NavID=61&WhitePaperID=80](http://www.dmreview.com/master.cfm?NavID=61&WhitePaperID=80). Acesso em 30 jan. 2002.
- [8] TKACH, D. Text mining technology: turning information into knowledge. A white paper from IBM. In: IBM Business Intelligence Solutions CD. USA, 1998. CD-ROM.
- [9] TEXT MINING RESEARCH GROUP AT THE UNIVERSITY OF WAIKATO. Text mining. Disponível em [www.cs.waikato.ac.nz/~nzdl/textmining](http://www.cs.waikato.ac.nz/~nzdl/textmining). Acesso em 30 jan. 2002.
- [10] WEB INTELLIGENCE SYSTEMS S.A. Documentos de registro da tecnologia AlphaThemis. INPI, 2001.
- [11] WOHL, A. D. Intelligent text mining creates business intelligence. In: IBM Business Intelligence Solutions CD. USA, 1998. CD-ROM.



# Aplicando Raciocínio Baseado em Casos para Representar o Conhecimento de Decisões Tributárias em Sistemas de Recuperação de Informação

Alexandre Serra Barreto  
fciclone@bol.com.br

Tânia Cristina D'Agostini Bueno, M.Sc.  
bueno@eps.ufsc.br

Hugo Cesar Hoeschl, Dr.  
digesto@digesto.net

Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, SC, Brasil, 88040-000

## ABSTRACT

*The Secretaria da Receita Federal (SRF) has needing for information systems so as to deal with the large amount of information generated by its public activities. Among these activities is comprehended the administrative judgment of taxpayer consults. One of the consequences of this judgment process is the possibility of different administrative decisions about the same topics all over the country. These decisions have to be pacified by a specific division at the federal capital. In some cases these decisions could have been avoided if the civil servant had accessed well-succeeded specific judgments before. So as to allow this access, this paper presents a case based model which goal is the retrieval and indexing of tax decisions. Despite its conception to deal with consult solutions evolving a specific Social Contribution, its concepts can be adjusted to any taxes and decisions emitted by SRF.*

## 1. INTRODUÇÃO

Um dos momentos em que o Poder Executivo exerce funções assemelhadas às de outros poderes instituídos da República ocorre quando do julgamento de processos administrativos tributários, realizados pela Secretaria da Receita Federal (SRF), quando envolvem tributos de competência ou administrados pela União. No contexto desses julgamentos, muito esforço é empregado em situações complexas, na qual a experiência anterior de um especialista, no caso um servidor da SRF, pode ser decisiva para uma solução de qualidade nas novas demandas formuladas. Esse esforço é ainda maior quando duas decisões divergentes ocorrem em diferentes jurisdições fiscais, e a divergência tem de ser dirimida por repartição da SRF competente em Brasília. Isso, em certos casos, pode ser evitado se forem consultadas, previamente à decisão final, decisões similares bem sucedidas anteriormente.

O objetivo geral desse trabalho é, então, estudar e identificar princípios teóricos que possam ser a base de sustentação da recuperação inteligente de decisões tributárias no contexto dos julgamentos já citados. O objetivo específico é, por meio da análise de uma coleção de decisões tributárias emitidas pela SRF e pertinentes a um tributo específico eleito, estabelecer critérios, parâmetros e índices que possam servir de maneira adequada à recuperação inteligente de decisões, com base na similaridade entre elas. Adicionalmente, esse estudo pode servir de base para a agregação de conhecimento jurídico tributário a sistemas prévia ou futuramente existentes.

Esse artigo está estruturado em um *abstract* e quatro seções, sendo a primeira esta introdução. Na segunda seção, será apresentada a metodologia de pesquisa aplicada que norteou o presente trabalho. Posteriormente, a terceira seção aborda o desenvolvimento do estudo, no qual são abordadas as técnicas de Inteligência Aplicada empregadas no desenvolvimento da modelagem proposta. A quarta seção apresentará

as conclusões do trabalho, seguindo-se das referências bibliográficas.

## 2. METODOLOGIA

### 2.1 Inteligência Aplicada e RBC

Existem diversas técnicas de Inteligência Aplicada em franco desenvolvimento no mundo científico atual. Dentre elas, o RBC tem obtido flagrante sucesso quando presentes determinadas condições no problema estudado. De acordo com Bueno et al [1], as situações favoráveis a essa solução são a formatação textual homogeneizada em grau relevante, a estrutura retórica padronizada e a necessidade de simplificação da linguagem do domínio. De acordo com Slade (*apud* Vargas [2]), RBC é um modelo para a geração de raciocínio fundamentado na visão de que uma porção significativa da resolução de problemas humanos envolve a recuperação de experiências anteriores. A idéia central do RBC é, portanto, ajustar soluções que foram usadas para resolver antigas situações e utilizá-las na resolução de situações novas (Azevedo, Brasil e Oliveira [3]). A abordagem do RBC tem sido especialmente útil nas situações em que é muito difícil o estabelecimento de regras *a priori*.

### 2.2 RBC Aplicado ao Mundo Jurídico

No contexto jurídico os sistemas baseados em RBC têm encontrado fértil seara para desenvolvimento de suas metodologias de aplicação e com resultados amplamente satisfatórios. Dentre esses sistemas pode-se relacionar os desenvolvidos pelo grupo de estudos jurídicos da Universidade de Pittsburgh, como o projeto SMILE desenvolvido por Bruninghaus e Ashley [4] e o sistema CATO, desenvolvido por Vincent Aleven (*apud* Bruninghaus e Ashley [4]), na Universidade Carnegie Mellon. No Brasil os avanços nessa área encontram-se no topo do estado da arte do RBC em contexto jurídico, e pode-se

exemplificar como sinônimos de sucesso os sistemas Olimpo e Jurisconsulto, dentre outros, desenvolvidos por pesquisadores do Instituto Jurídico de Inteligência e Sistemas (Ijuris) desta Universidade Federal de Santa Catarina.

### **2.2.1 O Sistema Olimpo**

O sistema Olimpo, desenvolvido por Hoeschl et al [5], tem sua aplicação centrada na combinação das técnicas de RBC e de recuperação de informações textuais de forma automatizada, a partir de uma adequada representação do conhecimento do domínio focado, no caso o das resoluções do Conselho de Segurança (CS) da ONU. O objetivo central do sistema é recuperar documentos textuais, resoluções do CS, com características similares à consulta fornecida pelo usuário em linguagem natural. Também foi proporcionada pelo sistema a inclusão automática de novos documentos na base de casos. As resoluções do CS configuram-se em ambiente propício ao RBC, tendo em vista que possuem as características favoráveis já relacionadas na subseção anterior.

Na fase de representação do conhecimento é que se destaca uma inovação proposta por Hoeschl et al [5], a Representação do Conhecimento Contextualizada Dinamicamente (RC2D). O RC2D é ferramenta poderosa para o tratamento do problema de indexação e o Olimpo o utiliza para a identificação de expressões indicativas que se tornarão indexadores do sistema. Por expressões indicativas, entenda-se uma palavra ou expressão relevante considerando-se o padrão argumentativo típico de um determinado domínio. O RC2D é basicamente composto de etapas iterativas, que vão dinamicamente identificando expressões indicativas.

### **2.2.2 O Sistema Jurisconsulto**

O Sistema Jurisconsulto, segundo Bueno et al [1], teve por objetivo modelar e aplicar a teoria do RBC à área do direito criminal, visando a recuperação de jurisprudências com carac-

terísticas semelhantes e, além disso, permitindo a inclusão de novos documentos automaticamente na base de conhecimento.

Nesse contexto, de acordo com Bueno et al [1], a procura por decisões jurisprudenciais passadas pode ser bastante complicada. Por isso, técnicas mais robustas e inteligentes foram modeladas pelo estudo da autora de forma a minimizar o tempo de procura e a qualidade da resposta recuperada.

Em relação ao problema de indexação, ele foi tratado no Jurisconsulto com base numa abordagem retórico-formal. A parte formal dessa denominação advém dos requisitos formais da sentença penal e foram utilizados na medida de sua relevância para a recuperação. A parte retórica surge por conta do conhecimento do domínio focado, incluído no sistema sob a forma de um dicionário de termos normativos e de um vocabulário controlado, que são cotejados com os índices - expressões indicativas - dos termos da nova situação submetida ao sistema.

Ressalte-se que o Jurisconsulto também realiza a extração automática da informação a partir de textos relativos ao domínio abordado, de três formas distintas, a saber: por localização; por meio do uso de um vocabulário controlado e de um dicionário de termos normativos; e por inferência.

A extração por localização procura extrair informações que se situam sempre no mesmo espaço ou campo de documento. A extração por meio do vocabulário controlado e do dicionário de termos normativos é aplicável nos casos em que certas informações relevantes de um texto não podem ser identificadas por localização.

Portanto, partes do texto são percorridas e comparadas com o glossário do sistema, o qual relaciona expressões relevantes para o domínio representado. Já a extração por

inferência é aquela que extrai informação relevante a partir da identificação, no texto, de citações a artigos ou dispositivos legais. Essa citação legal é, então, traduzida, por inferência, em expressões índices, por meio de um código ou diploma legal regulador do domínio.

## **2.3 Os Sistemas Tributários**

Para a consecução dos julgamentos tributários aqui tratados, uma gama de sistemas de informação é desenvolvida pela SRF de forma a automatizar e qualificar seus processos decisórios. Dentre eles, o sistema de emissão de decisões automatiza de forma bem sucedida a emissão de decisões, e incorpora em si uma ferramenta de pesquisa, na qual pode-se optar pela busca por tributos, por ano, por jurisdição fiscal e por texto, dentre outras opções.

Observe-se que esse sistema não é dotado de opção de pesquisa na qual o julgador possa, a partir de uma solicitação em linguagem natural, ser apresentado a decisões formuladas em situações consideradas de escopo similar àquela que ele se propõe a resolver. Isso justifica, então, o desenvolvimento de um módulo de busca, fundamentado na representação do conhecimento típico desse domínio, que faça a recuperação de decisões de escopo similar à situação demandada, a partir de uma solicitação em linguagem natural.

Consubstanciados pela situação da pesquisa relatada nesta seção, propõe-se, então, a aplicação da metodologia já utilizada em sistemas como o Jurisconsulta, Olimpo, dentre outros, na modelagem aqui tratada, visto que, como vai ser mais detalhado em seguida, as decisões tributárias apresentam estrutura textual homogeneizada em grau relevante e estrutura retórica padronizada, como as apresentadas por decisões judiciais já enfocadas com sucesso por RBC.



## **3 MODELAGEM INTELIGENTE PARA RECUPERAÇÃO DE DECISÕES TRIBUTÁRIAS**

### **3.1 Representação Retórico-Formal dos Casos de Consulta**

Para a fundamentação da representação do conhecimento típico tributário, elegeu-se um tributo específico para o início de estudos nesse sentido, tendo em vista que, se tomados em conjunto os tributos federais, a variedade e complexidade teórica, retórica e formal do domínio tratado pela SRF não seriam comportadas por um típico trabalho final de disciplina.

Por isso, aqui, ater-se-á às Soluções de Consultas (SC) emitidas pelas dez regiões fiscais da SRF, em razão de consultas de contribuintes sobre a correta aplicação da legislação de regência da Contribuição Social sobre o Lucro Líquido (CSLL) instituída pela Lei nº 7.689, de 15 de dezembro de 1988. A CSLL foi escolhida, em primeiro lugar, por ser um dos tributos mais complexos de nosso sistema impositivo em termos de aplicação de normas e administração tributária. Além disso, um dos autores deste trabalho atua já há algum tempo na aplicação e interpretação das normas regentes dessa contribuição. O escopo desse estudo inicial engloba, como universo de estudo, todas as decisões sobre a CSLL emitidas pelas regiões fiscais brasileiras, nos anos de 2000, 2001 e até o dia 20/08/2002.

Para a representação do conhecimento formado por esse microcosmo tributário, inicialmente uma análise da padronização textual formal apresentada pelos documentos é de extrema valia para a identificação de índices que formarão a futura estrutura de um caso tributário.

A estrutura formal de uma decisão apresenta o aspecto sinteticamente exposto na Tabela 1.

**Cabeçalho:** o cabeçalho é formado pelo número do processo administrativo formalizado em consulta, pelo número identificador e data de emissão da Solução de Consulta e pela identificação do interessado por meio de razão social, domicílio fiscal e CNPJ/CPF.

**Ementa:** a ementa de uma SC possui as características gerais de toda ementa jurídica, quais sejam, a conclusão final que foi delineada ao longo do desenvolvimento argumentativo da SC e o rol de dispositivos legais que justificam tal conclusão.

**Relatório:** este campo é o responsável por identificar claramente o que está sendo consultado pelo contribuinte e, a exemplo do item anterior, é potencial fonte de índices formados por expressões indicativas.

**Fundamentos Legais:** este extenso campo tem por função registrar o desenvolvimento de toda argumentação formulada pelo especialista, e é fonte de índices por meio das expressões indicativas que permeiam toda a argumentação nele detalhada.

**Conclusão:** este campo apresenta, ordinariamente, informação similar à já relatada pela ementa.

**Ordem de Intimação:** este campo é desprovido de interesse para fins de indexação de SC's.

### **Tabela 1. Estrutura fundamental de uma Solução de Consulta**

Em relação ao campo cabeçalho, vale ressaltar que as informações nele expostas são potenciais fontes de pesquisa e elas prestam-se muito bem à extração de uma SC, porém quando já se sabe qual delas se vai consultar. Por isso, o que se deseja, recuperação de Soluções de Consulta similares, dificilmente possa ser logrado apenas por esses índices de cabeçalho, pois os valores particulares de cada SC, como CNPJ do interessado, número de processo etc não são genéricos e por isso não se prestam para mensurar a similaridade entre documentos.

Por externar tanto a conclusão da SC, quanto os dispositivos legais que a embasaram, a ementa deve ser fonte fértil de indexadores para as SC, pois nela encontrar-se-ão fontes de índices que permitirão a extração do conhecimento tanto por inferência, quanto por meio de um vocabulário de comparação entre expressões indicativas. Também, o campo relatório, por apresentar em linhas gerais a questão que se formula à SRF, é fonte de expressões indicativas para a

recuperação de decisões de escopo similar à nova pendência. E ambos os campos, relatório e ementa, são importantes, pois enfatizam, respectivamente, a dúvida do contribuinte e a conclusão da SC. Assim, num primeiro momento, a ementa e relatório podem ser suficientes para a comparação de SC's com qualidade.

### **3.1.1 Índices para recuperação**

Para que se possa acessar o conhecimento representado em um sistema inteligente, assim como organizar os novos casos submetidos à base, deve-se enfrentar o já citado "problema da indexação" em RBC. Como se está propondo uma representação do conhecimento retórico-formal, pode-se estabelecer um sistema de indexação com base em informações constantes do cabeçalho da SC, juntamente com índices mais complexos baseados em expressões indicativas e dispositivos legais que são encontrados na ementa e relatório da SC. Vale dizer que a pesquisa por informações do cabeçalho já é realizada eficientemente pelo sistema decisões, contudo um diferencial seria possibilitar a sinergia dessa pesquisa com o modelo aqui proposto, que agrega a indexação por expressões indicativas e por dispositivos legais.

Em relação às expressões indicativas, pode-se, nesse momento, ressaltar importante conclusão, a partir de Hoeschl et al [5], que informa que tanto expressões ou palavras que ocorrem com muita frequência, quanto expressões muito raras ou ocasionais, não apresentam interesse à indexação, por conferirem pouca ou nenhuma capacidade de diferenciação de documentos à modelagem.

Partindo-se dessa orientação, observa-se que a ementa e o relatório apresentam vários elementos que podem fundamentar uma indexação com base em expressões indicativas. De fato, como visto na Tabela 1, a ementa apresenta tanto a conclusão expendida pela decisão, quanto os dispositivos legais que a fundamentaram. É interessante notar que a parte

da ementa que relaciona os dispositivos legais presta-se para a comparação de similaridade entre os casos que já estão na base.

Contudo, ao proceder a uma consulta ao sistema, o usuário, em geral, o faz em linguagem natural, sem ter idéia, num primeiro momento, de qual legislação vai se utilizar para fundamentar sua solução. Por isso é que são tão relevantes as expressões indicativas. Por meio delas é que se interconectam expressões indicativas, submetidas pelo usuário ao sistema, com a estrutura sistêmica de representação de conhecimento retórico-formal, possibilitando a comparação de casos. A Tabela 2 propõe uma abordagem inicial para a indexação de um caso na base de dados de referência.

<b>Índices</b>	<b>Descrição</b>	<b>Importância</b>
Número e data da SC	Número identificador da SC e data de emissão	Importante da para delimitar a recuperação a certas regiões fiscais e anos de emissão. Também tem importância na recuperação de SC's especificamente já identificadas pelo usuário.
Expressões Jurídicas da Ementa	Resumo da argumentação tributária emitida e conclusão da consulta	A ementa é importante para a indexação e recuperação de decisões similares, pois nela reside a conclusão final da SC, que é fonte de diversas expressões indicativas de interesse à comparação a ser efetuada pelo sistema.
Dispositivos Legais	Relaciona os dispositivos legais que embasaram a ementa	São importantes para a comparação de casos já incorporados à base, tendo em vista que uma consulta com vistas à resolução de nova situação tributária, em geral, não se formula com a citação de dispositivos legais.
Expressões Jurídicas do Relatório	Relata em linhas gerais a consulta formulada	O relatório é importante para a indexação e recuperação de decisões similares, pois nele reside a questão formulada à SRF, que se bem relatada é fonte de expressões indicativas de interesse à recuperação de consultas similares anteriores.

**Tabela 2. Descrição dos índices propostos**

### 3.1.2 Conhecimento do domínio tributário específico

Nesse ponto em que já se sabe o que analisar com o fito de comparar documentos é preciso atenção especial às peculiaridades de cada domínio estudado. No caso brasileiro, país de dimensões continentais, em que a SRF se projeta em dez regiões fiscais, os especialistas que emitem decisões tributárias utilizam-se de argumentação padronizada em certo grau pela norma legal e infralegal, porém essa padronização não é capaz de mitigar os efeitos sociais, pessoais e regionais que impactam sobre a linguagem utilizada no domínio. Dessa forma, a indexação e recuperação de casos têm de levar em consideração que a linguagem disposta na norma nem sempre é a utilizada pelo que decide, embora ele esteja utilizando tal norma como referência.

### 3.1.3 Vocabulário controlado e dicionário de termos normativos

Devido às variações de vocabulário regionalmente identificadas, é preciso que um vocabulário controlado e um dicionário de termos normativos sejam desenvolvidos, nos moldes dos já referenciados na seção 2 deste trabalho. Assim, é por meio desses léxicos que se pode inferir que "entidades sem fins lucrativos", "fundação de direito privado de caráter filantrópico" e "sociedade civil sem fim lucrativo", todos trechos extraídos de SC's, são similares entre si e correspondem às diversas variações encontradas na linguagem do domínio para o disposto no art. 15 da Lei nº 9.532, de 10 de dezembro de 1997, que se refere a "*instituições de caráter filantrópico, recreativo, cultural e científico e as associações civis que prestem os serviços para os quais houverem sido instituídas e os coloquem à disposição do grupo de pessoas a que se destinam, sem fins lucrativos*". Para a elaboração desses glossários, um rol básico da legislação de regência do domínio deve ser identificado, de forma a subsidiar a pesquisa, por parte do especialista, de expressões e termos que os comporão.

### 3.1.4 Analisador sintático

Uma das formas de potencializar o uso do vocabulário controlado e do dicionário de termos normativos, em um ambiente em que a legislação de regência é tão esparsa, como o da CSLL, é o uso combinado de um analisador sintático. Um analisador sintático já foi utilizado com sucesso em projetos como o SMILE, que se fundamenta em padrões proposicionais. Um padrão proposicional, de acordo com Bruninghaus e Ashley [4], é obtido a partir da análise da função sintática de certas palavras que se repetem comumente. Assim, uma vez identificado um padrão em um domínio, um analisador sintático pode extrair conhecimento de novos textos simplesmente identificando a ocorrência desse padrão, e, a partir dele, por exemplo, seu sujeito ou objeto direto.

No mundo jurídico da CSLL, o especialista tem de estar a par de um sem número de dispositivos emanados das normas legais e da legislação complementar, e dentre esta uma vasta gama de instruções normativas, atos declaratórios normativos e interpretativos, e os regulamentos e manuais emitidos pela SRF. A identificação exaustiva de termos normativos nessas normas pode se tornar tarefa infundável e por isso, o balizamento proporcionado por padrões proposicionais pode ser altamente útil. Ao se consultarem as normas tributárias verifica-se que, embora os termos e denominações para coisas similares se alterem, principalmente em nível de detalhamento, de uma norma para outra, os verbos e substantivos utilizados para dispor sobre as ações reguladas via de regra são os mesmos. Por exemplo, "deduzir", "dedutibilidade", e "computadas" são palavras-chave tributárias que são empregadas de forma padronizada em quaisquer normas, sejam elas legais ou infralegais. Com a utilização de padrões proposicionais a partir dessas palavras-chave, fica mais fácil identificar os termos e expressões indicativas distintas para coisas similares, tanto na fase de representação do conhecimento, quanto na de consulta. Para tal, basta que se procure, nas normas e textos de pesquisa, analisar sintática-

mente seus conteúdos com base em poucos padrões criteriosamente estabelecidos, para que se identifique se há ou não similaridade entre certos nomes e expressões relevantes. Por exemplo, uma norma superior dispôs que "a entidade sem fins lucrativos deve entregar anualmente...", assim, a busca automática de sinônimos para o vocabulário controlado, nas instruções normativas e atos declaratórios supervenientes à norma pode se dar com base no padrão proposicional "entregar" e na identificação de seu sujeito e objeto direto.

### **3.2 Extração Automática da Informação Textual**

Este tópico trata da situação em que uma nova decisão é emitida e, portanto, deve ser agregada à base de casos do modelo proposto. Como já visto, as SC apresentam formato padrão, o qual contribui para a extração de índices por localização. Esse é justamente o caso do primeiro e do terceiro índices na Tabela 2, que podem ser facilmente extraídos da SC por mera localização, extração e registro do índice por uma simples rotina adequada. Já no caso da ementa e do relatório, a extração por localização vai até onde é possível, ou seja, o posicionamento da rotina de extração no campo exato. A partir daí, já se viu que uma ampla gama de ferramentas utilizando o vocabulário de termos normativos e de expressões indicativas tem de ser acionada para a indexação de novos documentos.

O funcionamento das ferramentas, nesses casos, é similar ao já descrito nas subseções anteriores, buscando a identificação de expressões indicativas e dispositivos legais que possam contribuir para a posterior recuperação e comparação do novo documento. Essa possibilidade de agregação de novas decisões é que, por somar novas expressões indicativas à base de conhecimento, proporciona a capacidade de aprendizado por parte do sistema, que pode assim ser considerado inteligente. Um exemplo desse aprendizado pode se dar por meio do vocabulário controlado e dos padrões proposicionais, que, ao permitirem a identificação de novas

palavras ou expressões cumprindo as mesmas funções sintáticas em padrões que se repetem em dois documentos com alto grau de similaridade, conferem sensibilidade ao sistema para perceber, ou alertar, que duas palavras ou expressões distintas significam a mesma coisa naquele domínio de conhecimento específico, enriquecendo dinamicamente, portanto, seu vocabulário controlado.

### **3.3 Ajuste da situação**

A situação típica da utilização do modelo proposto neste artigo é a de um especialista da SRF que, ao ser demandado sobre determinada matéria tributária complexa, submete uma sentença, em linguagem natural ao sistema, com vistas a ser apresentado a situações antigas similares já solucionadas. O sistema vai ajustar a demanda à estrutura de indexação de casos, nela identificando automaticamente expressões e dispositivos legais relevantes para a máquina de comparação. Suponhamos que sentença submetida ao sistema seja "pessoa jurídica não computou a despesa com brindes na base de cálculo da CSLL" e que um trecho de um caso da base mencione "pessoa jurídica não computou a despesa de representação na base de cálculo da CSLL". Por meio do padrão proposicional previamente estabelecido "computar", o sistema pode rapidamente identificar a presença do padrão e verificar qual o seu objeto direto em ambas as sentenças. Com isso, pode-se rapidamente estabelecer se "despesas com brindes" e "despesas com representação" são similares ou não, com base em um vocabulário controlado interno ao sistema. Deve-se ratificar a utilidade do vocabulário controlado, na medida em que um especialista tributário, ao demandar o sistema com uma nova consulta em linguagem natural, provavelmente utilizará estruturas sintáticas que seguem tais padrões, tendo em vista ser ele um conhecedor do linguajar comumente utilizado nos domínios tributários. Isso é vantajoso, pois uma das características de um sistema inteligente é que, por ocasião do ajuste da situação, quanto mais informado ele for



textualmente a respeito de uma demanda, maior será a qualidade da resposta, diferentemente dos sistemas tradicionais, que em geral são vencidos por consultas longas e com muitos conectivos.

A partir do ajuste da situação, o próximo passo da metodologia de RBC é a recuperação de casos similares, com base em métrica previamente estabelecida, ao ajustado, na base de conhecimento.

## **4. CONCLUSÕES**

No contexto dos julgamentos tributários existem situações complexas, na qual a experiência anterior de um especialista pode ser decisiva para uma solução de qualidade ao processo. A complexidade do domínio tributário pode ser aferida pela relativa freqüência com que ocorrem duas decisões divergentes emitidas por regiões fiscais distintas, a respeito do mesmo assunto. Isso, em certos casos, pode ser evitado se forem consultadas, previamente à decisão final, decisões tributárias similares bem sucedidas anteriormente.

Observou-se que o sistema utilizado para auxílio às decisões tributárias não é dotado de opção de pesquisa na qual o julgador possa, a partir de uma solicitação em linguagem natural, ser apresentado a decisões formuladas em situações consideradas de escopo similar àquela complexa a qual ele se propõe a resolver. Isso justifica o desenvolvimento de um módulo de busca fundamentado na representação do conhecimento típico desse domínio, que faça a recuperação de decisões de escopo similar à situação demandada, a partir de uma solicitação em linguagem natural.

Por todo o exposto na metodologia, verificou-se que a aplicação de RBC ao caso em análise é viável tendo em vista que as decisões tributárias possuem formatação textual

homogeneizada em grau relevante e uma estrutura retórica padronizada, como as apresentadas por situações jurídicas já tratadas com sucesso por RBC.

Utilizando-se de RBC, foram estabelecidos, então, os pressupostos de uma modelagem aplicável ao domínio em estudo, resultando no estabelecimento de um plano de indexação para a recuperação de casos na base de conhecimento; na especificação do ferramental requisitado pelo modelo, compreendendo a utilização, dentre outras ferramentas, de vocabulário controlado e de termos normativos; e na proposição de um modelo de recuperação e indexação que por utilizar-se da abstração do conhecimento genérico, a partir de casos particulares, presta-se à representação do conhecimento sejam quais forem as espécies de decisões e de documentos formais envolvidos.

Ressalte-se, finalmente, que o modelo aqui proposto tem de passar, após ser materializado, por um processo de validação de seus pressupostos, de forma a se avaliar sua capacidade de indexação e recuperação de casos tributários.

## 5. REFERÊNCIAS

- 1 BUENO, T. C. D.; WANGENHEIN, C. G.; MATTOS, E. S.; HOESCHL H. C.; BARCIA, R. M. JurisConsulta: retrieval in jurisprudential text bases using juridical terminology. Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 1999.
- 2 VARGAS, J. E. Knowledge based systems. Chapter 15. Singapore, World Scientific, 1997.
- 3 AZEVEDO, F. M.; BRASIL, L. M.; OLIVEIRA, R. C. L. Redes neurais., Florianópolis, Brazil, Visual Books, 2000.
- 4 BRUNINGHAUS S.; ASHLEY K. D. Improving the representation of legal case texts with information extraction methods. University of Pittsburgh, 2001.

- 5 HOESCHL, H. C.; BARCIA, R. M. Sistema Olimpo: tecnologia da informação jurídica para o Conselho de Segurança da ONU. Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 2001.



# **Pesquisa Contextual Estruturada para o Conselho de Segurança da ONU**

Hugo Cesar Hoeschl, Dr.  
metajur@eps.ufsc.br

Tânia Cristina D'Agostini Bueno, M.Sc.  
bueno@eps.ufsc.br

Andre Bortolon  
bortolon@eps.ufsc.br

Eduardo da Silva Mattos  
emattos@eps.ufsc.br

Marcelo Stopanovski Ribeiro  
marcelos@eps.ufsc.br

Ricardo Miranda Barcia, Ph.D.  
barcia@eps.ufsc.br

Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, SC, Brasil, 88040-000

## **RESUMO**

Este trabalho apresenta um modelo genérico de metodologia que enfatiza o emprego de métodos de extração da informação combinados com a técnica de Inteligência Artificial denominada Raciocínio Baseado em Casos - RBC. Esta metodologia permite a indexação automática do conhecimento humano em sistemas baseados no conhecimento. Este tipo de representação compatibiliza a linguagem do usuário com a linguagem dos documentos na

base de conhecimento do sistema, retornando ao usuário respostas mais adequadas a solução do seu problema. Nesta pesquisa é descrito o sistema Olimpo, um sistema baseado no conhecimento, que permite a recuperação informação de documentos textuais com características similares à questão fornecida pelo usuário em linguagem natural. Para este desenvolvimento foram indexadas 300 resoluções do Conselho de Segurança da ONU disponíveis na internet.

## 1. INTRODUÇÃO

Em alguns complexos e específicos domínios, um sistema de recuperação de informação necessita mais que uma boa tecnologia para buscar documentos em amplas bases textuais. É necessário uma boa representação do conhecimento também.

O Conselho de Segurança - CS da ONU foi o escolhido como campo de aplicação do Sistema Olimpo, tendo em vista algumas características do *órgão* e das suas *resoluções*. Devido a amplitude dos documentos, repetições e similaridades de muitos títulos - uma pergunta nebulosa sobre o Secretário-Geral, por exemplo, pode recuperar centenas de documentos no sistema de busca disponível no site da ONU<sup>1</sup>. Por esta razão, o desenvolvimento de um sistema eficiente e efetivo para a recuperação da informação dos documentos do CS, usando técnicas na área da Inteligência Artificial (IA) é importante para todos os organismos da ONU e fornecerá significativo benefícios para a solução de conflitos internacionais.

Esta abordagem permite a recuperação de textos com características similares à informação fornecida pelo usuário em linguagem natural. Novos documentos são automaticamente incluídos na base de conhecimentos através da extração de informações relevantes, por meio da técnica denominada RC2D [8] [9]. Conceitos de Raciocínio Baseado em Casos (RBC) [1]

---

<sup>1</sup> [www.un.org](http://www.un.org).

[2] e técnicas de recuperação de informação forma aplicados para uma melhor performance do sistema, denominando uma metodologia chamada - Pesquisa Contextual Estruturada - PCE.

Este trabalho está estruturado da seguinte maneira: a) no item 2 descreve-se características do domínio e dos documentos do Conselho de Segurança, bem como o processo de análise destes documentos; b) no item 3 a metodologia de representação do conhecimento; no item 4 e 5 é apresentado o sistema e a avaliação de sua performance; no item 6 descreve-se os trabalhos futuros; e no 7 os resultados.

## **2. TRABALHOS RELACIONADOS**

Esta abordagem é orientada, especificamente, para a recuperação e indexação automática de documentos em amplas bases textuais. A metodologia de extração automática da informação teve como referência o trabalho desenvolvido no Sistema JurisConsulta [5] [6] [7]. O JurisConsulta é um sistema baseado em casos que auxilia os profissionais do Direito a utilizar as fontes de pesquisa jurídica (jurisprudência) de uma forma mais eficiente, através da recuperação de informações numa ampla e distribuída base de dados. No JurisConsulta, o conhecimento de domínio é empregado na construção do sistema, permitindo a extração automática de índices dos documentos, tornando a busca mais rápida e a interface amigável.

No JurisConsulta a representação do conhecimento é apoiada pelo uso de um vocabulário controlado e um dicionário de termos normativos, definidos com base numa estrutura jurídico-teórica. Embora essa forma de representação permita que o sistema estabeleça um grau de similaridade com maior precisão, aumenta o tempo da etapa de aquisição do conhecimento e conseqüentemente os custos de implemen-

tação do sistema. Além disso, a nossa abordagem inclui também - além da medida de semelhança global - o cálculo da similaridade local estimada para cada índice. Isto melhorou consideravelmente a recuperação, de uma forma efetiva e eficiente.

### **3. A REPRESENTAÇÃO DO CONHECIMENTO CONTEXTUALIZADA DINAMICAMENTE**

O Sistema Olimpo tem a sua performance centrada na combinação de aspectos derivados do Raciocínio Baseado em Casos - RBC e da recuperação de informações textuais, somados a uma adequada organização do conhecimento referente ao assunto focado (no caso específico, as resoluções do Conselho de Segurança da ONU). Esta organização é que enseja a RC2D, uma metodologia que propicia a possibilidade de comparação entre o contexto dos documentos, e não simplesmente entre as palavras ou atributos.

De um modo geral, o funcionamento do sistema apresenta um fluxo similar a outros sistemas baseados em casos [3] [5], veja Figura 1, segundo o qual uma entrada manual sofre um processo de ajuste e, após, é submetida à base de documentos, dentre os quais, através do cálculo da similaridade, são escolhidos os melhores.

O Sistema Olimpo, após uma modelagem refinada da base de dados, armazena as Resoluções, observando suas características e atributos *centrais* (p. ex., temas principais, assuntos relacionados, países envolvidos), *periféricos* (p. ex., outras resoluções relacionadas, outros órgãos da ONU referenciados) e *superficiais* (p. ex., datas, números e nomes). Tal estruturação enseja a atribuição de pesos (variáveis) aos atributos, o que permite uma pesquisa contextualizada e mais precisa.





**Figura 1. Processo de busca**

Uma vez definido isso, teve início o processo de captura dos documentos, via web, para a formação da base de conhecimento do sistema, através do site da ONU.

### **3.1 Processo de Análise da Estrutura Retórica**

A estrutura retórica do sistema, composta pelas expressões indicativas utilizadas para a realização dos comparativos, foi, pela primeira vez, realizada de forma dinâmica. Até então, era escolhida uma lista de indexadores proveniente de uma fonte externa ao grupo de pesquisa (p. ex., índices da biblioteca do tribunal "x").

Esta lista de indexadores, cuja escolha atendia a critérios de aproximação com o contexto do sistema em construção, praticamente não era trabalhada. Resolvemos, então, montar

uma lista absolutamente particularizada e específica para a tarefa do sistema, que estivesse em real alinhamento com as questões efetivamente tratadas nas resoluções e que, por outro lado, atendessem ao contexto documental da entidade gestora dos documentos. Para isso, fizemos leitura detalhada de resoluções, consultas no sistema de Bando de Dados - BD web da ONU e debates com o grupo de pesquisa, para coletar um conjunto de expressões. Depois, estas foram testadas e analisadas estatisticamente, de forma a que se analisasse sua real função como elemento de referência para recuperação e indexação documental. Expressões que ocorrem com uma frequência muito alta, próxima a 100% das resoluções ("Security Council"), têm pouca utilidade para diferenciá-las. Por outro lado, aquelas que aparecem somente em um ou dois documentos, salvo se forem extremamente caracterizadoras, também são pouco úteis ao estabelecimento de contexto. Procurou-se eliminar os extremos, selecionando o conjunto de expressões que possuam significância na comparação documental.

Este processo assumiu contornos dinâmicos, pois foi feito e refeito diversas vezes, e expressões iam sendo incluídas ou excluídas, conforme seu desempenho estatístico. Deste procedimento foi elaborado uma rotina, com a seguinte seqüência: 1. Uma equipe fez leitura das resoluções e extraiu expressões (palavras ou grupo de palavras) consideradas relevantes, formando a "pré-lista 1"; 1.1. outra equipe fazia consultas no banco de dados web da ONU, e, com base nos resultados, catalogava as expressões que apresentavam resultados satisfatórios, formando a "pré-lista 2"; 2. Somadas as pré-listas, chegou-se à primeira lista provisória; 3. A lista provisória foi utilizada "rodada" sobre os casos, a fim de que se verificasse qual a incidência das expressões ali contidas; 3.1. com base na lista provisória, novas consultas foram sendo feitas dentro do texto das resoluções; 3.2 eram registradas, em ambas as atividades, as novas expressões detectadas, as duas etapas do passo "3", iam sendo registradas as novas as

expressões detectadas, que formavam as listas adicionais "1" e "2".

Estas listas adicionais somavam-se à lista provisória, gerando a pré-lista, que era novamente submetida ao passo "3". Este procedimento, representado na Figura 2, foi repetido diversas vezes, de forma a que a lista final pudesse representar um conjunto de índices que espelhasse eficientemente a estrutura retórica genérica das resoluções, materializando uma representação de conhecimento contextualizada dinamicamente.



**Figura 2. Análise da estrutura retórica e definição dos índices**

## **3.2 Características da Documentação da Área de Aplicação**

Quanto à estrutura do documento, as resoluções do Conselho de Segurança da ONU possuem características que facilitam a aplicação da tecnologia aqui referenciada, quais sejam:

1. Formatação textual homogênea - formato padrão para todos os documentos;
2. Estrutura retórica padronizada - o vocabulário do domínio é bem específico e restrito, definido pelas atribuições do organismo presentes na Carta das Nações Unidas;
3. Necessidade de simplificação da linguagem do domínio para uma mais usual - as resoluções são documentos oficiais e públicos (acessados pelo público em geral e necessita uma linguagem usual para uma recuperação efetiva da informação).

## **3.3 Pesquisa Contextual Estruturada - PCE**

A pesquisa é considerada "contextual" e "estruturada" pelas seguintes razões: 1. É levado em consideração o contexto dos documentos armazenados quando da formação de estrutura retórica do sistema; 2. Este contexto norteia o processo de ajuste da entrada bem como da comparação e seleção dos documentos; 3. Quando da elaboração da consulta, a entrada não está limitada a um conjunto de palavras, ou à indicação de atributos, podendo assumir o formato de uma questão estruturada pelo conjunto de um longo texto somado à possibilidade de acionamento de atributos específicos, que funcionam como "filtros" e fazem uma seleção preliminar dos documentos a serem analisados.

Além disso, o controle do nível de profundidade da pesqui-

sa (*search level*) permite uma seleção de documentos de acordo com a maior ou menor ocorrência de expressões indicativas dentro de cada resolução, antes de se efetuar a comparação.

Este processo, descrito na Figura 3, faz com que o espaço de busca seja reduzido de forma mais eficiente, não se tratando de uma simples seleção de documentos de acordo com as suas características superficiais, mas de uma comparação preliminar orientada pelo contexto relacionado à questão de entrada.



**Figura 3. Pesquisa Contextual Estruturada**

Após efetuado este procedimento, o resultado textual é a formação de uma lista de expressões indicativas referente às resoluções, a qual foi novamente passada sobre cada uma delas, gerando um registro individualizado sobre a ocorrência de cada uma das expressões dentro dos respectivos textos.

Estes registros permitem a realização das comparações e a aplicação da métrica de similaridade globalizada.

Além das expressões indicativas, o processo de extração automática dos atributos foi preparado para detectar e extrair o assunto, a data, o número da resolução, as siglas, os países, o número de "decides" e partes dos trechos que contém as expressões com maior ocorrência no texto (resultado). Os dois últimos atributos não estão presentes no protótipo já implementado.

Após formada uma base contendo o total de resoluções escolhidas (100), bem como formadas as listas de expressões indicativas, países e siglas encontradas no corpo do texto das resoluções, já é possível a análise do comportamento estatístico dos dados coletados, como será visto em seguida.

#### **4. FUNCIONAMENTO E PERFORMANCE DO SISTEMA OLIMPO**

As principais características do Sistema Olimpo são a conjugação de técnicas de recuperação de informação textual com RBC e a possibilidade de entrada de uma consulta textual ampla.

Isto faz com que ele apresente uma performance diferenciada em termos de recuperação de informações. Porém, como o sistema realiza suas comparações com base em uma estrutura retórica previamente fornecida, o seu melhor funcionamento está vinculado à maior proximidade desta estrutura retórica quando da elaboração da consulta de entrada. Diante de tal cenário, pode-se afirmar que o desempenho do sistema vai sendo gradativamente mais consistente quanto maior for a proximidade da linguagem da consulta com a da estrutura identificada nos documentos que geraram a base de conhecimento do sistema. Veja a Figura 4.



**Figura 4. Performance do sistema**

A coleta dos dados que preenchem o sistema e possibilitam o seu funcionamento geram números passíveis de serem apresentados e analisados.

Primeiramente, é necessário dizer que todas as resoluções tiveram um monitoramento constante com relação ao número de expressões indicativas que iam apresentando ao longo da estruturação da base de resoluções.

A média de expressões indicativas por resolução é de 55,18, número bastante expressivo tendo em vista que, na metodologia utilizada anteriormente, esta média jamais superaria a casa de um dígito. Apesar disso, acreditamos que esta média ainda deve aumentar bastante quando do desenvolvimento integral do sistema, com todas as resoluções.

Dentre aquelas que apresentaram o melhor desempenho, selecionamos os dez maiores números de expressões encontradas, totalizando 12 resoluções que superaram 100 ocorrências. Já tendo sido explicado o funcionamento do sistema, bem como seu processo de formação e respectiva análise estatística, podemos passar à avaliação da sua performance.

Por último, verifica-se a quantidade de siglas diferentes que ocorrem dentro das resoluções, estando arroladas as dez resoluções que mais apresentam siglas no corpo de seu texto, significando um peso maior para elas.

Estas avaliações foram fundamentais para estabelecer a interface do Sistema Olimpo, para permitir que o usuário possa extrair conhecimento com o cruzamento das informações relevantes extraídas de dados não extruturados.

## 4.1 O Sistema Olimpo

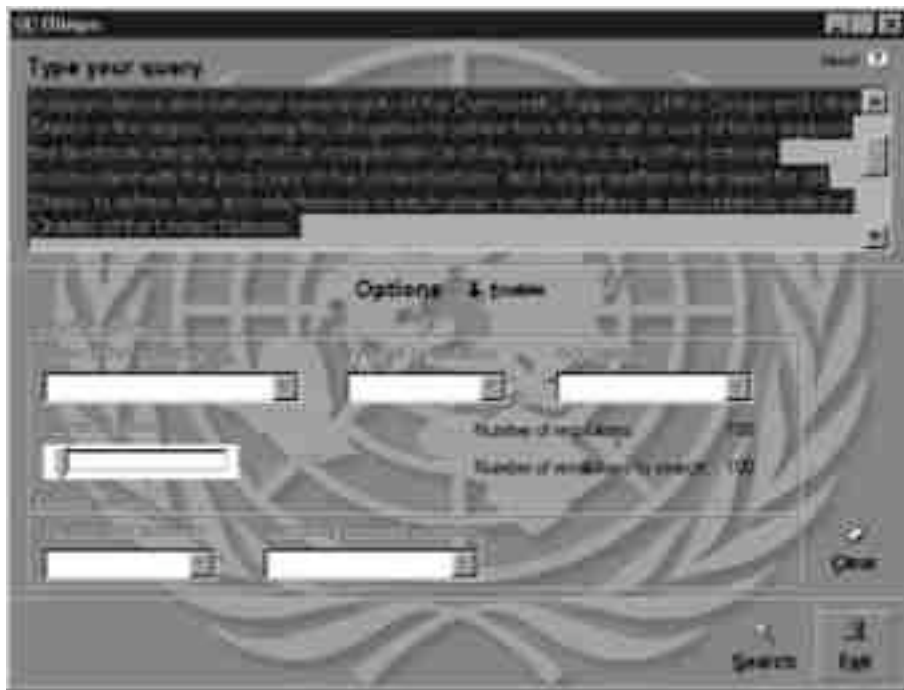
As informações dos documentos são representados em forma de casos, que consistem no documento original e um conjunto de 8 índices em forma de pares atributo-valor: *matéria, data, número da resolução, encontro, país, siglas, decisões e expressões indicativas*. São estes índices que dão forma e integram a interface do sistema, como é possível observar na Figura 5.

## 5. COMPARATIVO BD ONU X OLIMPO

Quando é feita uma entrada aleatória, mediante digitação, as chances de uma recuperação efetiva são inferiores à de uma entrada baseada em um texto cuja linguagem seja coincidente com aquela já identificada, como, por exemplo, uma outra resolução do Conselho de segurança. Para entradas aleatórias, um sistema usual de banco de dados tem desempenho



superior, quadro que vai sendo alterado à medida em que as consultas passam a ter maior relação com os documentos contidos na base de dados do sistema.



**Figura 5. Interface do Sistema Olimpo**

Diante de tal constatação, é proveitoso um comparativo entre o Olimpo e um sistema de banco de dados que opera em uma base similar, o que se verá na sequência.

Tal sistemática também permite a extração de informações sobre o regime de produção do Conselho, relacionando-o a temas, nomes, números e datas. Um interessante exemplo, visualizável preliminarmente, é o crescimento exponencial da produção de resoluções nos anos 90. Até o mês de agosto de 1990, o Conselho emitira 661 Resoluções. Porém, de lá até 1999, o número subiu para 1.248. Verifica-se que nos anos 90

o órgão produziu o mesmo número de decisões que havia produzido em toda a sua história.

Tendo em vista tais características, a realização de um rápido comparativo entre o Sistema Olimpo e o sistema de consulta disponibilizado no site oficial da ONU, estruturado sobre as tradicionais e corriqueiras técnicas de bancos de dados, com operadores lógicos, é tarefa bastante pertinente, e foi executada, produzindo os seguintes resultados, contidos na Tabela 1:

<b>Expressão pesquisada</b>	<b>Sistema Olimpo</b>	<b>Banco de dados da ONU (web)</b>
"Weapons"	10 registros (todos com 100%)	62 registros
"Nuclear weapons"	10 registros (o primeiro com 100%; os demais com 50%)	5 registros
"Children nuclear weapons"	10 registros (os três primeiros com 66,70%; os demais com 33,30%)	0 registros ("exact frase"; "all the words")
Texto com 30 linhas retirado de dentro do corpo da Resolução nº 1.234	10 registros (o primeiro, com 100%, é a própria Resolução 1.234; o segundo com 71,40%; o terceiro com 64,30%; e o quarto com 57,10%)	Não permite esta possibilidade

**Tabela 1. Comparativo entre o banco de dados da ONU e o Sistema Olimpo**

A sequência da pesquisa foi registrada de acordo com as respostas emitidas por ambos os sistemas. O sistema da ONU utilizado foi aquele disponível em seu site oficial.

A possibilidade de entrada de uma consulta com uma grande quantidade textual, 20 ou 30 linhas, ou mais, produz um significativo fator de diferenciação do Sistema Olimpo, conforme pode ser visualizado na Figura 5. Este tipo de entrada propicia um desempenho bastante detalhado do sistema, tendo em vista que ele está preparado para realizar

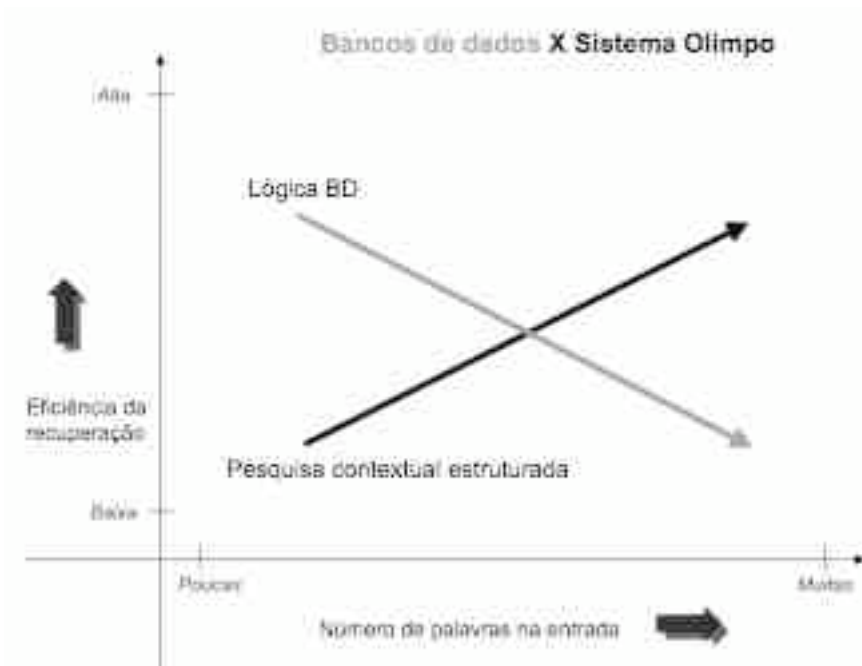
um comparativo pormenorizado entre o texto de entrada e aquele contido nas resoluções.

O resultado do comparativo é apresentado em forma de saída, onde se pode verificar a posição daquela resolução na ordem de importância de resultados, o percentual de similaridade entre a resolução e o texto de entrada, o número da resolução, o número do encontro, sua data, e um resumo sobre o tema da resolução.

Ao longo da realização do comparativo, vai brotando uma sensação de que o aumento da complexidade faz o sistema de banco de dados esmorecer, enquanto que, em sentido contrário, este mesmo aumento de complexidade faz o Olimpo efervescer, fazendo com que ele apresente um desempenho cada vez melhor, quanto mais estruturada for a entrada, e quanto maior for o texto utilizado como consulta. Este fenômeno está espelhado na Figura 6.

Assim, pode-se concluir que a eficiência dos sistemas de bancos de dados, como o analisado, tem uma relação inversamente proporcional à elevação da complexidade da consulta formulada, principalmente com relação ao número de palavras fornecido, isto é, quanto maior o número de palavras, pior o desempenho do sistema.

No caso do Olimpo, a relação é oposta, isto é, a eficiência tem relação diretamente proporcional à elevação de complexidade da consulta formulada.



**Figura 6. Eficiência do Sistema Olimpo em relação aos bancos de dados**

## 6. TRABALHOS FUTUROS

O trabalho corporificado na prototipagem do Sistema Olimpo ainda está muito longe de sua teleologia originária, e o relato até aqui apresentado tem a função de demonstrar a viabilidade da sua implementação e os aspectos a ela inerentes, sob as óticas empírica e teórica. Ainda há muito a ser feito.

Com relação à aplicação específica, uma contextualização mais aprofundada, com melhor avaliação do funcionamento do órgão envolvido, bem como das pessoas que atuam diariamente nas tarefas de documentação dos fenômenos a ele inerente, além daquelas que utilizam seus documentos constantemente,

gerará um incremento qualitativo e quantitativo das listas e dos atributos definidos, bem como um significativo enriquecimento do processo de RC2D.

No tocante às questões operacionais do próprio sistema, alguns aspectos técnicos podem gerar também uma sensível melhora de sua performance global. São eles:

- Implementação de técnicas já dominadas em termos de bancos de dados, no que diz respeito ao reconhecimento de plurais, erros de ortografia e aplicação de trigramas;
- Criação de um dicionário de sinônimos adequado ao contexto do sistema, também orientado pela RC2D;
- Incorporação destes aspectos à métrica de similaridade.

Além disso, um melhor aproveitamento de informações já disponíveis no presente estudo, mas ainda não incorporadas ao sistema, como a frequência de cada expressão indicativa dentro de cada uma das resoluções, o número de "decides" que cada resolução contém, e ainda os trechos de texto que envolvem aquelas expressões que possuem maior frequência dentro de uma resolução, também vão gerar elevação da acuracidade do sistema.

## **7. CONCLUSÕES**

Diante de todo o exposto, pode-se concluir, que:

- A técnica denominada "representação de conhecimento contextualizada dinamicamente" - RC2D, enseja significativo incremento de performance em sistemas estruturados em conhecimento;
- A adequada conjugação de características de bancos de dados textuais, raciocínio baseado em casos e da representação de conhecimento contextualizada dinamicamente ensinam a eficiente aplicação da "pesquisa

contextual estruturada" - PCE, empregada para a construção do Sistema Olimpo;

- Na construção e consolidação do Sistema Olimpo, a RC2D representa a sua característica preponderante;
- O adequado emprego da RC2D, combinado com outras técnicas, pode produzir um sistema com a tendência de apresentar eficiência gradativamente superior aos bancos de dados textuais na medida em que as consultas vão ficando mais contextualizadas, detalhadas e aprofundadas, e com maior volume textual.

## 8. BIBLIOGRAFIA

- 1 AMONDT, A., PLAZA, E. Case-based reasoning: fundamental issues, methodological variations, and system approaches. *AI Communications*, 17 (1), 1994.
- 2 BENCH-CAPON, T. J. M. Some observations on modeling case based reasoning with formal argument models. In: *Proceedings of The Seventh International Conference on Artificial Intelligence and Law*, pp. 36-42. Oslo, Jun. 14-18, 1999. 220 p.
- 3 BRUNINGHAUS, S.; ASHLEY, K. D. Toward adding knowledge to learning algorithms for indexing legal cases. In: *Proceedings of The Seventh International Conference on Artificial Intelligence and Law*, pp. 9-17. Oslo, Jun. 14-18, 1999. 220 p.
- 4 BUENO, T. C. D. O uso da Teoria Jurídica para recuperação em amplas bases de textos jurídicos. 1999. Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 1999.
- 5 BUENO, T. C. D.; HOESCHL, H. C.; MATTOS, E. S.; BARCIA, R. M.; WANGENHEIM, C. G. JurisConsulta: retrieval in jurisprudential text bases using juridical terminology. In: *The Seventh International Conference on Artificial Intelligence and Law*. Oslo, 1999. *Proceedings of the Conference*. New York, ACM, 1999. vol. 1. pp. 147-155.
- 6 BUENO, T. C. D.; HOESCHL, H. C.; MATTOS, E. S.; WANGENHEIM, C. G.; BARCIA, R. M. Uso da teoria jurídica para recuperação em

amplas bases de textos jurídicos. In: Encontro Nacional De Inteligência Artificial. Rio de Janeiro, 1999. Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação. Rio de Janeiro, Edições EntreLugar, 1999. vol. 4, pp. 107-120.

- 7 BUENO, T. C. D.; HOESCHL, H. C.; MATTOS, E. S.; BARCIA, R. M.; BORTOLON, A.; WANGENHEIM, C. G. Jurisconsulto. Florianópolis, Brazil, 1999. Software registrado.
- 8 HOESCHL, H. C. Sistema Olimpo: Tecnologia da informação jurídica para o Conselho de Segurança da ONU. 2002. Tese (Doutorado). Programa de Pós-Graduação em Engenharia de Produção Universidade Federal de Santa Catarina, Florianópolis, Brazil, 2002.
- 9 HOESCHL, H. C.; BARCIA, R. M.; BUENO, T. C. D.; MATTOS, E. S.; BORTOLON, A.; DONATTI, F. T. Sistema Olimpo. Florianópolis, Brazil, 2000. Software registrado.
- 10 HOESCHL, H. C.; BUENO, T. C. D.; MATTOS, E. S.; BORTOLON, A.; BARCIA, R.; M. Olimpo: Contextual Structured Search to improve the representation of UN security Council resolutions with information extraction methods. In: The 8<sup>th</sup> International Conference on Artificial Intelligence and Law. St. Louis, USA, 2001. Proceedings of the Conference. New York, ACM, 2001. vol. 1. pp. 271-218.





# UNL como Elemento Redutor das Barreiras Culturais nos Sistemas CSCW

Damián Rodríguez Sanchez<sup>1,2</sup>  
damian@eps.ufsc.br

Hugo Cesar Hoeschl, Dr.<sup>2</sup>  
metajur@eps.ufsc.br

Vinícius Medina Kern<sup>2</sup>  
kern@eps.ufsc.br

<sup>1</sup> Núcleo de Estudos em Sistemas de Informação da Faculdade Cenecista Presidente Kennedy, Campo Largo, PR, Brasil, damian@presidentekennedy.br

<sup>2</sup> Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, SC, Brasil, 88040-000

## RESUMO

Atualmente, os trabalhos de pesquisa para web à procura de mecanismos que permitirão não só a representação de informação na internet mas também o "entendimento" da mesma pelos computadores vêm centrando a atenção na Web Semântica e tecnologias como XML e RDF. Nesse contexto, mas visando mais precisamente o entendimento da informação pelas próprias pessoas, apresenta-se neste trabalho uma análise das diversas diferenças culturais que dificultam o adequado uso dos sistemas de CSCW, dando ênfase no relativo à barreira da língua. Finalmente, apresenta-se o projeto UNL e propõe-se como solução natural para essa problemática.

## 1. INTRODUÇÃO

O Trabalho Cooperativo por Meio do Computador (*Computer Supported Cooperative Work* - CSCW) inclui uma grande

variedade de aplicações que vão desde editores cooperativos até sistemas de engenharia concorrente e sistemas de teleconferência. Estas aplicações favorecem a integração dos usuários e permitem um aumento do rendimento na execução das tarefas, em relação ao trabalho individual isolado.

Além do mais, a contínua evolução tecnológica (redes de alta velocidade, web, suporte multimídia etc) tem projetado o trabalho cooperativo no mundo das aplicações em redes geograficamente distribuídas, oferecendo a possibilidade de intercâmbio de idéias e de execução de tarefas em conjunto para um número maior de pessoas.

A globalização está aumentando a demanda de sistemas CSCW que suportem grupos de trabalho geograficamente distribuídos nos quais os usuários possam trabalhar em tempo real. A criação deste tipo de sistema representa novos desafios para os desenvolvedores.

Um dos desafios mais importantes é o de evitar os problemas de usabilidade provocados pela diversidade de usuários de diferentes culturas. Nestes casos, são de se esperar falhas, conforme apontado em Bourges-Waldegg [1]. Nesse trabalho se argumenta que abordagens tradicionais são inapropriadas. Como exemplo cita-se a culturalização que diferencia e agrupa os usuários levando em consideração aspectos culturais para produzir diversas adaptações para cada "cultura objeto".

Em contraste com os sistemas de um único usuário, os usuários de sistemas compartilhados precisam interagir não só com o sistema em si, mas também com os outros usuários. Por este motivo, versões dos sistemas que atendam às diferentes culturas não seriam suficientes. Precisa-se também de uma linguagem comum que permita o entendimento entre os participantes de uma sessão cooperativa.

Existem muitas formas nas quais os aspectos de usabilidade para o software de um único usuário têm sido abordados. A maioria destas recomendações gerais se mantém para os sistemas de CSCW mas elas não são suficientes para garantir uma boa usabilidade de grupo em aplicações de CSCW. Existem assuntos específicos que os projetistas de aplicações precisam considerar. Por exemplo, as formas de cooperação e interação entre os participantes, onde o idioma joga um papel importante.

O presente trabalho pretende analisar as diferenças culturais nos sistemas CSCW, assim como apresentar uma nova tecnologia, que pode vir a solucionar uma parte fundamental desta problemática: o idioma. Trata-se da proposta de interlíngua UNL (*Unified Networking Language*). Não se pretende entrar em detalhes de soluções técnicas que permitam o casamento de ambas tecnologias, apenas apresentar uma visão geral do que pode vir a se desenvolver numa importante linha de pesquisa.

## **2. CSCW E A BARREIRA DA LÍNGUA**

CSCW pode ser definido como a disciplina de pesquisa para o estudo das técnicas e metodologias de trabalho em grupo e das formas como a tecnologia pode auxiliar este trabalho [2].

Na literatura existente, os aspectos culturais dos sistemas CSCW normalmente se centram na identificação e classificação de problemas específicos de usabilidade causados por diferenças culturais. Em Scrivener et al [3] classificam-se esses problemas em seis grandes áreas:

1. Convenções sócio-lingüísticas: culturas diferentes têm diferentes conceitos sobre o que consideram apropriado no uso da linguagem.

2. Convenções tipográficas: cada cultura tem formatos diferentes para a representação dos números, datas, hora etc.
3. Ícones e símbolos: a interpretação de símbolos e ícones varia de uma cultura para outra.
4. Instituições: existem diferentes sistemas econômicos, monetários e legais em cada cultura.
5. Convenções psicossociais: pessoas de diferentes culturas podem apreciar de forma diferente o que consideram humorístico ou tocante.
6. Idioma: se os usuários não falam ou escrevem fluentemente a mesma língua existe o perigo de aparecerem dificuldades na comunicação.

Segundo Bourges-Waldegg [1], esta classificação é útil para reconhecer os problemas que acontecem quando usuários de diferentes culturas interagem através de sistemas CSCW. Porém, é importante que estes aspectos sejam levados em consideração pelos desenvolvedores de sistemas para tentar minimizar os problemas que estas diferenças produzem.

Em Kern [10] se menciona a hipótese de Sapir-Whorf que propõe que a estrutura da língua nativa influencia fortemente a visão de mundo que uma pessoa adquire enquanto aprende a língua. Por esse motivo quando a comunicação acontece na língua nativa da pessoa existe uma melhor compreensão do contexto analisado já que se faz referência aos conceitos que naturalmente a pessoa foi construindo quando acontecia a aprendizagem natural da língua.

O problema da comunicação por motivos da língua sempre existiu. Foi isto que motivou o lingüista polaco Lázaro Zamenhof a criar uma linguagem universal que se conheceu

com o nome de esperanto. Zamenhof projetou esta linguagem para que as pessoas que habitavam seu povoado, uma mistura de russos, alemães e poloneses, tivessem uma ferramenta para poder se comunicar ultrapassando as barreiras impostas pelos diferentes idiomas.

Na atualidade temos um panorama parecido, só que o tamanho do povoado aumentou, assim como o número de pessoas que mediante o uso das novas tecnologias tentam se comunicar para resolver seus problemas. A globalização tem impulsionado o aumento da demanda pelas aplicações CSCW. A cooperação nas empresas torna-se imperativa, principalmente porque muitas têm seus escritórios em diferentes partes do mundo. É igualmente inevitável que o trabalho cooperativo represente uma mudança de paradigma uma vez que modificam-se as formas tradicionais de trabalho e o ambiente de comunicação organizacional [4].

Segundo Dornelas e Vargas [5], esta mudança de paradigma é resultante da convergência de uma série de fenômenos. Dentre os principais:

1. as oportunidades decorrentes da disponibilidade e ampliação de uso das redes de computadores;
2. o desejo de atender a pequenos grupos de trabalhos estabelecidos (workgroups), simultaneamente com requintes de comunicação e interação;
3. a execução de pesquisas e desenvolvimento de aplicações para prover suporte a decisão a grupos de executivos e gerentes;
4. a conjunção de telecomunicações e computadores, eliminando limitações geográficas e proporcionando utilização de recursos e mídias de representação antes inimagináveis;
5. a globalização dos mercados levando as organizações a novas formulações em termos de estruturas e parcerias e a redefinição do ambiente de negócios.

No entanto, apesar dos sistemas CSCW serem uma ferramenta eficaz para juntar idéias e desenvolver pesquisas e trabalhos em colaboração entre profissionais de diferentes cantos do mundo, juntando conhecimentos e experiências que podem contribuir de forma significativa ao desenvolvimento das nações, um dos principais problemas enfrentados por seus usuários encontra-se na barreira da língua. O inglês, na maioria das vezes, é a língua predominante nesse meio, apesar de ser a língua materna de uma pequena porcentagem da população mundial. Na Europa Ocidental, por exemplo, somente 15% das 387 milhões de pessoas têm o inglês como língua nativa, enquanto 28% têm um certo nível de conhecimento deste idioma [6].

Por isso, pelo menos uma das partes deve estabelecer a comunicação utilizando uma segunda língua, o que pode prejudicar a dinâmica da interação assim como a exatidão das informações trocadas.

Tem-se ainda o problema da redução do número de usuários potenciais deste tipo de sistemas. Muitas organizações e indivíduos resolveriam problemas do seu dia-a-dia em colaboração com seus pares espalhados pelo mundo se não fosse pela barreira do idioma, seja pelo desconhecimento da outra língua, seja pelo esforço adicional que representa seu uso para elaborar idéias.

### **3. APARECE A UNL COMO UMA ESPERANÇA DE SOLUÇÃO**

A proposta de linguagem de representação universal conhecida como UNL [11] (*Universal Networking Language*) é um projeto financiado pela Universidade das Nações Unidas (*The United Nations University* - UNU), mais particularmente pelo Instituto de Estudos Avançados (*Institute of Advanced Studies* -

IAS/UNU), com sede em Tóquio, que pode vir a solucionar em parte as limitações que a tecnologia CSCW tem em relação à problemática das diferenças culturais e, especificamente, do idioma.

Já existem diversos sistemas de tradução automática. Alguns, como o Systrans ([www.systrans.com](http://www.systrans.com)), até diretamente acessíveis por internet, o que permite traduzir automaticamente um documento encontrado na rede. Mas todas essas ferramentas (baseadas no uso de dicionários que levaram anos de trabalho) permitem tratar apenas duas línguas por vez. Pode-se ter, por exemplo, 6 dicionários que permitam realizar traduções do inglês para seis outros idiomas (francês, alemão, espanhol, italiano, japonês e russo), mas não entre o espanhol e o alemão. Seria necessário possuir 15 dicionários adicionais para cobrir o conjunto das combinações possíveis com 7 idiomas.

Segundo Delbecq [7], seriam necessários mais de 50.000 dicionários para administrar todas as combinações das línguas faladas na Terra.

O novo projeto de língua coloca o problema de forma radicalmente diferente. A UNL está sendo construída por 17 equipes de pesquisadores como uma língua livre de todas as ambigüidades das línguas tradicionais, podendo servir como intermediário de toda tradução de uma língua para outra, quaisquer que sejam. Os seus desenvolvedores estão em processo de definição de dicionários constituídos de um conjunto de palavras universais conhecidas como UW (*Universal Words*) escolhidas essencialmente do inglês e acompanhadas de comandos que lhes permitem descrever o contexto e as relações com outras UWs.

Os pronomes e artigos das línguas ocidentais não são levados em consideração. Ao invés de representar a especificidade de cada língua, tenta-se representar aquilo que é comum entre

elas, sendo no final traduzido de uma língua para outra, apenas o sentido do enunciado inicial. A UNL não é capaz de representar todas as sutilezas de uma língua (pelo menos na sua concessão atual). Ela apenas transporta grosseiramente o sentido da frase. Grandes textos literários ou tratados internacionais que não suportariam uma análise "grosseira" estarão, portanto, excluídos da UNL por um bom tempo, talvez para sempre.

A internet, principal lugar para a implantação da UNL e também principal meio de comunicação para usuários de sistemas CSCW espalhados geograficamente, está composta de um conjunto de computadores que contêm informações e que estão interconectados entre si. Quando os provedores das informações incorporarem o sistema UNL, a base documental ficará escrita na linguagem UNL. Esta linguagem permite conectar o computador com a informação a ser transmitida com um componente de sistema denominado servidor de linguagem. Teremos servidores de linguagens em todos os idiomas do sistema. Desta forma os usuários receberão a informação na linguagem nativa.

A escrita de documentos na linguagem UNL é feita através de um componente chamado de forma genérica de *editor UNL*. Seu elemento essencial é um conjunto de programas chamados de *enconverter*, fornecido pelo Instituto de Estudos Avançados da Universidade das Nações Unidas, o qual se encarrega de codificar ou manipular as regras de funcionamento de um idioma. Em cada idioma as regras e o dicionário são desenvolvidos por equipes de pesquisa dos diferentes países participantes. No dicionário as palavras de cada linguagem estão associadas à UW correspondente.

Um sistema UNL deverá ter, por último, um *visualizador UNL* do lado do usuário interessado em consumir a informação originalmente criada na língua estrangeira. Como resultado de um processo de decodificação, o visualizador mostraria a



mesma informação na nova língua, utilizando para isso um *deconverter*.

Surge então uma ferramenta que promete eliminar os problemas causados pela barreira lingüística nas aplicações que envolvam envio e recebimento de texto entre pessoas de diferentes línguas com vantagens sobre as abordagens atuais. Estas aplicações têm muito em comum com as aplicações de CSCW, em cujos ambientes de trabalhos, como veremos, reúnem-se diversos recursos computacionais similares.

#### **4. SUPORTANDO UNL NO SISTEMA CSCW**

A comunicação entre indivíduos, ou entre indivíduos e grupos, e vice-versa, constitui um ponto chave para que a cooperação ocorra e que um grupo de usuário ou de aplicações dispersas geograficamente possa utilizar vários recursos computacionais dentro de um ambiente que permita a resolução de problemas conjuntamente e de forma mais eficiente.

Vários exemplos destes ambientes cooperativos podem ser citados: *Basic Support for Cooperative Work* ([www.bscw.gmd.de](http://www.bscw.gmd.de)), *Cu-SeeMe* ([www.cu-seeme.net](http://www.cu-seeme.net)), *AulaNet* ([www.les.inf.puc-rio.br/aulanet](http://www.les.inf.puc-rio.br/aulanet)), *Web Course Tools - WebCT* ([www.webct.com](http://www.webct.com)), cada um com sua especificidade e funcionalidade. No geral, porém, encontramos um conjunto básico de características para a comunicação e armazenamento de objetos que se complementam na tarefa de facilitar a cooperação. Podemos resumi-las em Moraes e Zorzo [8]:

1. Armazenamento das atividades dos alunos
2. Whiteboard
3. Quadro de avisos
4. Correio eletrônico
5. Lista de discussão
6. Suporte a áudio

7. Suporte a vídeo
8. Conferência
9. Chat
10. Armazenamento de objetos

A incorporação de suporte a UNL nos sistemas CSCW seria facilmente realizada mediante a transformação em editores e visualizadores UNL daqueles recursos computacionais compartilhados que não envolvem a transmissão de mídias contínuas. Teria-se então a aplicação da tecnologia relacionada à UNL de forma quase direta aí onde a informação trocada for puro texto. Os processos envolvidos de codificação e decodificação teriam no entanto que ser realizados em tempo real para permitir a interatividade natural dos processos colaborativos.

Por outro lado, durante os últimos 10 anos tem se observado um progresso significativo nas tecnologias de reconhecimento de fala. As taxas de erro de palavras caem por um fator de 2 a cada dois anos, os sistemas são cada vez mais independentes do locutor, as falas mais contínuas e o vocabulário maior. Vários fatores têm contribuído para isto, mas o fundamental é o fato de se ter alcançado um certo grau de maturidade no desenvolvimento das ferramentas mais utilizadas como são os modelos ocultos de Markov (HMM) e as redes neurais artificiais (ANN) [9].

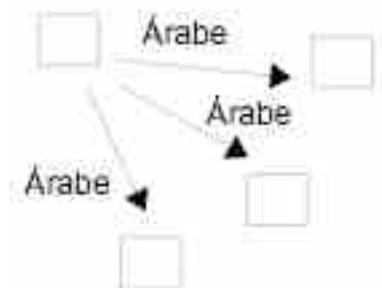
O estado atual desta tecnologia, assim como a de geração de fala (menos complexa tecnologicamente), acrescentariam então mais uma etapa nesta seqüência de processamento: transformar a fala em informação textual antes de realizar sua codificação/decodificação, permitindo um ambiente de cooperação completo quase totalmente transparente do ponto de vista linguístico. O futuro dirá até que ponto isto será viável.

Como dito anteriormente, a UNL não é capaz de representar todas as sutilezas de uma língua. Cada palavra pode ter inúmeras acessões dependendo do contexto, e estas, obviamente, seriam diferentes de uma língua para outra. Para

facilitar a junção da tecnologia relativa à UNL com a dos sistemas CSCW num futuro mais imediato, o uso de dicionários específicos à área de conhecimento na qual se pretenda colaborar será imperativo. Isto é uma prática natural já levada em consideração no desenvolvimento atual de UNL que permite a materialização de uma tecnologia que de outra forma seria inviável devido à enorme complexidade da linguagem natural. A criação de dicionários específicos centram a atenção nas UWs específicas à área do conhecimento em questão.

## 5. VISUALIZANDO O PROCESSO

Para ilustrar os processos de codificação e decodificação que estariam envolvidos numa sessão de CSCW com UNL podemos imaginar um simples *broadcast* de uma mensagem de texto para os diferentes usuários participantes. Sem UNL, cada usuário receptor receberia obrigatoriamente a mensagem na língua em que foi criada, digamos árabe. Esta situação está representada na Figura 1.



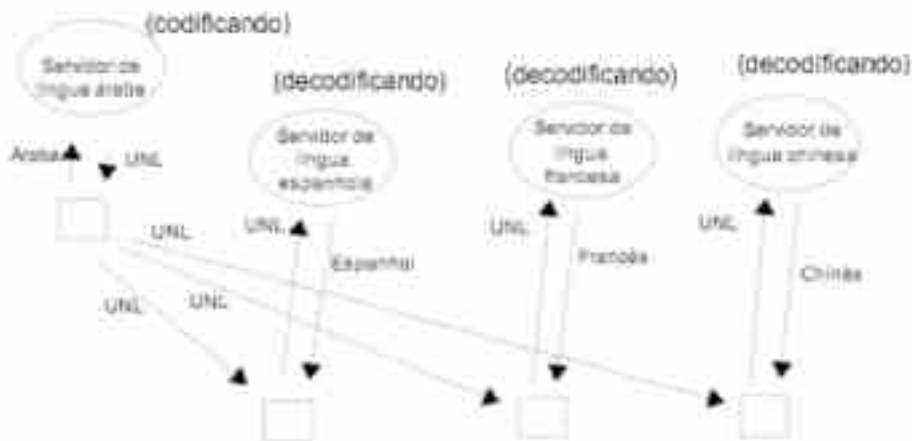
**Figura 1. Broadcast sem UNL**

Com UNL teríamos os processos de codificação e decodificação já descritos sendo oferecidos por servidores de linguagens residentes no domínio da internet. Eles são respectivamente os responsáveis pela tradução do texto da língua original para UNL, e de UNL para qualquer outra língua.

Imaginemos agora que o idioma utilizado pelo originador do nosso *broadcast* continue sendo o árabe, enquanto que os nossos receptores só entendem espanhol, francês e chinês respectivamente.

Sendo assim, o editor da mensagem em árabe reconheceria o conteúdo como árabe e enviaria uma requisição para o servidor de língua árabe para codificar o texto. Uma vez codificado para UNL, o servidor de língua árabe envia o resultado de volta para ao editor UNL. O utilitário em uso no ambiente CSCW para o envio das mensagens fará então, na verdade, o *broadcast* da mensagem em UNL.

Uma vez recebida a mensagem, cada receptor enviará então, por sua vez, uma requisição ao servidor da língua de sua preferência (servidores de língua espanhola, francesa e chinesa no nosso exemplo). Isto fará com que o texto em UNL seja traduzido para a língua desejada em cada caso (decodificado) e mostrado dessa forma por cada visualizador. A Figura 2 mostra o processo descrito.



**Figura 2. Broadcast com UNL**

## 6. COMENTÁRIOS FINAIS

Neste trabalho foi descrito como a rápida disseminação das novas tecnologias, especialmente da internet, tem trazido cada vez mais uma maior diversidade de usuários de diferentes culturas querendo realizar trabalhos em ambientes de CSCW. As diferenças culturais e suas repercussões negativas em relação à usabilidade dos sistemas CSCW foram analisadas, e se mostrou o grande desafio que elas constituem para os desenvolvedores de sistemas CSCW. Obviamente, a língua constitui a principal barreira neste sentido.

Viu-se como o projeto UNL vem resolver esta questão e como o CSCW se adequa naturalmente a esta tecnologia.

Há, no entanto, grandes desafios neste casamento, para não dizer no desenvolvimento da própria UNL - que se encontra ainda nas primeiras etapas de vida e que deverá lidar com as enormes sutilezas da linguagem natural e a infinidade de significados que cada palavra pode ter numa única língua segundo o contexto no qual esteja inserido.

Devido à interação natural entre os usuários dos sistemas CSCW e ao grande volume de processamento envolvido nos processos de codificação e decodificação da UNL, maiores demoras acontecerão, fazendo com que problemas de sincronia em tempo real tenham que ser analisados e adequados ao novo panorama.

Considera-se o uso conjunto de UNL e CSCW como uma importante linha de pesquisa que nasce. Estimula-se então o desenvolvimento de dissertações de mestrado e teses de doutorado que contribuam para termos um dia ambientes de CSCW onde pessoas de diferentes línguas possam trabalhar juntas na solução de objetivos comuns, mesmo não conhecendo a língua dos outros.

## 7. REFERÊNCIAS

- 1 BOURGES-WALDEGG, L. Dealing with cultural differences in computer supported cooperative work. A one-day workshop on workspace models and collaboration. Department of Computer Science, Queen Mary & Westfield College, University of London, Apr 1999.
- 2 GREENBERG, S. Personalizable groupware: accomodating individual roles and group differences. ECSCW '91 European Conference of Computer Supported Cooperative Work, pp. 17-32. Amsterdam, Kluwer Academic Press, 1991.
- 3 SCRIVENER, S. A. R.; HARRIS, D.; CLARK, S.; ROCKOFF, T.; SMYTH, M. Designing at a distance via real-time designer-to-designer interaction. Design Studies, vol. 14, n. 3, pp. 261-282, 1993.
- 4 ROBINSON, M. Computer supported cooperative work: cases and concepts. Groupware '91. The Netherlands, Serc Publications, 1991.
- 5 DORNELAS, J. S.; VARGAS, L. M. Contribuições para a análise de informações sobre groupware a partir do uso de base de dados. Revista Read, 3<sup>a</sup> ed., v. 2, n. 2, nov. 1996.
- 6 SCHWARTZ, G. A internet ainda não mudou o mapa-múndi. Folha de S. Paulo, 26 set 1999. Disponível em [www.race.nuca.ie.ufrj.br/journal/s/s.htm](http://www.race.nuca.ie.ufrj.br/journal/s/s.htm). Acesso em maio 2002.
- 7 DELBECK, D. UNL, un nuevo "esperanto" para la web. Le Monde, 5 dez 1998. Tradução. Disponível em [166.114.106.9/~arteaga/INF-099/www.puc.cl/curso\\_dist/cbc/anexos/texto\\_a/natrad.html](http://166.114.106.9/~arteaga/INF-099/www.puc.cl/curso_dist/cbc/anexos/texto_a/natrad.html). Acesso em maio 2002.
- 8 MORAES, I.; ZORZO, A. F. Uma arquitetura genérica para aplicações colaborativas. Relatório técnico. FACIN, PUCRS, Campus Global, Rio Grande do Sul, Brasil, 2000.
- 9 ZUE, V.; COLE, R. Survey of the state of art in human language technology. 1996. Disponível em [cslu.cse.ogi.edu/HLTsurvey](http://cslu.cse.ogi.edu/HLTsurvey). Acesso em maio 2002.
- 10 KERN, V. M. Manutenibilidade da semântica de modelos de dados de produtos compartilhados em rede interoperável. 1997. Tese

(Doutorado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Brasil, 1997.

- 11 UNL CENTER. The Universal Language (UNL) specifications. Version 3, Edition 1. UNDL Foundation. Disponível em [www.ul.ias.unu.edu/unlsys/unl Specifications.htm](http://www.ul.ias.unu.edu/unlsys/unl%20Specifications.htm). Acesso em maio 2002.







**ENEMPRE**  
5º Encontro Nacional de Empreendedorismo  
(2003, Florianópolis)  
[www.ene.ufsc.br/enempre2003](http://www.ene.ufsc.br/enempre2003)



# Por Que as Pequenas Empresas Devem Tentar o e-Business?

Felipe Zurita Quadros, Esp.<sup>1</sup>  
felipe@autocommerce.com.br

Rafael Otacílio Nowascky, Esp.<sup>2</sup>  
rafael@outplan.com

Francisco Pereira da Silva, Dr.<sup>3</sup>  
francisco@pfpassociados.com.br

Hugo Cesar Hoeschl, Post Doc.<sup>4</sup>  
digesto@digesto.net

Maria Marta Leite, M.Sc.<sup>5</sup>  
marta@inf.ufsc.br

## RESUMO

Este artigo visa aprofundar as definições que permeiam o desenvolvimento de estratégias para implantação de plataformas de comércio eletrônico entre empresas (B2B), com ênfase nos benefícios envolvidos neste tipo de prática para pequenas empresas. Dentro do universo de novos empreendimentos, busca-se identificar as melhores práticas e estratégias aplicadas a este tipo de processo através da análise e observação de casos ocorridos no Brasil e no mundo recentemente, além de dados e estatísticas acerca deste mercado. Através da segmentação de alguns tipos de negócios eletrônicos busca-se melhor delinear este trabalho, enumerando-se as vantagens alcançadas

---

<sup>1</sup> Mestrando em Engenharia de Produção e Sistemas pela UFSC e especialista em e-Commerce pela UNESA-RJ.

<sup>2</sup> Especialista em e-Commerce pela UNESA-RJ.

<sup>3</sup> Doutor em Engenharia de Produção e Sistemas pela UFSC.

<sup>4</sup> Pós-Doutor em Engenharia de Produção e Sistemas pela UFSC.

<sup>5</sup> Doutoranda em Engenharia de Produção e Sistemas pela UFSC.

através da implantação destas plataformas para todos os envolvidos neste tipo de processo de negócios.

## 1. INTRODUÇÃO

O panorama mundial de gestão corporativa está sendo alvo de inúmeros conceitos, siglas, estratégias e inovações tecnológicas ultimamente, tendo como foco principal o relacionamento das corporações com seus fornecedores, clientes, parceiros, colaboradores, e até mesmo seus próprios concorrentes. E isto atinge em cheio todos os níveis empresariais.

A realidade dinâmica da nova economia mostra que o posicionamento e as decisões estratégicas que são tomadas por quem está à frente de um negócio envolve cada vez mais uma quantidade enorme de variáveis. Isto torna mais crítica todas as atividades de gestão integrada de uma empresa, indo desde a administração de compras e logística, e indo até o setor de marketing e de relacionamento com clientes, por exemplo.

Quando e como tomá-las, quais negócios executar e como estruturar a sua organização, são algumas das questões em pauta. Nunca uma estratégia altamente focada foi tão crítica para o sucesso dos negócios. Penzias (2001), ganhador do premio Nobel de Física em 1978, descreve acerca de variados tipos de redes e como elas devem servir as empresas, colocando que as redes deverão ir muito além das fronteiras corporativas, criando o que se chama de "empresa expandida".

Em 2003, segundo previsões da E-Consulting Corp., todas as operações de comércio eletrônico no Brasil envolvendo *EmpresaXEmpresa* (B2B) e *EmpresaXConsumidor* (B2C) no Brasil movimentarão US\$ 8.7 bilhões, sendo só o segmento B2B responsável por 78% desse valor. A princípio trata-se de um valor considerável, mas se comparado ao montante envolvido neste tipo de operação em países como os Estados

Unidos, por exemplo, os números do Brasil ainda deixam muito a desejar. Afinal, só o mercado americano de B2B atingirá aproximadamente US\$ 819 bilhões em 2003 e espera-se que atinja US\$ 1,8 trilhões em 2005.

Esta grande oportunidade trazida pela crescente ascensão dos mercados eletrônicos e ferramentas de colaboração entre empresas mostra tal fato não é apenas uma tendência, mas sim que vieram para revolucionar a forma de trabalhar das empresas, principalmente as pequenas e médias, pois tendem a colocá-las em pé de igualdade com grandes. Analisando o nível utilização e os benefícios que estas práticas estão trazendo para empresas do mundo inteiro também se percebe facilmente que o caminho a ser percorrido pelo Brasil é grande, porém extremamente promissor.

## 2. ESTRATÉGIA ELETRÔNICA

Os fundamentos básicos na tomada de decisão para empreendedores ainda são aplicados: identificar necessidades e desejos dos clientes, definir o valor único agregado (*unique selling proposition*) a ser oferecido, e como apresentar esse valor para obter uma vantagem competitiva capaz de sustentar estratégias. E esta é a grande vantagem encontrada nos processos de comércio eletrônico quando aplicados à gestão de negócios, seja na área de compras, distribuição, ou vendas por exemplo.

Estes são preceitos básicos. Porém, esses fundamentos estratégicos são muito mais desafiadores e difíceis de serem definidos e executados num mundo onde as leis de negócios estão sendo reescritas a todo o momento e o cenário competitivo está em constante mutação. A estratégia nesses casos é a de buscar cada vez mais um caminho para a integração dos mais variados processos. Para Porter (1996), criar encaixe entre atividades de uma empresa é o que deverá

definir o sucesso de uma estratégia. Os encaixes deverão ser cada vez mais numerosos e, conseqüentemente, mais complexos. Sem estes encaixes entre as atividades, não haverá uma estratégia específica, e sua sustentabilidade deverá ser muito pequena. Além disso, também se deve tomar como base um posicionamento baseado em necessidades, completa o autor, sendo que sempre há a necessidade de um maior compartilhamento de conhecimento entre os parceiros. Cunningham (2001) confirma estas tendências ao afirmar que o comércio eletrônico tende a se tornar rapidamente o pilar principal de sustentação da estratégia, das operações e do sistema de informação nas organizações, independente de seu porte.

Todas as empresas, sejam pequenas, médias ou grandes, já passaram ou estão passando pela revolução da qualidade de produtos e serviços, principalmente os baseados em tecnologia. Segundo Albrecht (2001) é preciso prestar atenção em outra revolução que está a caminho, a "revolução da qualidade da informação". Esta deve levar à sistematização de formas de evitar custos elevados resultantes dos defeitos ou erros de manipulação das informações. O segredo para o sucesso de uma plataforma B2B é senão estar baseado num posicionamento que vise prover acesso de forma rápida e eficaz a todos os possíveis envolvidos no processo, sem distinção. Para Porter (1996) isto reflete um *posicionamento baseado no acesso*. Tal atitude objetiva segmentar, e posteriormente agrupar, consumidores com acessibilidades diferentes. Desta forma, dá a cada empresa a chance de negociar em pé de igualdade com todas as outras, tendo em vista que o acesso à informação é irrestrito, e as políticas de negociação extremamente claras e bem definidas. O acesso pode ser função da geografia ou da escala dos consumidores, neste caso das próprias empresas, ou em função de qualquer coisa que exija um conjunto diferente de atividades para alcançar da melhor forma todos os envolvidos no processo. Isto possibilita, por exemplo, pequenas empresas que atuem em sua microrregiões, a atuar em

outras regiões geográficas que não as suas, tendo em vista a capacidade de quebrar fronteiras físicas que a internet possui.

### **3. A PEQUENA EMPRESA E O BUSINESS TO BUSINESS (B2B)**

A primeira fase do mercado de operações B2B (negócios eletrônicos entre empresas) foi dominada por iniciativas locais e por ações capitaneadas por empresas (com seus fornecedores, revendedores, varejistas, etc). Agora grandes conglomerados internacionais, que representam segmentos verticais, tais como varejo, mineração, químico e automotivo, roubam a cena do e-commerce no mundo e no Brasil. Citando alguns nomes temos o *Transora*, megaportal que reúne gigantes como Unilever, Coca-Cola e Procter & Gamble para compras conjuntas. Outro bom exemplo é o portal *Covisint*, controlado pelas montadoras General Motors, Ford, Daimler-Chrysler, Renault e Nissan para compras conjuntas de insumos. O Covisint já opera no Brasil há cerca de dois anos. Além de ser o maior portal de compras automotivo do mundo, está avançando a passos largos. O site já realizou diversas compras para a GM no Brasil e está acertando com a Ford e com a Daimler-Chrysler os detalhes para fazer o mesmo. Só para se ter uma idéia, as três companhias detêm 81% das ações do site e, em conjunto com a Renault e a Nissan, pretendem comprar em 2005 US\$ 240 bilhões em pneus, parafusos e uniformes no mundo inteiro, através do portal. No ano passado esse total foi de US\$ 45 bilhões, dos quais apenas US\$ 150 milhões no Brasil. Mas o interessante de todos estes números é que atualmente as pequenas empresas estão desvendando este universo. Já é possível para qualquer empreendedor facilmente se cadastrar dentro de portais desta natureza. Isto possibilita para a pequena empresa tornar-se um fornecedor de um grande grupo ou conglomerado, ou ainda participar de processos de compra conjunta com grandes empresas, obtendo assim grandes descontos graças a um

gigantesco ganho de escala na hora da compra, sem falar na integração de processos com seus fornecedores e clientes. E o mais interessante: este ganho é repartido para todos os envolvidos no processo, seja ele o pequeno empreendedor ou o grande empresário.

#### **4. A TECNOLOGIA EM FAVOR DOS NEGÓCIOS**

Diferentemente de companhias online, focadas nos consumidores (varejistas online) - que muitos administradores desenvolvem com conhecimento industrial ou experiência mínima - as operações B2B demandam uma profundidade de conhecimento industrial ou conhecimento de processos especializados, bom relacionamento com compradores e fornecedores da respectiva área industrial e entendimento de como os ambientes digitais podem tornar as transações mais fáceis para todos, pequenas, médias e grandes empresas.

As soluções B2B são classificadas por várias formas de interação entre empresas, que vão desde a integração de processos, incluindo ERP, CRM ou SCM, até portais de compra e venda de uma empresa específica (e-procurement), portais setoriais/verticais, como de uma determinada indústria e portais matriciais de transação entre várias empresas (e-marketplaces). Os modelos de transações comerciais do B2B são muitos. Entre eles podemos citar: leilão, leilão reverso ou clubes de compra, envolvendo desde sobras de estoque até compras corporativas.

A internet tem um papel fundamental e seu impacto na condução de negócios é inquestionável. Afinal, só através da criação dos locais de negociação eletrônica ou *marketplaces* foi possível a chegar neste estágio de realização de negócios eletronicamente (e-commerce). Os benefícios para os compradores são a agilidade e dinamismo das compras, redução expressiva de custos, menos burocracia, decisões de



compra mais rápidas. O fato de comprar e vender remotamente em qualquer hora e lugar altera produtos, processos e relações entre clientes, empresas, fornecedores e intermediários, afirma Turban (1999). Para os vendedores, as vantagens são a ampliação da carteira de clientes, maior exposição através da internet, redução de custos na administração de vendas.

Destaca-se com certo vigor dentre as soluções B2B o e-procurement, uma ferramenta que dá às empresas acesso ilimitado a informações e transações realizadas dentro de um ambiente conectado eletronicamente, inter ou intra-organizacionalmente. Compradores corporativos são atraídos para soluções de e-procurement e-marketplaces porque dessa forma podem encontrar produtos de marcas concorrentes e assim comparar preços, condições, características e serviços antes de efetivar a decisão de compra, trazendo assim benefícios reais e palpáveis como economia, tempo e transparência nos processos. E isto em boa parte graças às possibilidades que estes mecanismos fornecem para a democratização das operações de negócios entre as empresas, tornando possível, por exemplo, que uma pequena empresa se torne fornecedora de uma grande rede graças à sua capacidade de atender pedidos no prazo ou por possuir uma política de preços diferenciada, por exemplo.

Os mercados eletrônicos (*marketplaces*), por sua vez, proporcionam uma vasta oferta de produtos, serviços e conteúdo, além de proporcionar um local para múltiplos tipos de negociações. Como afirmado anteriormente, e-marketplaces (mercados eletrônicos) freqüentemente incluem trocas como parte de seus serviços. Por exemplo, um mercado pode oferecer catálogos, negociações e trocas adicionalmente a vários serviços dentro da cadeia de suprimento para várias empresas localizadas em várias partes do mundo simultaneamente.

Complementarmente, e-marketplaces proporcionam conteúdo com alto valor agregado e ambiente para formação de comunidades a seus participantes. Dentro do conteúdo podem estar incluídas notícias especializadas e artigos sobre o setor. Também podem figurar, dentro dos e-marketplaces, treinamentos online que contenham tutoriais sobre as mais recentes soluções do setor. Vale lembrar que estas informações são muito importantes para o pleno desenvolvimento de pequenas empresas, as quais freqüentemente não dispõem de dinheiro em caixa para ações de pesquisa e desenvolvimento.

De acordo com essa visão, o e-procurement é a utilização da internet para a automatização dos processos de compra e gestão de bens e serviços necessários à atividade da empresa. O e-procurement pode ser utilizado como a base para uma revisão estratégica do processo de compras, eliminando as compras fora de contratos e melhorando o relacionamento com todos os fornecedores.

## **5. CONCLUSÃO**

Vale ressaltar que a internet é mola propulsora no desafio estratégico de otimizar esforços e custos no implacável mundo dos negócios. Os avanços na comunicação são pontos relevantes para o sucesso das estratégias corporativas, especialmente no âmbito das pequenas empresas. Fechar os olhos para a tecnologia da informação nesse momento é o mesmo que jogar os resultados à própria sorte.

As empresas não dispõem de tempo para filosofar sobre velhos paradigmas. Reinventar e aplicar tais conceitos no mundo moderno pode ser visto muitas vezes até como uma tarefa suicida. O mercado exige soluções passíveis de serem rapidamente implementadas, que sejam pouco invasivas, e o mais importante, que efetivamente sirvam ao propósito de alavancar o empreendimento de forma a garantir o crescimen-

to da empresa. Neste caso, o empreendedor terá de flexibilizar suas ações no sentido de aproveitar ao máximo as oportunidades que surgem todos os dias no plano tecnológico, pois a grande questão trazida à tona pelas novas tecnologias aplicadas ao e-business não reside no fato de ter-se criado uma nova tecnologia em si. Não se trata de reinventar-se a roda! O que está ocorrendo é a descoberta de novas aplicações a tecnologias já conhecidas e disponíveis no mercado, cada qual focada em um propósito específico. E a todo o momento surgem pequenas empresas, que após adotarem uma postura pró-ativa do ponto de vista tecnológico, ao adotar determinadas práticas de e-business, encontram o caminho do sucesso e do rápido crescimento. Isto mostra que abrir seus horizontes e repensar seu modelo de negócio, além de não ser uma tarefa fácil, é algo capaz de selar o destino de algumas corporações.

Toda esta reviravolta dentro das corporações reflete perfeitamente o que é tratado por Drucker (2001) como a "Revolução da Informação". Segundo o autor, esta revolução está atualmente no mesmo patamar em que a revolução industrial estava no início da década de 1820. Naquele cenário os tecidos caíam de preço, assim como vem acontecendo com os microchips. Mas em seus primeiros 50 anos, o único fato relevante talvez tenha sido a mecanização da produção de mercadorias que já existiam havia muito tempo. Sua explosão neste caso viria mais tarde.

O *business to business* (B2B) consolida-se cada vez mais como uma necessidade para o mercado, principalmente para as pequenas e médias empresas. A nova forma das empresas fazerem negócios entre si traduz-se numa criação ou integração de processos e operações que usa meios, sistemas ou plataformas tecnológicas, permitindo por exemplo que empresas da mesma área ou de áreas diferentes, parceiros ou concorrentes, fornecedores ou clientes, possam comprar, vender, trocar, gerenciar e comunicar-se eletronicamente, dividindo informações e recursos com o objetivo de facilitar,

maximizar e/ou intensificar transações através de ganhos de escala até então inimagináveis há alguns anos atrás. E isto, sem dúvida, representa uma mudança radical e sem precedentes na forma de se fazer negócios em todo o mundo.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALBRECHT, K. A terceira revolução da qualidade. A TI no dia-a-dia: como aproveitá-la externamente. Coletânea HSM Management e-Business e Tecnologia. São Paulo, Publifolha, 2001.
- ALVES, G. Entrevista com o coordenador do processo de supply chain e e-procurement da Intelbras S.A. São José, out. 2002.
- CUNNINGHAM, M. B2B: how to build a profitable e-commerce strategy. Cambridge, Perseus Publishing, 2001.
- DRUCKER, P. Além da revolução - O que a internet realmente representa. Coletânea HSM Management e-Business e Tecnologia. São Paulo, Publifolha, 2001.
- E-CONSULTING® CORP. Internetização. Disponível em [www.e-consultingcorp.com.br/insider\\_info/indicadores.shtml](http://www.e-consultingcorp.com.br/insider_info/indicadores.shtml). Acesso em 22 set. 2003.
- ISTO É DINHEIRO. A peça que faltava. Revista. Disponível em [www.terra.com.br/istoedinheiro/260/ecommerce/260\\_peca\\_faltava.htm](http://www.terra.com.br/istoedinheiro/260/ecommerce/260_peca_faltava.htm). Acesso em 19 set. 2003.
- LAMBERT, D. M.; STOCK, J. R.; ELLRAM, L. M. Fundamentals of logistics management. USA, Irwin McGraw-Hill, 1998.
- PENZIAS, A. Lições de um Prêmio Nobel: o peso da TI e das redes na economia. Coletânea HSM Management e-Business e Tecnologia. São Paulo, Publifolha, 2001.
- PORTER, M. E. What is strategy? Harvard Business Review. Boston, v. 74, n. 6, pp. 61-68, Nov.-Dec. 1996.
- TURBAN, E.; LEE, J.; KING, D.; CHUNG, H. M. Electronic commerce: a managerial perspective. New Jersey, Prentice-Hall, 1999.