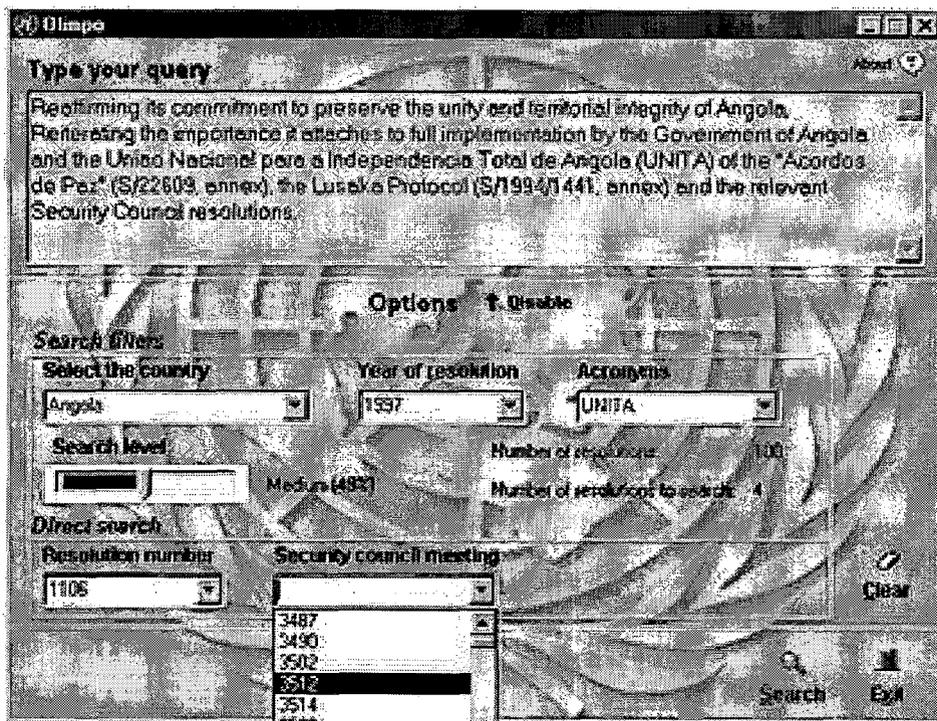


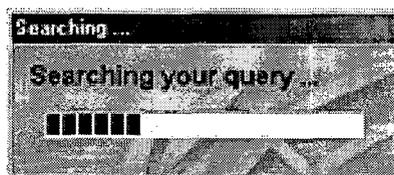
SISTEMA OLIMPO: (BU)



TECNOLOGIA DA INFORMAÇÃO JURÍDICA PARA O CONSELHO DE SEGURANÇA DA ONU



03447203



Universidade Federal de Santa Catarina
Programa de Pós-graduação em
Engenharia de Produção e Sistemas

**SISTEMA OLIMPO:
TECNOLOGIA DA INFORMAÇÃO JURÍDICA
PARA O CONSELHO DE SEGURANÇA DA ONU**

Hugo Cesar Hoeschl

Defesa de tese de doutoramento, apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina como requisito para obtenção do título de Doutor em Engenharia de Produção e Sistemas

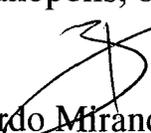
Florianópolis
2001

SISTEMA OLIMPO:
TECNOLOGIA DA INFORMAÇÃO JURÍDICA
PARA O CONSELHO DE SEGURANÇA DA ONU

Candidato: Húgo Cesar Hoeschl

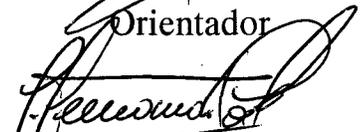
Esta tese foi julgada e aprovada para a concessão do título de doutor em engenharia de produção e sistemas pela Universidade Federal de Santa Catarina

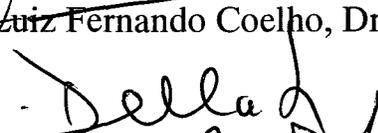
Florianópolis, 09/03/2001


Prof. Ricardo Miranda Barcia, PhD
Coordenador do Curso

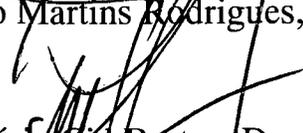
BANCA EXAMINADORA


Prof. Ricardo Miranda Barcia, PhD
Orientador


Prof. Luiz Fernando Coelho, Dr


Prof. Tarcisio Guido Della Senta, PhD


Prof. Alejandro Martins Rodrigues, Dr.


Prof. Rogério Cid Bastos, Dr.

Prof. Aires José Rover, Dr (moderador)

Para Tânia, Oriana, Milena e o próximo *serzinho* que está chegando

Agradecimentos

À Universidade Federal de Santa Catarina
Ao Prof. Ricardo Miranda Barcia, orientador
Aos Professores do PPGEP

Aos colegas pesquisadores da área de tecnologia da informação jurídica
Aos amigos do Ijuris – instituto jurídico de inteligência e sistemas

.....

.....

A todos que direta ou indiretamente contribuíram para a realização desta
pesquisa

“ A informática não tem mais nada a ver com computadores. Tem a ver com a vida das pessoas”

Nicholas Negroponte

SUMÁRIO

Lista de figuras, p. ix

Lista de tabelas, p. xi

Lista de gráficos, p. xii

Lista de reduções, p. xiii

Resumo, p. xiv,

Apresentação, p. 1

Capítulo I – Aspectos teóricos da tecnologia da informação jurídica, p. 6

Introdução, p. 6

1) Aspectos teóricos, p. 8

2) Elementos para o desenvolvimento de aplicações inteligentes no domínio jurídico, p. 11

A) IA X IN, p. 17

B) Figuras de raciocínio, p. 18

3) A técnica escolhida para desenvolver a presente aplicação, p. 23

4) Revisão bibliográfica, “estado da arte”, p. 32

Capítulo II - Aspectos empíricos da tecnologia da informação jurídica, p. 41

Introdução, p. 41

1) Descrição das experiências do grupo, p. 41

A) Digesto, p. 44

B) Metajuris, p. 48

C) Themis, p. 57

D) “P1”, p. 65

E) Prudentia, p. 73

F) Jurisconsulto, p. 74

2) Análise comparativa, p. 80

A) Olimpo X Jurisconsulto, p. 86

Capítulo III – Aplicação: Sistema Olimpo, p. 90

Introdução, p. 90

1) Identificação do universo, p. 90

2) Estrutura de um caso, p. 96

3) O Processo de formação e funcionamento do sistema, p. 99

A) Processo de análise da estrutura retórica, p. 102

B) Formação dos registros e cálculo da similaridade, p. 105

4) Estatísticas do sistema, p. 111

5) Funcionamento e performance do sistema, p. 118

A) Comparativo BD ONU X Olimpo, p. 119

B) Tarefas futuras, p. 124

Conclusões, p. 126

Bibliografia, p. 127

Anexos, p. 134

LISTA DE FIGURAS

- Figura 1: Sistema Olimpo, p. 04
- Figura 2: Livro eletrônico: tecnologia da informação jurídica, p. 07
- Figura 3: Tecnologia da informação jurídica, p. 10
- Figura 4: Raciocínio analógico, p. 19
- Figura 5: Resposta emitida pelo Sistema Olimpo, p. 29
- Figura 6: RBC, p. 28
- Figura 7: Digesto, site web, p. 45
- Figura 8: Exemplo de resultado de busca do Digesto, p. 47
- Figura 9: Complexidade do tratamento textual, p. 50
- Figura 10: Metajuris, p. 52
- Figura 11: Resultado emitido pelo Metajuris (primeira versão), p. 53
- Figura 12: Script do Metajuris, p. 53
- Figura 13: Html do Metajuris, p. 54
- Figura 14: Resultado emitido pelo Metajuris (versão em Perl), p. 56
- Figura 15: Sistema Themis (primeira interface), p. 58
- Figura 16: Sistema Themis (interface com as listas), p. 59
- Figura 17: Pesos dinâmicos no Sistema Themis, p. 62
- Figura 18: Modelo de resultado emitido pelo Sistema Themis, p. 64
- Figura 19: Resultado emitido pelo Sistema Themis (versão 99), p. 64
- Figura 20: Acesso às Súmulas dentro do Sistema Themis, p. 65
- Figura 21: Edição de casos no protótipo "P1" (seleção), p. 66
- Figura 22: Edição de casos no protótipo "P1", p. 67
- Figura 23: Caso 11766 no protótipo "P1" (primeira parte), p. 68
- Figura 24: Caso 11766 no protótipo "P1" (segunda parte), p. 68
- Figura 25: Pesos no protótipo "P1", p. 69
- Figura 26: Lista de regras no protótipo "P1", p. 70
- Figura 27: Exemplo de regra no protótipo "P1", p. 71
- Figura 28: Outro exemplo de regra no protótipo "P1", p. 71
- Figura 29: Caso de entrada no protótipo "P1", p. 72
- Figura 30: Resultado da consulta no protótipo "P1", p. 72
- Figura 31: Topografia lógica do Sistema Jurisconsulto, p. 75
- Figura 32: Tela de entrada do Sistema Jurisconsulto, p. 76
- Figura 33: Área de consulta do Sistema Jurisconsulto, p. 76
- Figura 34: Resultado emitido pelo Sistema Jurisconsulto, p. 77
- Figura 35: Acórdão na íntegra, p. 78
- Figura 36: Organograma ONU, p. 91

- Figura 37: Conselho de Segurança da ONU, p. 92
- Figura 38: Documentos da ONU, p. 93
- Figura 39: Centro de documentação web da ONU, p. 94
- Figura 40: Resolução 1.284, em html, p. 95
- Figura 41: Processo de consulta do Sistema Olimpo, p. 100
- Figura 42: Consulta com filtros, p. 101
- Figura 43: Consulta com nível de profundidade, p. 102
- Figura 44: Análise da estrutura retórica, p. 105
- Figura 45: Performance do Sistema Olimpo, p. 119
- Figura 46: Base de dados web da ONU, p. 121
- Figura 47: Interface de entrada do Sistema Olimpo, p. 122
- Figura 48: Interface de saída do Sistema Olimpo, p. 123
- Figura 49: Comparativo BD ONU X Sistema Olimpo, p. 124

LISTA DE TABELAS

- Tabela 1: Rendimento do Sistema Olimpo, p. 25
Tabela 2: 1º comparativo, p. 80
Tabela 3: 2º comparativo, p. 82
Tabela 4: 3º comparativo, p. 83
Tabela 5: 4º comparativo, p. 85
Tabela 6: Modelo de representação de uma Resolução do CS, p. 106
Tabela 7: Expressões contidas nos casos, p. 111
Tabela 8: Casos com mais expressões, p. 113
Tabela 9: Tabela de países/casos, p. 114
Tabela 10: Tabela de países, total, p. 115
Tabela 11: Tabela de siglas/casos, p. 116
Tabela 12: Tabela siglas, total, p. 117
Tabela 13: Comparativo BD ONU(web) X Sistema Olimpo, p. 120

LISTA DE GRÁFICOS

- Gráfico 1: Inteligência artificial, p. 15
- Gráfico 2: Avaliação dos tribunais, p. 44
- Gráfico 3: “P1” e Prudentia, p. 81
- Gráfico 4: Digesto e Metajuris, p. 81
- Gráfico 5: Themis e Jurisconsulto, p. 81
- Gráfico 6: Sistema Olimpo, p. 82
- Gráfico 7: Expressões/Resoluções, p. 113
- Gráfico 8: Países mais frequentes, p. 116
- Gráfico 9: Siglas mais frequentes, p. 117

LISTA DE REDUÇÕES

AIL: Associação internacional de “artificial intelligence and law”
BD: Banco de dados
CS: Conselho de segurança da ONU
IA: Inteligência artificial
ICAAIL: International conference on artificial intelligence and law
ICCBR: International conference on case-based reasoner
IN: Inteligência natural
ONU: Organização das Nações Unidas
P1: Primeiro protótipo
PCE: Pesquisa contextual estruturada
PPGEP: Programa de pós-graduação em engenharia de produção
RBC: Raciocínio baseado em casos
RD²C: Representação de conhecimento contextualizada dinamicamente
STF: Supremo Tribunal Federal
STJ: Superior Tribunal de Justiça
TJPB: Tribunal de Justiça da Paraíba
TRTSC: Tribunal Regional do Trabalho de Santa Catarina
TJDF: Tribunal de Justiça do Distrito Federal
TJSC: Tribunal de Justiça de Santa Catarina
TSE: Tribunal Superior Eleitoral
TJRJ: Tribunal de Justiça do Rio de Janeiro
TRF2: Tribunal Regional Federal da 2^a Região
TRF4: Tribunal Regional Federal da 4^a Região
UFSC: Universidade Federal de Santa Catarina
WEB: world wide web, interface gráfica da internet

RESUMO

Hoeschl, Hugo Cesar. Sistema Olimpo: tecnologia da informação jurídica para o conselho de segurança da ONU. Florianópolis, 2001. Tese de doutoramento (Doutorado em Engenharia de Produção e Sistemas), PPGEP, UFSC, 2001.

Resumo informativo: A tecnologia da informação jurídica baseia-se na representação do conhecimento jurídico para potencializar sistemas de recuperação de informações textuais. Com base na experiência adquirida na construção de sistemas anteriores (Digesto, Metajuris, Metalex, Themis, Prudentia e Jurisconsulta), alguns dos quais inteligentes, bem como na avaliação de serviços tecnológicos oferecidos por tribunais brasileiros, os pesquisadores do grupo de inteligência aplicada ao direito do PPGEP aceitaram o desafio de estruturar um novo sistema para aplicação nas resoluções do Conselho de Segurança da ONU. Esta empreitada teve como produto a construção do Sistema Olimpo, cujas características serão descritas no presente trabalho, assim como seu processo de consolidação, bem como seus mecanismos de funcionamento. Sua tarefa é recuperar, de forma inteligente, as resoluções mais semelhantes ao texto oferecido como entrada.

Palavras-chave: Tecnologia da informação jurídica, inteligência aplicada, sistemas baseados em conhecimento

SISTEMA OLIMPO

TECNOLOGIA DA INFORMAÇÃO JURÍDICA PARA O CONSELHO DE SEGURANÇA DA ONU

APRESENTAÇÃO

O Sistema Olimpo, assim como todos os trabalhos do grupo de inteligência aplicada ao direito do PPGEP/UFSC, é produto de uma feliz característica da Engenharia de Produção e Sistemas, qual seja, a multidisciplinariedade. Foi ela quem permitiu que profissionais e pesquisadores de diferentes áreas pudessem trabalhar em conjunto, planejando, modelando e construindo os sistemas que serão apresentados no curso deste trabalho.

Objetivos

Os objetivos desta tese são os seguintes:

Geral: apresentar a tecnologia da informação jurídica, a pesquisa contextual estruturada – PCE – e a representação do conhecimento contextualizada dinamicamente – RC²D – como instrumentos úteis à solução de problemas na busca computacional inteligente de informações textuais jurídicas.

Específicos:

- Apresentar elementos teóricos para a sedimentação epistêmica da tecnologia da informação jurídica, principalmente no que diz respeito ao desenvolvimento de aplicações inteligentes no domínio jurídico;
- Descrever as experiências realizadas pelo autor e pelo grupo Ijuris, no âmbito do desenvolvimento de ferramentas (algumas das quais inteligentes);
- Descrever, de forma mais detalhada, a aplicação desenvolvida pelo autor, especificamente para o presente trabalho, qual seja, o Sistema Olimpo, bem como suas técnicas particularizadas de representação do conhecimento.

Organização (estrutura)

A trabalho está dividido em três feixes centrais. No capítulo “I”, abordaremos aspectos teóricos da tecnologia da informação jurídica, com a análise de um panorama geral e posterior apresentação de elementos para o desenvolvimento de aplicações inteligentes no domínio jurídico, bem como um apanhado teórico sobre as técnicas escolhidas para o desenvolvimento da principal aplicação apresentada neste trabalho, além de uma revisão bibliográfica atinente ao objeto da pesquisa.

No capítulo “II”, descreveremos os procedimentos empíricos já desenvolvidos na área pelo autor e pelo grupo Ijuris, e respectivas características, com destaque para os sistemas “Themis”, P1”, “Prudentia” e “Jurisconsulto”, além do “Digesto” e do “Metajuris”.

No capítulo “III”, descreveremos o processo de constituição do Sistema Olimpo, e a forma de aplicação das técnicas sobre a base de resoluções, com a respectiva seleção dos casos, a formação das listas e a calibragem final do protótipo. Faremos comparativos estatísticos sobre o funcionamento do sistema e suas características,

bem como um comparativo de desempenho entre o Sistema Olimpo e os tradicionais sistemas de bancos de dados textuais.

Quanto ao modelo redacional, adotamos a linha seguida por Leite [64], no tocante à estruturação, divisão e numeração das respectivas partes do texto (“I”, “1”, “A”, “§”, “a”).

Metodologia de trabalho

O trabalho compreendeu aspectos empíricos e teóricos. Do ponto de vista teórico, a pesquisa documental e bibliográfica foi fundamental para o mapeamento do cenário e a prospecção de alternativas, seja no intuito de selecionar as fontes de orientações doutrinárias (tanto na área jurídica como na área tecnológica), seja no de buscar alternativas de implementação.

As atividades empíricas consistiram, basicamente, na captação de documentos e dados, no desenvolvimento de métodos de extração e representação de conhecimento, na capacitação para a operação com softwares e linguagens de programação, bem como treinamento para utilização de ferramentas para modelagem de bases de dados e desenvolvimento de aplicações inteligentes.

Justificativa

A importância deste tipo de trabalho reside na necessidade de se obter a melhor informação, da forma mais rápida e qualificada possível. Utilizar experiências anteriores para resolver situações atuais é uma tradição jurídica milenar e mundial, que depende, fundamentalmente, de se conhecer situações anteriores, ou de se poder localizá-las com agilidade.

Por outro lado, a modelagem computacional do raciocínio analógico tem sido potencializada, atualmente, pelos avanços registrados nas técnicas de inteligência artificial, bem como o constante aumento da capacidade de processamento dos computadores pessoais, o que facilita a utilização de tais aplicações em larga escala.

Utilizando tais referenciais, construímos o Sistema Olimpo, const ante da figura 1.

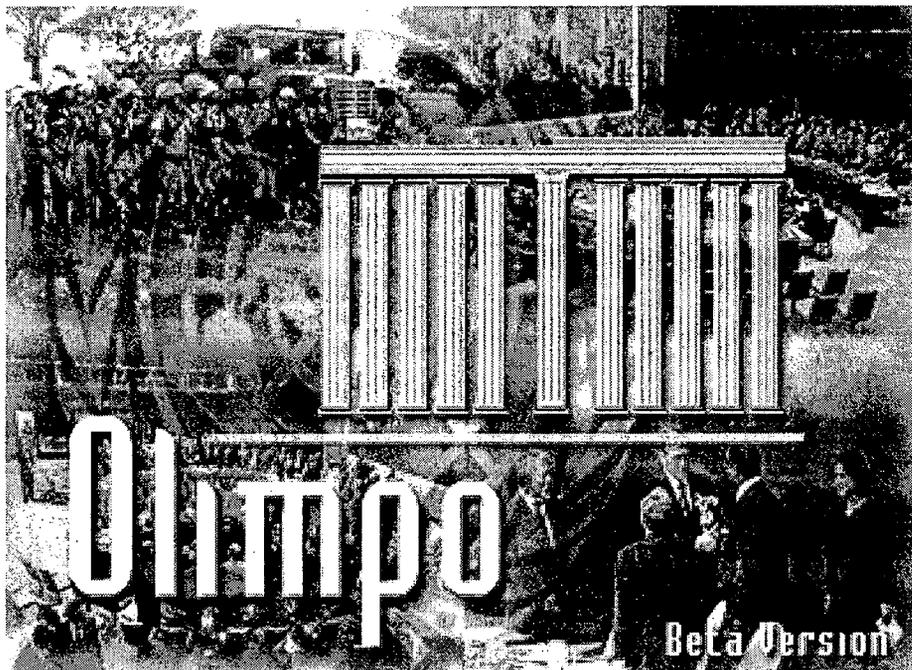


Figura 01

Problemas e Hipóteses

O principal problema aqui discutido é a geração de alternativas para o adequado desenvolvimento de ferramentas inteligentes no domínio jurídico, principalmente a efetiva construção de sistemas cuja tarefa seja a recuperação computacional de informações jurídicas textuais. Este tipo de informação geralmente está contida em

grandes textos, cuja leitura é demorada. Os bancos de dados tradicionais, calcados na busca por palavras, minimizam o problema, mas não o resolvem.

Sabe-se que sistemas inteligentes podem desempenhar a tarefa de seleção de documentos de forma rápida e eficaz. As hipóteses aqui apresentadas são a construção de tais sistemas de forma tal a que utilizem o *contexto* e a *representação do conhecimento* para adquirir incremento de desempenho.

Na hipótese primária, afirmamos a viabilidade da aplicação do raciocínio analógico nos dois campos (jurídico e tecnológico), materializada pela identidade conceitual existente entre ambos, bem como pela intersecção entre a inteligência artificial – IA – e a inteligência natural – IN.

Na primeira hipótese derivada, a pesquisa contextual estruturada, PCE, como produto de conjugação de características de outras técnicas, permite a realização prática desta aplicação, com a construção de sistemas que avançam no sentido da solução do problema identificado

Na segunda hipótese derivada, a representação do conhecimento contextualizado dinamicamente, RC²D, um dos componentes centrais da PCE, adequadamente combinada com outras técnicas, é apresentada como peça estratégica na produção de sistemas inteligentes que apresentam a tendência de serem mais eficientes do que os bancos de dados tradicionais em tarefas de recuperação textual.

CAPÍTULO I

ASPECTOS TEÓRICOS DA TECNOLOGIA DA INFORMAÇÃO JURÍDICA

INTRODUÇÃO

As técnicas que constituem a tecnologia da informação, principalmente a telemática e a internet, a inteligência artificial e a realidade virtual, oferecem a possibilidade de desenvolvimento de diversas ferramentas, processos e métodos que vão facilitar as tarefas diárias de formação e aplicação do direito, materializando interessante aplicação da Engenharia de Produção. A instalação de redes, a emissão de sinais, a comunicação à distância, o desenvolvimento de "*softwares*" específicos, a aplicação da telepresença, a manipulação de grandes massas de dados, entre outras atividades, estão entre as muitas a serem desenvolvidas. Dentro deste leque, situaremos a foco escolhido para o desenvolvimento do presente trabalho.

A adequada apresentação do fenômeno "tecnologia da informação jurídica" requer o desenvolvimento de dois aspectos básicos, *teóricos e empíricos*. O autor e a equipe de pesquisa da qual faz parte, possuem experimentos e formulações teóricas nesta área, que serão apresentados no decorrer do texto.

Porém, destaca-se neste contexto a edição do *ebook* – livro eletrônico -, intitulado "tecnologia da informação jurídica", uma coletânea dos trabalhos do grupo Ijuris e

dos alunos da disciplina com o mesmo nome (oferecida no PPGEp em 2000/2), lançado pela Rocket Library, ilustrado na figura 2.

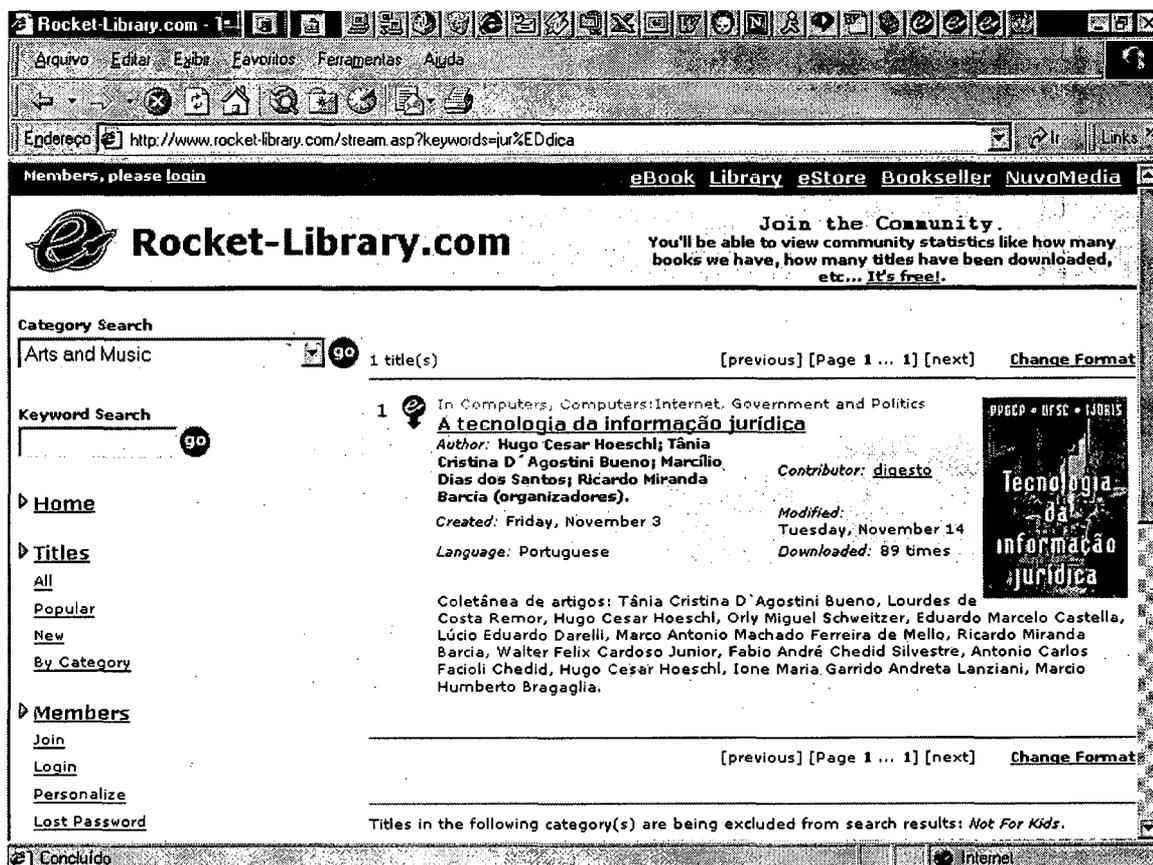


Figura 2

Na seqüência, passaremos ao exame mais detalhado dos aspectos teóricos da pesquisa.

1. ASPECTOS TEÓRICOS

A “tecnologia da informação jurídica” pode ser definida como a aplicação *contextualizada* dos recursos da tecnologia da informação ao universo jurídico. Por isso este nome foi escolhido, e não simplesmente “tecnologia da informação e direito”, o que sugere somente uma aproximação de institutos. Um importante conclusão de nosso grupo de pesquisa, aprovada perante “The Seventh International conference on ARTIFICIAL INTELLIGENCE AND LAW”, é que um dos maiores problemas na aplicação de recursos tecnológicos na área jurídica é a interpretação da linguagem e teoria jurídicas. Acrescentamos, agora, que os procedimentos, rotinas e métodos também são outro grande problema. Assim, a *contextualização* referida na conceito apresentado exerce um diferencial importante na formação do conceito da “tecnologia da informação jurídica”.

Assim os recursos oferecidos pela tecnologia da informação podem envolver, de forma exemplificativa:

1. a captação de documentos e formação de bases de dados;
2. O emprego de tecnologias de comunicação (redes, comunicação sem fio, videoconferência, *workgroup*, *wap*, *web*);
3. O desenvolvimento de aplicações inteligentes, que foi o caminho escolhido para este trabalho.

No que diz respeito ao item “3”, ficam mais destacadas as particularidades do domínio jurídico já referidas, principalmente tendo em vista a organização do

conhecimento necessária para a boa construção de sistemas inteligentes, e sua expressiva influência na performance dos mesmos (quanto melhor representado o conhecimento jurídico, mais satisfatório o desempenho do sistema).

A “artificial intelligence and law” é considerada, no plano internacional, como uma área da “information technology and the law”, conforme a associação internacional de “artificial intelligence and law” [82]. Fazendo as devidas adaptações, lingüísticas e culturais, a “inteligência aplicada ao direito” é, então, um dos ramos da “tecnologia da informação jurídica”.

O paradigma conceitual fixado pela AIL¹ - artificial intelligence and law – tem seu escopo orientado para a constituição de um fórum internacional para a disseminação de pesquisas interdisciplinares nas seguintes áreas [82]:

- Estudos teóricos ou empíricos em inteligência artificial, psicologia cognitiva, jurisprudência, lingüística ou filosofia, cujo objetivo é o desenvolvimento de modelos formais ou computacionais do conhecimento jurídico, raciocínio ou produção de decisões;
- Estudos aprofundados sobre sistemas inovadores em inteligência artificial que estão sendo usados no domínio jurídico;
- Estudos destinados a avaliar as implicações éticas e sociais no campo da inteligência artificial e direito.

¹ O autor é membro da AIL

Embora o autor e o grupo Ijuris estejam também desenvolvendo pesquisas na área do “impacto social da tecnologia da informação”², o presente trabalho fica limitado aos dois primeiros pontos, principalmente o segundo.

No âmbito das aplicações inteligentes, pode -se efetuar uma nova divisão: aplicações orientadas pelo aprendizado (redes conexionistas); 2. Sistemas estruturados logicamente; 3. Sistemas estruturados sobre casos, que também foi o caminho que escolhemos. Dentro desta opção, a Analogia, o Raciocínio Analógico, a Interpretação extensiva, o Silogismo e o Raciocínio Baseado em Casos são figuras com importância destacada, e serão objeto de aprofundamento específico.

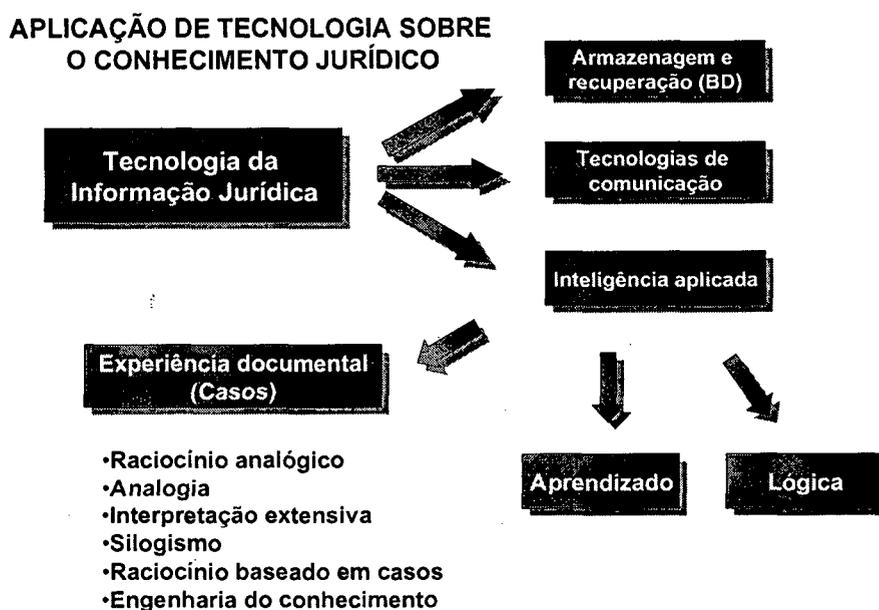


Figura 3

² Disciplina oferecida no PPGE, 2000/3

2) Elementos para o desenvolvimento de aplicações inteligentes no domínio jurídico³

Serão dedicadas algumas linhas aqui às possibilidades oferecidas pelo desenvolvimento de aplicações inteligentes, e algumas técnicas que permitem tal interação, comparando-as, quando possível, a figuras tradicionais do raciocínio jurídico, como, por exemplo, a analogia.

Veja-se, então, uma visão, introdutória e superficial, sobre como possa ser definida a inteligência artificial [74]:

"Inteligência artificial - artificial intelligence

O campo da ciência da computação que busca aperfeiçoar os computadores dotando-os de algumas características peculiares da inteligência humana, como a capacidade de entender a linguagem natural e simular o raciocínio em condições de incerteza.

Muitos pesquisadores da inteligência artificial admitem que a IA falhou em alcançar seus objetivos, e os problemas que impedem seu avanço são tão complexos que as soluções podem demorar décadas - ou até séculos. Ironicamente, as aplicações da Inteligência artificial que, antes, eram consideradas as mais difíceis (como programar um computador para jogar xadrez ao nível dos grandes

mestres) acabaram sendo produzidas com razoável facilidade, e as aplicações consideradas, a princípio, como mais tranquilas (como a tradução de Idiomas) têm-se mostrado extremamente complicadas.

Contudo, as tentativas de dotar os computadores de inteligência foram, sob certos aspectos, compensadoras: elas comprovaram a quantidade inacreditável de conhecimentos que os seres humanos utilizam em suas atividades cotidianas, como decodificar o significado de uma frase falada. Douglas Lenat, pesquisador de inteligência artificial que está tentando transportar para o computador uma boa parte de seus conhecimentos de vida, assinala que o computador não consegue decodificar plenamente nem trabalhar com frases como 'Sr. Almeida está em São Paulo' sem antes registrar uma infinidade de informações como 'Quando uma pessoa está numa cidade, seu pé esquerdo também está na cidade'. Se, algum dia, você já acordou preocupado com a possibilidade de que os computadores estivessem ficando mais Inteligentes que os seres humanos, este exemplo servirá para tranquilizá-lo".

Esta definição pode, evidentemente, ser alvo de diversas críticas, sob inúmeros aspectos, mas, em nosso entender, cumpre seu objetivo, que é o de iniciar o pensamento sobre a inteligência artificial em um dado contexto, aquele permeado por questões conceituais tidas como “questões abertas”, marcadas pela polissemia.

³ A parte central deste capítulo foi publicada como capítulo do livro “A tecnologia da informação jurídica”.

Importantes aspectos da IA, segundo Rabuske[76] são os seguintes, entre outros:

- desenvolvimento de métodos heurísticos para solução de problemas;
- representação de conhecimento;
- tratamento de linguagem natural;
- aquisição de conhecimento;
- raciocínio artificial;
- lógicas e ferramentas.

Dentre suas principais aplicações [76], temos as seguintes:

- Sistemas especialistas;
- Processamento de linguagem natural;
- Reconhecimento de padrões;
- Robótica;
- Bases de dados inteligentes;
- Prova de teoremas;
- Jogos

Existem diversas outras aplicações, muitas delas sendo amplamente estudadas atualmente no PPGEP-UFSC, que, embora não citadas pelo ilustre autor, devem figurar neste seleto e qualificado rol, como, por exemplo, algoritmos genéticos, redes neurais, conjuntos difusos, mineração de dados e reconhecimento de padrões.

Muita coisa pode ser e vem sendo discutida pela comunidade científica internacional sobre inteligência artificial, e muitos enfoques podem surgir para definir critérios

para o estabelecimento de um conceito de IA. Cotejá-los e explorá-los não é o objetivo do presente trabalho. Estamos entendendo a natureza epistêmica da IA como o conjunto acumulado de conhecimentos em várias áreas, dentre as quais destacam-se as seguintes:

- **Computação**, mais especificamente a engenharia de software, peça chave na estruturação dos mecanismos digitais de atuação da IA;
- **Filosofia**, com enfoque para a lógica formal e a epistemologia, fundamentais nas tarefas de representação do conhecimento humano;
- **Psicologia**, em especial a fisiologia cerebral, relevante no que diz respeito aos modelos de raciocínio e controle do conhecimento pelo cérebro.

Subsidiariamente, muitos outros ramos do conhecimento podem marcar presença neste contexto, como a matemática, a estatística, a física, a lingüística, a química, a medicina, o direito, e outros ainda, mas estes três são os centralizadores epistêmicos, representados no gráfico 1.

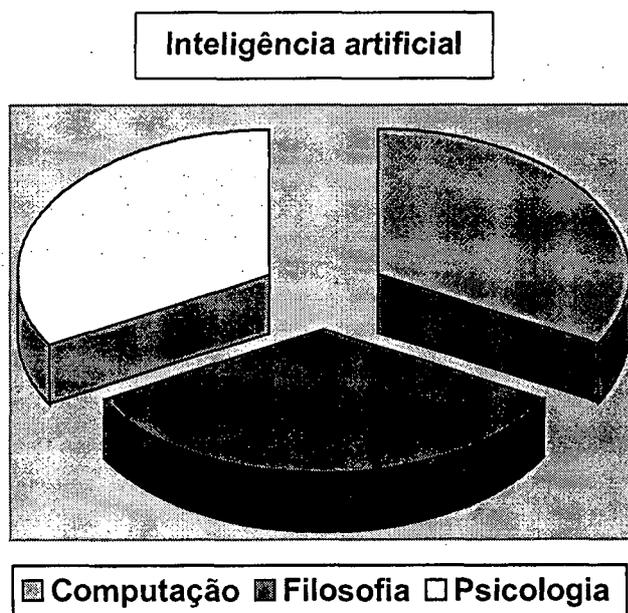


Gráfico 1

Porém, fazer uso de técnicas inteligentes e tentar desenvolver ferramentas computacionais - dotadas de lógica ou estruturadas em casos ou ainda construídas com técnicas de aprendizado -, para auxiliar na tarefa do estudo de dados jurídicos, envolve um trabalho dificultoso, qual seja, analisar a forma escolhida pelo homem para se comunicar e materializar suas normas: a codificação da palavra em símbolos abstratos e rigorosas regras gramaticais.

Tal sistemática é relativamente recente, levando -se em conta a existência humana, e, até a idade média, ainda estava limitada aos padres e eruditos. Eles entendiam a codificação, e a maioria das pessoas era analfabeta, conforme Wolkmer [98]. Nos dias de hoje, o número de analfabetos ainda é grande e ainda é relativamente restrito - embora não tanto quanto antes - o universo daqueles que realmente dominam a técnica da escrita.

Vale lembrar que "*a invenção e a difusão da técnica da escritura, somada à compilação de costumes tradicionais, proporcionam os primeiros códigos da Antigüidade, como o de Hamurábi, o de Manu, o de Sólon e a Lei das XII Tábuas*", no dizer de Wolkmer [98].

Naqueles tempos, no surgimento das primeiras codificações, estava em curso a maior mudança do direito ao longo de sua história, quando se passou a considerar a escrita um mecanismo superior à memória das pessoas para a armazenagem das normas, conforme Summer Maine (*apud* Wolkmer [98]), pelo simples fato de ser uma técnica mais segura. Isso modificou profundamente o direito e as formas de organização social, e os grandes sábios, líderes e tiranos deixaram de ser a fonte do direito, passando a ser intérpretes.

Ou seja, o surgimento de uma nova técnica de comunicação e registro de informações foi o responsável pela maior mudança até então registrada no universo jurídico, e não, ao contrário do que se possa imaginar inicialmente, a discussão de novos temas que foram surgindo com o passar dos anos.

Então, dada a posição atual da escrita nas formas de estruturação e armazenagem dos comandos do direito, o estudo e desenvolvimento de qualquer sistemática de tratamento automático e inteligente das informações jurídicas envolve, basicamente, duas tarefas: *1. o tratamento da linguagem natural; 2. a busca de novas técnicas de armazenagem.*

Na primeira, é necessária a estruturação de um mecanismo que faça uma leitura de textos e, devidamente orientado, identifique uma série de características relevantes para o utilizador, em algumas etapas específicas. Deve buscar referências superficiais e estáticas, como datas, nomes, números, etc. Deve identificar assuntos,

temas e subtemas. Deve, igualmente, detectar conclusões e lições, destacando -as. Além, é claro, de outras funções.

Na segunda tarefa, cabe indagar sobre o retorno às origens da linguagem. Explicando: as primeiras formas de escrita eram pictográficas, e, no âmbito computacional, o desenvolvimento de linguagens e interfaces está nos permitindo o uso de ícones (formas pictográficas) [31], um meio de comunicação mais confortável e prático do que a ortografia.

Isso está nos permitindo idealizar um avanço significativo na comunicação, segundo o qual "*textos escritos vão dar lugar a imagens mentais que apresentam tanto objetos reais quanto simbólicos e enfatizam a interação e a experiência em detrimento do aprendizado passivo.*" [31] (Destacado do original).

A) Inteligência artificial X inteligência natural

No sentido de se buscar, no plano prático, essa evolução anunciada, tem -se um poderoso referencial: a interseção entre a inteligência natural - IN - e a inteligência artificial - IA -, onde é possível tentar conciliar a velocidade de processamento da segunda e a sofisticação da primeira, como apontou Epstein [26].

A inteligência artificial, dentro do contexto ora delimitado - sem prejuízo da definição já apresentada - pode ser entendida também, em uma ótica ainda bastante primária, como "*o conjunto de técnicas utilizadas para tentar realizar autômatos adotando comportamentos semelhantes aos do pensamento humano*", como apontou Morvan (*apud* Epstein [26]).

Sabemos que a IN perde para a artificial na capacidade de busca e exame de opções, mas é superior em tarefas refinadas e perceptivas, como fazer analogias e criar metáforas.

Assim, um mecanismo que combine técnicas de IN e IA, buscando uma adequada manipulação da linguagem natural, permite a identificação de idéias dentro de um texto jurídico.

Porém, é importante enfatizar que um passo no sentido de se buscar, no corpo de um escrito, aquilo que uma pessoa "pensou", ou seja, suas idéias e conclusões, está teleologicamente ligado ao desejo de se buscar aquilo que uma pessoa realmente "sentiu" ao analisar o tema sobre o qual escreveu.

B) Figuras de raciocínio

Vale frisar que a inteligência artificial é uma figura típica da tecnologia da informação, praticamente moldada por ela. Para o delineamento da interseção apontada, vamos destinar breve atenção a algumas figuras ligadas à inteligência natural, como o raciocínio analógico, pré-existente aos computadores [74]:

“Raciocínio analógico – analogical reasoning

Uma forma de conhecimento na qual a dinâmica de um fenômeno do mundo real – como a aerodinâmica de um avião que se pretende construir – é compreendida a partir do estudo de um modelo do fenômeno. Uma das maiores contribuições da informática foi reduzir o custo (e aumentar a conveniência) do raciocínio analógico.

"O raciocínio analógico era comum antes do computador, conforme atesta o uso de maquetes de aviões em túneis de vento. Como reduzem muito o custo do raciocínio analógico, os computadores provocaram uma verdadeira explosão de descobertas analógicas – e, a propósito, no tempo certo. Os cientistas admitem, cada vez mais, que a maioria dos fenômenos do universo não se caracteriza pelas simples relações do tipo $f=ma$ que distinguem as grandes descobertas da física; pelo contrário, os sistemas complexos – como o sistema imunológico humano, as sociedades humanas, a ecologia, o clima do mundo e a interação das estruturas cosmológicas de grande escala – se caracterizam por um comportamento não-linear e caótico, que não pode ser descrito por equações simples. Esses sistemas não podem ser entendidos por outros meios que não o raciocínio analógico. Ao permitir que a humanidade crie modelos analógicos de abrangência sem precedentes, os computadores possibilitaram o surgimento de uma nova ciência: a ciência da complexidade".

A figura 4 apresenta uma boa descrição visual do raciocínio analógico.

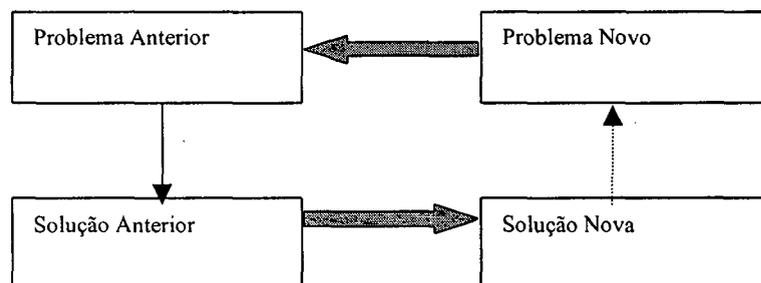


Figura 4

Tal instituto, anterior aos computadores - como já dito - foi adequadamente incorporado pela tecnologia da informação, assim como o raciocínio baseado em casos. É claro, sabemos, o raciocínio baseado em algum caso é algo quase tão velho quanto o hábito humano de "andar para a frente". Porém, aqui se trata de uma ferramenta da inteligência artificial que utiliza tal nomenclatura, podendo ser definida como uma "metodologia", que tem como característica básica buscar em experiências passadas a melhor solução para uma situação atual, aplicando o conhecimento já consolidado e cuja eficácia já foi validada.

Tais procedimentos, derivados da tecnologia da informação, possuem seme lhança evidente com uma tradicional figura do raciocínio jurídico, a *analogia*, um dos mais eficazes e pertinentes instrumentos de integração dos comandos do direito.

Segundo Bobbio, a analogia pode ser assim definida [5]:

"Entende-se por 'analogia' o procedimento pelo qual se atribui a um caso não-regulamentado a mesma disciplina que a um caso regulamentado semelhante. .

.....
"A analogia é certamente o mais típico e o mais importante dos procedimentos interpretativos de um determinado sistema, normativo: é o procedimento mediante o qual se explica a assim chamada tendência de cada ordenamento jurídico a expandir-se além dos casos expressamente regulamentados." (Destacado do original) .

A noção é de utilidade indubitável, e a delimitação da análise da semelhança, ponto de contato entre os casos, é necessária, conforme o mesmo autor [5]:

"Para que se possa tirar a conclusão, quer dizer, para fazer a atribuição ao caso não-regulamentado das mesmas conseqüências jurídicas atribuídas ao caso regulamentado semelhante, é preciso que entre os dois casos exista não uma semelhança qualquer, mas uma semelhança relevante, é preciso ascender dos dois casos a uma qualidade comum a ambos, que seja ao mesmo tempo a razão suficiente pela qual ao caso regulamentado foram atribuídas aquelas e não outras conseqüências." (Destacado do original).

Outras figuras assemelham-se ao contexto apresentado, como a *interpretação extensiva* e o *silogismo*, com as quais a analogia não pode ser confundida. O *silogismo* possui um mecanismo vertical de obtenção de conclusões, enquanto a *analogia* e a *interpretação extensiva* se valem de um recurso horizontal. Mas, mesmo que próximas e horizontalizadas, *analogia* e *interpretação extensiva* possuem significativa diferença entre si, apontada também pelo mesmo autor [5]:

"Mas qual é a diferença entre analogia propriamente dita e interpretação extensiva? Foram elaborados vários critérios para justificar a distinção. Creio que o único critério aceitável seja aquele que busca colher a diferença com respeito aos diversos efeitos, respectivamente, da extensão analógica e da interpretação extensiva: o efeito da primeira é a criação de uma nova norma jurídica; o efeito da

segunda é a extensão de uma norma para casos não previstos por esta".

Esta sutil diferença provoca um forte impacto sobre a atividade de construção e modelagem de sistemas inteligentes na área jurídica, visto que a proposta não é a construção de sistemas que *gerem normas*, mas que *facilitem a sua aplicação* (pelo menos por enquanto).

A comparação dos institutos nos demonstra a importância da análise dos processos lógicos estruturados em torno do raciocínio de uma área específica, e nos demonstra, também, que a lógica tem muita contribuição a oferecer à inteligência artificial – *mesmo em sistemas baseados em casos* –, residindo justamente aí um dos mais fortes aspectos favoráveis da interseção apontada entre IA e IN.

Este o comparativo tem a finalidade de demonstrar que a aproximação destes institutos tende a produzir bons resultados, bem como a viabilidade das interseções, tanto da IA com a IN, como das figuras de raciocínio derivadas da tecnologia da informação com aquelas particulares ao universo jurídico.

Por fim, é certa a necessidade de atenção à produção de ferramentas, enfatizando que tal atividade - que gerará novos métodos e técnicas de armazenamento e manipulação de informações - vai provocar fortes reflexos sobre o direito e a justiça, como a escrita o fez.

3) A técnica escolhida para desenvolver a presente aplicação

O enfoque está centrado no modelo teórico de raciocínio descrito no item anterior, referente a experiências passadas contidas em registros identificáveis, figura que encontra identidade teleológica principalmente na *analogia*, sob o aspecto jurídico, e no *raciocínio baseado em casos*, na ótica da inteligência aplicada.

Porém, no caso em questão, a tarefa do tipo de sistema que se pretende construir e aperfeiçoar, tem uma função um pouco diversa tanto da analogia quanto do RBC, eis que, em ambos os casos, o processo de aplicação da técnica termina geralmente com a apresentação de uma solução, que pode gerar uma nova norma ou um novo caso. O que se pretende com a presente pesquisa é a modelagem de sistemas que não realizem esta tarefa, mas somente recuperem a melhor informação disponível, em termos de precedentes, como forma de suporte à solução do problema atual. Assim, as técnicas já bastante usuais de recuperação de informação textual em bancos de dados passam a ser um componente relevante no contexto desejado, eis que o melhor funcionamento dos sistemas pretendidos é por elas influenciado.

Levando em consideração que o RBC não é uma “tecnologia”, mas sim uma “metodologia”, conforme Wattson [94], estamos tomando a liberdade de “pegar emprestado” uma parte desta metodologia, para mesclá-la com técnicas de recuperação de informação textual, com o intuito de gerar um tipo de sistema que possa combinar parte de ambas e apresentar um desempenho superior aos bancos de dados tradicionais, em matéria de pesquisa textual, e que também permita uma forma mais fácil de realizar consultas, calcada naquilo que se chama de “pesquisa contextual estruturada” – PCE. A PCE, como se verá adiante, compreende, além das tradicionais técnicas de extração e representação de atributos, um processo dinâmico

de análise do contexto geral que envolve o problema enfocado, que ora se define como “representação de conhecimento contextualizado dinamicamente” - RC²D, .

A PCE é *pesquisa* no sentido de configurar um sistema que pesquisa informações, ou as busca, ou as recupera. É *contextua* porquê o conhecimento por ela utilizado é representado contextualmente, e a pesquisa é realizada na base levando em consideração o contexto apresentado quando da solicitação da pesquisa. É *estruturada* em função de assim analisar tanto o contexto da solicitação quanto aquele dos documentos nos quais a pesquisa é realizada. Para que esta técnica funcione adequadamente, é fundamental a representação do conhecimento seja contextualizada, e que isto ocorra dinamicamente.

A RC²D consiste no processo de representação do conhecimento e respectivo ajuste, de forma dinâmica, a fim de que o funcionamento do sistema seja potencializado. Na construção do tipo de sistema nos quais temos trabalhado, a definição das expressões relevantes em termos de recuperação de conhecimento é um ponto muito importante. Normalmente temos construído um vocabulário controlado, com base em expressões consideradas relevantes pelos usuários do sistema. Geralmente, tratando-se de um sistema com perfil institucional, existem repositórios que contém listas com estas expressões. Porém, a simples inclusão destas listas na base de um sistema caracteriza uma limitação. Decidimos, então, *construir* uma base de conhecimento, estruturada sobre um conjunto de expressões, utilizando diferentes referenciais, de forma a tratar com o máximo possível de personalização estas expressões. Este processo de *construção* é chamado de RC²D.

Um exemplo de sua eficiência é constatado com a expressão “stability pact”. Segundo os métodos tradicionais, uma busca textual – estruturada com pesos e atributos – somente registraria a presença desta expressão se ela constasse desta

mesma forma. Este fato geralmente é verificável mediante uma regra de produção do tipo “se...” “então...”, onde “se ‘stability pact’ está presente no texto”, “então lance a pontuação ‘x’ na cálculo de similaridade”.

Ocorre que as derivações da expressão, “stability” e “pact”, somente seriam identificadas se fossem tratadas de forma específica e isolada. Porém, se tratarmos a expressão com um todo, decompondo -a e abrangendo seus termos derivados, o desempenho do sistema tende a adquirir incremento. A expressão exemplificativa, decomposta, passa a gerar três referenciais (“stability pact”, “stability” e “pact”). Nesse caso, três situações são possíveis: 1. Encontrar a expressão integral; 2. Encontrar as duas palavras, porém separadas; 3. Encontrar somente uma das palavras. Cada um dos itens representa 1/3, e somente na primeira situação teríamos uma recuperação com similaridade integral. No segundo caso, teríamos uma recuperação com um valor de 2/3, e, no terceiro, 1/3.

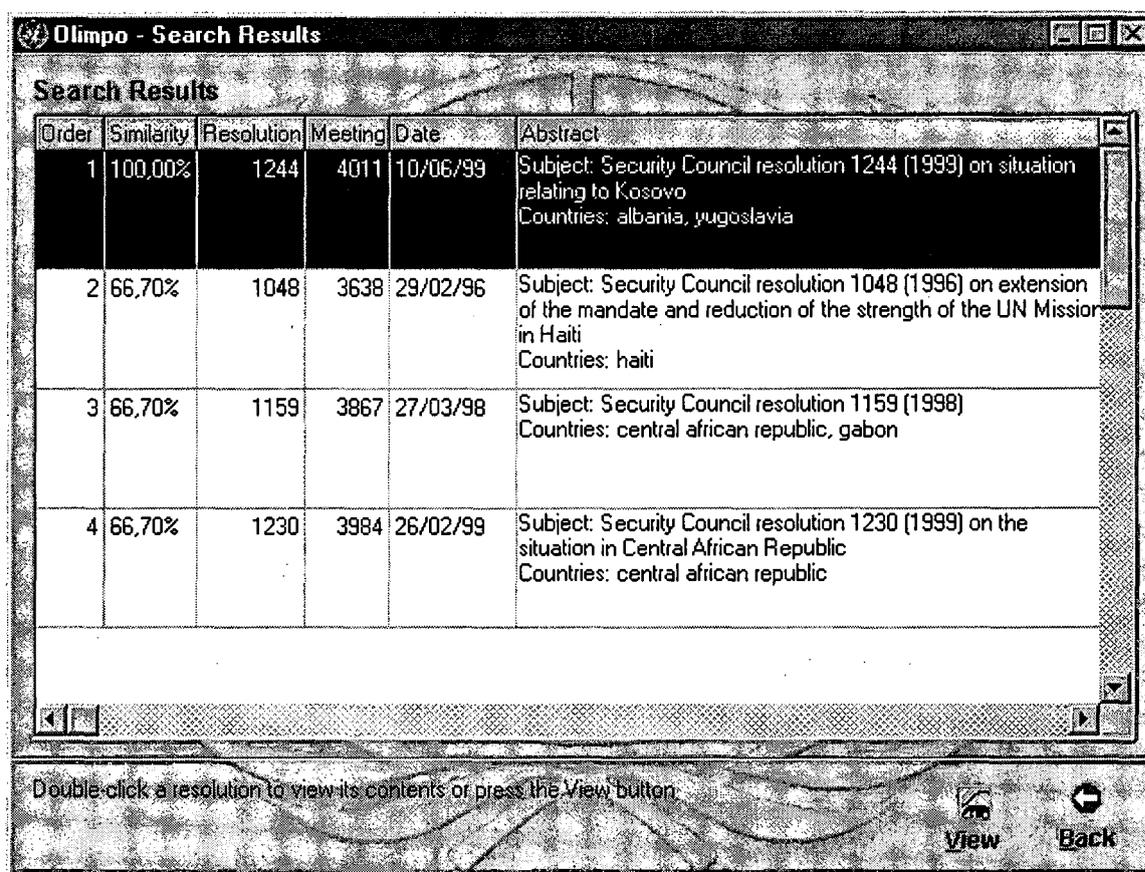
Um teste do sistema demonstra o adequado rendimento da técnica, nesta situação, com as expressões exemplificadas. Buscando -se pela expressão “stability pact” no Sistema Olimpo , o resultado da consulta é o seguinte, contido na tabela 1:

Classificação	Número da Resolução	Percentual de similaridade
1º.	1.244	100%
2º.	1.048	66,70%
3º.	1.159	66,70%
4º.	1.230	66,70%
5º.	1.265	66,70%
6º.	1.007	33,30%
7º.	1.014	33,30%

8°.	1.029	33,30%
9°.	1.054	33,30%
10°.	1.063	33,30%

Tabela 1

A figura 5 apresenta o resultado emitido pelo próprio sistema, de acordo com os critérios referenciados.



The screenshot shows a window titled "Olimpo - Search Results" with a table of search results. The table has columns for Order, Similarity, Resolution, Meeting, Date, and Abstract. Below the table, there is a footer with instructions and navigation buttons.

Order	Similarity	Resolution	Meeting	Date	Abstract
1	100,00%	1244	4011	10/06/99	Subject: Security Council resolution 1244 (1999) on situation relating to Kosovo Countries: albania, yugoslavia
2	66,70%	1048	3638	29/02/96	Subject: Security Council resolution 1048 (1996) on extension of the mandate and reduction of the strength of the UN Mission in Haiti Countries: haiti
3	66,70%	1159	3867	27/03/98	Subject: Security Council resolution 1159 (1998) Countries: central african republic, gabon
4	66,70%	1230	3984	26/02/99	Subject: Security Council resolution 1230 (1999) on the situation in Central African Republic Countries: central african republic

Double-click a resolution to view its contents or press the View button.

View Back

Figura 5

O processo de RC²D envolve uma etapa teórica e outra empírica:

- Etapa teórica: diz respeito à análise de documentos e textos referentes ao assunto enfocado, e posterior processo dialético envolvendo especialistas na área de recuperação documental e/ou no assunto abordado pelo sistema, a fim de identificar quais expressões são relevantes e caracterizadoras dos assuntos tratados nos documentos. No caso específico, foram analisados os principais documentos orgânicos da ONU (Carta, Regimento do CS, relatórios, listas oficiais – Países -, resoluções, etc).
- Etapa empírica: consiste na experimentação feita com as expressões extraídas do processo teórico sobre os documentos que farão parte do sistema, bem como análise numérica sobre a ocorrência das expressões . *In casu*, tanto a base de casos da ONU quanto os documentos selecionados para a construção do sistema foram objeto do processo empírico. Também foram levadas em consideração as estatísticas sobre incidência das expressões nos documentos, dado utilizado para inclusão/ampliação de índices, ou supressão de alguns deles, ou decomposição ou, ainda, agrupamento.

Este *cross over* realizado com os dados obtidos de ambas as etapas do processo é que confere especial dinâmica ao processo de RC²D.

Diante deste contexto, a PCE assume uma autonomia axiológica em relação ao RBC e à IR. Não se pode afirmar que esta técnica consiste em mera recuperação de informação (IR), eis que está qualificada por procedimentos inteligentes. Da mesma forma, não se pode afirmar que o trabalho ora sob análise está restrito aos círculos do RBC, eis que vai além dele. Decorrencia lógica é que não se pode enquadrá-lo como RBC textual. Seu perfil está mais próximo de um sistema inteligente de gerenciamento de bases de dados (IDBMS).

Algumas observações sobre a PCE e o RBC merecem melhor aprofundamento. O RBC é uma metodologia que possui etapas específicas, constantes da figura 6.

As etapas de “revisão” e “reutilização” não fazem parte do processo da PCE, da forma como preconizadas pelo RBC, pois as novas consultas, adequadamente respondidas, não se transformam em novos “casos”, o que só acontece quando um novo registro, emitido pela fonte oficial, é incorporado à base.

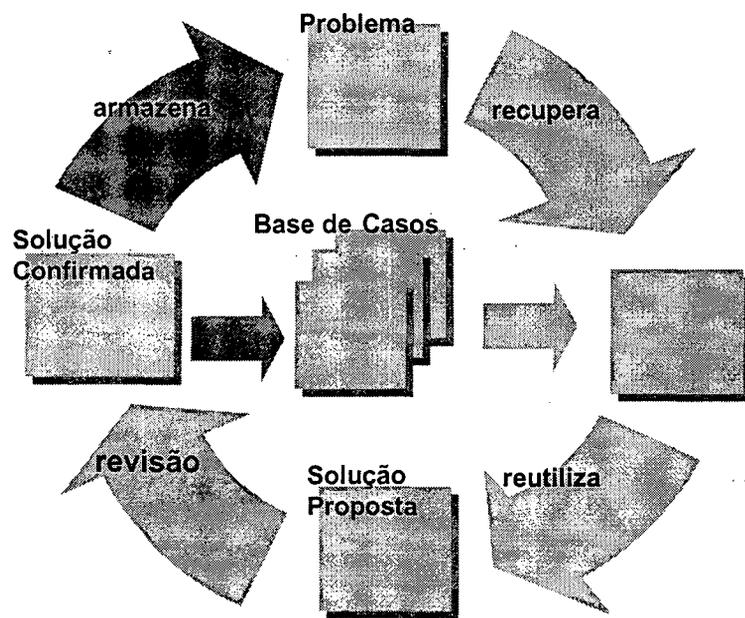


Figura 6

Um importante aspecto a ser evidenciado é que aquele trabalho inicial, oriundo da formulação dos sistemas “Themis” e do protótipo “P1”, e que posteriormente gerou o protótipo “Prudentia”, estava centrado exclusivamente nas técnicas de RBC (embora a proposta contida no Projeto Themis não estivesse restrita a isso). Posteriormente, houve um processo de desmembramento de linhas de atuação, o

qual ocorreu no plano epistêmico, principalmente por uma questão axiológica, qual seja, uma parte do grupo (a que permaneceu desenvolvendo as pesquisas aqui descritas) acreditava, e continua acreditando, que a intimidade com as questões afetas ao domínio jurídico, e o estudo de importantes aspectos de sua teoria argumentativa – como fez Bueno [11] em sua dissertação de mestrado -, era ponto chave para a adequada evolução do sistema, enquanto que a outra corrente, que deixou grupo, entendia de forma diferenciada. Este fato, no plano científico, fez as pesquisas até ali desenvolvidas tomarem rumos diferenciados, de forma tal a que a nossa linha de atuação não tenha ficado centrada nos problemas da técnica RB C, ou na técnica em si mesma, mas nas questões afetas à melhor forma de representar o conhecimento do domínio jurídico, tendo em vista que pequenas vírgulas modificam grandes coisas em termos jurídicos, e que “fundamentos” e “argumentos” não são, e nunca serão, a mesma coisa.

No trabalho aqui descrito, esta preocupação se reflete principalmente no trabalho de análise e “calibragem” da estrutura retórica utilizada para as comparações estruturadas, conforme se verá adiante.

Os pontos fortes do Sistema Olimpo, descrito no capítulo III, são a conjugação de técnicas e a metodologia de organização do conhecimento contido no sistema. Este detalhe faz dele também um trabalho típico da Engenharia de Produção, tendo em vista que esta é caracterizada como uma engenharia de “métodos e de procedimentos”, conforme Silva [83].

Ou, indo ainda mais longe, um trabalho típico da Engenharia do Conhecimento, assim referenciada por Rabuske [76]:

“Uma das principais áreas de pesquisa deverá ser a engenharia do conhecimento, aparecendo cursos universitários com orientação particular para este tema. As formas atuais de reunir conhecimento, para armazená-lo em computador, certamente não são os modelos do futuro. Por outro lado, há um imenso acervo de informação que a humanidade produziu durante milênios, que está armazenado em escritos. Por outro lado, há uma clara tendência de que o meio eletrônico seja uma forma importante de armazenar informações. Tanto uma quanto a outra forma de armazenar informações não deixam a comunidade científica tranqüila quanto ao acesso e aproveitamento pleno das mesmas. A linguagem natural é um seríssimo problema a ser superado”.

No caso concreto aqui teorizado e implementado, a tarefa está muito próxima disso, eis que nossa proposta é *reunir conhecimento, armazená-lo em meio digital que suporte processamento*, enfatizando que este conhecimento está contido em *textos formatados em linguagem natural*. Assim, independentemente da “sopa de letrinhas” técnica que envolve o presente estudo (PCE, RC²D, IA, RBC, IR, DBMS, etc), na realidade o que se está fazendo é a construção de uma base de dados inteligente. Uma adequada definição de base de dados inteligente pode ser a seguinte (bastante adequada as características apresentadas por este estudo), também oferecida por Rabuske [76]:

“A maioria dos sistemas de informação em uso envolve grandes bases de dados, sob o controle de um sistema de Administração de Bases de Dados (DBMS), do inglês

'DataBase Management System' . Contudo, se for associada ao DBMS uma base de conhecimento capaz de fazer raciocínios, produzindo resultados impossíveis de serem obtidos de outra forma, ter-se-á, então, uma base de dados inteligente. A principal razão do interesse em interações entre IA e base de dados é o aumento da produtividade e funcionamento dos sistemas de informação, o que requer o tratamento da informação como se fosse conhecimento, além de acréscimo de técnicas de IA para raciocinar e para resolver problemas ”.

Embora o autor tenha constatado o bom desempenho de sistemas estruturados com dados contidos em simples arquivos textuais (sem qualquer sistema de gerenciamento de registros em bancos de dados), em pelo menos duas situações (inclusive com forte e surpreendente incremento de performance), não resta dúvida que, independente deste fato, as técnicas aqui relatadas (principalmente PCE, RC²D e RBC) tem a função de melhor organizar a informação que está contida em registros textuais que precisam ser armazenados. E, conforme Feigenbaum e MacCorduck (*apud* Rabuske [76]), as tarefas de *coleccionar e seleccionar entre* constituem, sem dúvida, comportamento inteligente.

4) Revisão bibliográfica – “estado da arte”

A revisão de bibliografia concernente ao presente trabalho está centralizada nas atividades da comunidade científica internacional de inteligência artificial e direito, principalmente no que diz respeito ao desenvolvimento da aplicação.

Complementarmente, alguns trabalhos oriundos dos círculos do RBC integram a presente revisão, também no tocante ao desenvolvimento da aplicação.

Como a maioria dos sistemas aqui descritos, desenvolvidos pelo grupo e pelo autor, foram voltados ao direito brasileiro, fez-se necessário um suporte bibliográfico a ele referente.

Além disso, também foi necessário respaldo bibliográfico para fundamentar questões transitórias e de suporte aos pontos centrais enfocados.

Dentre estes aglomerados bibliográficos, a melhor atenção descritiva será dispendida aos artigos internacionais sobre inteligência artificial e direito, eis que tangenciam a aplicação em si, ponto mais relevante da pesquisa aqui descrita.

Assim, a revisão é centrada no principal foco de produção científica na área de inteligência artificial e direito, que é a conferência internacional de inteligência artificial e direito – ICAIL, cuja última edição, realizada em Oslo, Noruega, teve um trabalho do grupo Ijuris selecionado para apresentação e posterior publicação. O ICAIL 99 teve 32 trabalhos aprovados, dos quais 23 eram artigos e 9 eram resumos de pesquisa. Nosso trabalho foi aceito na qualidade de artigo, para apresentação presencial. Complementarmente, analisaremos a bibliografia procedente do “Journal

of Artificial Intelligence and Law”, publicado pela associação internacional de inteligência artificial e direito.

Neste contexto, destacam-se os seguintes trabalhos:

- *Toward adding knowledge to learning algorithms for indexing legal cases*, Bruninghaus [9]. Neste artigo, os autores ratificam importantes conceitos básicos no âmbito do desenvolvimento de sistemas inteligentes para o domínio jurídico - principalmente aqueles estruturados sobre casos – como, por exemplo o alto custo e a lentidão da representação manual de situações concretas. Através do Sistema Smile, construído pelos autores, é utilizada uma classificação baseada em abordagem, para localizar situações pontuais em textos jurídicos, de forma automática. Outra importante conclusão é a de que a integração, ao sistema, de um vocabulário controlado e de informações lingüísticas, com um algoritmo de aprendizagem de máquina, podem ser importantes aliados na tarefa de superar as tradicionais dificuldades de linguagem.
- *Bankruptcy case law: a hybrid IR-CBR approach*, onde Elhadi e Tibor [25] descrevem um trabalho de combinação entre recuperação da informação (*information retrieval, IR*) e RBC, com uma modelagem baseada nos procedimentos segundo os quais os operadores jurídicos efetuam suas pesquisas técnicas como parcela da dialética processual, sugerindo, especificamente, que as experiências acumuladas pelos profissionais do domínio jurídico sejam aproveitadas na modelagem do sistema.
- *Some observations on modeling case based reasoning with formal argument models*, onde o genial Bench-Capon [4] adverte a comunidade internacional

sobre a importância de se conciliar *casos* e *normas* na modelagem de um sistema para este tipo de aplicação, afirmando categoricamente que “a verdade sobre o assunto é que ambos são essenciais”.

- *The evaluation of legal knowledge based system*, texto no qual Stranieri e Zeleznikow [86] lançam importantes premissas sobre a avaliação de sistemas baseados no conhecimento jurídico, afirmando que avaliações estratégicas de sistemas que operam em um domínio tão complexo como o jurídico são difíceis em razão da especificidade dos sistemas, propondo um modelo de avaliação com redes neurais, referenciado pelos binômios “estrutural/funcional” e “qualitativo/quantitativo” para avaliar as respostas emitidas por um sistema.
- *Dialectic semantic for argumentation frameworks*. Este texto de Jakobovits e Vernier [56] propõe um estudo formalista de diálogos, onde um diálogo é visto com um embate entre duas pessoas, iniciado pelo proponente que defende a tese proposta. Realizando um exame dos critérios argumentativos e dos diferentes tipos de diálogos, os autores propõem prover uma correspondência semântica alusiva ao diálogo travado.
- *Precedent, deontic logic and inheritance*. Horty [55] propõe estabelecer algumas conexões entre raciocínio baseado em precedentes e as lógicas deôntica e monotônica. Sustenta que a lógica deôntica atua como raciocínio sensível prioritário em conflito de normas, reformulando e simplificando um modelo de raciocínio baseado em precedentes de Ashley, de acordo com a estrutura da lógica deôntica.

- *AI techniques for modeling legal negotiation.* Neste texto, Bellucci and Zeleznikow[3] afirmam que negociação é um processo de produção cooperativa de decisões entre as partes envolvidas em uma disputa. O objetivo da negociação é desenvolver uma solução que seja aceita por ambas as partes. Neste artigo, os autores centralizam o estudo na introdução do desenvolvimento de um algoritmo que usa métodos de IA para suportar a produção de uma decisão negocial, utilizando uma forma de mapeamento cognitivo denominada “*bidirected fuzzy cognitive maps*”.
- *The integration of retrieval, reasoning and drafting for refugee law: a third generation legal knowledge based system.* Segundo Yearwood e Stranieri [100], a identificação de argumentos é básica no funcionamento de um sistema de auxílio na construção de argumentos e esboço de deliberações legais. Os autores desenvolveram uma estrutura para construção de argumentos que inclui mais de 200 argumentos genéricos, em colaboração com a *Refugee Review Tribunal of Austrália*. O processo de construção destes argumentos requer a integração de recuperação de informação textual (IR) com raciocínio.
- *Justice: A judicial search tool using intelligent concept extraction.* Osborn e Sterling [72] apresentam a modelagem de um sistema baseado em conhecimento jurídico chamado “Justice”, o qual tem o objetivo recuperar casos anteriores. Ele pode identificar representações heterogêneas de conceitos, utilizando uma representação do conhecimento inspirada no senso comum, derivada do processo humano de identificação e posicionamento de conceitos. Utiliza conceitos de recuperação conceitual de informação, sumarização conceituada, análise estatística automatizada e conversão de documentos informais para representações semi-estruturadas formalizadas.

- *A demonstration of a legal reasoning system based on teleologic al analalogies*, texto no qual Kakuta e Haraguchi[57] apresentam um sistema de raciocínio jurídico baseado em uma abordagem teleológica, para interpretar normas jurídicas usando um exemplo atual. Utilizando uma estrutura denominada GDA (*Goal-Dependent Abstraction*), os autores analisam uma situação exemplar julgada pela Suprema Corte do Japão, diante de duas consistentes correntes doutrinárias, demonstrando como o sistema pode se tornar um importante auxílio na tarefa de avaliar e revisar interpretações de no rmas positivadas.
- *Agents and norms: how to fill the gap ?*, onde Conte, Falcone e Sartor [21] afirmam que duas abordagens estruturais específicas importam ao trabalho descrito: 1. A da teoria jurídica e aplicações computacionais relacionadas, especialmente nas áreas de sistemas especialistas jurídicos, raciocínio normativo e diagnóstico jurídico; e 2. A da teoria dos sistemas multiagentes (*multi-agents systems*, MAS) e aplicações computacionais relacionadas, especialmente nas áreas do trabalho cooperativo informatizado (*computer supported cooperative work*, CSCW). No cotejo de ambas as abordagens, os aspectos mais importantes são os seguintes: a linguagem e seu formalismo; teorias de referência (filosofia jurídica e filosofia deontica, cotejada com a teoria do s agentes); objetivos (modelos de instituições jurídicas, sistemas de informação jurídica); filosofia “retórica” e o conceito de norma. Os autores acreditam que uma abordagem sobre as normas e a capacidade operativa dos agentes pode gerar um resultado eficiente, o que requer a integração dos domínios jurídico e de multi -agentes. Para isso, os autores apresentam um resumo sobre as principais abordagens existentes em ambas as áreas, formulam algumas questões genéricas e ponderam

que suas soluções derivam da síntese de todas as abordagens, e apresentam a tentativa de resposta para algumas das questões formuladas.

- *Norms in artificial decision making*, texto no qual Boman [7] efetuou um estudo sobre agentes artificiais autônomos programados para representar e avaliar informações vagas ou imprecisas, com a caracterização de agentes com método de produção de decisão em tempo real, na presença de risco ou incerteza, terminando por apresentar um modelo de “*constraining action*” usando normas. Os procedimentos identificados auxiliam na difícil transição da grande quantidade de decisões produzidas para a análise qualitativa dos agentes artificiais.
- *Prescribed mental attitudes in goal-adoption and norm-adoption*. Neste artigo, com enfoque teórico, Castelfranchi [18] afirma que as normas não querem somente conformidade comportamental, mas também pretendem que esta conformidade produza aculturação e reafirme a autoridade das normas, tendo em vista que elas requerem submissão (sob a ótica de Castelfranchi).
- *Approximate syllogisms, on the logic of everyday life*, texto no qual Philips [75] afirma que, desde Aristóteles, é reconhecido que um silogismo válido não pode ter duas premissas específicas, preconizando que algumas regras podem ser estabelecidas para silogismos aproximados com premissas particulares, afirmando que os leigos estão corretos se for considerado que estes silogismos não tem validade estrita, mas aproximada, tendo em vista que as premissas disponíveis no cotidiano são tipicamente particulares.

- *Two examples of decision support in the law*, artigo no qual Borgulya [8] propõe dois sistemas empregando IA no direito para suporte de decisões. O primeiro, sobre indenizações na área de propriedade imaterial, é resultado da combinação de uma rede neuronal com um sistema especialista, no qual a rede neuronal fornece os primeiros patamares da indenização, enquanto o sistema especialista vai refinando as saídas da rede. O segundo estuda casos de acidentes de trânsito, utilizando métodos de lógica difusa para calibrar as decisões, atualmente fornecidas por juízes. Os sistemas estão conectados pelo fato de que ambos fornecem novas informações para decisões com base em precedentes similares.

Além destes, outros trabalhos, que envolvem aplicações em outras áreas, também são relevantes para o presente contexto. Não analisaremos o RBC, globalmente, tendo em vista que a pesquisa aqui relatada somente leva em consideração alguns aspectos desta metodologia. Porém, alguns trabalhos da área merecem especial atenção, eis que tratam de peculiares aspectos do RBC que importam à nossa pesquisa. São eles:

- *Using machine learning for assigning indices to textual cases*, onde Brunighaus e Ashley [10] reportam suas conclusões preliminares no trabalho de desenvolvimento de métodos automáticos para indexar casos descritos textualmente, de forma que um sistema de RBC possa raciocinar com eles. Este trabalho faz parte da construção do sistema CATO.
- *What you saw is what you want: using cases to seed information retrieval*, texto no qual Daniels e Rissland [22] propõem a construção de um sistema híbrido, composto de RBC e IR (recuperação de informação), chamado SPIRE.

- *Applying case-based reasoning to automated deduction*, artigo no qual Fuchs e Fuchs [28] afirmam que o uso do RBC tem sido bastante pro ficuo em muitas áreas da IA, mas que aplicar RBC para automatizar deduções é um problema muito complexo, e que as premissas tradicionais do RBC não são satisfatórias para automatizar deduções.
- *A scalable approach for questions based indexing of encycloped ic texts*, artigo no qual Wisdo [97] descreve uma ferramenta desenvolvida para auxiliar na análise de índices textuais para uso em uma particular forma de hipermídia chamada de ASK System. A ferramenta auxilia o analista a implementar um banco de modelos para representar tipos de questões.
- *Using case-based reasoning for argumentation with multiple viewpoints*, onde Karacapilidis, Trousse e Papadias [58] afirmam que a integração do RBC com outras técnicas de solução de problemas atrai interesse de pesquisa na área de sistemas de suporte à decisão (decision support systems – DSS). Os autores propõe uma estrutura onde RBC e raciocínio baseado na argumentação auxiliam agentes em processos de produção de decisões.
- *Cbr is not a technology, it's a methodology*, trabalho no qual o genial Wattson [94] questiona o RBC como uma tecnologia nivelada às redes neuronais, raciocínio baseado em regras ou algoritmos genéticos, afirmando que o RBC é, na realidade, uma metodologia para resolução de problemas, que pode ser usada com outra tecnologia apropriada. Wattson descreve quatro aplicações de RBC (*vizinho mais próximo, indução, lógica difusa e SQL*) para mostrar que o RBC é uma metodologia, e não uma tecnologia.

Independente da importância de que se reveste a tarefa de revisão da bibliografia referente às áreas descritas, é importante registrar que nenhuma tarefa superou, em termos de vibração e sentimento de descoberta, a análise dos documentos, o planejamento da aplicação e a projeção das variantes de implementação, sendo este, em nosso entendimento, o principal aspecto a ser registrado neste texto, em função da sua singularidade.

CAPÍTULO II

ASPECTOS EMPÍRICOS DA TECNOLOGIA DA INFORMAÇÃO JURÍDICA

INTRODUÇÃO

Neste capítulo, examinaremos os procedimentos empíricos adotados pelo autor e pelo grupo Ijuris, relatando as experiências realizadas em termos de tecnologia da informação jurídica, principalmente sobre a primeira e a terceira modalidades constates da figura 3.

Dentre estas experiências, destacam -se os sistemas de recuperação de informação textual e os inteligentes, bem como o trabalho de avaliação dos sites dos tribunais brasileiros, que serão brevemente descritos a seguir.

1) Descrição das experiências

O Grupo de inteligência aplicada ao direito vem desenvolvendo pesquisas e implementando protótipos, envolvendo tecnologia da informação jurídica. Entre os sistemas já desenvolvidos pelo grupo, destacam -se os seguintes:

- **Digesto**, uma ferramenta de busca na web, com representação de conhecimento e particionamento triplo de sua base de dados, com fundamento no tipo de documento indexado, características ainda exclusivas em termos de portais

jurídicos, e que conferem ao Digesto, atualmente, aproximadamente 1.000 visitas diárias;

- **Metajuris**, uma ferramenta de metabusca na web, que funciona acoplada ao Digesto, e faz pesquisa simultânea de jurisprudência nos principais tribunais brasileiros, e que, quando foi concebida, era absolutamente inédita na web brasileira;
- **Themis**, um sistema (protótipo) de armazenagem e recuperação inteligente das súmulas dos tribunais superiores brasileiros. Este sistema é o pioneiro, a idéia inicial, que gerou as demais pesquisas do grupo, embora somente tenha sido efetivamente implementado posteriormente;
- **“P 1”**, o primeiro protótipo realmente implementado pelo grupo, com vinte e cinco casos, desenvolvido como trabalho final da disciplina de raciocínio baseado em casos/96, cujas conclusões foram aprovadas no ICCBR/97, e posteriormente teve sua experiência e funcionamento relatada em monografia de curso de especialização de um dos integrantes do grupo de pesquisa;
- **Prudentia**, a evolução do “P 1”, protótipo implementado e cujas conclusões foram aprovadas no GWCBR/98, e, posteriormente, gerou uma tese de doutoramento da monitora da cadeira de RBC quando as pesquisas tiveram início, então integrante do grupo de pesquisa;
- **Jurisconsulta**, novo sistema prototipado em função das novas descobertas do grupo, em 98 e 99, principalmente a nova métrica de similaridade e os primeiros estudos com o sistema de entrada textual na consulta. Suas conclusões e

procedimentos foram aprovados perante o ICAIL/99, e, posteriormente, foram apresentadas em dissertação de mestrado de um dos integrantes do grupo, bem como fazem parte do objeto de outra dissertação em andamento no PPGEF, devendo ser apresentada em breve;

- **Olimpo**, novo sistema prototipado pelo grupo, com fundamento nas novas evoluções e descobertas, relatadas ao longo do presente trabalho, aplicadas às resoluções do conselho de Segurança da ONU

Além deles, o grupo idealizou e implementou outros sistemas e atividades na mesma área:

- **Sectra**, um sistema especialista sobre crimes contra a organização do trabalho, construído por um integrante do grupo, com auxílio de dois pesquisadores externos;
- **Habeas Corpus na Madrugada e Júri Digital**, dois jogos/simuladores, sobre procedimentos jurídicos, disponibilizados na web;
- **Avaliação de sites dos tribunais**, um método desenvolvido e aplicado pelo grupo para avaliar as páginas dos principais tribunais brasileiros, cuja divulgação foi referenciada por importantes veículos da imprensa nacional, como a Gazeta Mercantil, o Estado de São Paulo e o Jornal de Ciência do CNPQ. A experiência, com respectivas metodologia e resultados finais (gráfico 2), está disponível na web.

Resultado final da avaliação de sites

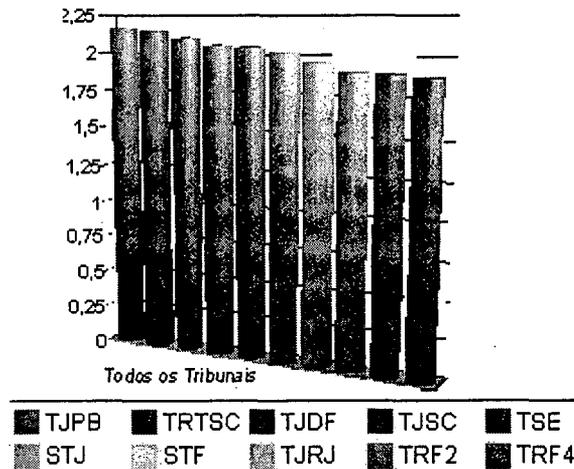


Gráfico 2

Dentre estas experiências, sete foram escolhidas para descrições mais detalhadas:

A) Digesto

O Digesto [44] (figura 7), é a ferramenta mais básica desenvolvida pelo autor, com a colaboração de alguns integrantes do grupo de pesquisa em IA e direito do PPGEP - UFSC. Trata-se de uma ferramenta de busca de arquivos html na internet -www, que evoluiu, agregando conteúdo, para se transformar em um site de tecnologia e conteúdo. Sua tecnologia está nos robôs de busca (o próprio Digesto, o Metajuris e o Metalex), e seu conteúdo está em textos html e livros eletrônicos disponibilizados pelo site, na área de direito e tecnologia da informação (“Direito Digital”), além de informações para o público em geral (“Direito Fácil”) e dos JurisJogos “Habeas Corpus na Madrugada” e “Juri Digital”.

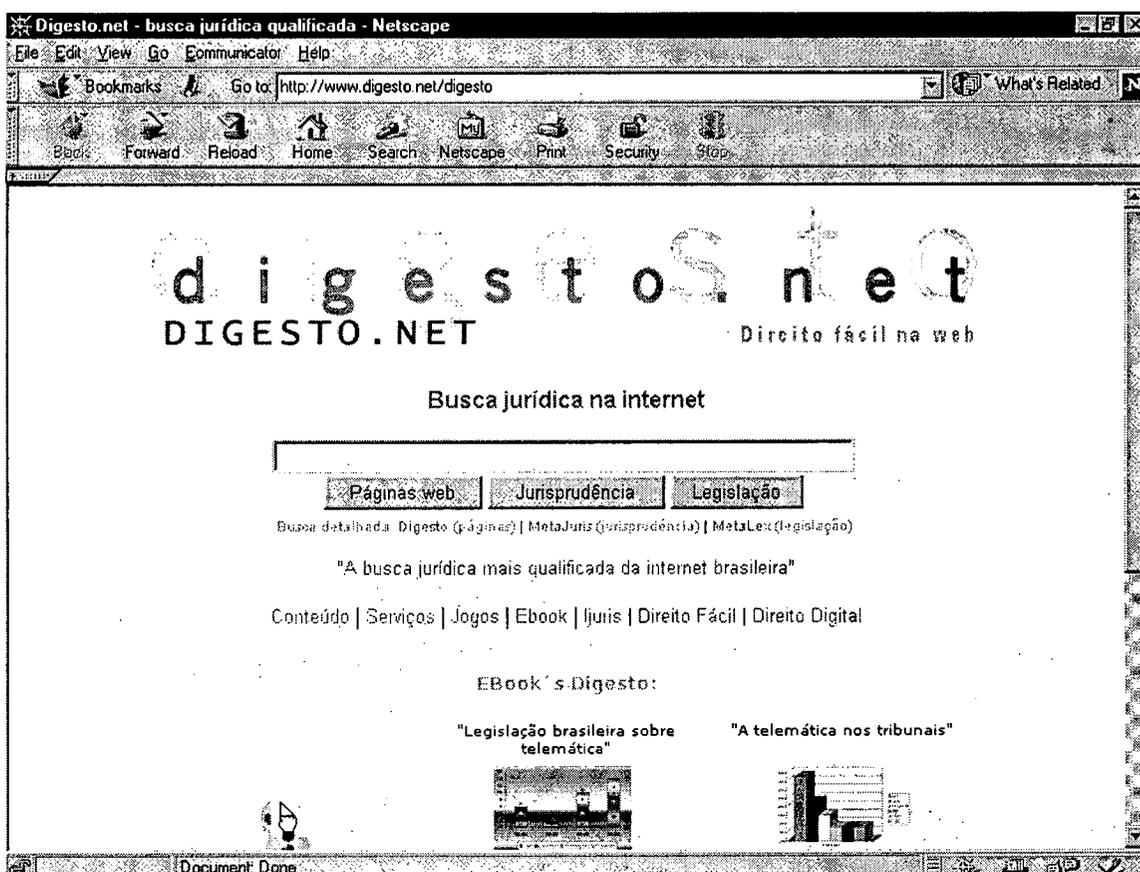


Figura 7

Porém, o que nos interessa no momento não é exatamente o perfil do portal, mas sim a ferramenta de busca central, o Digesto propriamente dito.

Dotada de particularidades que a diferenciam qualificadamente das demais ferramentas de busca (em geral) e sites jurídicos (em especial), a base de dados do Digesto apresenta uma organização que qualifica seu funcionamento, pelas seguintes razões:

- 1- *Indexação artesanal dos registros*. Cada um dos registros recuperados em uma busca no Digesto passou por um processo personalizado de avaliação, com relação ao assunto, temas, subtemas e tipo de documento;

- 2- *Base tripartida*. Sua base de dados tem um formato tripartido, de acordo com a natureza do documento indexado (“páginas”, “artigos” e “acórdãos”), o que potencializa as tarefas de pesquisa e armazenagem. Isso significa que o usuário terá sempre em suas mãos um resultado de pesquisa pontual, sem o tradicional ruído que as ferramentas de busca nos apresentam comumente;
- 3- *Elementos adicionais*. Além daquelas informações tradicionalmente coletadas pelas ferramentas de busca, o Digesto apresenta informações adicionais referentes ao assunto, temas e subtemas pesquisados, bem como os dispositivos legais referidos ao tema (números das leis e de seus artigos, bem como da Constituição Federal), o que qualifica os registros e pode gerar uma recuperação inesperada de alta qualidade;
- 4- *Busca e metabusca temáticas agregadas*. O funcionamento simultâneo dos robôs de busca (Digesto) e metabusca (Metajuris e Metalex) confere ao site uma característica extremamente peculiar, agregando os efeitos positivos da pesquisa tematizada, a qual possui tendência de maior eficiência nas respostas.

Um exemplo de resultado de busca está na figura 8.

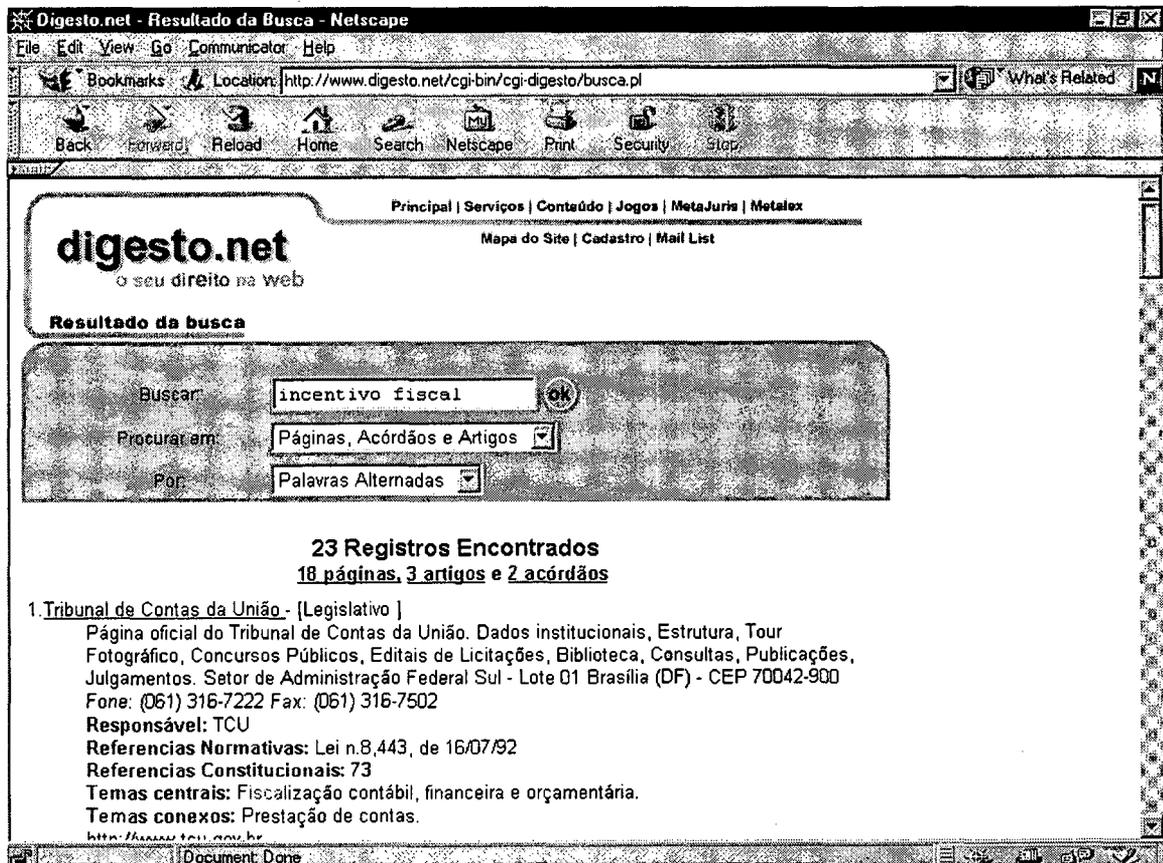


Figura 8

Desta forma, o site “www.digesto.net” vem tendo um crescente incremento no número de acessos. Cabe frisar que o acesso ao site - tanto aos serviços tecnológicos, como ao conteúdo - é absolutamente gratuito.

As experiências adquiridas no desenvolvimento do Digesto, no tocante à busca textual e à organização e particionamento de sua base de dados, foram vitais para o desenvolvimento do Sistema Olimpo, pois comprovaram que uma base bem organizada, com registros bem triados, pode, simplesmente por estes motivos, superar o desempenho de potentes ferramentas tecnológicas.

B) Metajuris

Os recursos agregados em torno das técnicas e soluções oferecidas pelo hipertexto estão ensejando soluções para mineração de dados em bases disponibilizadas na internet-*www*. Mecanismos de pesquisa e busca estão se propondo a realizar a tarefa de obtenção da melhor informação para seus usuários e para a comunidade de internautas em geral. No que tange ao universo jurídico, existem diversos tipos de dados e informações disponíveis nas bases web.

Quanto aos tribunais, aproximadamente uma centena deles já disponibiliza suas decisões para consulta “on line”, mediante mecanismos de bancos de dados, baseados em busca de palavras e expressões específicas. Este número aumenta diariamente. Tendo em vista que existem semelhanças de padrões de pesquisa em alguns deles, é possível a construção de mecanismos que realizem consultas simultâneas, poupando o trabalho de entrar e sair repetidamente de cada uma das páginas dos tribunais.

O objetivo do mecanismo proposto neste trabalho é minerar bases remotas de jurisprudência, disponíveis na internet-*www*, visando à obtenção de expressões jurídicas, contidas nas estruturas textuais literais de decisões judiciais, que expressem lições peremptórias específicas, as quais podem servir de fundamentos e argumentos para solução de novas situações sob discussão.

Para a execução da tarefa proposta, três aspectos precisam ser especificados: 1. A noção de *data mining*, voltada para conhecimento textual; 2. A análise dos níveis de complexidade da pesquisa em estruturas textuais; 3. A forma e os recursos técnicos adotados para a implementação das soluções.

1. Existem diversas definições de *data mining* aceitas pela comunidade científica internacional, desde *obtenção de dados e informações até extração direcionada de conhecimento profundo*. De um modo geral, pode-se afirmar que a noção é relativa, isto é, depende diretamente das noções de *informação* e *conhecimento*, estritamente vinculadas à pessoa e ao contexto interessados. Aquilo que propicia conhecimento para uma determinada pessoa, em razão de sua atividade e formação pessoal, pode nada ou pouco significar para as demais.

Pode-se afirmar, então, que existem dois cenários possíveis para a definição. Um, primário, genérico e mais simplificado, *data mining lato sensu*, uma definição *light*. Trata-se de obtenção de informações “um passo adiante” dos tradicionais mecanismos de bancos de dados. Outro, complexo e estrito, no qual se está aludindo à extração de conhecimento específico e determinado, com a obtenção de lições pontuais e conclusões definidas, diante da análise de uma massa de dados. Uma noção mais pesada. No momento atual, a proposta deste trabalho (Metajuris) situa-se no âmbito da primeira noção, primária e genérica, de *data mining*, isto é, apresenta, por enquanto, apenas uma superação das atuais tarefas realizadas pelos mecanismos de bancos de dados remotos disponíveis em sua área de atuação (textual).

2. Procurar informações em textos modelados em linguagem natural não é uma tarefa fácil. Acreditava-se, antes, que a inteligência artificial possuía objetivos *díficeis* e *fáceis*. *Difícil* era programar um computador para jogar xadrez ao nível dos grandes mestres internacionais. *Fácil* era traduzir idiomas e interpretar a linguagem. Hoje verifica-se uma inversão desta avaliação, constatando-se que o *fácil* era muito mais complexo do que se imaginava. Porém, esta constatação é de grande relevância, na medida em que explicita a “*quantidade inacreditável de conhecimentos que os seres humanos utilizam em suas atividades cotidianas, como decodificar o significado de uma frase*” [Pfaffemberger, 74].

De fato, ao se tentar programar qualquer dispositivo para realizar a leitura e interpretação de textos, surge uma gama de dificuldades, a saber: polissemia, vagueza da linguagem, dubiedades, metáforas e figuras de linguagem, manifestações de conhecimento inconsciente, entrelinhas, mensagens cifradas, lições indiretas, identificação de contextos, etc. Estes problemas ocorrem quando os textos são preenchidos com linguagem livre, sem controle de vocabulário.

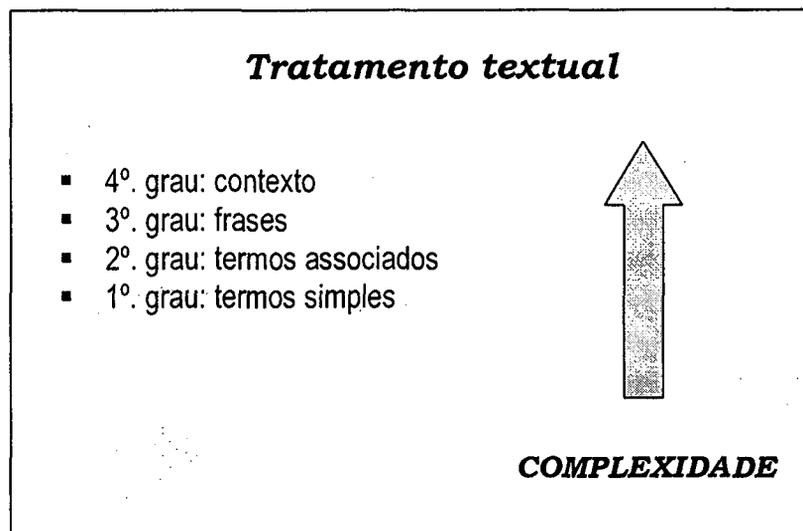


Figura 9

É possível graduá-los. Iniciam-se na palavra, a unidade de um texto, e terminam (por ora) na mensagem global, ou contexto geral. São visíveis quatro graus de complexidade: 1º e mais simples, o tratamento de termos isolados; 2º, o tratamento de termos associados; 3º a análise de frases; 4º e mais complexo, a análise do contexto. O vetor de complexidade é o mostrado na figura 9.

A pesquisa por termos simples é um problema resolvido pelas técnicas computacionais básicas. Os métodos de busca por termos associados estão disponíveis na maioria dos mecanismos de bancos de dados no mercado mundial,

com variações específicas, e a internet é o melhor exemplo disto. Deste ponto em diante, a problema assume dimensões de grande porte. Uma frase não é somente uma lista de palavras, pois tem, geralmente, estrutura e finalidade, com componentes sintáticos específicos, e, principalmente, uma mensagem. A IBM [Lent & Agrawal, 65] propôs uma aplicação, intitulada PatentMiner, cujo objetivo é descobrir tendências em bancos de dados textuais com base em análise de frases, associando - se a elas o histórico da frequência de suas ocorrências, verificando se est a aumenta ou diminui durante um certo período.

Com relação á complexidade semântica, a solução aqui proposta está situada nos dois primeiros níveis, quais sejam, o tratamento das expressões, isoladas ou somadas. *Metajuris* utiliza o reduzido grau de dificuldade linguística contido na pesquisa por termos simples como forma de poder realizá -la simultaneamente em bases remotas diversas, gerenciadas por mecanismos diferentes. O objetivo é rumar no sentido progressivo nas duas escalas, visando, por um lado, ati ngir o mais elevado grau, a pesquisa contextual, no âmbito da complexidade de pesquisa textual, e, por outro, no âmbito da definição de mineração de dados textuais, alcançar os efeitos do sentido forte que a expressão enseja.

3. O objetivo do *Metajuris* é fundamentar a construção de um sistema minerador que, mediante uma única postagem de um formulário de dados, realize a pesquisa, segundo os mesmos critérios, em bases remotas distintas, retornando, individualizadamente, os resultados de cada uma delas, não i mportando a irregularidade dos respectivos tempos de resposta.

Utilizando-se recursos de HTML, Javascript e PERL, foi construído o protótipo de pesquisa, o qual agrupou bases de acordo com os temas a serem pesquisados, da seguinte forma: Tribunais Superiores, Tribunais de Justiça dos Estados, Tribunais

Regionais do Trabalho, Tribunais Regionais Eleitorais, Tribunais de Contas, informações Governamentais e Bases Jurídicas de Língua Inglesa.

Por ora, será examinado, de forma mais detida, o módulo central, referente aos Tribunais Superiores. Sua primeira interface (HTML) é a constante da figura 10.

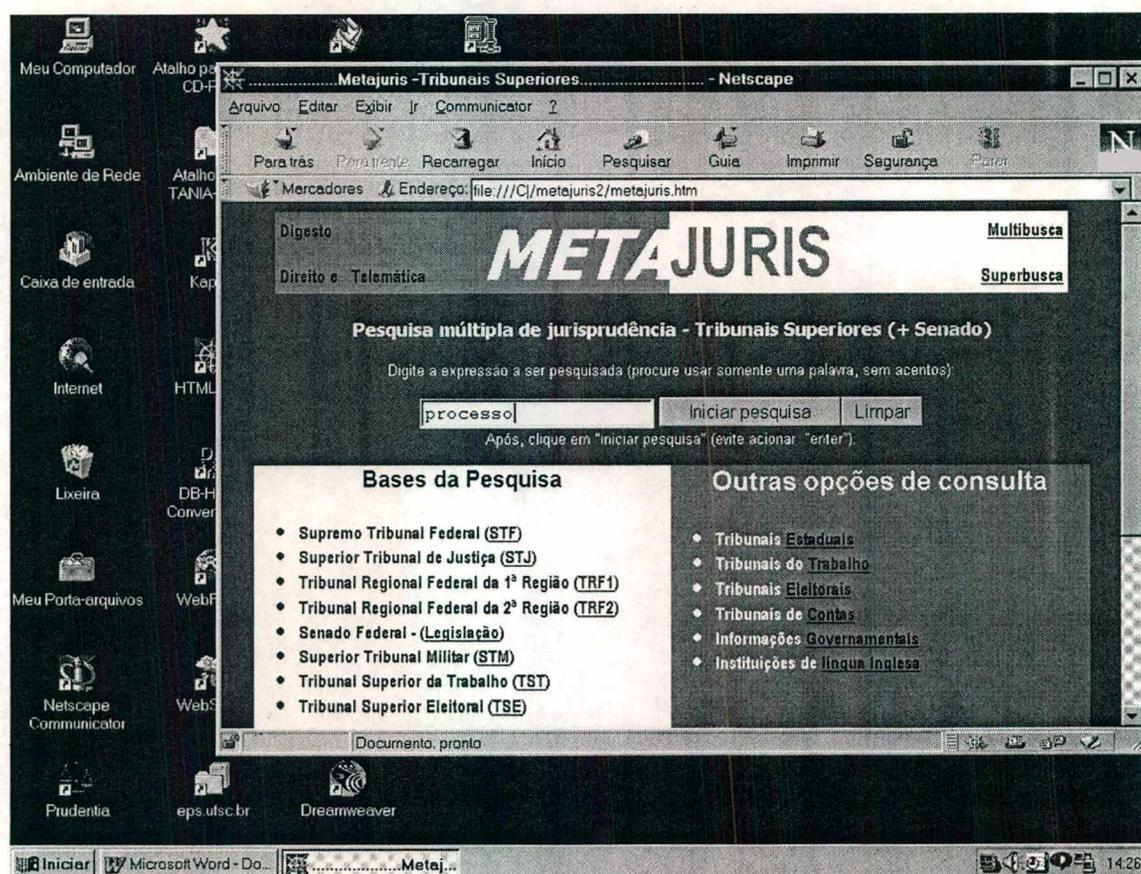


Figura 10

Como demonstra a figura 10, a interface é descomplicada. Contém uma caixa de diálogo, onde é digitada a expressão que se deseja ver pesquisada. Após, os botões de acionamento para o início da pesquisa ou limpeza do campo, em caso de erro de digitação. Uma vez digitada uma expressão e acionado o botão de pesquisa, inicia -se

o processo de busca, com o retorno, em poucos segundos, das informações desejadas, apresentadas de acordo com a disposição demonstrada na figura 11:

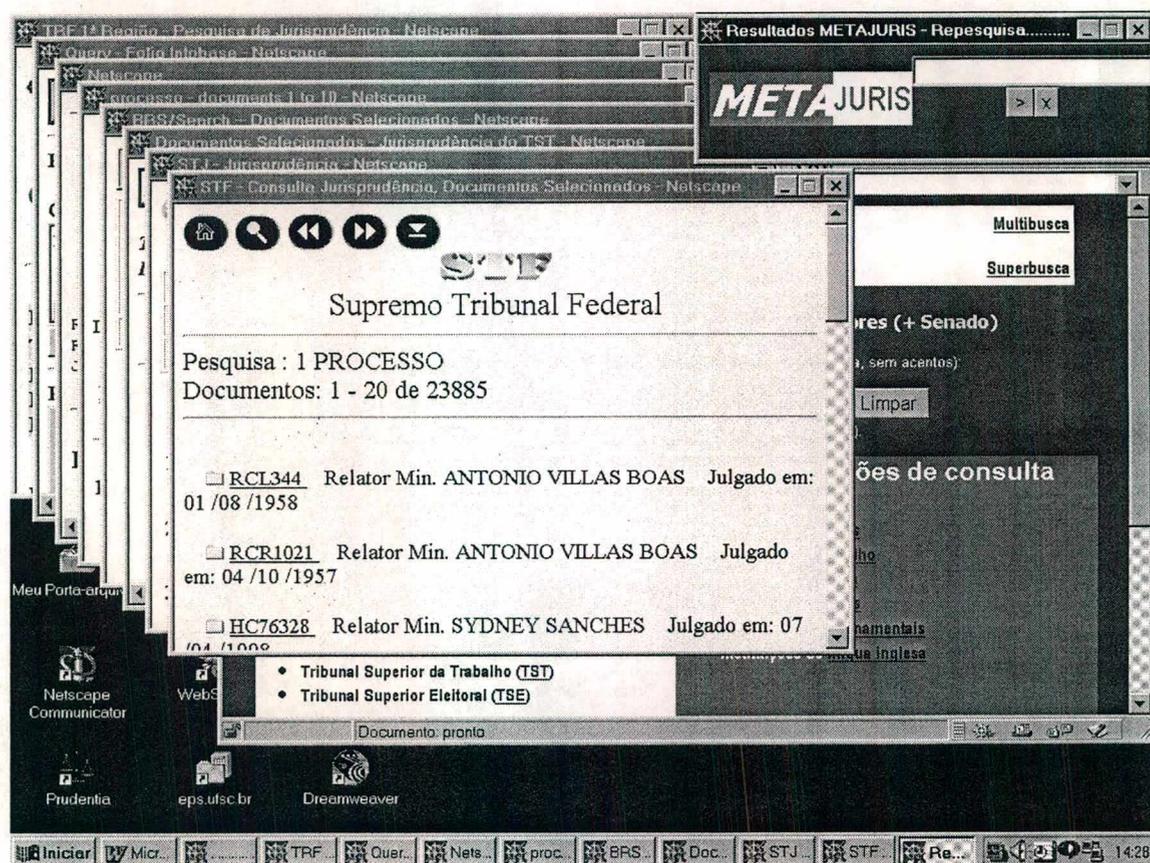


Figura 11

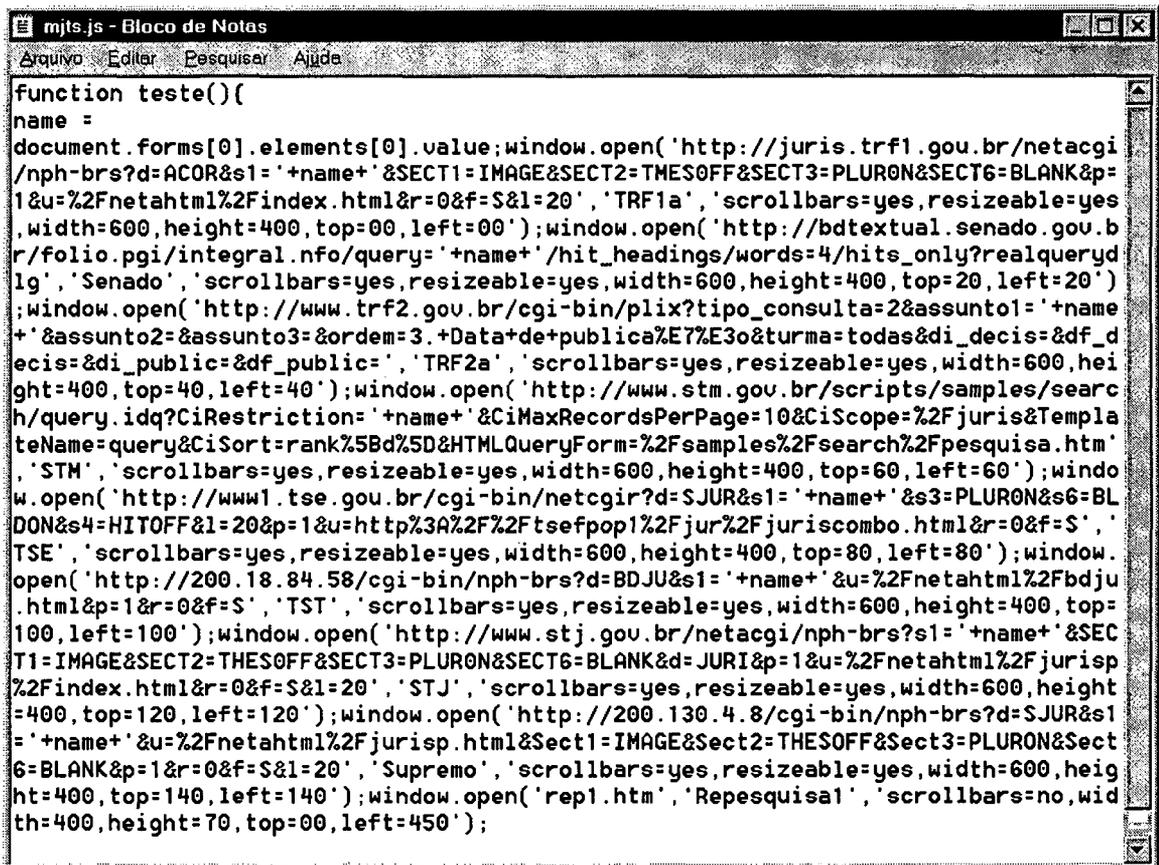
A realização da atividade até aqui descrita, na primeira versão, é levada a efeito através da combinação de técnicas de HTML⁴ e Javascript⁵, mediante a combinação de recursos do sistema operacional e do programa de navegação⁶. O script descreve a realização de uma função, designada “*teste*” a qual será assimilada pelo arquivo

⁴ Linguagem de marcação de hipertexto, utilizada para a criação dos hiperdocumentos disponíveis na world wide web.

⁵ “Javascript é uma linguagem script para o HTML e para o browser...” e seus scripts “são pequenos programas que interagem com o browser e o conteúdo HTML de uma página”, conforme McCOMB, na obra “JavaScript™ Sourcebook”, p. 4.

⁶ A Programação descrita atinge funcionamento pleno nos “browsers” versões 4.x ou superiores.

HTML e executada pelo programa de navegação. A estrutura dos comandos está visível na figura 12.



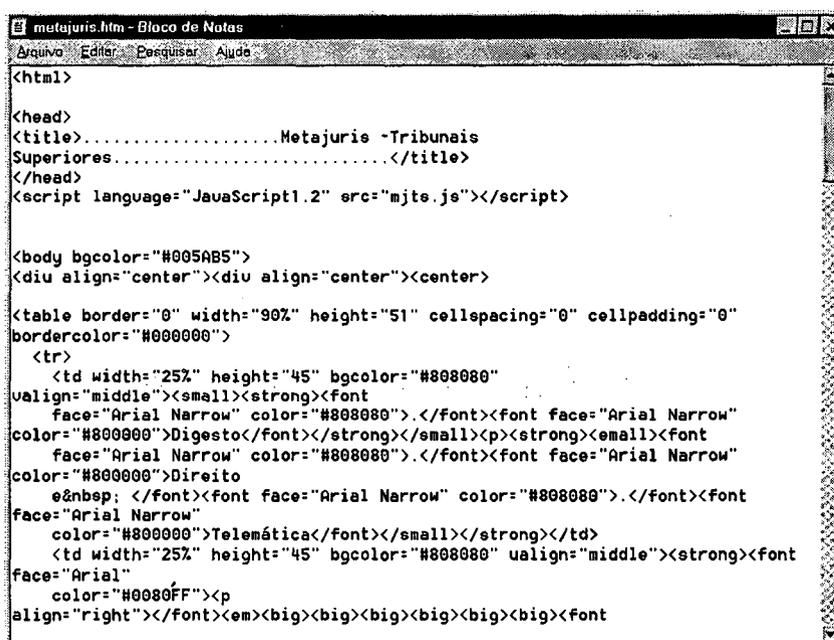
```
function teste(){
name =
document.forms[0].elements[0].value;window.open('http://juris.trf1.gou.br/netacgi/nph-brs?d=ACOR&s1='+name+'&SECT1=IMAGE&SECT2=THESOFF&SECT3=PLURON&SECT6=BLANK&p=1&u=%2Fnetahtml%2Findex.html&r=0&f=S&l=20','TRF1a','scrollbars=yes,resizeable=yes,width=600,height=400,top=00,left=00');window.open('http://bdtextual.senado.gov.br/folio.cgi/integral.nfo/query='+name+'/hit_headings/words=4/hits_only?realqueryd lg','Senado','scrollbars=yes,resizeable=yes,width=600,height=400,top=20,left=20');window.open('http://www.trf2.gou.br/cgi-bin/plix?tipo_consulta=2&assunto1='+name+'&assunto2=&assunto3=&ordem=3.+Data+de+publica%E7%E3o&turma=todas&di_decis=&df_decis=&di_public=&df_public','TRF2a','scrollbars=yes,resizeable=yes,width=600,height=400,top=40,left=40');window.open('http://www.stm.gou.br/scripts/samples/search/query?CiRestriction='+name+'&CiMaxRecordsPerPage=10&CiScope=%2Fjuris&TemplateName=query&CiSort=rank%5Bd%5D&HTMLQueryForm=%2Fsamples%2Fsearch%2Fpesquisa.htm','STM','scrollbars=yes,resizeable=yes,width=600,height=400,top=60,left=60');window.open('http://www1.tse.gou.br/cgi-bin/netcgir?d=SJUR&s1='+name+'&s3=PLURON&s6=BLDON&s4=HITOFF&l=20&p=1&u=http%3A%2F%2Ftsefpop1%2Fjur%2Fjuriscombo.html&r=0&f=S','TSE','scrollbars=yes,resizeable=yes,width=600,height=400,top=80,left=80');window.open('http://200.18.84.58/cgi-bin/nph-brs?d=BDJU&s1='+name+'&u=%2Fnetahtml%2Fbdju.html&p=1&r=0&f=S','TST','scrollbars=yes,resizeable=yes,width=600,height=400,top=100,left=100');window.open('http://www.stj.gou.br/netacgi/nph-brs?s1='+name+'&SECT1=IMAGE&SECT2=THESOFF&SECT3=PLURON&SECT6=BLANK&d=JURI&p=1&u=%2Fnetahtml%2Fjurisp%2Findex.html&r=0&f=S&l=20','STJ','scrollbars=yes,resizeable=yes,width=600,height=400,top=120,left=120');window.open('http://200.130.4.8/cgi-bin/nph-brs?d=SJUR&s1='+name+'&u=%2Fnetahtml%2Fjurisp.html&sect1=IMAGE&sect2=THESOFF&sect3=PLURON&sect6=BLANK&p=1&r=0&f=S&l=20','Supremo','scrollbars=yes,resizeable=yes,width=600,height=400,top=140,left=140');window.open('rep1.htm','Repesquisa1','scrollbars=no,width=400,height=70,top=00,left=450');
```

Figura 12

Cada um dos blocos contém as seguintes informações: 1. o conteúdo do texto digitado na caixa de diálogo está contido na *string* “name”; 2. é determinado ao browser que abra uma nova janela, procurando pelo endereço indicado, o qual se refere diretamente ao acionamento da pesquisa do banco de dados desejado; 3. a chave de pesquisa contém a expressão “name”, na qual está inserido o conteúdo textual digitado na caixa de diálogo do formulário apresentado na interface HTML; 4. Este acionamento provoca com o resultado a geração de uma página HTML com os dados de resposta; 5. A nova janela está programada para ser aberta sem que

sejam demonstrados alguns de seus elementos comuns (barras de navegação, estatus e menu, location), com um tamanho e uma posição específicos. Esta regulagem de posição permite que todas elas sejam apresentadas em cascata, com semi-sobreposição.

No âmbito do HTML, a programação segue os seguintes passos (excetuando -se a parte visual): 1. É informado ao browser a utilização de uma linguagem para a geração de scripts (script language="JavaScript1.2"); 2. É informada a existência de um script específico, fora do arquivo HTML, mas no mesmo diretório deste, que executará funções específicas (src="mjts.js"); 3. O formulário que absorve o texto de entrada recebe a orientação para executar a função contida no script quando o usuário deixar o formulário após provocar dentro do campo alguma alteração (onchange:"teste()"); 4. O texto digitado dentro da caixa do formulário fica armazenado dentro da expressão "name". O arquivo HTML do protótipo Metajuris está demonstrado na figura 13.



```

metajuris.htm - Bloco de Notas
Arquivo  Editar  Pesquisar  Ajuda

<html>
<head>
<title>.....Metajuris -Tribunais
Superiores.....</title>
</head>
<script language="JavaScript1.2" src="mjts.js"></script>

<body bgcolor="#005AB5">
<div align="center"><div align="center"><center>

<table border="0" width="90%" height="51" cellspacing="0" cellpadding="0"
bordercolor="#000000">
<tr>
<td width="25%" height="45" bgcolor="#808080"
valign="middle"><small><strong><font
face="Arial Narrow" color="#808080">.</font><font face="Arial Narrow"
color="#800000">Digesto</font></strong></small><p><strong><small><font
face="Arial Narrow" color="#808080">.</font><font face="Arial Narrow"
color="#800000">Direito
&nbsp;   </font><font face="Arial Narrow" color="#808080">.</font><font
face="Arial Narrow"
color="#800000">Telemática</font></small></strong></td>
<td width="25%" height="45" bgcolor="#808080" valign="middle"><strong><font
face="Arial"
color="#8080FF"><p
align="right"></font><em><big><big><big><big><big><big><font

```

Figura 13

Uma outra versão, em Perl, reúne todos os resultados em um único formulário, apresentando-os em uma janela única, como se pode ver na figura 14.

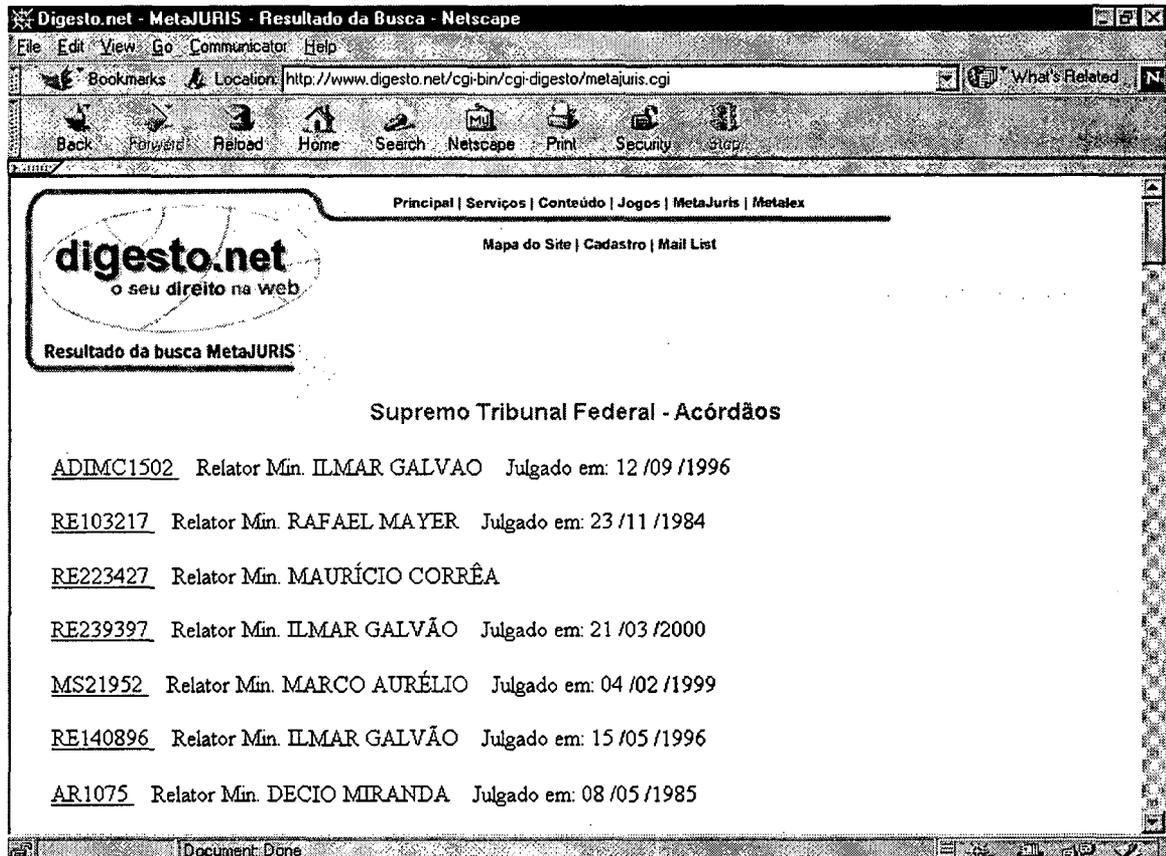


Figura 14

Cabe registrar que a sistemática apresentada proporciona um expressivo ganho de tempo na pesquisa, bem como gera um significativo aumento da sua amplitude lateral (diversidade de bases).

O ganho de tempo é obtido calculando -se o tempo médio destinado à operação de entrada e saída de cada uma das bases, que pode ser estimado, em média, em 2

minutos⁷. Assim, para realizar uma pesquisa através do *Metajuris*, utilizando-se o módulo “*Tribunais Superiores*”, seria gasto o tempo médio de 2 minutos, enquanto que, pelas vias normais, seriam gastos 16 minutos para consultar as mesmas bases.

C) Themis

O *Projeto Themis* é a gênese de todo o trabalho aqui apresentado, levando -se em consideração todos os protótipos e modelos descritos e comentados. Aquela idéia originária, apresentada nas primeiras linhas, continha as bases para os posteriores desdobramentos das pesquisas desenvolvidas pelo grupo e pelo autor. Sua idéia básica era a construção de um sistema de recuperação do conhecimento contido nas súmulas dos tribunais nacionais. Ou seja, partiu-se do *assunto* e não da *técnica*, sendo que esta foi uma consequência posterior, derivada da necessidade de se mesclar a pesquisa estruturada em texto e atributos, que pudesse ensejar a formação de um contexto.

A modelagem teórica conferida ao Sistema Themis (que somente foi implementado depois de outras experiências do grupo) lançou as bases para a construção do “P1” – adiante descrito – o que, posteriormente, ensejou a construção dos demais sistemas.

Sua idéia básica era tratar o conhecimento contido nas súmulas e seus respectivos enunciados (textos de aproximadamente três linhas, que contém importantes resumos sobre as interpretações judiciais consolidadas) de forma tal a que a interface

⁷ Estimativa realizada atendendo às seguintes premissas: 1. Cálculo de tempo necessário à digitação dos endereços; 2. Somatório do tempo para acesso ao mecanismo de consulta de jurisprudência; 3. Cômputo do tempo de digitação dos critérios de pesquisa; 4. Tempo médio de resposta, em dias normais e horários de tráfego mediano. Tendo em vista estas variáveis, bem como a velocidade de processamento, modem, navegador, etc., o tempo pode oscilar a maior ou a menor.

de do sistema permitisse ao usuário formular uma entrada na qual seria possível mesclar atributos fixos com texto livre, conforme a figura 15.

Raciocínio Baseado em Casos - [Procura por casos similares]

Arquivo Cadastro Pesquisa

Caso
legal impetração tribunal tese 100

Termos conexos 100

- ABSOLVIÇÃO
- ABUSO DE AUTORIDADE
- AÇÃO MANDAMENTAL
- AÇÃO ORIGINÁRIA
- ADMINISTRADOR PÚBLICO
- ADVOCACIA
- APOSENTADORIA
- ASSUNTOS EDUCACIONAIS
- ATO ABUSIVO

Termos relativamente conexos 100

- ABSOLVIÇÃO
- ABUSO DE AUTORIDADE
- AÇÃO MANDAMENTAL
- AÇÃO ORIGINÁRIA
- ADMINISTRADOR PÚBLICO
- ADVOCACIA
- APOSENTADORIA
- ASSUNTOS EDUCACIONAIS
- ATO ABUSIVO

Demais Informações

Tema primário: Ações Constitucionais 30

Tema secundário: 100

Tribunal: STF 25

Ramo do direito primário: Processual Civil 50

Ramo do direito secundário: 100

Ano inicial: 1900

Ano final: 1999 100

Filtro para busca

Limpa Consulta Sair

Figura 15

Além dessa, uma outra variante de interface foi desenvolvida, a fim de que se pudesse permitir ao usuário visualizar o maior número possível de informações, simultaneamente (inclusive boa parte do conteúdo das listas de indicadores temáticos), conforme a figura 16.

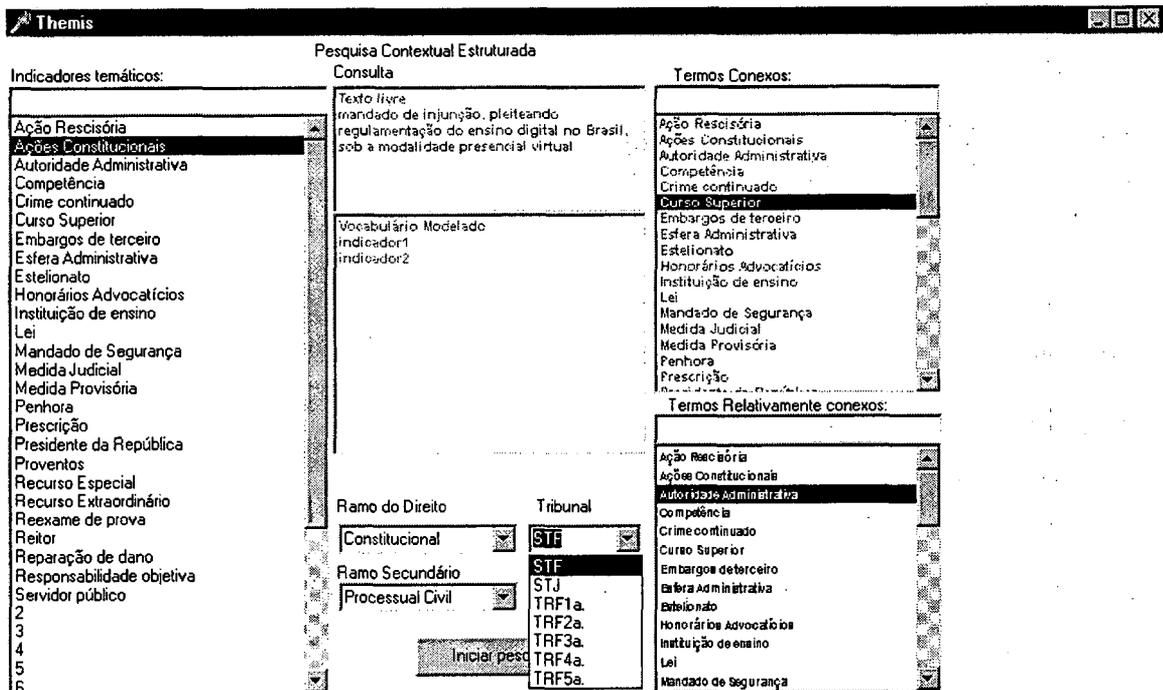


Figura 16

A apresentação das listas, levando -se em consideração os itens *tamanho da fonte*, *posição da lista e nomeclatura* (“indicadores temáticos”, “termos conexos” e “termos relativamente conexos”) sugerem a existência de uma hierarquia entre elas, que, de fato, existe, e produz efeitos sobre a métrica de similaridade operada pelo sistema para decidir qual enunciado das súmulas é mais semelhante à entrada formulada.

A operação do sistema Themis está fundamentada na extração de dez atributos provenientes dos enunciados das súmulas, com os seguintes valores a eles atribuíveis:

1. Tribunal (lista fechada, 7 itens, valor: 0,5 ou 0);
2. Ano (lista fechada, 34 itens, valor: 0,5 ou 0);
3. Ramo do direito (lista fechada, 35 itens, valor: 1 ou 0);
4. Ramo secundário (lista fechada, 35 itens, valor: 1 ou 0);
5. Indicador temático central (lista fechada, valor: 1 ou 0);
6. Indicador temático subsidiário (lista fechada, valor 1 ou 0);
7. Texto do enunciado da súmula (área de texto, valor 1 para cada palavra encontrada);
8. Termos fortemente conexos (área de texto, 3 a 5 termos, valor: 0,7 por palavra encontrada);
9. Termos relativamente conexos (área de texto, valor: 0,5 por palavra encontrada);
10. Número da súmula (não integra a métrica).

Elaboração de métrica de similaridade:

1. *Primeiro nível (Pn)*: leva em consideração os **atributos fixos**, contidos nos campos de “1” a “6”, que equivalem a 50% da métrica. Somados, perfazem o valor máximo de 5,0. Exemplo: de 6 atributos escolhidos, 4 foram encontrados (1, 2, 4 e 5). A pontuação total, neste caso, é de 3,0, ou 30% da similaridade total. Os valores do primeiro nível são fixos, ou seja, 5,0 sempre será igual a 50%, 3,5=35%, etc;
2. *Segundo nível (Sn)*: leva em consideração as **palavras informadas** na entrada, operando sobre os campos “7” a “9”, que equivalem aos outros 50% da métrica total. Cada palavra equivale a 1,0. Exemplo: se 7 palavras foram informadas, 7=50%. Na busca, duas não foram encontradas, uma foi encontrada no campo 9 (recebe o valor 0,5), duas foram encontradas no campo 8 (recebem o valor 0,7 cada uma delas), e as demais foram encontradas no campo 7 (recebem o valor

1,0 cada uma delas). O valor total é $0,5+1,4+2,0=3,9$, que corresponde a 27,85% da similaridade total (100%), tendo em vista que $7=50\%$.

O valor final da **similaridade** é $Pn + Sn$, o que, no exemplo citado, corresponderia a $30+27,5=57,85\%$.

O segundo nível (Sn) tem o seu valor máximo variável, dependendo de quantas palavras forem informadas para consulta. Cada palavra valendo 1,0, se forem informadas 2 palavras, $2=50\%$, se 4, $4=50\%$, se 9, $9=50\%$, se 20, $20=50\%$, a assim por diante.

O resultado é apresentado em uma interface de saída que apresenta os resultados hierarquizados de acordo com o percentual de similaridade operado entre a consulta e os enunciados das súmulas, conforme a figura 18.

Outro aspecto relevante do sistema são os *pesos dinâmicos*. Trata-se da possibilidade, contida na interface, de alteração da relevância dos atributos dentro do contexto pesquisado, calibrando -se, individualizadamente, os pesos de cada um dos atributos, dentro de uma escala de 0 a 100, conforme destaques contidos na figura 17.

Raciocínio Baseado em Casos - [Procura por casos similares]

Arquivo Cadastro Pesquisa

Caso
legal impetração tribunal lese 100

Termos conexos

ABSOLVIÇÃO 100
 ABUSO DE AUTORIDADE
 AÇÃO MANDAMENTAL
 AÇÃO ORIGINÁRIA
 ADMINISTRADOR PÚBLICO
 ADVOCACIA
 APOSENTADORIA
 ASSUNTOS EDUCACIONAIS
 ATO ABUSIVO

Termos relativamente conexos

ABSOLVIÇÃO 100
 ABUSO DE AUTORIDADE
 AÇÃO MANDAMENTAL
 AÇÃO ORIGINÁRIA
 ADMINISTRADOR PÚBLICO
 ADVOCACIA
 APOSENTADORIA
 ASSUNTOS EDUCACIONAIS
 ATO ABUSIVO

Demais Informações

Tema primário
 Ações Constitucionais 30

Tema secundário 100

Tribunal
 STF 25

Ramo do direito primário
 Processual Civil 50

Ramo do direito secundário 100

Ano inicial 1900 Ano final 1999 100

Filtro para busca

Pesos Dinâmicos

Limpa Consulta Sair

Figura 17

Além disso, o projeto inicial de interface de saída apresenta também um somatório visual das informações solicitadas, isto é, todos os mecanismos que o usuário ativou na realização da consulta, bem como os valores respectivamente atribuídos, estão ali consignados, conforme figura 18.

Themis - Resultado da Pesquisa

Estrutura da pesquisa solicitada:

<p>Texto livre:</p> <p>transferência professor "engenharia do conhecimento" universidade reitor federação publicidade "ensino público" titularidade</p>	<p>Vocabulário modelado:</p> <p>Mandado de Segurança Curso Superior Medida Provisória Honorários Advocatícios Prescrição Recurso Especial Servidor Público Reparação de Dano</p>	<p>Competência Autoridade Administrativa Instituição de Ensino</p>
<p>Opções:</p> <p>Ramo do Direito: Tributário Tribunal: STF Ramo Secundário: (nenhum) Ano: (nenhum)</p>		<p>Ações Constitucionais Estatuto Administrativo Uso Recomeço de Prova</p>

Resultado:

1. 82%	STF	47	1973	"Não cabe transferência imotivada em curso superior para servidor designado em outro órgão."
2. 80%	STJ	35	1992	"A universidade pública é obrigada a matricular o pesquisador recém-chegado do exterior."
3. 71%	TRF	07	1997	"Testes com software, na madrugada, torram a família e fazem o sono fugir."
4. 65%	STJ	12	1989	"Impugnações ao trabalho científico, por profissionais liberais, exigem fundamentação."

Mais Resultados Nova Pesquisa Carregar

Figura 18

O Projeto Themis virou sistema prototípico em 1999. A necessidade de modelagem manual de cada um dos registros, com hierarquização personalizada das listas e ajuste específico de sinônimos, tem dificultado a evolução do sistema, pois materializa tarefas caras e trabalhosas. Porém, várias conclusões importantes foram obtidas em estudos referentes ao seu desenvolvimento, e uma delas é a mescla de texto com atributos, posteriormente utilizada nos sistemas Jurisconsulto e Olimpo.

A versão implementada em 1999 (figura 15), conta com importantes inovações, como a possibilidade de alteração dinâmica dos pesos dos atributos, como se pode verificar na figura 17, onde foram ativados 6 itens de pesquisa (5 + o texto livre), dos quais 3 ("tema primário", "ramo do direito" e "tribunal") o foram com pesos

diferenciados (30%, 50% e 25%, respectivamente), gerando o resultado apresentado na figura 19.

Resultado da Consulta

Rank das Súmulas Recuperadas

Súmula	Tribunal	Grau de Similaridade
266	STF	0,432098765432099
2	STJ	0,320987654320988
343	STF	0,185185185185185
400	STF	0,185185185185185
621	STF	0,185185185185185

Texto da Súmula

Não cabe mandado de segurança contra lei em tese.

Ramo Primário: Tema Primário: Ano:

Ramo Secundário: Tema Secundário:

Termos Conexos

- ATO ABUSIVO
- AÇÃO MANDAMENTAL
- DESCONSTITUIÇÃO
- ILEGALIDADE

Termos Relativamente Conexos

- CABIMENTO
- CONCESSÃO
- LEGISLAÇÃO
- LIMINAR

Figura 19

Outro aspecto relevante do resultado do sistema apresentado na figura 19 é a apresentação dos atributos originariamente fixados para aquele registro específico, com respectivos valores atribuídos em cada um deles, a fim de que o usuário possa checar quais os atributos e valores influenciaram na escolha do registro.

Além disso, foi construído um módulo específico onde podem ser visualizados cada um dos enunciados das súmulas, diretamente, conforme figura 20.

Número	Descrição de súmula
7	A pretensão de simples reexame de prova não enseja recurso especial.
9	Em matéria previdenciária, torna-se desnecessário o prévio exaurimento da via administrativa, como condição de ajuizamento da ação.
13	Os servidores públicos federais civis e militares ainda não haviam implementado a condição temporal para a incorporação à sua remuneração do índice de reajuste de 84,32%, correspondente ao IPC de março de 1990, quando sobreveio a medida provisória n. 154, de 15 de março de 1990.
16	O reajuste dos servidores militares, estabelecido na Lei n. 8.237/91 não tem aplicação aos servidores civis.
29	Não cabe a exigência de estágio profissionalizante para efeito de matrícula em curso superior.
47	Reitor de Universidade não é livremente demissível pelo presidente da república durante o prazo de sua investidura
130	A empresa responde, perante o cliente, pela reparação de dano ou furto de veículo ocorridos em seu estacionamento.
266	Não cabe mandado de segurança contra lei em tese.
343	Não cabe ação rescisória por ofensa a literal dispositivo de lei, quando a decisão rescindenda se tiver baseado em texto legal de interpretação controvertida nos tribunais.
400	Decisão que deu razoável interpretação à lei, ainda que não seja a melhor, não autoriza recurso extraordinário pela letra a do art. 101, III, da Constituição Federal.
497	Quando se tratar de crime continuado, a prescrição regula-se pela pena imposta na Sentença, não se computando o acréscimo decorrente da continuação.
512	Não cabe condenação em honorários de advogado na ação de mandado de segurança.
521	O foro competente para o processo e julgamento dos crimes de estelionato, sob a modalidade da emissão dolosa de cheque sem provisão de fundos, é o do local onde se deu a recusa do pagamento pelo sacado.
621	Não enseja embargos de terceiro a penhora à promessa de compra e venda não inscrita no registro de imóveis.

Figura 20

O Projeto Themis materializa um tema tão relevante – tendo em vista que as súmulas representam o melhor e mais autorizado extrato da interpretação jurídica nacional – que seu desenvolvimento ainda será objeto de trabalhos futuros do autor.

D)“P1”

No segundo trimestre do ano de 1996, durante a disciplina de RBC, no PPGEF - UFSC, decidimos implementar um sistema prototípico envolvendo conhecimento

jurídico. Optamos pela utilização das técnicas de extração de atributos desenvolvidas para o Projeto Themis, para a construção de um sistema de RBC “puro”, hoje chamado de “P1”, voltado para a recuperação de acórdãos judiciais (textos deliberativos, geralmente com mais de duas páginas, podendo passar de dez), cujo tema eram casos de “Habeas corpus” (pedido judicial formulado para libertar alguém que se encontra injustamente preso ou t em sua liberdade civil constrangida ilegalmente) em crimes de “homicídio”.

Os 22 (vinte e dois) casos escolhidos para a elaboração do protótipo foram analisados e as características comuns entre eles, consideradas mais relevantes, foram relacionadas. Outras características importantes foram adicionadas à relação, denominada “lista de atributos”. A escolha da lista de atributos foi realizada após o estudo de todos os fatores relevantes para a concessão ou não do “habeas corpus” em crime de homicídio. Desta forma, poderão ser adicionadas à base de casos outros acórdãos referentes a área em questão, pois a intenção desta lista de atributos é atingir qualquer caso de *“habeas corpus” em crime de homicídio*.

A ementa na íntegra teve que ser dividida nos campos D_1 a D_6, devido a limitação de caracteres nos campos da plataforma ESTEEM (utilizada para a construção elaboração do protótipo). As características escolhidas são aquelas constantes das figura 21 e 22.

Edit Case

Case Name:

- C_17_11772
- C_18_10011
- C_19_10100
- C_20_10806
- C_21_10150

Figura 21

Case Base Definition Editor			
Current Case-Base: Habeas_2_CB		Current Similarity Definition: habeas_sim	
Feature Names	Feature Value Types	Feature Names	Feature Value Types
DataPublicacaoDJ	Text	Numero	Numeric
CaracterizacaoHomicidio	One of a List	Qualificacao	One of a List
StatusHomicidio	One of a List	CoAutoria	One of a List
Solicitacao	One of a List	FundamentoAcao	One of a List
FundSecAcao	One of a List	FundamentoReu	One of a List
FundSecReu	One of a List	RelacaoAcao	Text

Case Base Definition Editor			
Current Case-Base: Habeas_2_CB		Current Similarity Definition: habeas_sim	
Feature Names	Feature Value Types	Feature Names	Feature Value Types
FundSecReu	One of a List	RelacaoAcao	Text
RelacaoReu	Text	Observacoes	Text
Circunstancia	Text	Decisao	One of a List
Unanimidade	One of a List	D_1	Text
D_2	Text	D_3	Text
D_4	Text	D_5	Text

Figura 22

Na seqüência, pode-se observar que os atributos foram agrupados observando os seguintes critérios: os fundamentos da decisão referentes aos argumentos que sejam pertinentes ao réu ou ao processo (primária e secundariamente); existência de qualificadoras ou privilégio (simple, qualificado, privilegiado); *iter criminis* (consumado, tentado); e outros particulares. No âmbito dos fundamentos da

argumentação é mais elevada a densidade de elementos relevantes à análise, como a ausência de autoria e/ou materialidade, a segurança da instrução processual, os bons antecedentes, a tipicidade, a primariedade, a garantia da ordem pública, etc.

Os casos foram analisados, e tais figuras foram identificadas, objetivamente, e representadas conforme as figuras 23 e 24.

Case Editor		Save Case	Get Case
Current Case-Base: Habeas_2_CB		Case Name: C_1_11766	
Feature Names	Feature Values	Case Saved	Clear Case
DataPublicacaoDJ	09-09-94		
Numero	9070		
CaracterizacaoHomicidio	doloso		
Qualificacao	Qualificado		
StatusHomicidio	consumado		
CoAutoria	Sim		
Solicitacao	trancamento		
FundamentoAcao	SegurancaInstrucaoProcessual		

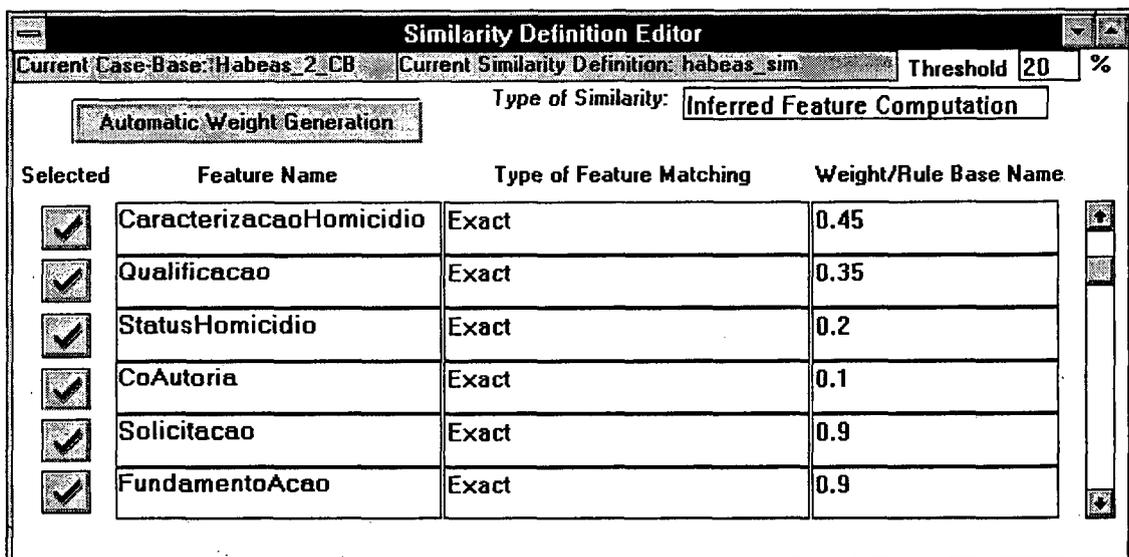
Figura 23

Case Editor		Save Case	Get Case
Current Case-Base: Habeas_2_CB		Case Name: C_1_11766	
Feature Names	Feature Values	Case Saved	Clear Case
FundSecReu	BonsAntecedentes		
RelacaoAcao			
RelacaoReu			
Observacoes	Caracterizacao do homicidio presumida.		
Circunstancia	irrelevância das condições pessoais favoráveis ao a		
Decisao	denegada		
Unanimidade	Sim		
D_1	Ementa : IIHABEAS CORPUS - PRISÃO TEMPORÁRIA		

Figura 24

Em seguida, foi estabelecida uma graduação referente ao relacionamento das características dos casos recuperados com o caso de entrada, conforme figura 25.

Tal graduação foi o referencial da métrica de similaridade, materializada em regras inseridas em uma base específica para tal. Cada grupo de atributos recebeu um peso específico para a análise de similaridade. Assim, o mecanismo, ao constatar a presença de identidade entre um atributo do caso de entrada e o de um dos casos da base, é orientado, pelos pesos, para estabelecer, numericamente, a importância da coincidência, de forma tal que se possa hierarquizar as semelhanças.



Selected	Feature Name	Type of Feature Matching	Weight/Rule Base Name
<input checked="" type="checkbox"/>	CaracterizacaoHomicidio	Exact	0.45
<input checked="" type="checkbox"/>	Qualificacao	Exact	0.35
<input checked="" type="checkbox"/>	StatusHomicidio	Exact	0.2
<input checked="" type="checkbox"/>	CoAutoria	Exact	0.1
<input checked="" type="checkbox"/>	Solicitacao	Exact	0.9
<input checked="" type="checkbox"/>	FundamentoAcao	Exact	0.9

Figura 25

Tal resultado enseja um incrível ganho de tempo ao pesquisador, eis que a análise dos casos recuperados parte de um referencial concreto, tornando a busca muito mais precisa.

A próxima etapa foi a elaboração de regras (figuras 26, 27 e 28) para os campos do fundamento referente à ação, fundamento referente aos argumentos apresentados pelo paciente (pessoa em nome de quem é impetrado o *Habeas Corpus*) e a relação entre estes dois campos. Esta etapa foi necessária pela existência de grau de similaridade entre algumas características destes campos, como, por exemplo, primariedade e bons antecedentes, isto é, se o usuário entrar com um caso cuja característica (uma delas) seja *bons antecedentes*, aqueles casos que estiverem na base do sistema e contenham a característica *primariedade* certamente deverão ter um alto grau de similaridade com a mesma, o mesmo não ocorrendo se o caso de entrada tivessê como característica *maus antecedentes*.

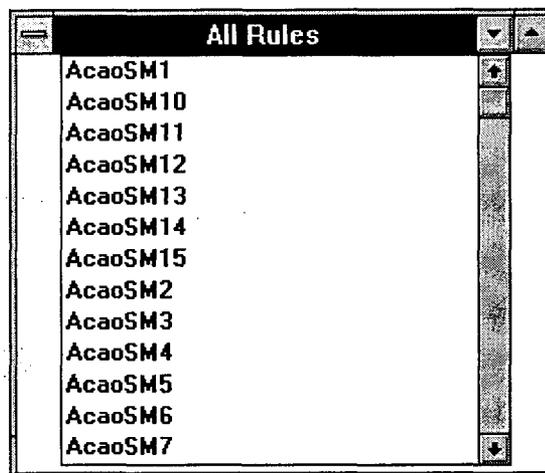


Figura 26

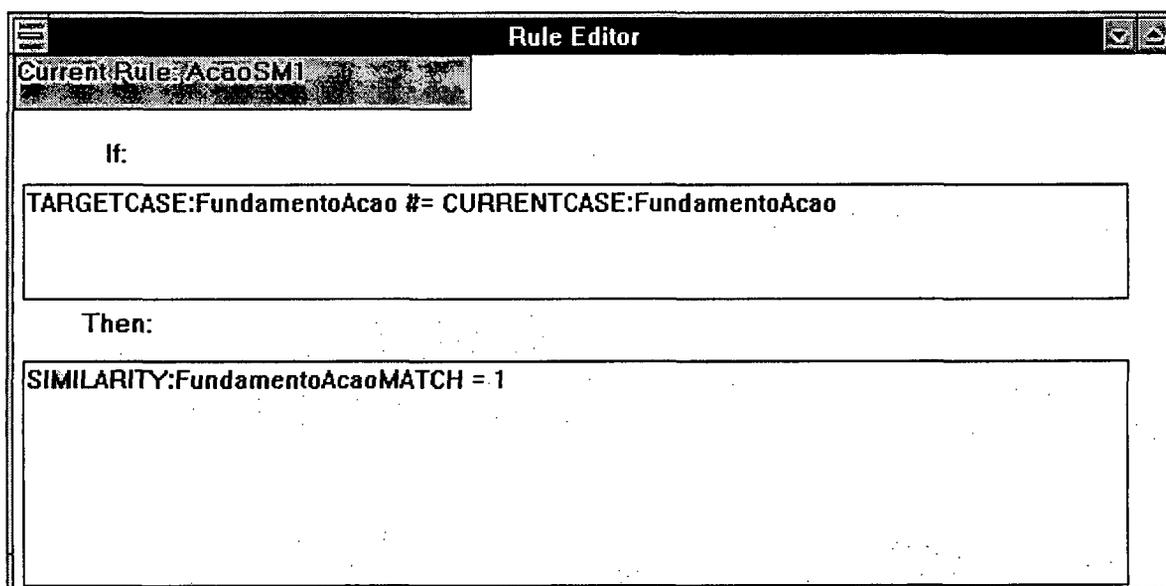


Figura 27

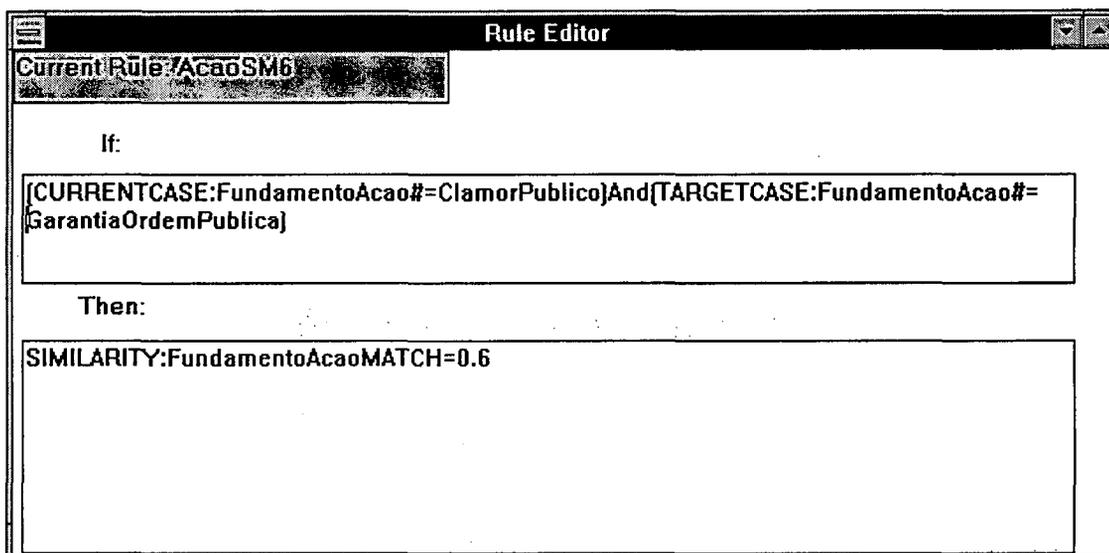


Figura 28

A próxima etapa é aquela na qual o usuário deflagra o funcionamento do sistema, para recuperar o caso mais semelhante. Ele fará isto através do preenchimento dos campos constantes da figura 29:

ESTEEM Application Interface	
File Help	
Change Retrieval Attributes Retrieve Adaptation Incorporate New Case Help Print Exit	
Enter Target Case	
CaracterizacaoHomicidio	culposo
Qualificacao	Simple
StatusHomicidio	consumado
CoAutoria	Sim
Solicitacao	trancamento
FundamentoAcao	FaltaJustaCausa

Figura 29

Os casos recuperados são apresentados de forma hierarquizada, podendo -se afirmar que o caso com o melhor resultado numérico é o mais similar, conforme a figura 30.

Retrieved Case List			
Score	Case Name	Solicitacao	Decisao
100	C_3_10476	trancamento	denegada
58	C_7_9819	trancamento	concedida
40	C_2_11109	trancamento	denegada
38	C_4_10436	trancamento	denegada
34	C_5_10292	trancamento	denegada
25	C_6_10150	trancamento	denegada
24	C_1_11766	trancamento	denegada
23	C_14_10272	revogacao	denegada
23	C_19_10100	revogacao	denegada

Figura 30

O usuário, desejando visualizar o teor processo, como representado, basta levar o cursor do mouse na linha correspondente ao registro desejado. Será acionada uma janela com as respectivas informações.

O protótipo “P1” apresenta importantes aspectos que tornam seu funcionamento bastante sofisticado, como a hibridação entre regras e casos, bem como a hierarquização conceitual e respectiva atribuição de *pesos bilaterais* entre institutos jurídicos (primariedade=>bons antecedentes= 0,6, bons antecedentes => primariedade = 0,8). Tendo em vista que a construção dos pesos bilaterais é uma tarefa artesanal e personalizada, a construção de um sistema de grande porte materializa, atualmente, tarefa significativamente trabalhosa, mas a construção do protótipo, além de gerar importantes conclusões a nível de pesquisa, comprova a viabilidade da edificação de um sistema de maior porte, o que, de fato, ocorreu posteriormente, porém com um grau de sofisticação menor, e sem a atribuição de *pesos bilaterais* (por enquanto).

O resultado final do “P1” superou as expectativas iniciais do grupo, a ponto de transformar-se em objeto central da monografia de especialização realizada por Bueno [12], vindo a fundamentar, no ano seguinte, o trabalho do grupo, aprovado para apresentação e publicação perante a conferência internacional de raciocínio baseado em casos, realizada em Providence, EUA.

E) Prudentia

O sistema Prudentia foi a evolução quantitativa natural do “P1”, materializado nos trabalhos de Mattos [69] e Webber [96]. Tratava-se da aplicação pura de técnicas de RBC, extraindo e comparando atributos, mediante uma métrica de similaridade operada entre eles. Porém, o grande aspecto qualificador do sistema - que era a atribuição de *pesos bilaterais* aos atributos - não foi incorporado ao Prudentia.

Importantes descobertas feitas pelo grupo de pesquisa, ao longo da elaboração do sistema, possibilitaram o incremento de sua performance. Uma delas é a

identificação da padronização textual e retórica dos documentos jurídicos, feita pelos juristas do grupo. Mostramos aos profissionais e pesquisadores das demais áreas que os acórdãos possuem uma estrutura uniformizada, o que permitiu a construção de mecanismos de extração automática de atributos e características dos textos, tarefa brilhantemente implementada por Mattos e Webber.

Porém, por trabalhar exclusivamente com atributos estáticos (“fundamento 1, 2 e 3”, “tema 1 e 2”, “artigo 1 e 2”, “recurso” e “tipificação”), o sistema não demorou a demonstrar suas limitações.

F) JurisConsulta

Embora o sistema Themis tenha sido idealizado anteriormente, o JurisConsulta foi o primeiro protótipo do grupo a realmente mesclar “busca textual” com “comparação de atributos”. Sua estruturação foi objeto de trabalho aprovado para apresentação – e posterior publicação – na sétima conferência internacional de inteligência artificial, organizada pela associação internacional de inteligência artificial e direito. Também foi objeto da dissertação de mestrado intitulada “O uso da teoria jurídica para recuperação em amplas bases de textos jurídicos”, defendida por Bueno “[11] perante o PPGEF-UFSC, onde se ressalta a importância das teorias ligadas ao direito para a adequada representação do conhecimento jurídico. Outra dissertação de mestrado, em andamento, realizada por Mattos, enfoca os aspectos computacionais referentes ao sistema. Bueno permanece pesquisando o assunto, e certamente deve nos brindar com novidades interessantes em sua tese de doutorado, em fase de consolidação.

O JurisConsulta apresenta-se como uma proposta inovadora em ambos os aspectos. Sob a ótica computacional, as técnicas de pré-consulta e da graduação do nível de

profundidade da pesquisa – entre outras – constituem significativos avanços em termos de sistemas inteligentes de recuperação de informação. Além disso, ele teve duas versões implementadas – uma em Delphi outra em Smalltalk – e foi objeto de relevantes descobertas para o grupo de pesquisa em IA e direito. Sob a ótica da representação do conhecimento, também se constitui em uma proposta absolutamente vanguardista, tendo em vista por exemplo, o aprimoramento das técnicas que permitem a diferenciação de crimes “tentado” e “consumado”, ou “culposos” e “qualificados” quando da elaboração da consulta. Não há proposta similar no cenário mundial atual de IA e direito.

Seu processo lógico de funcionamento está descrito na figura 31, e sua interface primária é apresentada na figura 32.

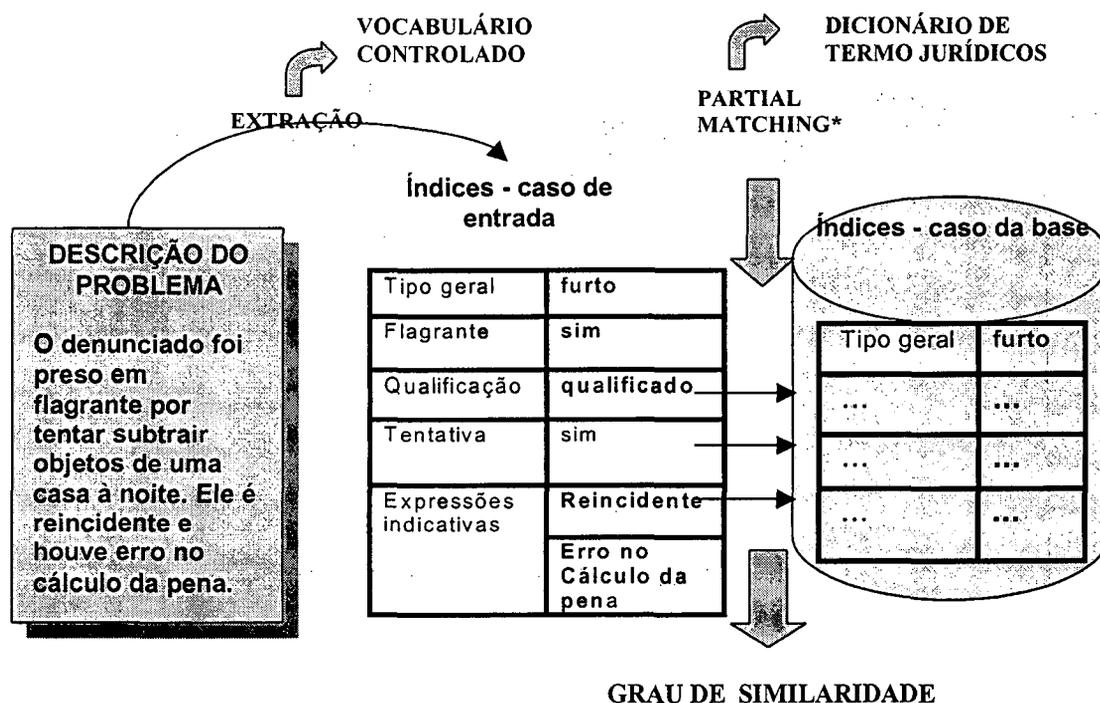


Figura 31



Figura 32

Acionando a opção “elaborar” o usuário ingressa na área de consulta (figura 33), onde vai descrever a questão para a qual deseja obter resposta.

Figura 33

A adequada descrição do sistema é tarefa de complexidade autônoma. Porém, vamos apresentar alguns elementos que demonstram a força do sistema.

Protocolamos a seguinte descrição de um caso, na área de consulta:

“testemunha ocular de crime de tráfico de entorpecentes, praticado com auxílio de menor, mediante uso de arma de fogo, com risco de vida e possibilidade de homicídio ”

Sem o acionamento de filtros, ou indicação de outros atributos, o sistema apresentou como resultado da consulta os casos descritos na figura 34.

JurisConsulta - Resultado		
Resultado da Consulta		
Posição	Similar.	Ementa
1º	72,70	Tóxico. Tráfico. Condenação. Apelação criminal com pedidos alternativos de: absolvição; desclassificação para o crime de uso; reconhecimento da atenuante da menoridade e redução da pena imposta. Desprovemento do ... Acórdão: 27465AC - Relator: Neuro Colleço - Localização: Taió - Data: 05/11/2091 - Resultado: Denegado
2º	72,70	CRIME CONTRA A SAÚDE PÚBLICA. TRÁFICO DE ENTORPECENTES (COCAÍNA). AUTORIA E MATERIALIDADE DEVIDAMENTE COMPROVADAS. LAUDO PERICIAL IDÔNEO PARA O EMBASAMENTO DO DECRETO CONDENATÓRIO. IRRELEVÂNCIA DE UM EVENTUAL DESENCONTRO, ENTRE A QUANTIDADE DE SUBSTÂNCIA TÓXICA APREENDIDA E A REMETIDA À PERICIA, MORMENTE SE MÍNIMA E INSIGNIFICANTE É ESSA ... Acórdão: 33173AC - Relator: Nereide Xavier de Amorante - Localização: Curitiba - Data: ...
3º	63,60	CRIME CONTRA A SAÚDE PÚBLICA - Narcotráfico - Sentença que desclassificou o delito imputado para o art. 16, da Lei n. 6.368/76 - Acusado viciado e dependente da droga em grau moderado - (Exame pericial anexo) - Ausência de prova robusta de narcotráficação - Decisão mantida - Recurso ... Acórdão: 26016AC - Relator: Vladimir d'Ivanenko - Localização: São José - Data: 31/08/2090 - Resultado: Denegado

Você poderá visualizar o acórdão na íntegra, através de um duplo clique, no caso desejado.

Acórdão Ajuda Elaborar

©1999 PFGEP

Figura 34

A partir da lista de acórdãos, é possível, com um clique duplo, acessar o teor integral do julgamento referenciado, conforme a figura 35.

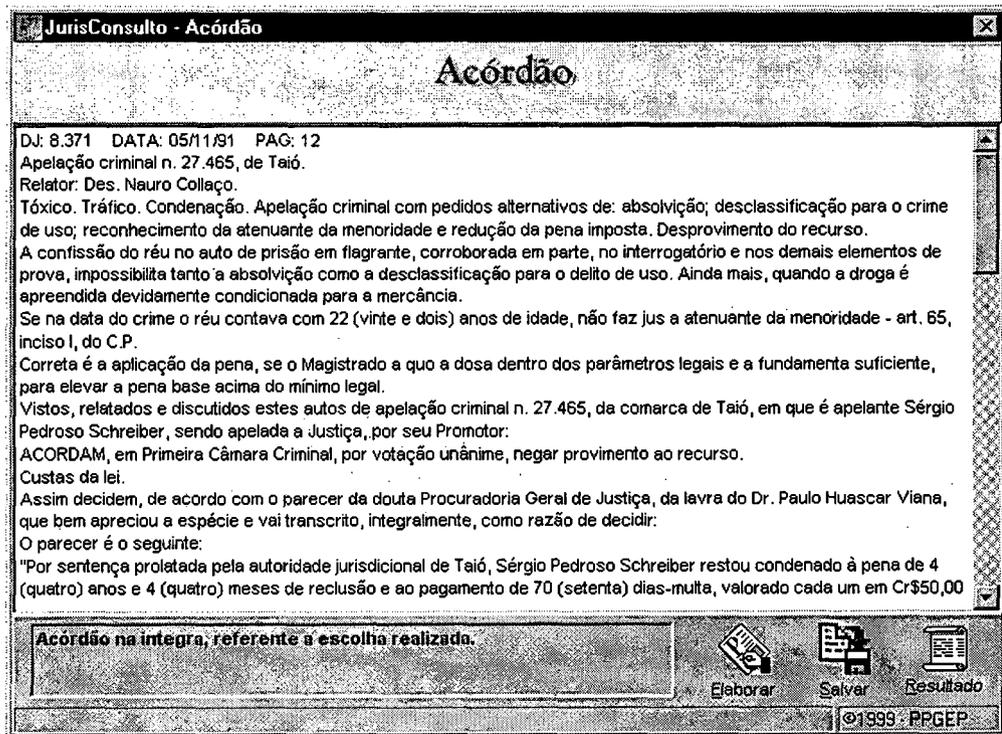


Figura 35

Porém, o JurisConsulta possui pontos frágeis que ainda o impedem de ser classificado com um sistema estável e com bom índice de recuperação, embora possua um potencial de desenvolvimento muito forte. Chegamos a realizar testes de consulta onde eram escritas até três linhas de texto como entrada, e o sistema não identificava elementos capazes de integrarem a métrica de similaridade. Isso não reduz a força inovadora deste sistema revolucionário, mas nos mostrou que ele deveria ser aperfeiçoado, o que está ocorrendo neste momento.

Cada um destes sistemas e protótipos, mesmo aqueles considerados sem comportamento inteligente, gerou conclusões e experiências extremamente úteis na confecção do Sistema Olimpo, de forma que se pode afirmar que ele é, também, o produto resultante de toda esta fase de experimentação e prototipagem. Do Digesto e

do Metajuris surgiram importantes conhecimentos sobre organização e funcionamento de bancos de dados textuais. Do “P1” e do Prudentia veio uma boa experiência no trabalho com atributos estáticos e sua respectiva representação. O Themis e o JurisConsulta nos geraram a capacidade de estruturação de mecanismos de pesquisa que hibridassem textos e atributos, bem como foram marcantes nas tarefas de identificação e extração de atributos estáticos.

Conhecendo os principais pontos fortes e fracos de cada um destes sistemas e protótipos, foi muito mais agradável partir para a construção de um sistema no qual o autor pôde conciliar suas concepções iniciais com as experiências já realizadas, idealizando um novo sistema que pudesse desempenhar, de forma estável, as tarefas já dominadas pelo grupo de pesquisa, e que não sofresse com os problemas já conhecidos, tendo como ponto forte uma reorganização metodológica das bases de conhecimento e uma nova forma de estruturação dos processo de representação do conhecimento. Nascia o Sistema Olimpo.

Um próximo e interessante passo é conciliar as técnicas consolidadas no Olimpo com as possibilidades oferecidas pelo Metajuris, de fazer pesquisas múltiplas e simultâneas em bases de dados remotas. Teríamos, então, um sistema ao mesmo tempo inteligente e com muita amplitude.

2) Análise comparativa

Os principais sistemas idealizados e construídos pelo grupo podem ser comparados de diversas formas, e sob critérios diferenciados. O primeiro comparativo leva em consideração os seguintes aspectos, constantes da tabela 2:

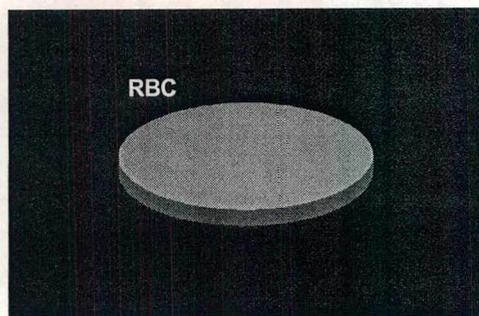
Sistema → ↓Característica	Digesto	Metajuris	Themis	PI	Prudentia	Jurisconsulto	Olimpo
Conhecimento representado	•	-	•	•	•	•	•
Inteligência	-	•	•	•	•	•	•
Cruzamento de técnicas ⁸	-	-	•	-	-	•	•
Alimentação automática	-	-	-	-	•	•	•
Comparação de documentos textuais	-	-	-	-	-	•	•

Tabela 2

Levando em consideração características genéricas, percebe-se que os sistemas OLIMPO e JURISCONSULTO possuem o melhor desempenho, que será melhor analisado em seguida.

Antes disso, convém analisar os sistemas sob a ótica das técnicas empregadas para a sua construção. Alguns deles foram construídos mediante o emprego de RB C puro ("P1" e Prudentia, gráfico 3).

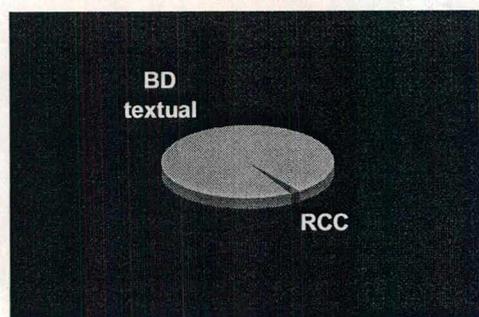
⁸ Embora o Digesto e o Metajuris utilizem uma pequena pitada de RC2D, por terem sido os primeiros objeto de experimento desta técnica, sua participação é muito pequena para que sejam considerados positivos no quesito "Cruzamento de técnicas".



**Somente RBC: "P1"
e Prudentia**

Gráfico 3

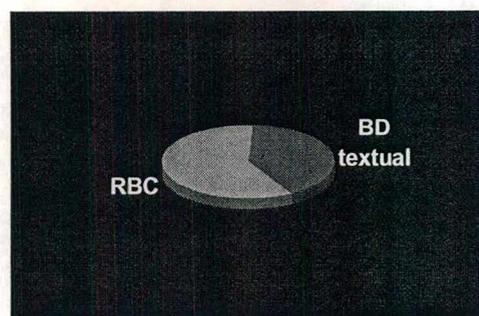
Outros utilizaram somente técnicas de bancos textuais com uma pitada de RC²D (Digesto e Metajuris, Gráfico 4).



**DB textual e RC²D:
Digesto e Metajuris**

Gráfico 4

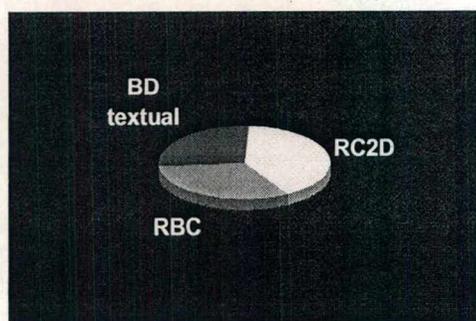
Alguns utilizaram BD textual e RBC (Themis e Jurisconsulto, Gráfico 5):



**BD textual e RBC:
Themis e
Jurisconsulto**

Gráfico 5

Já o Sistema Olimpo foi o primeiro a utilizar RC²D, RBC e BD textual (gráfico 6).



RC²D, BD textual e RBC: Olimpo

Gráfico 6

Cruzando características de bancos de dados textuais, raciocínio baseado em casos e representação de conhecimento contextualizada dinamicamente, com uma leve predominância desta última, principalmente no que diz respeito à performance do sistema, o Olimpo materializa a PCE anteriormente de finida.

Prosseguindo com o comparativo, deve-se analisar, mais detalhadamente, o comportamento dos sistemas. Antes, porém, deve ser colocado que estas análises estão restritas à prototipagem dos sistemas cujo desenvolvimento contou com a participação do autor, a qual deu-se de acordo com os registros contidos na tabela 3:

Sistema→ ↓Participação	Digesto	Metajuris	Themis	P1	Prudentia	Jurisconsulto	Olimpo	Habeas corpus na madrugada
Idealizou	X	X	X	-	-	-	X	X
Participou da idealização	-	-	-	X	X	X	-	-
Modelou	X	X	X	-	-	-	-	X
Participou da modelagem	-	-	-	X	X	X	X	-
Implementou	-	X*	-	-	-	-	-	X*
Participou da equipe de implementação	X	-	X	X	X	X	X	

Tabela 3

(*) Primeira versão

Analisando, agora, os sistemas, de acordo com critérios de modelagem e performance, é possível confrontar os protótipos, levando em consideração as tarefas da TIJ descritas na figura 3. Agrupando -se os sistemas enquadráveis na primeira e terceira tarefas, o resultado é o seguinte, contido na tabela 4:

Sistema→ ↓Característica	Digesto	Metajuris	Habeas corpus na madrugada	Themis	P1	Prudentia	Juriconsulto	Olimpo
Representação de conhecimento na base	Baixa	Não tem	Não tem	Alta	Alta	Média	Média	Média
Representação de conhecimento na interface	Não tem	Não tem	Alta	Média	Baixa	Média	Média	Média
Técnica	Recuperação textual	Metabusca textual	Frames com conhecimento	RBC/ Recuperação textual	RBC	RBC	RBC/ Recuperação textual	RBC/ Recuperação textual/ RD2C
Ambiente	Web	Web	Web	Desktop	Desktop	Desktop	Desktop	Desktop
Ferramenta de implementação	Cgi/Perl Javascript	Cgi/Perl	Html/ Javascript	Delphi	Esteem	Delphi	Delphi	Delphi
Comportamento inteligente	Não	Não	Não	Sim	Sim	Sim	Sim	Sim

Tabela 4

Este comparativo, orientado por aspectos que permitem uma distinção de acordo com o grau de complexidade de cada um deles, termina por apontar aqueles que possuem comportamento inteligente, eis que selecionam e hierarquizam registros, de forma comprovada e repetível, com base em conhecimento representado, entre outras características. São os seguintes (ordem cronológica):

- **Themis;**
- **P1;**
- **Prudentia;**
- **Jurisconsulto;**
- **Olimpo**

Estes sistemas, todos prototipados, derivaram da mesma idéia central, aquela contida no sistema Themis, qual seja, recuperar documentos provenientes de decisões judiciais utilizando representação de conhecimento. Assim, todos possuem algumas características em comum. Porém, possuem, também, diferenças entre si, bastante significativas. Os critérios utilizados para a realização de um comparativo entre os sistemas são os seguintes:

- **Forma de inclusão de registros**
- **Fórmula centralizada de similaridade**
- **Média de expressões indicativas contidas nos casos:**
- **Hierarquia entre as expressões indicativas**
- **Pré-consulta**
- **Entrada textual**
- **Nível de profundidade da consulta**
- **Acesso direto aos documentos**
- **Ferramenta de implementação**
- **Base de documentos**
- **Número de casos na base**
- **Interface**
- **Idioma**
- **Ajuste dinâmico de expressões**

- Natureza dos casos
- Comparação textual com alto número de palavras
- Combinação de técnicas diferenciadas
- Processo plúrimo na formação das listas
- Pesos dinâmicos para atributos

Comparando estas diferenças, chegamos ao quadro que segue (tabela 5):

Sistema→ Característica↓	Themis	PI	Prudentia	Jurisconsulto	Olimpo
Inclusão de registros	Manual	Manual	Automática	Automática	Automática
Média de expressões indicativas por caso	Alta	Baixa	Baixa	Baixa	Alta
Hierarquia entre as expressões indicativas	Sim	Sim	Não	Não	Não
Pré-consulta	Não	Não	Não	Não(*)	Sim
Entrada textual	Aceita	Não aceita	Não aceita	Aceita	Aceita
Nível de profundidade	Não	Não	Não	Não	Sim
Acesso direto aos documentos	Sim	Não	Não	Não	Sim
Ferramenta de implementação	Delphi	Esteem	Delphi	Delphi Smalltalk	Delphi
Base de documentos selecionada	Sim	Sim	Não	Não	Sim
Número de casos na base	37	25	300	2.400	100
Interface	Operacional	Pré-operacional	Operacional	Operacional	Pós-operacional
Idioma	Português	Português	Português	Português	Inglês
Ajuste dinâmico da lista de expressões	Média	Alta	Baixa	Baixa	Alta
Natureza dos casos	Súmulas selecionadas	Habeas corpus do TJSC	Apelações criminais e Habeas Corpus do TJSC	Apelações criminais e Habeas Corpus do TJSC	Resoluções do Conselho de Segurança da ONU
Comparação textual	Não	Não	Não	Sim	Sim
Combina diferentes técnicas	Sim	Não	Não	Sim	Sim
Processo plúrimo na formação das listas	Não	Não	Não	Não	Sim
Pesos dinâmicos	Sim	Não	Não	Não	Não

nos atributos				
---------------	--	--	--	--

Tabela 5

(*) A nova versão já conta com estes recursos

Após a análise dos comparativos, chega-se à conclusão de que os Sistemas Jurisconsulto e Olimpo aparentemente possuem muita semelhança, o que não se flexibiliza na última análise, a da tabela “5”, onde se percebe diversas diferenças entre ambos.

Inobstante, estes sistemas são responsáveis por uma valiosa inovação em termos de recuperação de informações textuais, que é a combinação de técnicas de recuperação textual pura com técnicas oriundas do Raciocínio Baseado em Casos, que, para este tipo de informação, materializam uma forma muito eficiente de recuperação, qual seja, a possibilidade de uma entrada baseada em um grande texto, ao invés de algumas palavras. Tomando-se como base a atual importância das ferramentas de busca em todas as partes do mundo, e levando-se em consideração que em sua grande maioria estas trabalham com pequenos números de palavras como formas de entrada, sabemos da relevância e importância desta constatação.

É importante salientar que as técnicas puramente textuais esbarram nas tradicionais limitações dos mecanismos de banco de dados. Por outro lado, as técnicas de RBC puro também esbarram em dificuldades, principalmente no que diz respeito às limitações impostas pela modelagem a recuperação através de atributos. O somatório das técnicas gerou uma característica inesperada e não planejada inicialmente, trata-se de uma peculiaridade descoberta quase que acidentalmente, pelo autor, dentro do grupo de pesquisa.

A) Olimpo X Jurisconsulto

Visto que estes sistemas apresentam significativa proximidade técnica, embora difiram em detalhes operacionais, é importante evidenciar as principais e mais significativas diferenças operadas entre ambos, que são as seguintes:

- **Lista de expressões indicativas.** Esta é a principal diferença entre eles. Não exatamente a lista em si, mas a forma de sua constituição e validação. No Jurisconsulto, a lista de expressões indicativas é formada por listas já consolidadas, provenientes de periódicos especializados. No caso do Olimpo, a lista foi concebida especial e personalizadas, conforme será visto na Parte II. O processo de consolidação da estrutura retórica do Sistema Olimpo é dinâmico e plúrimo, com constante validação das expressões. Outro aspecto significativo é a análise estatística das expressões constantes da lista, característica peculiar e exclusiva do Olimpo. Mais um aspecto relevante é o trabalho manual de ajuste de expressões. A grande diferença é uma lista pensada e testada.
- **Incidência de expressões indicativas nos casos.** O número médio de expressões indicativas encontrado nos casos incorporados à base do Jurisconsulto fica entre cinco e dez expressões por caso, dificilmente sendo superada a primeira dezena. No Olimpo, existem casos que chegam a ter uma centena de expressões indicativas. Tal fato otimiza a performance do sistema e intensifica incrivelmente sua qualidade, o que se reflete diretamente no índice de satisfatoriedade e dos resultados. Esta característica é uma consequência direta da anterior.
- **Organização metodológica.** O Olimpo passou por um processo de lapidação mais efetivo, teve seu projeto modelado previamente e uma implementação planejada, o que, entre outros aspectos, gerou expressivo ganho de tempo na execução do projeto. Mas isso, por si só, não caracteriza diferenciação significativa. O aspecto mais importante é que o Jurisconsulto apresenta grandes

potencialidades, mas, tendo em vista que sua organização metódica não é tão elevada quanto a do Olimpo, seu desempenho geral é irregular.

- **Tipificação.** Significa o enquadramento legal do caso. Este item é exclusivo do Jurisconsulto, tendo em vista que o direito penal, sua área de aplicação, facilita a tarefa, pois os casos estão previamente descritos em lei. Ela pode funcionar de forma semelhante a um filtro, ou ser incluída na métrica de similaridade. Porém, a representação de conhecimento necessária para a definição de suas regras de extração automática materializaram tarefa de elevada complexidade, fato que, sem dúvida, qualifica o sistema.
- **Idioma.** O Jurisconsulto foi feito em português (base e interface) e o Olimpo em Inglês;
- **Base selecionada.** Os documentos integrantes da base de casos do sistema Jurisconsulto foram escolhidos aleatoriamente, de acordo com a disponibilidade dos mesmos, sem aplicação de critério estatísticos. A base de resoluções do sistema Olimpo foi objeto de um preliminar planejamento de coleta, segundo o qual foram definidos, antecipadamente, os anos que integrariam a amostragem, bem como o número total de casos por ano.
- **Métrica de similaridade:** A métrica de similaridade do Jurisconsulto é mais sofisticada, e é apurada em dois níveis: *local* e *global*. *Local* é a similaridade operada entre os atributos estáticos (características do caso), comparado -se um a um. *Global* é soma desta com a similaridade operada levando -se em consideração as expressões indicativas encontradas. Ocorre que o grupo percebeu que poderia potencializar a métrica operada entre as expressões indicativas, dando-lhe maior ênfase, tendo em vista a maior densidade da estrutura retórica construída. Este fato permite que o Olimpo realize comparações documentais sem a necessidade de preenchimento de atributos estáticos. Nada impede que novas evoluções de ambos os sistemas, filhos da mesma tecnologia, incorporem características um do outro, mas é importante

ênfatizar que o Olimpo foi concebido com o objetivo de expandir a técnica de comparação entre as expressões indicativas, e com uma métrica de similaridade menos sofisticada, ele apresenta melhor eficiência em tarefas de recuperação, conclusão que será amplamente aproveitada em outros trabalhos da equipe.

- **Tipo de documento.** O Jurisconsulto possui uma base povoada por acórdãos do Tribunal de Justiça do Estado de Santa Catarina, enquanto o Olimpo opera com as resoluções do Conselho de Segurança da ONU. Estes documentos, entre si, possuem o mesmo tipo de formato, isto é, textual, mas possuem estruturas de representação retórica bastante diferenciadas.

Em seguida, passaremos ao capítulo III, com a descrição do processo de consolidação do Sistema Olimpo.

CAPÍTULO III

APLICAÇÃO: SISTEMA OLIMPO

INTRODUÇÃO

Nesta Segunda parte, será feita a identificação do universo escolhido para a aplicação das técnicas relatadas, com respectivo processo de captura dos documentos e estruturação das bases de dados, bem como o processo de consolidação e funcionamento do sistema. Posteriormente, analisaremos alguns dados estatísticos e a sua performance, com respectivos testes comparativos, principalmente com os sistemas tradicionais de banco de dados.

1. IDENTIFICAÇÃO DO UNIVERSO

A Organização das Nações Unidas possui seis órgãos especiais, de acordo com o artigo 7-1 de sua Carta. Todos eles, apresentados na figura 36, são emissores de documentos relevantes e cuja adequada recuperação se reveste de destacada importância.

Dentre eles, o Conselho de Segurança foi o escolhido como campo de aplicação do Sistema Olimpo, tendo em vista algumas características do *órgão* e das suas *resoluções*.

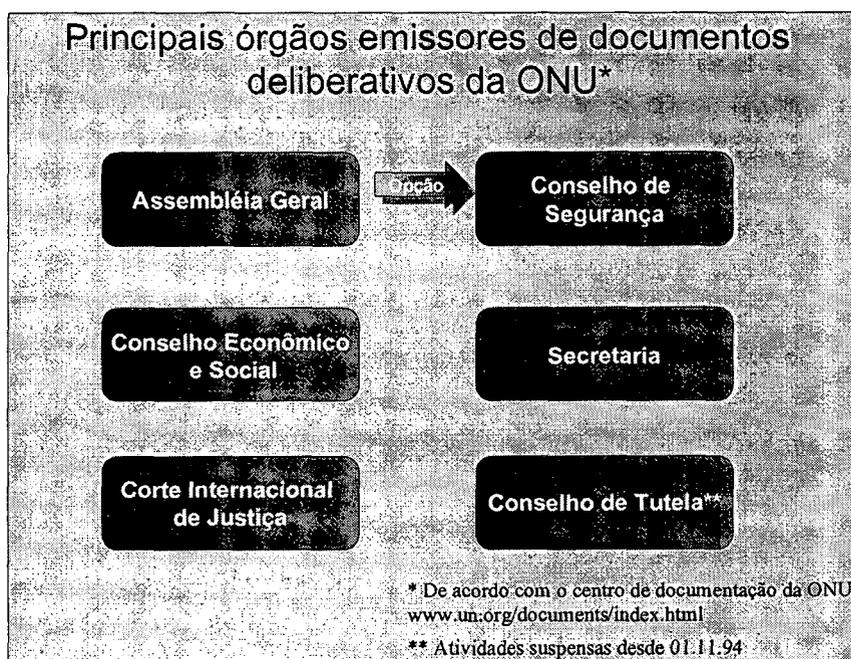


Figura 36

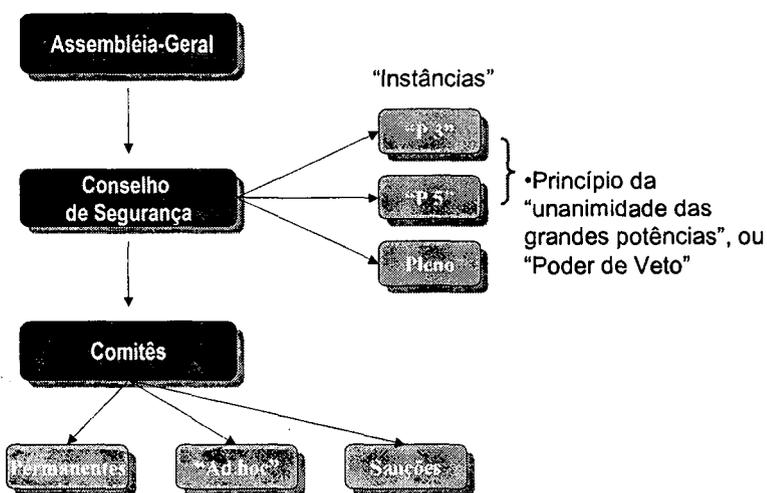
Quanto ao *órgão*, o Conselho de Segurança é descrito na Carta das Nações Unidas, inicialmente no artigo 7-1, onde é referenciado como um dos órgãos especiais das Nações Unidas. Recebe tratamento específico no Capítulo V da Carta, nos artigos 23 a 32. Conforme o artigo 24-1, sua função central é assumir “a principal responsabilidade na manutenção da paz e da segurança internacionais”.

É de se ressaltar que o CS possui um perfil jurídico e executivo. É jurídico porque detém o monopólio da violência legítima no âmbito internacional e julga a existência de fatos, aplica sanções sobre eles e determina quem será o executor de tais sanções, conforme Kelsen (*apud* Steinfus[85]). Tais características fazem dele um órgão jurídico. Este perfil enseja uma boa aplicação da “tecnologia da informação jurídica”, em especial da PCE e de seu particular método de análise da estrutura retórica de um dado contexto jurisdicional, baseada na estrutura de conhecimento que envolve o órgão, a qual, quando adequadamente modelada, maximiza a tarefa de recuperação inteligente de documentos. Tendo em vista que

possui feições políticas e que seu poder para constatar violações é discricionário, conforme Seitenfus [85], o CS também possui natureza executiva, o que faz dele um órgão jurídico-executivo.

O CS possui características e peculiaridades. Uma delas é a de ser, na atualidade, o órgão colegiado jurisdicional mais poderoso do planeta. Outra é a existência de instâncias internas informais, chamadas de “P 3” (Países Ocidentais permanentes) e “P 5” (todos os Países permanentes), conforme Steitenfus [85]. Outra é a existência de órgãos internos com delegação de poderes específicos para determinadas tarefas, podendo ser permanentes e “ad hoc”, assim como o comitê de sanções, conforme demonstrado na figura 37.

Posição topográfica do Conselho de Segurança



*Fonte: www.un.org

Figura 37

Dentre os tipos de documentos emitidos pelo conselho de segurança, seis possuem maior destaque, apresentados na figura 38. Dentre eles, as *Resoluções* foram escolhidas tendo em vista sua relevância e sua estrutura.

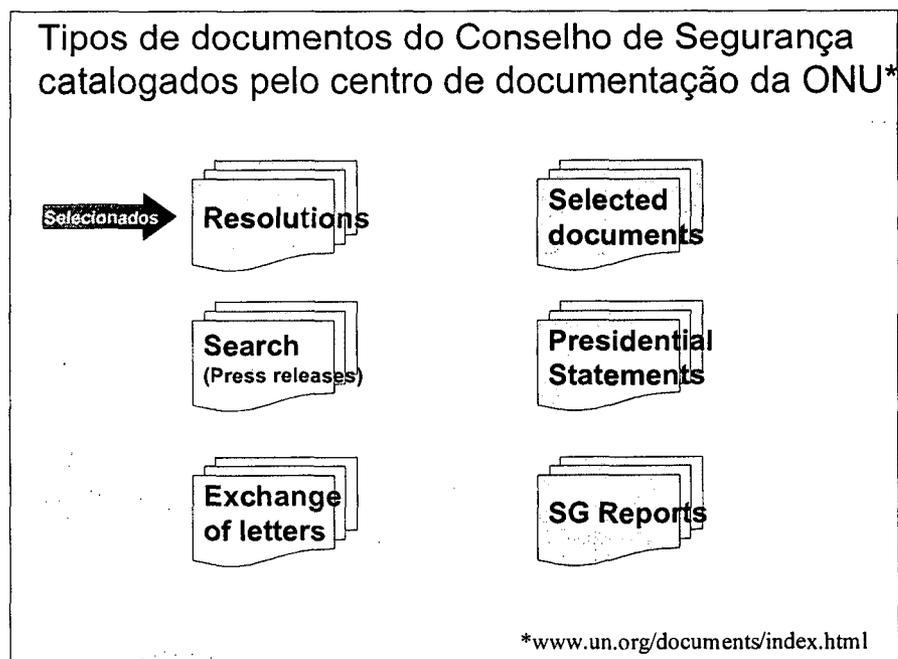


Figura 38

Quanto à estrutura do documento, as resoluções possuem características que facilitam a aplicação da tecnologia aqui referenciada, quais sejam: 1. formatação textual homogênea; 2. estrutura retórica padronizada; 3. esteriótipo formal bastante definido.

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.

Dentro do site oficial, existe um seção específica sobre os documentos, denominada "Centro de Documentação", que mantém documentos atualizados dos órgão especiais da ONU, entre eles o Conselho de Segurança (figura 39).

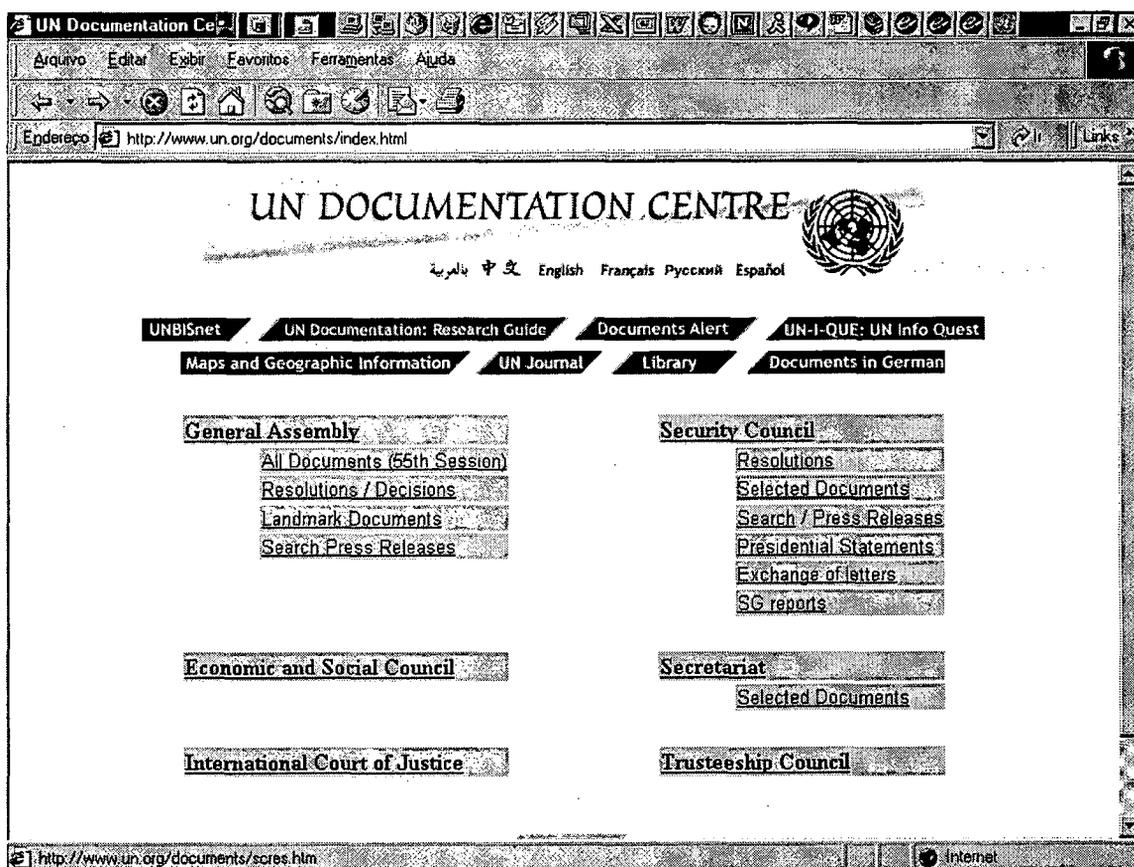
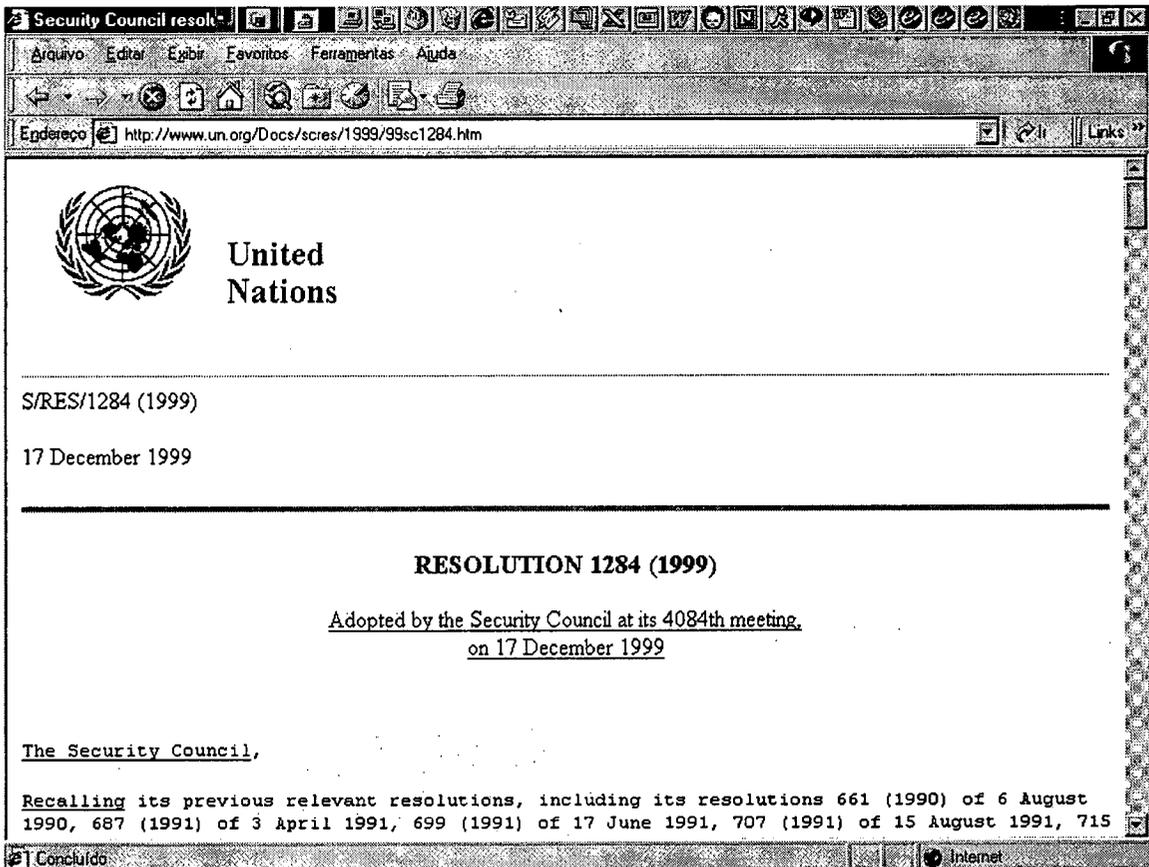


Figura 39

Uma interessante constatação é a de que o CS possui a maior variedade documental dentre os órgão ali elencados, o que ratifica a sua escolha confirma a relevância de se estruturar as informações dele provenientes. U ma vez definido isso, teve início o processo de captura dos documentos, via web, para a formação da base de conhecimento do sistema, conforme figura 40.



The image is a screenshot of a web browser window. The title bar reads "Security Council resol...". The menu bar includes "Archivo", "Editar", "Exibir", "Favoritos", "Ferramentas", and "Ajuda". The address bar shows the URL "http://www.un.org/Docs/scres/1999/99sc1284.htm". The main content area features the United Nations logo and the text "United Nations". Below this, it lists "S/RES/1284 (1999)" and "17 December 1999". A horizontal line separates this header from the main text, which begins with "RESOLUTION 1284 (1999)" in bold. This is followed by "Adopted by the Security Council at its 4084th meeting, on 17 December 1999". The text then reads "The Security Council," and "Recalling its previous relevant resolutions, including its resolutions 661 (1990) of 6 August 1990, 687 (1991) of 3 April 1991, 699 (1991) of 17 June 1991, 707 (1991) of 15 August 1991, 715". The browser's status bar at the bottom shows "Concluido" and "Internet".

Security Council resol...

Archivo Editar Exibir Favoritos Ferramentas Ajuda

Endereço http://www.un.org/Docs/scres/1999/99sc1284.htm Links

 United Nations

S/RES/1284 (1999)

17 December 1999

RESOLUTION 1284 (1999)

Adopted by the Security Council at its 4084th meeting,
on 17 December 1999

The Security Council,

Recalling its previous relevant resolutions, including its resolutions 661 (1990) of 6 August 1990, 687 (1991) of 3 April 1991, 699 (1991) of 17 June 1991, 707 (1991) of 15 August 1991, 715

Concluido Internet

Figura 40

2. ESTRUTURA DE UM CASO

O caso escolhido para ser apresentado é a Resolução n.º 1244, de 1999⁹, tendo em vista que ela é a resolução que apresentou o maior número de ocorrências de expressões indicativas, 137 no total.

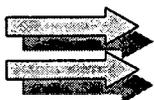
Transcreve-se parte de seu texto:

RESOLUÇÃO 1244/99



Security Council resolution 1244 (1999) on situation relating to Kosovo

United Nations



S/RES/1244 (1999)

10 June 1999

RESOLUTION 1244 (1999)

Adopted by the Security Council at its 4011th meeting, on 10 June 1999



The Security Council,
 Bearing in mind the purposes and principles of the Charter of the United Nations, and the primary responsibility of the Security Council for the maintenance of international peace and security,
 Recalling its resolutions 1160 (1998) of 31 March 1998, 1199 (1998) of 23 September 1998, 1203 (1998) of 24 October 1998 and 1239 (1999) of 14 May 1999,
 Regretting that there has not been full compliance with the requirements of these resolutions,
 Determined to resolve the grave humanitarian situation in Kosovo, Federal Republic of Yugoslavia, and to provide for the safe and free return of all refugees and displaced persons to their homes,
 Condemning all acts of violence against the Kosovo population as well as all terrorist acts by any party,
 Recalling the statement made by the Secretary-General on 9 April 1999, expressing concern at the humanitarian tragedy taking place in Kosovo,
 Reaffirming the right of all refugees and displaced persons to return to their homes in safety,

⁹ Somente uma parte dela.

Recalling the jurisdiction and the mandate of the International Tribunal for the Former Yugoslavia,
 Welcoming the general principles on a political solution to the Kosovo crisis adopted on 6 May 1999 (S/1999/516, annex 1 to this resolution) and welcoming also the acceptance by the Federal Republic of Yugoslavia of the principles set forth in points 1 to 9 of the paper presented in Belgrade on 2 June 1999 (S/1999/649, annex 2 to this resolution), and the Federal Republic of Yugoslavia's agreement to that paper,
 Reaffirming the commitment of all Member States to the sovereignty and territorial integrity of the Federal Republic of Yugoslavia and the other States of the region, as set out in the Helsinki Final Act and annexes 2,
 Reaffirming the call in previous resolutions for substantial autonomy and meaningful self-administration for Kosovo,
 Determining that the situation in the region continues to constitute a threat to international peace and security,
 Determined to ensure the safety and security of international personnel and the implementation by all concerned of their responsibilities under the present resolution, and acting for these purposes under Chapter VII of the Charter of the United Nations,
 1. Decides that a political solution to the Kosovo crisis shall be based on the general principles in annex 1 and as further elaborated in the principles and other required elements in annex 2;
 2. Welcomes the acceptance by the Federal Republic of Yugoslavia of the principles and other required elements referred to in paragraph 1 above, and demands the full cooperation of the Federal Republic of Yugoslavia in their rapid implementation;

...

(.....)

...

10. Suspension of military activity will require acceptance of the principles set forth above in addition to agreement to other, previously identified, required elements, which are specified in the footnote below.¹ A military-technical agreement will then be rapidly concluded that would, among other things, specify additional modalities, including the roles and functions of Yugoslav/Serb personnel in Kosovo:

Withdrawal

- Procedures for withdrawals, including the phased, detailed schedule and delineation of a buffer area in Serbia beyond which forces will be withdrawn;

Returning personnel

- Equipment associated with returning personnel;
 - Terms of reference for their functional responsibilities;
 - Timetable for their return;
 - Delineation of their geographical areas of operation;
 - Rules governing their relationship to the international security presence and the international civil mission.

Notes

¹ Other required elements:

- A rapid and precise timetable for withdrawals, meaning, e.g., seven days to complete withdrawal and air defense weapons withdrawn outside a 25 kilometer mutual safety zone within 48 hours;
 - Return of personnel for the four functions specified above will be under the supervision of the international security presence and will be limited to a small agreed number (hundreds, not thousands);
 - Suspension of military activity will occur after the beginning of verifiable withdrawals;
 - The discussion and achievement of a military-technical agreement shall not extend the previously determined time for completion of withdrawals.

-----"

As setas de destaque indicam, respectivamente, os itens “assunto”, “número da resolução”, “data” e o início do texto, de onde são extraídas as expressões indicativas.

3. O PROCESSO DE FORMAÇÃO E FUNCIONAMENTO DO SISTEMA

O Sistema Olimpo tem a sua performance centrada na combinação de aspectos derivados do RBC e da recuperação de informações textuais, somados a uma adequada organização do conhecimento referente ao assunto em foco (no caso específico, as resoluções do Conselho de Segurança da ONU). Esta organização é que enseja a RC²D, a qual se constitui na principal inovação detectada pelo grupo de pesquisa e pelo autor, de forma tal a que propicie outro importante aspecto inovador, a possibilidade de comparação de documentos, e não simples palavras ou atributos.

De um modo geral, o funcionamento do sistema apresenta um fluxo similar a outros sistemas baseados em conhecimento, descrito na figura 41, segundo o qual uma entrada manual sofre um processo de ajuste e, após, é submetida à base de documentos, dentre os quais são escolhidos os melhores.

Este é o processo básico de consulta realizado pelo sistema. A pesquisa é considerada “contextual” e “estruturada” pelas seguintes razões: 1. É levada 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.

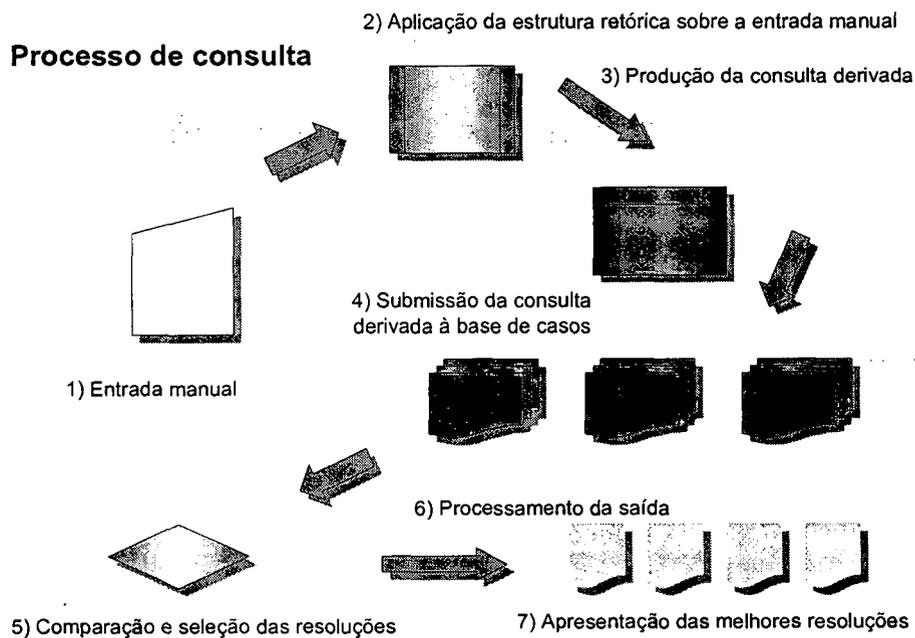


Figura 41

O processo de filtragem, descrito na figura 42, reduz o espaço de análise do sistema, a atua de forma tal a que não seja um simples particionamento da base de dados, pois trata de questões de “mérito”, e não de “forma”, ensejando, inclusive, uma visualização preliminar do universo da pesquisa, antes mesmo de a consulta ser deflagrada.

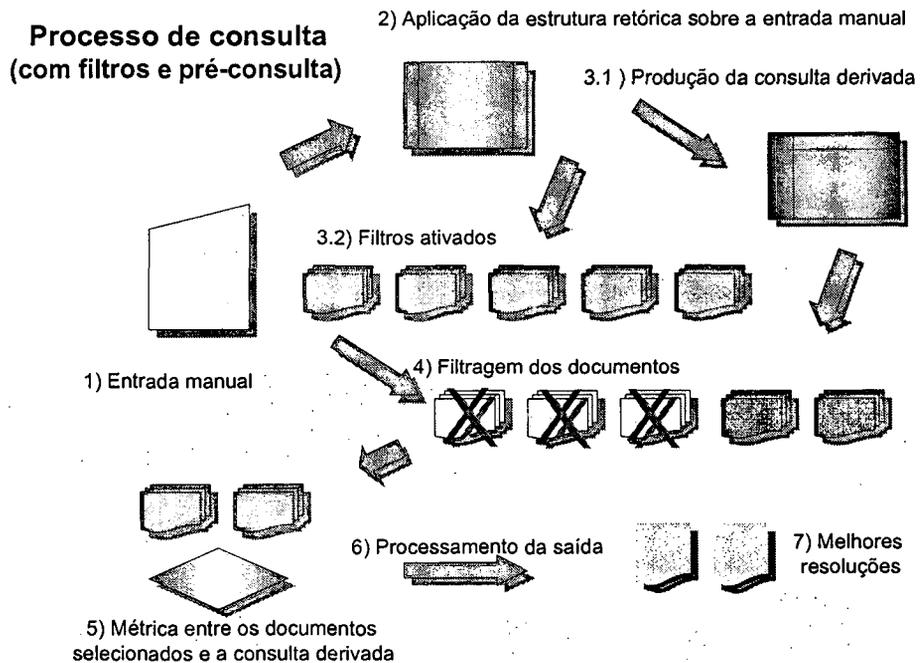


Figura 42

Além disso, o controle do nível de profundidade da pesquisa (“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 43, 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.

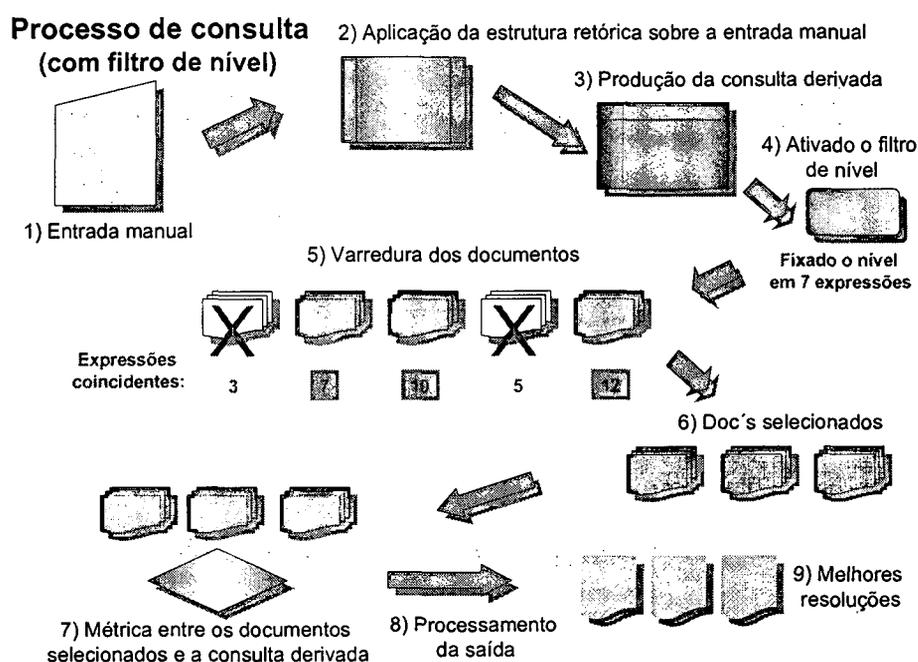


Figura 43

A) 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, e 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 BD web da UNO 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”), tem pouca utilidade para diferenciá-los. 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 possuía significância na comparação documental.

Este processo assumiu contornos dinâmicos, pois foi feito e refeito diversas vezes, e as expressões iam sendo incluídas ou excluídas, conforme seu desempenho estatístico. Este procedimento foi rotinizado, com a seguinte seqüência:

1. Uma equipe fez a leitura das resoluções e extraiu as 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 desempenho satisfatório, formando a “pré-lista 2”;
2. Somadas as pré-listas, chegou-se à primeira lista provisória;
3. A lista provisória foi “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 dos textos das resoluções;
 - 3.2 em ambas as atividades eram registradas as novas expressões detectadas, para que formassem as listas adicionais “1” e “2”;
4. Estas listas adicionais somavam -se à lista provisória, gerando a pré-lista, que era novamente submetida ao passo “3”;
 5. Após algumas repetições de todo o procedimento, chegou -se à lista final.

Este procedimento, representado na figura 44, 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 - RC²D, figura inédita dentro dos trabalhos da equipe de pesquisa de IA e direito do PPGEP.



Figura 44

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.

B) Formação dos registros e cálculo da similaridade

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), conforme a tabela 6.

Index	Description	Importance of retrieval
Subject	Short description about the most important situation discussed in the resolution	Used to know the resolution subject
Date	Year of resolution Decisions	Determine a year specific of interest
Resolution Number	States the number supplied by de UN Security Council	Used to identify the resolution
Meeting	States de number of the meeting where the resolution is decided	Can be used to retrieve the resolution in reference of the meeting
Contries	States the contries involved in the question	Used to know the contries involved in the conflict
Decisions	Number of diferents matters of decisions in the resolution	Can be used to understand the extension and complexity of the resolution
Acronyms	Acronyms of official and unoffical institutions	Can be used to retrieve the institutions related in the resolution. Ex: OTAN
Indicative Expressions	Determines the most relevant information in the resolution and its redefinitions	Used to retrieval the resolutions based in the similar topics

Tabela 6

Como exemplo de um caso concreto, após a extração destas informações, apresentamos a já referida Resolução 1.244, a qual registrou a ocorrência de 137 expressões indicativas no bojo de seu texto:

79 ° Arquivo - 1244.txt

Assunto: Security Council resolution 1244
(1999) on situation relating to Kosovo
Data: 10 June 1999
Resolução: 1244
Encontro: 4011

Palavras:

1° - access to - 3
2° - acting in - 1
3° - acts of - 1
4° - acts of violence - 1
5° - administrative - 2
6° - affected by - 1
7° - Against the - 1
8° - agreement - 8
9° - aid organizations - 3
10° - air defense - 1
11° - arms - 1
12° - authorized to - 1
13° - authorizes the - 1
14° - Belgrade - 1
15° - Charter of the United Nations - 2
16° - civil mission - 2

17° - clearing minefields - 1
18° - Commissioner for Refugees - 1
19° - commitment of - 1
20° - complete withdrawal - 1
21° - compliance - 1
22° - conclusion of the meeting - 1
23° - constitute a threat - 1
24° - consultation - 2
25° - cooperate fully - 2
26° - cooperation - 4
27° - crisis - 7
28° - crisis region - 2
29° - delivery - 1
30° - demilitarization - 3
31° - democracy - 2
32° - deployment of - 3
33° - development - 6
34° - displaced person - 8
35° - displaced persons - 8
36° - economic development - 3
37° - economic prosperity - 2
38° - economic reconstruction - 1
39° - effective delivery - 1
40° - election - 1
41° - elections - 1
42° - end to - 1
43° - end to violence - 1

- 44° - ensuring the withdrawal - 1
 45° - establish - 17
 46° - federal republic of Yugoslavia - 13
 47° - first reports - 1
 48° - foreign - 2
 49° - foreign minister - 2
 50° - Former Yugoslavia - 2
 51° - framework - 2
 52° - framework agreement - 2
 53° - freedom of movement - 1
 54° - full compliance - 1
 55° - full cooperation - 2
 56° - G-8 - 2
 57° - G-8 Foreign Ministers - 2
 58° - government - 4
 59° - high commissioner - 1
 60° - hostilities - 1
 61° - human rights - 1
 62° - humanitarian aid - 4
 63° - humanitarian situation - 1
 64° - Implementation of - 5
 65° - imposed by - 1
 66° - inhabitants - 3
 67° - interim political - 2
 68° - international civil - 18
 69° - international peace - 2
 70° - international personnel - 1
 71° - International Tribunal - 2
 72° - kosovo crisis - 5
 73° - law and order - 1
 74° - leaders - 1
 75° - Leadership - 1
 76° - maintenance of - 1
 77° - military - 13
 78° - military forces - 4
 79° - military-technical agreement - 2
 80° - mine - 5
 81° - minefields - 1
 82° - Movement of - 1
 83° - national police - 1
 84° - normal life - 3
 85° - North Atlantic Treaty Organization - 1
 86° - offensive - 1
 87° - Office - 1
 88° - Office of the United Nations - 1
 89° - Pact - 2
 90° - paramilitary - 4
 91° - peace and security - 2
 92° - police forces - 1
 93° - political settlement - 2
 94° - political solution - 3
 95° - prohibition - 1
 96° - protection - 1
 97° - reconstruction - 3
 98° - renewed hostilities - 1
 99° - report to the Council - 1
 100° - requirements for demilitarization - 1
 101° - return of refugees - 1
 102° - right of all refugees - 1
 103° - safe return - 2
 104° - safety and security - 1
 105° - Secretary-General - 6
 106° - security presence - 17
 107° - security presences - 7
 108° - self-government - 4
 109° - Serb - 5
 110° - settlement - 4
 111° - Sovereignty - 3
 112° - sovereignty and territorial integrity - 3
 113° - special representative - 3
 114° - Stability - 4
 115° - Stability Pact - 2
 116° - supporting the consolidation - 1
 117° - territorial integrity - 3
 118° - terrorist acts - 1
 119° - the protection - 1
 120° - the sovereignty - 1
 121° - threat to international peace - 1
 122° - timetable - 4
 123° - transfer of authority - 1
 124° - transitional - 3
 125° - transitional administration - 3
 126° - under chapter - 2
 127° - unimpeded access - 3
 128° - unit - 8
 129° - United Nations High Commissioner - 1
 130° - use of - 1
 131° - Verifiable - 6
 132° - violence - 4
 133° - weapons - 1
 134° - weapons withdrawn - 1
 135° - withdrawal - 12
 136° - withdrawal from - 3
 137° - withdrawal from Kosovo - 3
- Signas:
 1° - Unit - 8
- Países:
 1° - Albania
 2° - Yugoslavia
- Total de decides: 6
- Resultado:
- ... x 2;
 5. Decides on the deployment in Kosovo, under United Nations auspices, of international civil and security presences, with appropriate equipment and personnel as required, a ...
- ... anner;
 7. Authorizes Member States and relevant international organizations to establish the international security presence in Kosovo as set out in point 4 of annex 2
- *****-----

Dentre estes oito atributos valorados automaticamente, três operam, efetivamente, a métrica de similaridade:

- Expressões indicativas;
- Países;
- Siglas.

Três funcionam com filtro:

- Países;
- Data (ano);
- Siglas.

Dois funcionam como acesso direto aos documentos:

- Número da resolução;
- Número do encontro.

O assunto atua como resumo do documento apresentado no resultado da consulta. O número de “decides”, embora já coletado, não está ativo na presente versão, devendo ser aproveitado futuramente.

Neste ponto, pode surgir uma interessante dúvida: como as siglas e países podem, ao mesmo tempo, atuar como filtro e integrar a métrica de similaridade ?

Resposta: tais itens são independentes. A métrica de similaridade, operada sobre todas expressões indicativas encontradas, considera países e siglas em igualdade de condições com outras expressões (sem pesos).

A métrica foi estruturada para considerar as expressões indicativas presentes no caso e na consulta, após a aplicação da estrutura retórica sobre a entrada textual e respectiva produção da consulta derivada, conforme passos “2” e “3” da figura 43.

Ou seja, é a consulta derivada, na realidade, que serve de referencial para a elaboração da métrica de similaridade.

Exemplificando: em um determinado caso, a aplicação da estrutura retórica sobre um dado texto resulta na produção de uma consulta derivada com um total de 50 expressões. Este conjunto de expressões é comparado com os registros, e o percentual de similaridade é estabelecido em função do número de expressões encontradas dentro dos registros individualizados. Se forem encontradas 43 expressões, a similaridade é de 86%, ou 72% para 36 expressões, 56% para 28 expressões, e assim sucessivamente. Este tipo de métrica é bastante singela, uma das mais simples que poderia ser aplicada nesta situação, mas funciona de forma bastante estável, e comporta, no futuro, incrementos bastante factíveis, quando da incorporação de novos mecanismos ao sistema, como, por exemplo, a aplicação dos trigramas, ou a atribuição de pesos internos para as palavras mais frequentes no texto das resoluções (dado já disponível).

Ocorre que, na realidade, não é a métrica que representa o grande ponto forte do sistema, mas sim a organização das expressões indicativas, de forma a fazerem com que a métrica tenha uma performance incrementada.

Uma situação bastante exemplificativa sobre esta peculiaridade do sistema é uma expressão com duas palavras, por exemplo, “United Nations”. Uma aplicação simples de similaridade por palavras isoladas apresentará um índice de 100% quando ambas as palavras estiverem presentes dentro de um texto (não importando a sua posição), ou 50% se apenas uma delas for encontrada. Porém, se utilizarmos uma indexação diferenciada, fazendo com que “United Nations” (palavras juntas e exatas) sejam uma expressão, “United” seja outra, e “Nations” seja uma terceira (todas com o mesmo peso, por enquanto), teremos uma situação diferenciada. Nesta situação, não basta ao sistema que as palavras, mesmo que dissociadas, estejam presentes no texto, elas precisam estar juntas, com o sentido exato. De acordo com este critério de representação, quando o sistema encontra apenas uma palavra, o índice de similaridade é de 33,33%, quando encontra as duas, dissociadas, ele passa a ser de 66,66%, e somente vai atingir 100% quando ambas estiverem presentes, porém juntas.

Um caso concreto de uma expressão que apresenta exatamente este comportamento é “stability pact”.

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. ESTATÍSTICAS 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.

O resultado final, caso a caso, é o contido na seguinte tabela 7:

Ordem	Número da Resolução	Ocorrências
1	1036	74
2	1039	28
3	1042	30
4	1045	82
5	1048	57
6	1051	40
7	1054	49
8	1057	11
9	1060	38
10	1063	70
11	970	44
12	973	28
13	977	9
14	981	74
15	984	26
16	988	47
17	992	35
18	995	17
19	1066	46
20	1069	13
22	1072	106

22	1075	124
23	1078	98
24	1081	11
25	1084	50
26	1087	119
27	1090	4
28	1092	96
29	1144	56
30	1142	31
31	1139	11
32	1136	57
33	1133	30
34	1130	9
35	1127	83
36	1124	111
37	1121	12
38	1118	66
39	1115	42
40	1112	13
41	1109	11
42	1106	36
43	1103	27
44	1100	50

45	1097	54
46	1094	45
47	1093	79
48	1145	86
49	1218	31
50	1215	34
51	1211	11
52	1207	39
53	1203	100
54	1199	112
55	1195	61
56	1191	5
57	1187	99
58	1183	62
59	1174	111
60	1171	34
61	1179	26
62	1164	65
63	1167	64
64	1159	97
65	1151	30
66	1155	49
67	1219	21
68	1147	64
69	1272	92
70	1276	11
71	1269	37
72	1279	88
73	1265	123
74	1262	30
75	1258	81

76	1255	83
77	1250	28
78	1247	113
79	1244	137
80	1241	5
81	1230	87
82	1237	86
83	1234	120
84	1227	36
85	1223	30
86	1220	23
87	1282	23
88	1284	106
89	1000	67
90	1004	45
91	1007	78
92	1009	58
93	1014	76
94	1017	40
95	1022	69
96	1026	36
97	1029	81
98	1032	65
99	1034	86
100	1035	28

Tabela 7

A média de expressões indicativas por resolução é de 55,18 expressões por resolução, 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 (tabela 8):

Ordem	Número da Resolução	N. de Ocorrências
1	1244	137
2	1075	124
3	1265	123
4	1234	120
5	1087	119
6	1247	113
7	1199	112
8	1124	111
9	1174	111
10	1072	106
11	1284	106
12	1203	100

Tabela 8

Estes dados podem ser representados graficamente, conforme o gráfico 7:

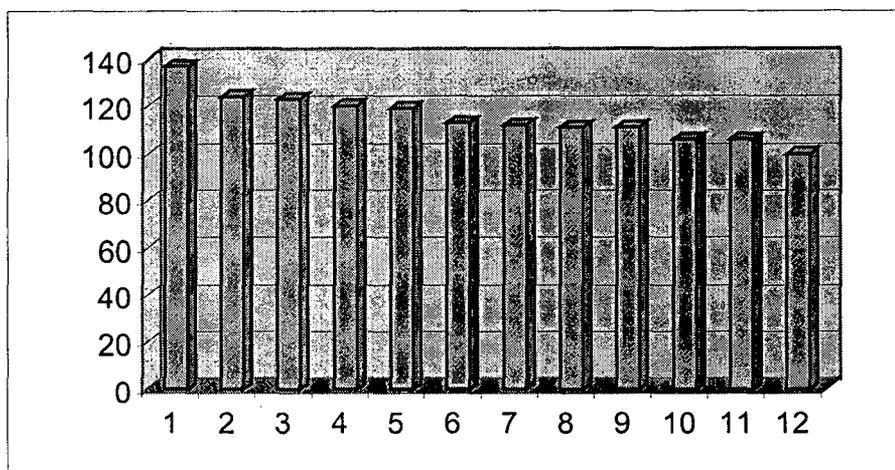


Gráfico 7

Analisando outra informação que foi capturada automaticamente, os *países* referidos ao longo das resoluções, encontramos o seguinte contexto, apresentado na tabela 9:

Resolução	Qtde Países
1199	11
1250	6
1191	5
981	4
1142	4
1174	4
970	3
988	3
1014	3
1022	3
1026	3
1054	3
1094	3
1136	3
1155	3
1207	3
1247	3
1255	3
1262	3
1272	3
977	2
992	2
1009	2
1017	2
1034	2
1036	2
1042	2
1060	2
1066	2
1075	2
1084	2
1087	2
1093	2
1112	2
1115	2
1124	2
1145	2
1147	2
1159	2
1183	2
1187	2
1203	2

1215	2
1227	2
1234	2
1237	2
1244	2
1258	2
1279	2
973	1
995	1
1000	1
1004	1
1007	1
1029	1
1032	1
1035	1
1039	1
1045	1
1048	1
1051	1
1063	1
1072	1
1078	1
1092	1
1100	1
1103	1
1106	1
1118	1
1127	1
1130	1
1133	1
1144	1
1151	1
1164	1
1167	1
1171	1
1179	1
1195	1
1218	1
1219	1
1220	1
1223	1
1230	1
1241	1

1265	1
1282	1
1284	1

Tabela 7

Outro dado que apresenta interess antes conclusões é a incidência de países nas resoluções, isto é, quantas vezes um determinado país foi referenciado ao menos uma vez em uma das resoluções, conforme a tabela 10:

Pais	Total		Total
Yugoslavia	22	Liberia	2
Bosnia and Herzegovina	15	Sierra Leone	2
Croatia	13	United kingdom of great Britain and northern Ireland	2
Angola	11	United states of America	2
Western Sahara	8	Australia	1
Cyprus	6	Burundi	1
Russian federation	6	Canada	1
Albania	5	China	1
France	5	Côte d'Ivoire	1
Morocco	5	Egypt	1
Central African republic	4	Eritrea	1
Gabon	4	Finland	1
Georgia	4	Ghana	1
Iraq	4	Greece	1
Rwanda	4	Guatemala	1
Congo	3	Japan	1
Democratic republic of the Congo	3	Luxembourg	1
Germany	3	Macedonia	1
Haiti	3	Mexico	1
Lebanon	3	Nigeria	1
Portugal	3	Romania	1
East Timor	2	Sudan	1
Ethiopia	2	Tajikistan	1
Indonesia	2	The former Yugoslav republic of	1
Ireland	2	United republic of Tanzania	1
Kuwait	2		

Tabela 10

Embora isto não seja um torneio europeu de basquete, a Iugoslávia sagrou-se campeã com vinte e duas aparições, fato explicável tendo em vista que o período

escolhido para a retirada da amostragem de documentos seja aquele compreendido entre 1995 e 1999, durante o qual este país esteve envolvido com questões securitárias internacionais de destaque. Os dez países com maior incidência estão contidos no gráfico 8.

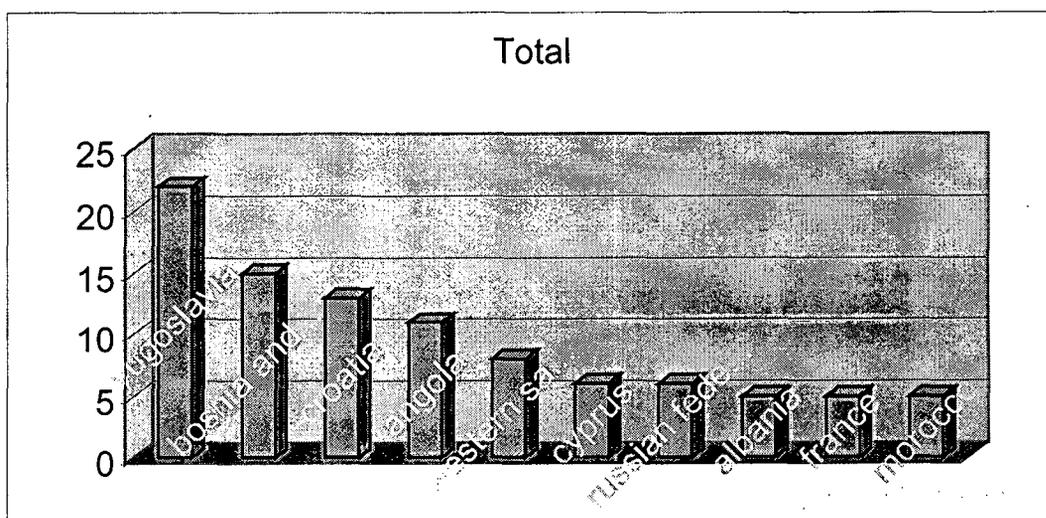


Gráfico 8

Analisando as siglas encontradas, chega-se à seguinte lista, contida na tabela 11:

Sigla	Total
Nato	11
Unita	11
Oau	10
Sfor	9
Minurso	8
Osce	6
Iptf	5
Unavem	5
Monua	4
Unficyp	4
Unhcr	4
Unomig	4
Icfy	3

Micivih	3
Misab	3
Unmih	3
Iaea	2
Ifor	2
Mnf	2
Sadc	2
Uncro	2
Unomil	2
Unprofor	2
Untaes	2
Unamir	1
Unmot	1
Unsmih	1

Tabela 11

Sua representação gráfica tem a seguinte forma (gráfico 9):

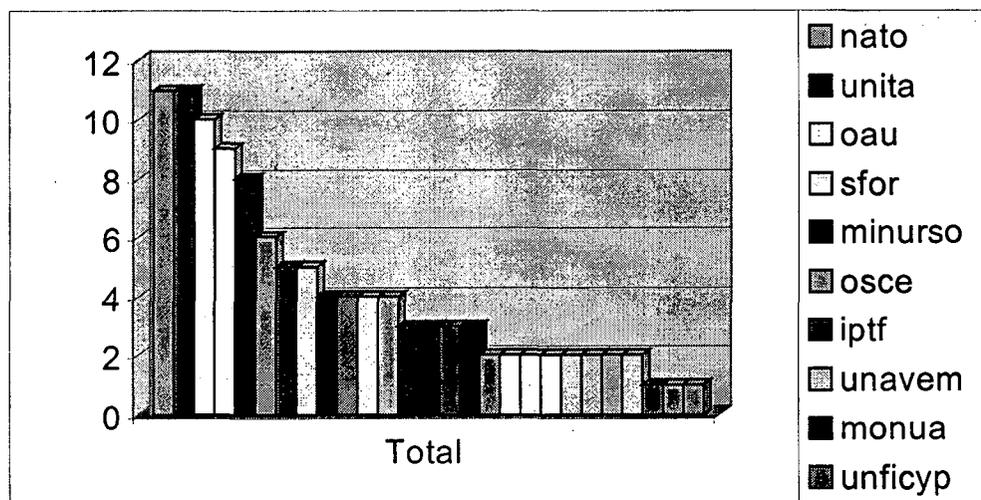


gráfico 9

Por último, verifica-se o número de siglas nas resoluções, isto é, quantas siglas diferentes ocorrem dentro das resoluções, estando arroladas na tabela 12 as dez resoluções que mais apresentam siglas no corpo de seu texto:

Resolução	Qtde Siglas
1075	6
1045	5
1063	5
1087	5
1118	5
1035	4
1048	4
1124	4
1127	4
1195	4

Tabela 12

Em seguida, passaremos à performance do sistema.

5. FUNCIONAMENTO E PERFORMANCE DO SISTEMA

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.

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, conforme ilustração contida na figura 45:

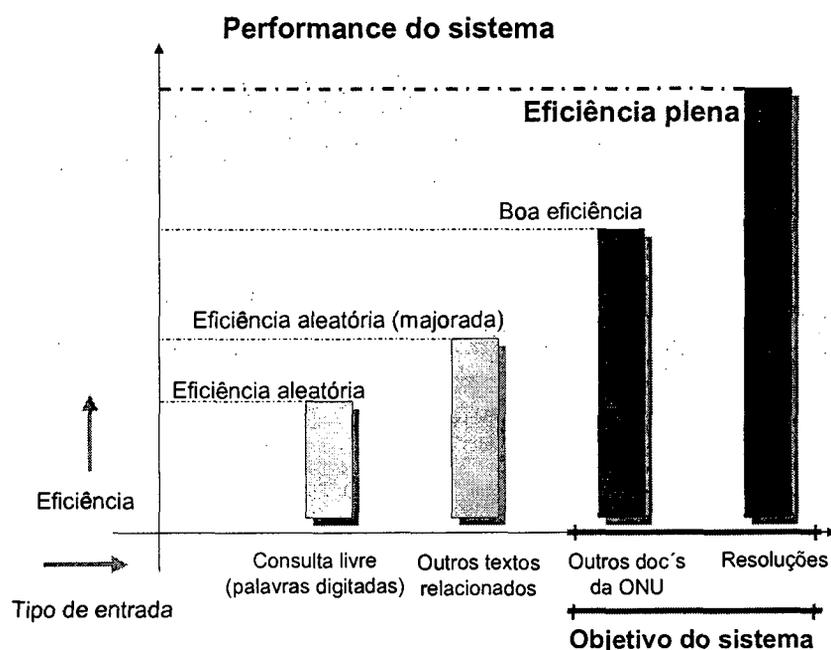


Figura 45

Assim quando é feita uma entrada aleatória, mediante digitação, as chances de uma recuperação efetiva são inferiores a 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 do sistema. 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 seqüência.

A) COMPARATIVO BD ONU X OLIMPO ¹⁰

¹⁰ Realizado no dia 23.11.2000

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 13:

Expressão pesquisada	Sistema Olimpo	Banco de dados da ONU (web)
“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 n. 1.234 ^a	10 registros O primeiro, com 100%, é a própria Resolução n. 1.234. O segundo com 71,40%, o terceiro com 64,30%, e o quarto com 57,10	Não permite esta possibilidade

Tabela 13

A seqüência 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, conforme figura 46.

The image shows a Netscape browser window titled "Security Council Search". The address bar contains the URL "http://www.un.org/documents/searchsc.htm". The page content includes navigation links for "UN Home", "Documentation Centre", and "About the UN". A "Select databases" section has five checkboxes, with "Security Council Resolutions" selected. A search form labeled "Find" contains the text "weapons" and three radio button options: "Match this phrase", "Match all of these words" (which is selected), and "Match any of these words". Below the form are "Begin search" and "Reset form" buttons. A note at the bottom states: "Posting of these documents is being done courtesy of the Department of Public Information as a result of expressed information needs of Member States. It will be done on a temporary basis until the Optical Disk System is made available to Member States."

Figura 46

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 47.

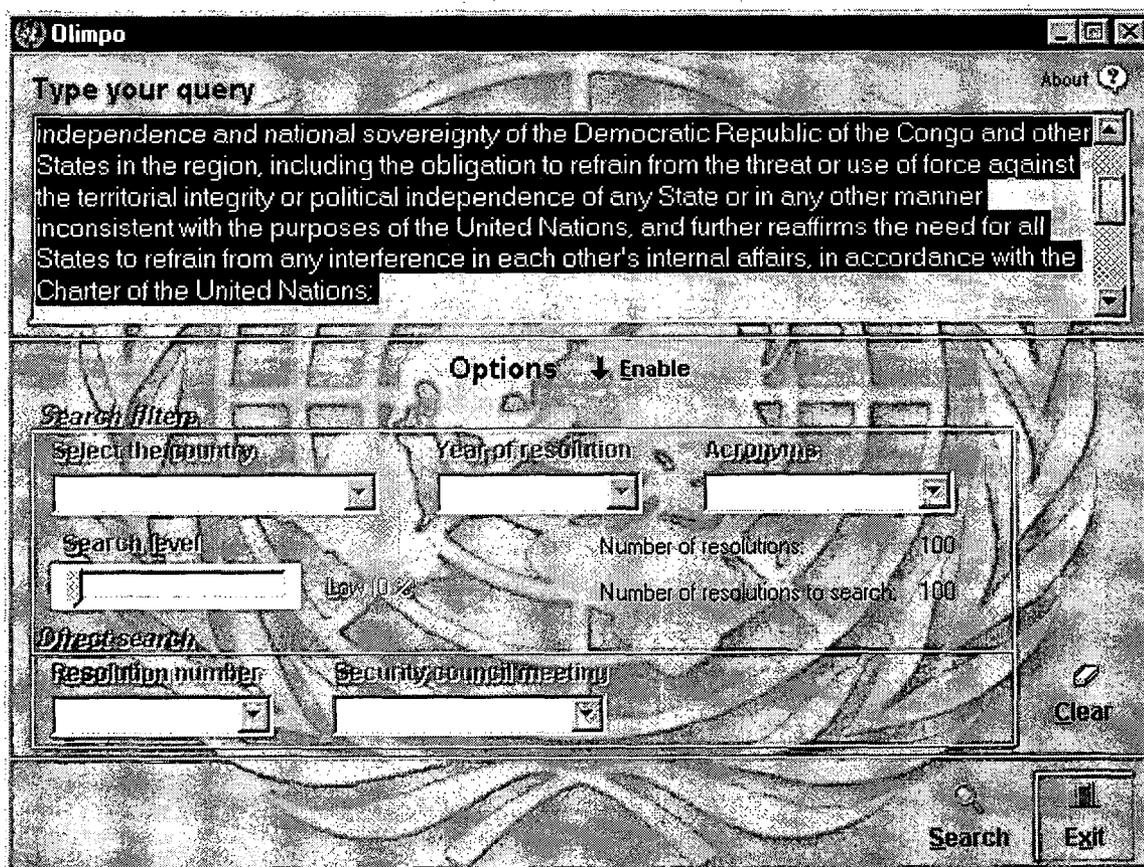


Figura 47

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, conforme a figura 48, 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.

The screenshot shows a window titled "Olimpo - Search Results". Inside, there is a table with the following columns: Order, Similarity, Resolution, Meeting, Date, and Abstract. The table contains four rows of data. Below the table, there is a "View the whole resolution" link and two buttons labeled "View" and "Back".

Order	Similarity	Resolution	Meeting	Date	Abstract
1	100.00%	1234	3993	09/04/99	Subject: Security Council resolution 1234 (1999) on the concerning the Democratic Republic of the Congo Countries: congo, democratic republic of the congo
2	71.40%	1199	3930	23/09/98	Subject: Security Council resolution 1199 (1998) Countries: albania, bosnia and herzegovina, canada, france, germany, ireland, japan, russian federation, united kingdom of great britain and northern ireland, united states of america, yugoslavia
3	84.30%	1187	3912	30/07/98	Subject: Security Council resolution 1187 Countries: georgia, russian federation
4	57.10%	1092	3728	23/12/96	Subject: Security Council resolution 1092 (1996) on the situation in Cyprus Countries: cyprus

Figura 48

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 49.

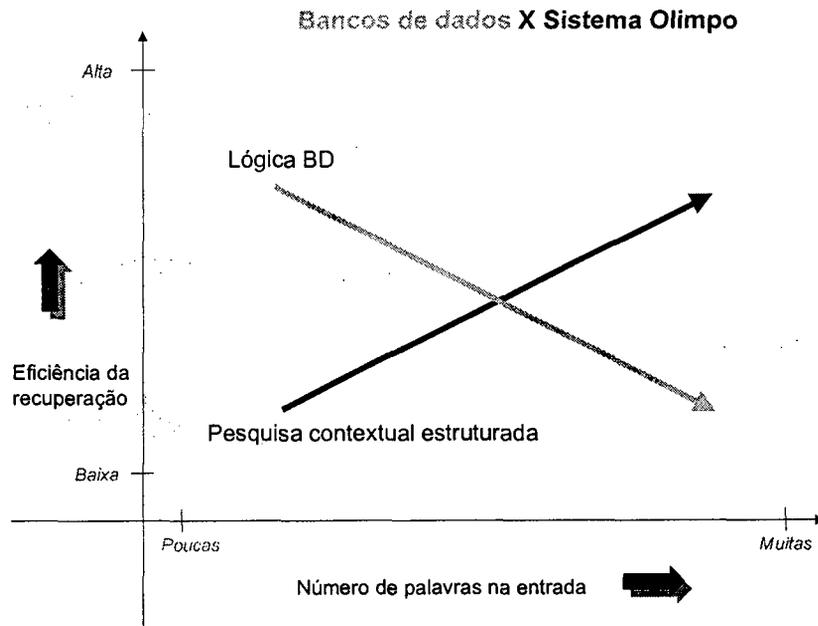


Figura 49

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 e 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.

B) Tarefas futuras

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.

CONCLUSÕES

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

- I. A técnica denominada “representação de conhecimento contextualizada dinamicamente” - **RC²D**, enseja significativo incremento de performance em sistemas estruturados em conhecimento;
- II. A adequada conjugação de características de bancos de dados textuais, raciocínio baseado em casos e da representação de conhecimento contextualizada dinamicamente ensejam a eficiente aplicação da “pesquisa contextual estruturada” – **PCE**, empregada para a construção do Sistema Olimpo;
- III. Na construção e consolidação do Sistema Olimpo, a **RC²D** representa a sua característica preponderante.
- IV. Que o adequado emprego da **PCE** e 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 densidade textual.

BIBLIOGRAFIA

- [1] ALEXANDRE, Silvio. Willian Gibson, o autor e sua obra, in *Neuromancer*, 2 ed., São Paulo: Aleph, 1991, p: 247.
- [2] BARRETO, Vicente. Bioética e a ordem jurídica, in *Revista da Faculdade de Direito*, v2, Rio de Janeiro, 1994, p. 443.
- [3] BELLUCCI, Emilia, ZELEZNIKOV, John. AI techniques for modeling legal negotiation. In proceedings of the seventh international conference on artificial intelligence and law, p. 108-116, Oslo: Norway, June, 14-18, 1999. 220 p.
- [4] 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, p. 36-42, Oslo: Norway, June, 14-18, 1999. 220 p.
- [5] BOBBIO, Norberto. *Teoria do ordenamento jurídico*. 4 ed., São Paulo: Unb, 1994.
- [6] BOBBIO, Norberto. *Teoria general del derecho*. Bogotá: Temis, 1987.
- [7] BOMAN, Magnus. Norms in artificial decision making. *Artificial intelligence and law*, Dordrecht: Netherlands, v 7, n. 1, p. 17-35, march, 1999.
- [8] BORGULYA, Istvan. Two examples of decision support in the law. *Artificial intelligence and law*, Dordrecht: Netherlands, v 7, ns. 2-3, p. 303-321, march, 1999.
- [9] BRUNINGHAUS, Stefanie, ASHLEY, Kevin D. Toward adding knowledge to learning algorithms for indexing legal cases. In proceedings of the seventh international conference on artificial intelligence and law, p. 9-17, Oslo: Norway, June, 14-18, 1999. 220 p.
- [10] BRUNINGHAUS, Stefanie, ASHLEY, Kevin D. Using machine learning for assigning indices to textual cases. In: LEAKE, David B., PLAZA, Enric (Eds.). *Case-based reasoning research and development*. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p, p. 303-314.
- [11] BUENO, Tânia Cristina D'Agostini. *O uso da teoria jurídica para Recuperação em Amplas Bases de Textos Jurídicos*. Dissertação de Mestrado, PPGEP/UFSC. Florianópolis: 1999.
- [12] BUENO, Tânia Cristina D'Agostini. *Recuperação inteligente de jurisprudência criminal*. Monografia de especialização, Curso de especialização em informática jurídica, Univali. Bigaçu: 1997.
- [13] BUENO, Tânia Cristina D'agostini, HOESCHL, Hugo Cesar, MATTOS, Eduardo da Silva, BARCIA, Ricardo Miranda, WANGENHEIM, Christiane Gresse Von. *JurisConsulto: Retrieval in Jurisprudencial Text Bases using Juridical Terminology*. In: THE SEVENTH

- INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 1999, Oslo. Proceedings of the Conference. New York: ACM, 1999. v.1. p.147-155.
- [14] BUENO, Tânia Cristina D'agostini, HOESCHL, Hugo Cesar, MATTOS, Eduardo da Silva, WANGENHEIM, Christiane Gresse Von, BARCIA, Ricardo Miranda. Uso da teoria jurídica para recuperação em amplas bases de textos jurídicos. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 1999, Rio de Janeiro. Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação. Rio de Janeiro: Edições EntreLugar, 1999. v.4. p.107-120.
- [15] BUENO, Tânia Cristina D'agostini, HOESCHL, Hugo Cesar, MATTOS, Eduardo da Silva, BARCIA, Ricardo Miranda, BORTOLON, André, WANGENHEIM, Christiane Gresse Von. Jurisconsulto. Florianópolis: 1999. Software sem registro
- [16] CARRAZA, Roque Antonio. Curso de direito constitucional tributário. 3 ed., São Paulo: RT, 1991.
- [17] CARVALHO, Paulo de Barros. Curso de direito tributário. 5 ed., São Paulo: Saraiva, 1991.
- [18] CASTELFRANCHI, Cristiano. Prescribed mental attitudes in goal-adoption and norm-adoption. *Artificial intelligence and law*, Dordrecht: Netherlands, v 7, n. 1, p. 37-50, march, 1999.
- [19] CASTORIADIS, Cornélius. O mundo fragmentado. Rio de Janeiro: Paz e terra, 1992.
- [20] CERQUEIRA, Tarcisio Queiroz. O direito do ciberespaço. Texto apostilado, 1995.
- [21] CONTE, Rosaria, FALCONE, Rino, SARTOR, Giovanni. Introduction: agents and norms: how to fill the gap?. *Artificial intelligence and law*, Dordrecht: Netherlands, v 7, n. 1, p. 1-15, march, 1999.
- [22] DANIELS, Jody J., RISSLAND, Edwina L. What you saw is what you want: using cases to seed information retrieval. In: LEAKE, David B., PLAZA, Enric (Eds.). *Case-based reasoning research and development*. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p, p. 315-336.
- [23] DARELLI, Lúcio Eduardo, HOESCHL, Hugo Cesar, BUENO, Tânia Cristina D'agostini. Estatuto do Instituto Jurídico de Inteligência e Sistemas - IJURIS. Florianópolis, 1999. Registrado junto ao Cartório do 1º Ofício de títulos e documentos de Florianópolis.
- [24] DINAMARCO, Candio Rangel. A reforma do CPC. 2Ed, São Paulo: Malheiros, 1995.
- [25] ELHADI, Mohamed T., VAMOS, Tibor. Bankruptcy case law: a hybrid IR-CBR approach. In proceedings of the seventh international conference on artificial intelligence and law, p. 134-135, Oslo: Norway, June, 14-18, 1999. 220 p.
- [26] EPSTEIN, Isaac. Cibernética. São Paulo: Ática, 1986.
- [27] FEATHER, Stephen. Javascript em exemplos, São Paulo, Makron books, 1997.
- [28] FUCHS, Marc, FUCHS, Matthias. Applying case-based reasoning to automated deduction. In: LEAKE, David B., PLAZA, Enric (Eds.). *Case-based reasoning research and development*.

Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p., p. 24-32.

- [29] GARCIA, Dinio de Santis. Informática jurídica
- [30] GIBSON, Willian. Neuromancer. 2 ed., São Paulo: Aleph, 1991.
- [31] HAMIT, Francis. A realidade virtual e a exploração do espaço cibernético. Rio de Janeiro: Berkley, 1993.
- [32] HOESCHL, Hugo Cesar, BARCIA, Ricardo Miranda, BUENO, Tânia Cristina D'Agostini, MATTOS, Eduardo da Silva, BORTOLON, André, DONATTI, Fabrício Tadeu. Sistema Olimpo. Florianópolis, 2000. Software registrado junto ao INPI.
- [33] HOESCHL, Hugo Cesar, BARCIA, Ricardo Miranda. A telemática e os direitos da sétima dimensão. Revista Trimestral de Jurisprudência dos Estados, São Paulo, v.174, p.9-14, 1999.
- [34] HOESCHL, Hugo Cesar, BORTOLON, André, BARCIA, Ricardo Miranda, BUENO, Tânia Cristina D'agostini, MATTOS, Eduardo da Silva, DONATTI, Fabrício Tadeu, DARELLI, Lúcio Eduardo, ROVER, Aires José, BUENO, Luis Gustavo D'agostini, SANTOS, Cristina Souza, DARELLI, Lúcio Eduardo. Web site do Instituto Jurídico de Inteligência e Sistemas. Florianópolis: Ijuris, 1999. www.digesto.net/ijuris
- [35] HOESCHL, Hugo Cesar, BORTOLON, André, DONATTI, Fabrício Tadeu. Metalex. Florianópolis, 2000. Software sem registro.
- [36] HOESCHL, Hugo Cesar, BUENO, Tânia Cristina D'Agostini, BARCIA, Ricardo Miranda. Tecnologia da informação jurídica. Material didático-instrucional de pós-graduação, Florianópolis: PPGE/UFSC, 2000. Disponível na internet: <http://digesto.net/ijuris/Tecnologia%20info%20juridica/index.htm>
- [37] HOESCHL, Hugo Cesar, BUENO, Tânia Cristina D'Agostini, BARCIA, Ricardo Miranda. O impacto social da tecnologia da informação. Material didático-instrucional de pós-graduação, Florianópolis: PPGE/UFSC, 2000. Disponível na internet: <http://digesto.net/ijuris/isti/index.htm>
- [38] HOESCHL, Hugo Cesar, BUENO, Tânia Cristina D'Agostini, BARCIA, Ricardo Miranda, SANTOS, Márcilio Dias dos. A Tecnologia da informação jurídica, Florianópolis: Rocket Library, 2000, 663 p.
- [39] HOESCHL, Hugo Cesar. A legislação brasileira sobre telemática, Florianópolis: Rocket Library, 2000, 115 p.
- [40] HOESCHL, Hugo Cesar. A liberdade de expressão na internet. Revista Trimestral de Jurisprudência dos Estados, São Paulo, v.160, p.13-18, 1997.
- [41] HOESCHL, Hugo Cesar. A telemática nos Tribunais, Florianópolis, Rocket library, 2000, 65 p.

- [42] HOESCHL, Hugo Cesar. Aspectos constitucionais da Lei 9.296/96. In ROVER, Aires J. (org) e outros. Direito Sociedade e informática. Florianópolis: Fundação Boiteux, 2000. 246. P. 105-113.
- [43] HOESCHL, Hugo Cesar. Banco de petições da Procuradoria da União em Santa Catarina. Florianópolis: 1996.
- [44] HOESCHL, Hugo Cesar. Digesto. Florianópolis: 1998. Software registrado sob n. 00032134, "Digesto". 29 de jun. de 2000 (Depósito); 29 de jun. de 2000 (Exame); 06 de jul. de 2000 (Concessão).
- [45] HOESCHL, Hugo Cesar. Habeas Corpus na Madrugada. Florianópolis: 1998. Software sem registro.
- [46] HOESCHL, Hugo Cesar. Juri Digital. Florianópolis: 2000. Software sem registro.
- [47] HOESCHL, Hugo Cesar. Metajuris. Florianópolis: 1998. Software registrado sob número 00032122, "Metajuris". 29 de jun. de 2000 (Depósito); 29 de jun. de 2000 (Exame); 06 de jul. de 2000 (Concessão).
- [48] HOESCHL, Hugo Cesar. O ciberespaço e o direito. Revista Trimestral de Jurisprudência dos Estados, São Paulo, v.167, p.15-28, 1998.
- [49] HOESCHL, Hugo Cesar. O relacionamento da telemática com o direito e seu tratamento jurídico no Brasil. Dissertação de mestrado junto ao CPGD da UFSC. Florianópolis: 1997.
- [50] HOESCHL, Hugo Cesar. REIS, Cassiano. Themis. Florianópolis: 1999. Software sem registro.
- [51] HOESCHL, Hugo Cesar. Sistema de controle de prazos processuais da Procuradoria da União em Santa Catarina. Florianópolis: 1996.
- [52] HOESCHL, Hugo Cesar. Sistemas especialistas para mediação. Revista Trimestral de Jurisprudência dos Estados, São Paulo, v.165, p.87-93, 1998.
- [53] HOESCHL, Hugo Cesar. Telemática e direito. Monografia de especialização junto ao curso de especialização em informática jurídica, Univali. Biguaçu: 1997.
- [54] HOESCHL, Hugo, BUENO, Tânia, BORTOLON, André, BARCIA, Ricardo Miranda, DONATTI, Fabrício T., BUENO, Gustavo, DARELLI, Lúcio Eduardo, MATTOS, Eduardo, SANTOS, Cristina de Souza. Avaliação dos sites dos tribunais brasileiros na web – 99. Florianópolis: Ijuris, 1999. Disponível na internet: http://digesto.net/ijuris/avaliacao/resultado_final.htm
- [55] HORTY, John F. Precedent, deontic logic and inheritance. In proceedings of the seventh international conference on artificial intelligence and law, p. 63-72, Oslo: Norway, June, 14-18, 1999. 220 p.
- [56] JAKOBOVITS, H., VERMEIR, D.. Dialectic semantic for argumentation frameworks. In proceedings of the seventh international conference on artificial intelligence and law, p. 53-62, Oslo: Norway, June, 14-18, 1999. 220 p.

- [57] KAKUTA, Tokuyasu, HARAGUCHI, Makoto. A demonstration of a legal reasoning system based on teleological analogies. In proceedings of the seventh international conference on artificial intelligence and law, p. 196-205, Oslo: Norway, June, 14-18, 1999. 220 p.
- [58] KARACAPILIDIS, Nikos, TROUSSE, Brigitte, PAPADIAS, Dimitris. Using case-based reasoning for argumentation with multiple viewpoints. LEAKE, David B., PLAZA, Enric (Eds.). Case-based reasoning research and development. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p., p. 541-552.
- [59] LAFER, Celso. A reconstrução dos direitos humanos: um diálogo com o pensamento de Hanna Arendt. São Paulo: Cia das Letras, 1988.
- [60] LAQUEY, Tracy e RYER, Jeanne C. O manual da internet. Rio de Janeiro: Campus, 1994.
- [61] LEAKE, David B., PLAZA, Enric (Eds.). Case-based reasoning research and development. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p.
- [62] LEE, R. W.; BARCIA, R.; PACHECO, R.; MARTINS, A.; HOESCHL, Hugo; BUENO, Tânia; COSTA, Márcio; RODRIGUES FILHO, Ilson. (1997). Representing Cases From Texts in Case-Based Reasoning. III Congresso Internacional de Engenharia Industrial e XVII ENEGEP, 6 a 9 de Outubro de 1997, Canela, RS, Brasil.
- [63] LEE, Rosina W. , HOESCHL, Hugo Cesar, RODRIGUES, Alejandro Martins, PACHECO, Roberto Carlos dos Santos, BUENO, Tânia Cristina D'agostini, COSTA, Márcio C da, RODRIGUES FILHO, Ilson W., BARCIA, Ricardo Miranda. A Large Case-Based Reasoner for Legal Cases. Organizado por LEAKE, David B., PLAZA, Enric. Case-Based Reasoning Research and Development. Berlin; Heidelberg; New York, 1997, v.1266, p.190-199.
- [64] LEITE, Eduardo de Oliveira. A monografia jurídica. Porto Alegre: Fabris, 1985.
- [65] LENT, AGRAWAL & SRIKANT. Discovering trends in text databases, IBM Almaden Research Center (complementar e ampliar).
- [66] LENZ, Mario, BURKHARD, Hans-Dieter. In: LEAKE, David B., PLAZA, Enric (Eds.). Case-based reasoning research and development. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p, p. 84-93.
- [67] MAcCOMB, Gordon. Javascript sourcebook, São Paulo, Makron books, 1997.
- [68] MARTIN, James. Hiper documentos e como criá-los. Rio de Janeiro: Campus, 1992.
- [69] MATTOS, Eduardo. Pesquisa jurisprudencial inteligente. Trabalho de conclusão de curso, Univali. Biguaçu: 1997.
- [70] MATTOS, Eduardo da Silva, BARCIA, Ricardo Miranda, LEE, Rosina W., HOESCHL, Hugo Cesar, BUENO, Tânia Cristina D'agostini, PACHECO, Roberto Carlos dos Santos,

- RODRIGUES, Alejandro Martins. Prudentia. Florianópolis, 1997. Software de domínio público, por força do disposto na Lei 9.609.
- [71] NEGROPONTE, Nicholas. A vida digital. São Paulo: Cia das letras, 1995.
- [72] OSBORN, James, STERLING, Leon. Justice. A judicial search tool using intelligent concept extraction. In proceedings of the seventh international conference on artificial intelligence and law, p. 173-181, Oslo: Norway, June, 14-18, 1999. 220 p.
- [73] PASOLD, Cesar Luiz. O advogado e a advocacia. Florianópolis: terceiro milênio, 1996.
- [74] PFAFFENBERGER, Bryan. Dicionário dos usuários de micro computadores. Rio de Janeiro: Campus, 1993.
- [75] PHILIPS, Lothar. Approximate syllogisms – on the logic of everyday life. Artificial intelligence and law, Dordrecht: Netherlands, v 7, ns. 2-3, p. 227-234, march, 1999.
- [76] RABUSKE, Renato Antonio. Inteligência Artificial. Florianópolis: Ed. Ufsc, 1995.
- [77] RADBRUCH, Gustav. Filosofia do direito. Coimbra: Armênio Amado, 1979.
- [78] RAMALHO, José Antonio. Iniciando em HTML, São Paulo, Makron Books, 1996.
- [79] ROUANET, Sérgio Paulo. As razões do iluminismo. São Paulo: Cia das Letras, 1987.
- [80] ROVER, Aires José (organizador). Informática, Direito e Sociedade. Fundação Boiteux: Florianópolis, 2000, 251 p.
- [81] SAVOLA, Tom. Usando HTML, Rio de Janeiro, Campus, 1996.
- [82] SARTOR, Giovanni, CONTE, Rosaria, FALCONE, Rino. Artificial intelligence and law, an international journal. Dordrecht: Kluwer, 1999.
- [83] SILVA, Edna Lúcia da, ESTERA, Muszkat Menezes. Metodologia da pesquisa e elaboração de dissertação: Laboratório de ensino à distância da UFSC, 2000, 118p.
- [84] SILVA, José Afonso da. Curso de direito constitucional positivo. São Paulo: RT, 1991.
- [85] STEINFUS, Ricardo. Manual da organizações internacionais. Porto Alegre: Livraria do Advogado, 1997, 352p.
- [86] STRANIERI, Andrew, ZELEZNIKOV, John. The evaluation of legal knowledge based system. In proceedings of the seventh international conference on artificial intelligence and law, p. 18-24, Oslo: Norway, June, 14-18, 1999. 220 p.
- [87] SUPERIOR TRIBUNAL DE JUSTIÇA. <http://www.stj.gov.br>
- [88] SUPREMO TRIBUNAL FEDERAL. <http://www.stf.gov.br>
- [89] TEMER, Michel. Elementos de direito constitucional. 9 ed., São Paulo: Malheiros, 1992.
- [90] TOLEDO, Francisco de Assis. Princípios Básicos de direito penal. São Paulo: Saraiva, 1991.
- [91] VARGAS, José Israel. Um caminho para a internet, in Informática Exame, nº 109, São Paulo: Abril, abril de 1995, p. 142.
- [92] WARAT, Luis Alberto. Introdução geral ao direito. Porto Alegre: Fabris, 1994.
- [93] WARAT, Luis Alberto. Por quien cantan las sirenas ? Florianópolis, edição eletrônica, 1996.

- [94] WATTSON, Ian. CBR is a methodology not a technology. University of Salford, Salford.
- [95] WEBER, R.; MARTINS, A.; MATTOS, E.; BUENO, T., HOESCHL, H.; PACHECO, R.; BARCIA, R. (1998). Reusing Cases to the Automatic Index Assignment from Textual Documents. 6th German Workshop on Case-Based Reasoning - Foundations, Systems, and Applications. Berlin, March 6-8, 1998.
- [96] WEBER, Rosina. Intelligent jurisprudence research: a new concept. In proceedings of the seventh international conference on artificial intelligence and law, p. 164-172, Oslo: Norway, June, 14-18, 1999. 220 p.
- [97] WISDO, Christopher. A scalable approach for questions based indexing of encyclopedic texts. In: LEAKE, David B., PLAZA, Enric (Eds.). Case-based reasoning research and development. Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Milan; Paris; Santa Clara; Singapore; Tokyo: Springer, 1997. 648 p., p. 200-210.
- [98] WOLKMER, Antônio Carlos. Fundamentos da História do direito (complementar).
- [99] WOLKMER, Antônio Carlos. Pluralismo Jurídico. São Paulo: Alfa Ômega, 1994.
- [100] YEARDWOOD, John, STRANIERI, Andrew. The integration of retrieval, reasoning and drafting for refugee law: a third generation legal knowledge based system. In proceedings of the seventh international conference on artificial intelligence and law, p. 117-137, Oslo: Norway, June, 14-18, 1999. 220 p.
- [101] ZAFFARONI, Eugênio Raul. Em busca das penas perdidas, Rio de Janeiro:Revam, 1991.