

## **AlphaThemis<sup>®</sup> - Do texto ao conhecimento**

Hugo Cesar Hoeschl, Tânia Cristina D'Agostini Bueno, Andre Bortolon, Eduardo Mattos, Marcelo Stopanovski Ribeiro

Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Campus Universitário, Trindade, Florianópolis, Santa Catarina, Brasil.

CEP: 88040-900

{metajur, bueno, bortolon, emattos, marcelos@eps.ufsc.br}

**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.

**Palavras-Chaves:** Inteligência Artificial, Sistemas Baseados em Conhecimento, Engenharia do Conhecimento, Mineração de Textos, Tecnologia da Informação Jurídica

**Submetido a:** IDEIA 2002

**Tópico:** Sistemas Baseados em Conhecimento

# AlphaThemis<sup>®</sup> - Do texto ao conhecimento

Hugo Cesar Hoeschl, Tânia Cristina D'Agostini Bueno, Andre Bortolon, Eduardo Mattos, Marcelo Stopanovski Ribeiro

Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Campus Universitário, Trindade, Florianópolis, Santa Catarina, Brasil.

CEP: 88040-900

{metajur, bueno, bortolon, emattos, marcelos@eps.ufsc.br}

**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®, 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][Tabela. 1]

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.”

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

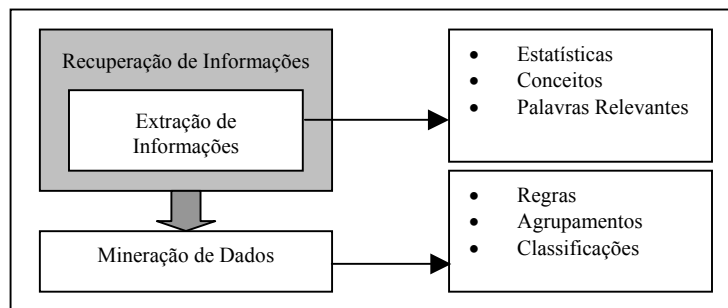
	Mineração de Dados	Mineração de Textos
Objeto de Investigação	Dados categóricos e numéricos	Textos
Estrutura de objeto	Bases de dados relacionais	Textos em formato livre
Objetivo	Prever resultados de situações futuras	Recuperar informações relevantes, purificar o significado, categorizar o resultado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

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]

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]



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

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 Hugo Cesar Hoeschl, Tânia Cristina D'Agostini Bueno, Andre Bortolon, Eduardo Mattos, Marcelo Stopanovski Ribeiro

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.

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

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

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 da 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 disto, 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 a 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 a Web.

As ferramentas de análise de texto, em conexão com o mecanismo de busca avançado ou com as ferramentas de acesso a 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 ensinar 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.[Fig. 3]

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.

Posição	Relevância	Tribunal	Texto	Data
1	100,00%	STJ	131 Nas ações de desapropriação incidem-se no cálculo de verba indenizatória os parcelos reais nos juros compensatórios e moratórios, devidamente corrigidos.	24/01/995
2	100,00%	STJ	54 Os juros moratórios fluem a partir do evento danoso, em caso de responsabilidade extracotratual.	18/08/992
3	100,00%	STJ	79 Os juros moratórios, na desapropriação direta ou indireta, contam-se desde o trânsito em julgado da sentença.	4/01/983
4	100,00%	STJ	102 A incidência dos juros moratórios sobre os compensatórios, nas ações expropriatórias, não constitui anatocismo vedado em lei.	26/5/994
5	83,30%	STJ	189 Os juros moratórios, na restrição ao inibido, são devidos a partir do trânsito em julgado da sentença.	23/6/997
6	83,30%	STJ	12 Em desapropriação, são cumuláveis juros compensatórios e moratórios.	5/11/990
7	79,20%	STJ	186 Nas indenizações por ato ilícito, os juros compostos somente são devidos por aquele que praticou o crime.	24/01/997
8	79,20%	STJ	176 É nula e cláusula contratual que supõe o devedor à taxa de juros divulgada pelo ANBICJETP.	5/11/996
9	79,20%	STJ	113 Os juros compensatórios, na desapropriação direta, incidem a partir de início na posse, calculados sobre o valor de indenização, corrigido monetariamente.	11/01/994
10	79,20%	STJ	114 Os juros compensatórios, na desapropriação indireta incidem a partir da ocupação, calculados sobre o valor de indenização, corrigido monetariamente.	11/01/994

**Figura. 2.** 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

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

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<sup>®</sup> 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<sup>®</sup> 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<sup>®</sup> 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<sup>®</sup> 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**

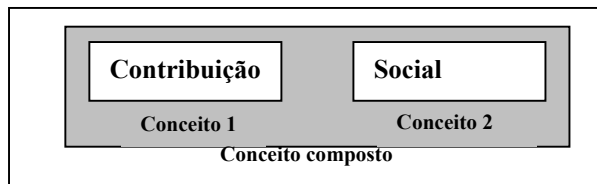
Ressalta-se que um dos pontos fortes dessa argumentação é a tarefa executada pelo programa que permite extrair conceitos subjetivos não diretamente explicitados no texto.

Conforme demonstrado pela figura 4, a junção de várias palavras podem 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

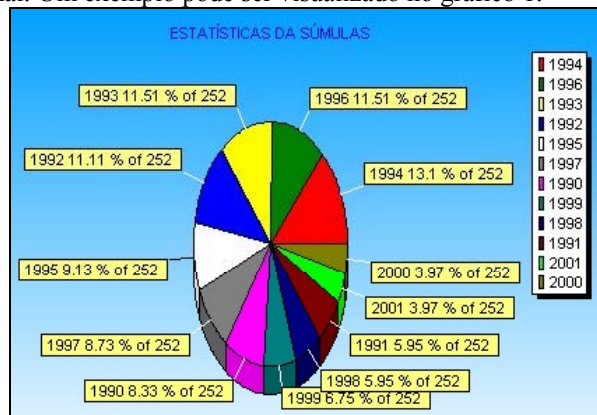
partir dos termos existentes na mesma. Verifica-se, então, em qual ramo do Direito estes conceitos se encaixam.

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.



**Figura 3.** Identificação de conceitos

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® 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 sua tecnologia flexível para a 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.

## Referências

- [1]ANANYAN, Sergei. **Text Mining Applications and technologies**. Disponível em: <[http://www.megaputer.com/down/tm/Text\\_Mining.pps](http://www.megaputer.com/down/tm/Text_Mining.pps)>. Acesso em: 30/01/2002.
- [2]BUENO, Tania Cristina D'Agostini. **Recuperação da Informação Jurídica em Sistema Baseado em Casos**. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis.
- [3]BUENO, Tania Cristina D'Agostini; HOESCHL, Hugo César. **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. EUA: 1998. CD-ROM.
- [5]NAHM, Un Yong; MOONEY, Raymond 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: <<http://www.semio.com/products/overview.html>>. Acesso em: 30/01/2002.
- [7]SEMIO CORPORATION. **Text Mining and the Knowledge Management Space**. Disponível em: <<http://www.dmreview.com/master.cfm?NavID=61&WhitePaperID=80>>. Acesso em: 30/01/2002.
- [8]TKACH, Daniel. **Text Mining Technology: Turning Information Into Knowledge. A White Paper from IBM**. In: IBM Business Intelligence Solutions CD. EUA: 1998. CD-ROM.
- [9]TEXT MINING RESEARCH GROUP AT THE UNIVERSITY OF WAIKATO. **Text Mining**. Disponível em: <<http://www.cs.waikato.ac.nz/~nzdl/textmining/>>. Acesso em: 30/01/2002.
- [10]Web Intelligence Systems S.A..**Documentos de registro da Tecnologia AlphaThemis**. INPI, 2001.
- [11]WOHL, Amy D. **Intelligent Text Mining Creates Business Intelligence**. In: IBM Business Intelligence Solutions CD. EUA: 1998. CD-ROM.