

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**“RBC-TEXT” - MODELO PARA  
TRATAMENTO DE DOCUMENTOS TEXTUAIS  
E VALIDAÇÃO NO DOMÍNIO JURÍDICO**

**EDUARDO DA SILVA MATTOS**

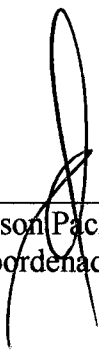
Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Mestre em Engenharia de Produção

FLORIANÓPOLIS  
2002

**EDUARDO DA SILVA MATTOS**

**“RBC-TEXT” - MODELO PARA  
TRATAMENTO DE DOCUMENTOS TEXTUAIS  
E VALIDAÇÃO NO DOMÍNIO JURÍDICO**


Esta dissertação foi julgada e aprovada para a  
obtenção do título de **Mestre em Engenharia de  
Produção no Programa de Pós-Graduação em  
Engenharia de Produção da**  
Universidade Federal de Santa Catarina.



---


Prof. Edson Pacheco Paladini, Dr.  
Coordenador do Curso

**BANCA EXAMINADORA**



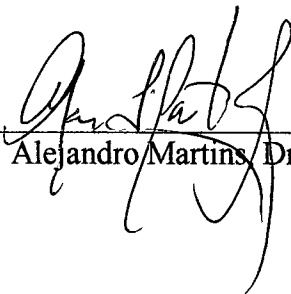
---

Prof. Roberto C. S. Pacheco, Dr.  
**Orientador**



---

Prof. Hugo Cesar Hoeschl, Dr.  
**Co-Orientador**



---

Prof. Alejandro Martins, Dr.

## AGRADECIMENTOS

Desejo manifestar minha gratidão às pessoas que contribuíram direta e indiretamente para a realização deste trabalho.

Ao meu orientador, Professor Roberto C. dos S. Pacheco, pela sua orientação precisa, proporcionando a conclusão desta pesquisa.

Aos Professores e amigos, Hugo Cesar Hoeschl e Tânia C. D'Agostini Bueno, pela amizade e pelos ensinamentos jurídicos, que foram fundamentais neste trabalho e para minha formação pessoal.

Aos amigos do IJURIS, pela colaboração, apoio e contribuição, que foram de extremo valor durante toda a realização da pesquisa.

Aos amigos do Laboratório Stela, que acompanharam e cooperaram para o desenvolvimento do trabalho.

A toda a minha família, que sempre me apoiou nos momentos mais difíceis.

E, especialmente, aos meus pais, que sempre acreditaram em mim, torceram e continuam torcendo pelo meu sucesso.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1. JUSTIFICATIVA .....	2
1.2. OBJETIVO GERAL .....	5
1.3. METODOLOGIA DE PESQUISA .....	6
1.4. ESTRUTURA .....	8
1.5. ESCOPO E DELIMITAÇÕES DO TRABALHO .....	9
<b>2. DIREITO LEGAL E INFORMÁTICA</b> .....	<b>10</b>
2.1. INTRODUÇÃO .....	10
2.2. TEORIA DA JURISPRUDÊNCIA .....	10
2.3. RELEVÂNCIA DA JURISPRUDÊNCIA NO DIREITO .....	12
2.4. A INFORMÁTICA PARA O DIREITO .....	12
2.5. CONSULTA JURISPRUDENCIAL .....	13
2.5.1. Consultas em banco de dados .....	14
2.6. INTELIGÊNCIA ARTIFICIAL NO DIREITO E RECUPERAÇÃO TEXTUAL .....	15
2.7. CONSIDERAÇÕES FINAIS .....	18
<b>3. RACIOCÍNIO BASEADO EM CASOS (RBC)</b> .....	<b>20</b>
3.1. INTRODUÇÃO .....	20
3.2. HISTÓRICO DO RACIOCÍNIO BASEADO EM CASOS .....	23
3.3. REPRESENTAÇÃO DO CASO .....	28
3.4. ESTRUTURA DO RACIOCÍNIO BASEADO EM CASOS .....	28
3.4.1. Representação do Caso .....	29
3.4.2. Recuperação dos Casos .....	33
3.4.3. Adaptação .....	35
3.4.4. Aprendizagem .....	35
3.5. VALIDAÇÃO .....	35
3.6. APLICAÇÕES .....	36
3.7. FERRAMENTAS .....	40
3.8. RBC E OUTRAS TÉCNICAS DE IA .....	41
3.9. RESUMO DAS CARACTERÍSTICAS DO RBC .....	42
3.10. CONSIDERAÇÕES FINAIS .....	43
<b>4. “RBC-TEXT” - MODELO PARA TRATAMENTO DE DOCUMENTOS TEXTUAIS E VALIDAÇÃO NO DOMÍNIO JURÍDICO</b> .....	<b>44</b>
4.1. INTRODUÇÃO .....	44
4.2. ANÁLISE DO DOMÍNIO .....	45
4.2.1. Soluções informatizadas disponíveis .....	47
4.2.2. Entrevistas com o especialista .....	47
4.2.3. Atributos/Índices na Área do Problema .....	47

4.2.4.	Vocabulário de Domínio .....	48
4.2.5.	Exemplos de Raciocínio Analógico .....	48
4.3.	TRANSFORMAÇÃO EM CASOS PARA RBC.....	49
4.4.	APLICAÇÃO “RBC-TEXT” .....	50
4.4.1.	Construção da Base de Casos.....	50
4.4.2.	Novos Casos.....	52
4.4.3.	Teste .....	53
4.4.4.	Refinamento e evolução .....	54
4.5.	CONSIDERAÇÕES FINAIS .....	54
<b>5.</b>	<b>ESTRUTURA DO SISTEMA .....</b>	<b>55</b>
5.1.	INTRODUÇÃO .....	55
5.2.	MODELAGEM EM BANCO DE DADOS RELACIONAL .....	55
5.2.1.	Dicionário de Dados.....	57
5.3.	FLUXO DE DADOS - GERAL .....	63
5.4.	FLUXO DE DADOS – AQUISIÇÃO DE NOVOS CASOS .....	67
5.5.	ALGORITMOS .....	69
5.6.	CONSIDERAÇÕES FINAIS .....	70
<b>6.</b>	<b>APLICAÇÃO .....</b>	<b>71</b>
6.1.	INTRODUÇÃO .....	71
6.2.	MÓDULO DE MEMÓRIA.....	72
6.2.1.	Extração dos dados.....	73
6.3.	MÓDULO DE RECUPERAÇÃO .....	76
6.3.1.	Indexação .....	76
6.3.2.	Avaliação da similaridade .....	79
6.4.	MÓDULO DE INTERFACE.....	82
6.4.1.	Elaboração da consulta.....	82
6.4.2.	Resultado da Consulta .....	86
6.4.3.	Acórdão na integra .....	88
6.5.	CONSIDERAÇÕES FINAIS .....	89
<b>7.</b>	<b>CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS.....</b>	<b>90</b>
<b>8.</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>92</b>
<b>ANEXO A .....</b>	<b>95</b>	
<b>ANEXO B.....</b>	<b>95</b>	
<b>ANEXO C .....</b>	<b>112</b>	

## LISTA DE FIGURAS

Figura 1 - Metodologia do trabalho.....	7
Figura 2 - Evolução da consulta jurisprudencial .....	14
Figura 3 - Pesquisa jurisprudencial (www.trtrio.gov.br).....	15
Figura 4 - Ciclo do RBC – (Aamodt et al., 1994) .....	29
Figura 5 - Modelo "RBC-Text" .....	44
Figura 6 - Metodologia para aplicação do modelo .....	46
Figura 7 - Modelagem do JurisConsulta .....	57
Figura 8 - Fluxo da informação .....	58
Figura 9 - Campos da entrada .....	58
Figura 10 - Campos da saída principal .....	60
Figura 11 - Campos da saída do acórdão.....	60
Figura 12 - Casos do caso na base.....	61
Figura 13 - Fluxo dos dados.....	62
Figura 14 - Fluxo dos dados para a aquisição de novos casos.....	67
Figura 15 - Acórdão .....	71
Figura 16 - Acórdão na base .....	72
Figura 17 - Atualização da base .....	73
Figura 18 - Valoração dos atributos - por localização.....	74
Figura 19 - Vizinho mais próximo .....	80
Figura 20 - Medida de similaridade .....	81
Figura 21 - Elaboração da consulta .....	83
Figura 22 - Nível de restrição.....	84
Figura 23 - Caso de entrada.....	85
Figura 24 - Resultado da consulta .....	87
Figura 25 - Acórdão na integra .....	88

**LISTA DE TABELAS**

Tabela 1 - Relação de trabalhos jurídicos (Souza, 2000) .....	3
Tabela 2 - Histórico do RBC nos EUA .....	24
Tabela 3 - Histórico do RBC na Europa.....	25
Tabela 4 - Exemplo de casos para diagnóstico e reparo automobilístico .....	33
Tabela 5 - Aplicações utilizando RBC .....	39
Tabela 6 - Exemplo de Atributos para "Cadastro Urbano" .....	51
Tabela 7 - Atributos x Forma de extração .....	74
Tabela 8 - Expressões para Denegado/Concedido .....	75
Tabela 9 - Valores do índice Tipificação .....	78
Tabela 10 - Algumas expressões precedentes .....	78
Tabela 11 - Algumas fontes .....	78
Tabela 12 - Caso de entrada x Caso na base .....	81

## RESUMO

O escopo deste trabalho está delimitado à aplicação da tecnologia de RBC na recuperação de informações constantes em documentos não estruturados. Para que esta recuperação seja possível, apresenta-se um modelo para a extração do conhecimento contido nos documentos. O propósito do modelo será o de considerar documentos textuais que apresentem as seguintes características:

- base tecnológica no raciocínio analógico;
- vocabulário e regras de domínio próprio; .
- casos descritos na forma textual, mas com normatização de formação.

A validação do modelo será realizada através da sua aplicação no domínio do Direito, mais precisamente nos documentos legais, acórdãos, que descrevem as decisões do Tribunal de Justiça do Estado de Santa Catarina (TJSC), uma vez que estes preenchem o requisitos supracitados.



## ABSTRACT

The aim of this work is focused on applying CBR technology on the retrieval of information contained in non-structured documents. Due to this goal is presented a model for document knowledge extraction, for which a document is considered a textual document when the following characteristics are present:

- technological base in analogic reasoning;
- vocabulary and rules for a specific domain;
- cases described in textual form, but with normalized organization.

Model validity is accomplished through its application in the legal domain, specifically in legal documents known as “acórdãos”, which describe decisions of the Justice Court of the State of Santa Catarina (TJSC), as they contain all the characteristics of textual documents presented above.

## INTRODUÇÃO

A massificação no uso da informática trouxe e continuará fornecendo benefícios aos mais diversos ramos da sociedade. Seja através das vantagens fornecidas por um sistema bancário totalmente informatizado, através da modernização das lojas comerciais, ou da comodidade proporcionada por uma eleição eletrônica, o importante é que o cidadão, de uma maneira geral, vem sendo beneficiado pela evolução tecnológica.

A partir do momento em que a sociedade começou a usufruir e a constatar as facilidades proporcionadas por tais benefícios, fica difícil imaginarmos algum ramo profissional fora deste contexto. E mais do que apenas colocar as informações nos computadores, tornou-se, também, de suma relevância tornar as informações públicas na *internet*.

Praticamente todos os segmentos da sociedade já estão adaptados com o uso dos computadores na resolução das suas tarefas diárias. Na área jurídica não é diferente, embora os profissionais do Direito, por tradição, relutem um pouco mais em aceitar novas tecnologias como instrumentos de apoio em suas atividades, atualmente estão cientes da necessidade e da importância dos aplicativos jurídicos na sua profissão (Deocleciano Júnior et al., 2000).

O sinal de que a utilização da tecnologia no campo do Direito é um caminho sem volta é observado através da grande variedade de aplicativos, publicações e endereços eletrônicos destinados aos operadores jurídicos. Podem ser citados os softwares: Lexli96 - Biblioteca digital, LIS – Legislação Informatizada Saraiva entre outros. Entre os diversos *sites* jurídicos exemplifica-se o *site* do Supremo Tribunal Federal ([www.stf.gov.br](http://www.stf.gov.br)), do Tribunal Superior do Trabalho ([www.tst.gov.br](http://www.tst.gov.br)), do Tribunal Superior Eleitoral ([www.tse.gov.br](http://www.tse.gov.br)), além dos demais *sites* institucionais, de revistas jurídicas, dos escritórios de advocacia, etc (Souza, 2000).

Apesar do grande número de opções computacionais direcionadas ao campo do Direito, verifica-se que nem sempre os recursos empregados para auxiliar o trabalho dos juristas são os mais adequados. Isto se evidencia mais nos mecanismos de buscas, onde o jurista encontra uma certa dificuldade em encontrar o documento mais adequado para auxiliá-lo em um determinado problema. É sobre este problema que esta dissertação está enfocada, ou

seja, objetiva-se apresentar uma alternativa inteligente para buscar as informações mais adequadas para ajudar na resolução do problema em questão.

Este trabalho apresenta o desenvolvimento de um protótipo utilizando os seguintes recursos: a técnica de Inteligência Artificial (IA), denominada *Raciocínio Baseado em Casos* (RBC), Programação Orientada a Objetos (Delphi) e Banco de Dados Relacionais (SQL Server).

O protótipo proposto, JurisConsulta, retorna as jurisprudências mais significativas para um usuário, mediante uma requisição realizada pelo mesmo. Tal requisição é comparada com uma base de casos (jurisprudências) resultando um conjunto contendo as mais similares. A recuperação destas jurisprudências objetiva auxiliar advogados e juízes na fundamentação dos casos. Possibilitando que tais profissionais designem menos tempo nas suas consultas jurídicas, porém analisem um número maior de questões.

### **1.1. Justificativa**

O campo das jurisprudências foi eleito por contribuir fortemente para a utilização da técnica de RBC e por serem relevantes para os juristas em seu processo de tomada de decisão. Conforme será visto no corpo deste trabalho, o formato dos documentos das jurisprudências preenche alguns pré-requisitos elencados por aplicações que fazem uso da técnica de Raciocínio Baseado em Casos. A relevância das jurisprudências no ramo do Direito é avalizada por Bueno, 1999, quando a autora comenta que as jurisprudências são de suma relevância para os juristas, auxiliando-os com uma valiosa fonte de interpretação da lei. Diante desta importância, jurisprudência torna-se um nicho a ser atacado pelos profissionais da informática. Isto é confirmado pela quantidade de softwares, baseados em Banco de Dados, lançados no mercado direcionados a esta área (Vasconcelos, 1996).

Na Tabela 1 encontram-se alguns trabalhos desenvolvidos para área jurídica.

Aplicativo/Site	Descrição
LIS-Legislação Informatizada Saraiva	Traz toda a legislação federal vigente, organizada em ordem cronológica e classificadas por áreas do Direito e de interesse.
JUIS-Jurisprudência Informatizada Saraiva	Traz o ementário fornecido pelo STF,STJ,TJ-SP,1o TAC-SP e 2º TAC-SP e Tribunal de Alçada Criminal-SP.
Lexli96	Biblioteca digital - softwares sobre textos de jurisprudência e legislação em mídia.
Juris Consult	Sistema de jurisprudência informatizada, coletânea da Lex Editora S/A.
Revista Eletrônica de jurisprudência Trabalhista	Jurisprudência dos principais tribunais do trabalho, com atualização bimestral.
Manual de Sentenças e Despachos	Modelos de sentenças e despachos judiciais.
www.stf.gov.br	Oferece informações institucionais, pesquisa de processos, jurisprudências, o Diário da Justiça da União e "links"(conexões) para outros servidores jurídicos.
www.tst.gov.br	Apresenta acompanhamento do processo, pesquisa jurisprudencial, acórdão inteiro teor, entre outras informações.
www.stj.gov.br	Oferece informações institucionais, acesso aos processos, pesquisa por jurisprudências, acórdão inteiro teor, outros links jurídicos, revista, etc.

Tabela 1 - Relação de trabalhos jurídicos (Souza, 2000)

Particularmente em jurisprudências, a principal deficiência dos sistemas existentes é seu embasamento em Banco de Dados Relacionais. Isto faz com que o sistema de buscas convencionalmente implementado, via palavras-chave, dificulte o encontro de casos realmente similares ao caso avaliando.

O aplicativo desenvolvido nesta dissertação apresentará na interface uma caixa de texto, na qual o usuário descreverá o seu problema. Após realizada a descrição e

disparado o processo de recuperação, o sistema apresentará as jurisprudências mais semelhantes ao problema descrito.

Os resultados alcançados na presente dissertação serão de suma relevância para a prática do Direito, pois estaremos aliando a modelagem, indexação e recuperação de informações textuais utilizando a técnica de RBC. Desta forma os operadores do Direito, ávidos por soluções que lhes proporcionem um maior aproveitamento do tempo, poderão fazer uso de uma ferramenta que acelere a obtenção de dados úteis. Haja vista que os sistemas atuais, por estarem fundados em pesquisas por palavras-chave, além de retornarem um número excessivo de informações, muitas vezes pecam na qualidade nos resultados obtidos.

No Brasil, profissionais do Direito têm duas fontes para pesquisar por casos legais anteriores para um terreno de novas decisões: livros e sistemas de base de dados.

Segundo Vasconcelos (1996), as pesquisas em literaturas são realizadas no que existe de informações disponíveis nas inúmeras prateleiras espalhadas por todo o escritório. Legislação, jurisprudência, e todos os tópicos encontram-se classificados e arquivados por prateleira. Algumas vezes demoram-se dias para localizar a legislação de um determinado assunto, ou a jurisprudência que possa ser utilizada no caso. Nem sempre o que se possui como fonte de informação está atualizado.

Raciocínio Baseado em Casos surge, neste contexto, como um excelente paradigma para a implementação de buscas inteligentes. Sua forma e essência de trabalho permitem a contextualização da busca muito além da simples comparação por palavras-chave. Técnicas de indexação fundamentadas na utilidade e relevância dos casos tornam esta ferramenta muito poderosa com relação aos sistemas de Banco de Dados disponíveis.

A pesquisa em jurisprudência é um exemplo de raciocínio analógico, onde a resolução de um problema está fundamentada em resoluções anteriores de problemas similares. Este fator viabilizou a escolha deste campo para ser o enfoque do emprego da técnica de RBC, uma vez que RBC também está fundamentado no raciocínio analógico.

Através dos estudos realizados para a concretização da presente dissertação, pode-se constatar a possibilidade de utilizar os conceitos nesta empregados, também em outros domínios. Atualmente problemas como: suporte ao cliente (Help-Desk), suporte inteligente ao comércio eletrônico, apoio ao gerenciamento de redes utilizam a técnica de RBC no tratamento dos seus problemas.

## 1.2. Objetivo Geral

Desenvolver sistema inteligente que utilize a técnica de RBC como base de busca, análise e resolução em problemas com conhecimento textual e com metodologia de solução baseada na analogia.

### Objetivos Específicos

- aplicar o sistema proposto em raciocínio jurisprudencial;
- disponibilizar ferramenta a promotores, advogados e juizes, capaz de reduzir o tempo despendido na pesquisa jurisprudencial;
- incluir técnicas recentes de formação de casos a partir de textos jurídicos, permitindo alimentar a base de casos a partir de descrição textual;
- realizar levantamento teórico em informática jurídica, em particular, soluções disponíveis em pesquisa jurisprudencial;

### 1.3. Metodologia de Pesquisa

Conceitualmente o presente trabalho está dividido em três passos: 1) o estudo do domínio do problema (Direito), utilizado na validação do modelo, e sua relação com a informática; 2) a apresentação da técnica de IA, RBC, sendo utilizada como base para o modelo; 3) a apresentação do modelo para tratamento de documentos textuais e a sua validação através do protótipo desenvolvido. Esta metodologia pode ser vista na Figura 1.

No primeiro passo destaca-se o domínio do problema: o Direito e a informática, contendo três subdivisões, a saber:

- a) *Direito Legal & Informática*: apresenta-se o campo do Direito e os recursos tecnológicos empregados neste, com o intuito de proporcionar um melhor uso das informações jurídicas;
- b) *Soluções/Natureza dos sistemas encontrados e buscas previstas nos mesmos*: apresenta-se os sistemas tradicionais encontrados na literatura direcionados a área do Direito, bem como a natureza e formas de recuperação de informações contidas nestas aplicações;
- c) *Jurisprudências*: apresentam-se conceitos, informações gerais e os sistemas relacionados ao campo específico das jurisprudências;

No segundo passo descreve-se sobre a técnica de RBC, contendo duas subdivisões, a saber:

- a) *Raciocínio Baseado em Casos*: apresenta-se a arquitetura da técnica de IA, RBC, utilizada no desenvolvimento do trabalho;
- b) *RBC em Direito/Exemplos encontrados na literatura*: apresentam-se os estudos do uso da técnica de RBC à área do Direito, e exemplos de aplicações desenvolvidas utilizando a referida técnica.

No terceiro e último passo apresenta-se o modelo proposto, bem como a implementação prática do estudo realizado, contendo três subdivisões, a saber:

- a) *Modelo proposto*: apresenta-se o estudo realizado na elaboração do modelo proposto;
- b) *Aplicação*: apresenta-se o processo de funcionamento da aplicação desenvolvida;
- c) *Conclusões & futuros desenvolvimentos*: apresenta-se as conclusões do estudo realizado e proposições para trabalhos futuros.

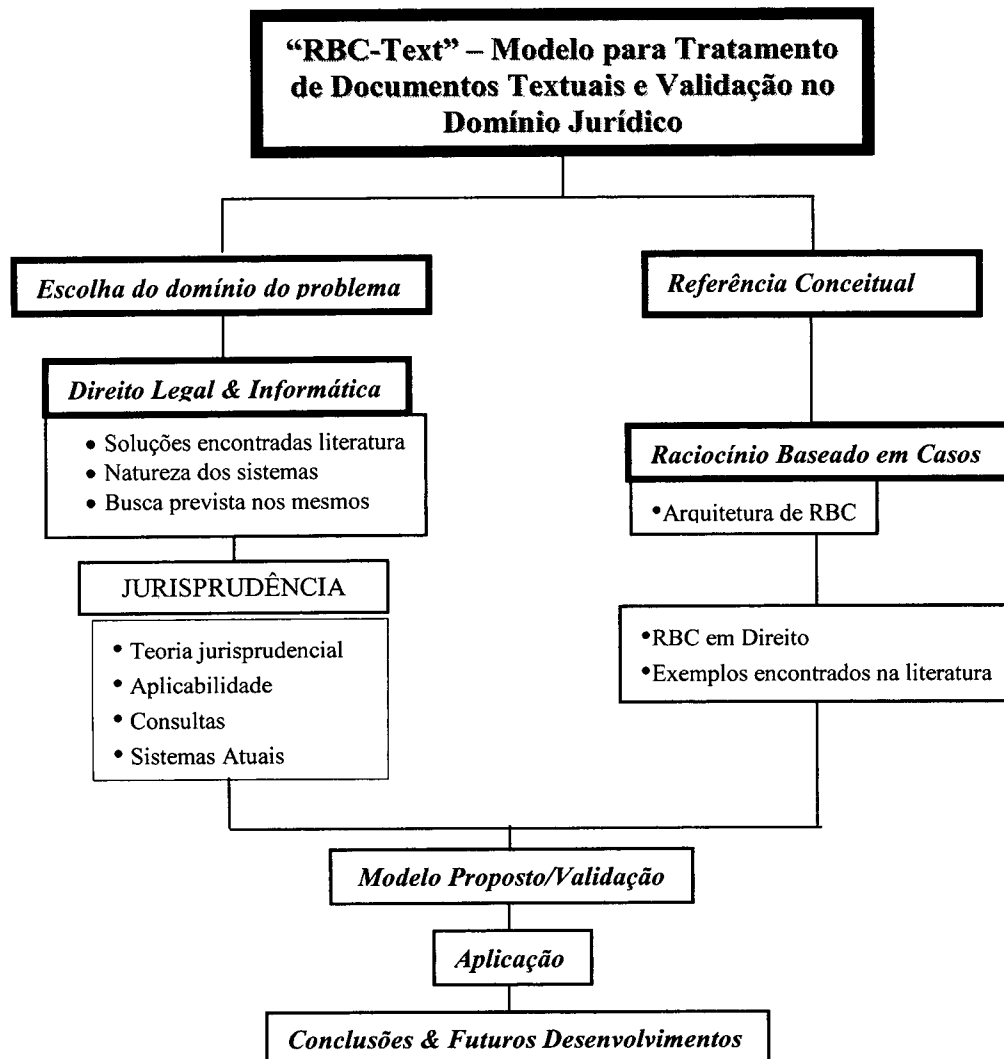


Figura 1 - Metodologia do trabalho



## 1.4. Estrutura

O trabalho estrutura-se da seguinte forma:

**Capítulo 1** – Contexto e justificativa do trabalho. Este capítulo apresenta os objetivos do trabalho bem como o domínio e os aspectos tecnológicos que serão utilizados na implementação do problema abordado.

**Capítulo 2** – Domínio da jurisprudência, da informática jurídica e os benefícios que a computação pode trazer ao Direito.

**Capítulo 3** – Traz o estudo da ferramenta de IA, Raciocínio Baseado em Casos, particularmente dos pontos relacionados ao trabalho.

**Capítulo 4** – Modelagem e estrutura do modelo. No capítulo 4 apresentar-se-á toda a análise realizada na elaboração e estruturação do modelo de uma aplicação utilizando a técnica de RBC em bases textuais exemplificada em documentos jurídicos.

**Capítulo 5** – Estrutura do Sistema. Aborda-se a modelagem do sistema de validação do modelo, bem como os algoritmos de extração de informações dos documentos.

**Capítulo 6** – Sistema desenvolvido. Este capítulo refere-se a apresentação do protótipo, na qual, explana-se os módulos da aplicação.

**Capítulo 7** – Conclusões e perspectivas futuras. Neste capítulo serão apresentadas as considerações finais com relação a pesquisa realizada nesta dissertação, bem como a apresentação de propostas futuras que nortearão a seqüência deste estudo.

**Capítulo 8** – Referências bibliográficas. No capítulo final constará as referências que suportaram esta obra.

## 1.5. Escopo e Delimitações do Trabalho

O escopo deste trabalho está delimitado à aplicação da tecnologia de RBC na recuperação de informações constantes em documentos não estruturados. Para que esta recuperação seja possível, apresenta-se um modelo para a extração do conhecimento contido nos documentos. O propósito do modelo será o de considerar documentos textuais que apresentem as seguintes características:

- base tecnológica no raciocínio analógico;
- vocabulário e regras de domínio próprio;
- casos descritos na forma textual, mas com normatização de formação.

A validação do modelo será realizada através da sua aplicação no domínio do Direito, mais precisamente nos documentos legais, acórdãos, que descrevem as decisões do Tribunal de Justiça do Estado de Santa Catarina (TJSC), uma vez que estes preenchem o requisitos supracitados.

Os registros do TJSC tem cerca de 90.000 trabalhos com descrições completas dos casos legais. Elas descrevem as experiências que formarão a base de conhecimento a ser explorada, ou seja, serão os casos do sistemas de RBC. Outra característica relevante neste trabalho é a alimentação automática da base de conhecimento, isto é, novos documentos são inseridos na base de casos sem a necessidade de uma avaliação prévia por parte do especialista. O protótipo foi elaborado considerando 2300 acórdãos.

A realização da etapa de engenharia do conhecimento, que é responsável pela modelagem dos casos que representarão os acórdãos, bem como pelas regras que possibilitarão valorar os atributos dos pares Atributo-Valor, viabiliza a inserção de novos casos na base sem a necessidade de uma avaliação prévia por parte do especialista.

## DIREITO LEGAL E INFORMÁTICA

Neste capítulo serão feitas abordagens ao domínio da jurisprudência, à informática jurídica e os benefícios que a computação pode trazer ao domínio legal. Observando as soluções encontradas na literatura, a natureza dos sistemas e da busca prevista nos mesmos.

### 1.6. Introdução

Saraiva (1994) observa que o Direito, em sua dimensão de ciência e em sua dimensão de realidade social, de *praxis*<sup>1</sup>, passa por um momento de crise em todo o planeta, gerando reações de toda ordem naquele que com ele lidam profissionalmente, da perplexidade à desilusão, da euforia à passividade. Cada vez mais, as demandas da sociedade dirigida aos operadores jurídicos se tornam complexas, o ensino jurídico tradicional não consegue atender às modernas necessidades dos trabalhadores jurídicos, a gama de informações necessárias ao exercício profissional atropela os atribulados operadores do Direito, e os modelos estatais de solução de conflitos caminham céleres<sup>2</sup> rumo ao esgotamento.

Verifica-se que a capacidade humana de raciocínio e compreensão torna-se impotente diante do grande volume de informações embutidas no domínio do processo. Sendo que o profissional jurídico sente a necessidade de elaborar respostas rápidas e com qualidade no intuito de satisfazer o anseio a sociedade, vê-se no domínio do Direito uma área fértil para aplicações baseadas na inteligência.

Literaturas de IA, vide capítulo 1.11. , indicam o emprego da técnica de RBC para os problemas que envolvem recuperação de documentos textuais. Haja vista que o domínio para a validação do modelo desta dissertação é a recuperação jurisprudencial, e sendo as jurisprudências documentos textuais, julga-se condizente a utilização do RBC no trabalho de recuperação das experiências armazenadas nos documentos legais.

### 1.7. Teoria da Jurisprudência

---

<sup>1</sup> Praxe - O que habitualmente se pratica; uso estabelecido.

Neste tópico serão apresentados alguns conceitos concernentes à teoria da Jurisprudência.

No artigo de Mirra, (1994), observa-se as seguintes definições:

- conjunto das decisões que promanam<sup>3</sup> dos tribunais, ao proclamarem o Direito, aplicando a lei aos casos concretos, ou seja, no sentido de orientação dominante e uniforme dos tribunais na decisão de casos semelhantes;
- representa *a forma viva do Direito* porque ela revela a maneira pela qual o Direito está sendo aplicado quotidianamente às relações humanas, à vida social. Tão importante se tornou a jurisprudência no estudo do Direito que ela passou a ser reconhecida inclusive com *fonte subsidiária do Direito*.

Em SILVA (1999) encontra-se a seguinte definição:

"JURISPRUDÊNCIA. derivado do latim *jurisprudencia*, de jus (Direito, Ciência do Direito) e *prudencia* (sabedoria), entende-se literalmente que é a ciência do Direito vista com sabedoria. Os romanos definiam-na, segundo Ulpiano, como o conhecimento das coisas divinas e humanas e a ciência do justo e do injusto: *divinarum atque humanum rerum notia, justii atque insjusti scientia*. E, segundo Demangeat, assim se exprimindo, Ulpiano quis mostrar que o verdadeiro jurisconsulto deve conhecer não somente a natureza divina, mas a natureza e o destino do homem, porque, para distinguir o justo do injusto, o moral do imoral, é preciso partir dos altos problemas filosóficos, que nos dão conhecimento das coisas divinas e humanas. Modernamente, é jurisprudência aplicada também no sentido de Ciência do Direito. Mas, como já anotavam os comentadores romanos, traz consigo um sentido subjetivo e outro objetivo, de modo que não significa simplesmente a noção científica das leis, ligada às capacidades de aplicação aos casos concretos, mas compreende um sistema de doutrinas, que tem por objeto os direitos e as obrigações (Cluck). É claro o sentido literal: o Direito aplicado com a sabedoria. Assim é que se entende a jurisprudência como sábia interpretação e aplicação das leis a todos os casos concretos que se submetem a julgamento da justiça. Ou seja, o hábito de interpretar e aplicar as leis aos fatos concretos, para que, assim, se decidam as causas. Desse modo, a jurisprudência não se forma isoladamente, isto é, pelas decisões isoladas. É necessário que se firme por sucessivas e uniformes decisões, constituindo-se em fonte criadora do Direito e produzindo um verdadeiro *jus novum*. É necessário pelo hábito, a interpretação e explicação das leis a venham formar. Os romanos sempre a consideram como a fonte do Direito,

<sup>2</sup> Ligeiro, veloz.

<sup>3</sup> Brotam, derivam.

designando-a com *auctoritas rerum perpetuo similiter judicatarum*, embora Justiniano aconselhasse que não se lhe desse uma autoridade exagerada, *cum non exemplis sed legibus judicandum sit*. Aliás, é firmado hoje que a jurisprudência somente obriga a espécie julgada, não sendo, propriamente, fonte de Direito. Mas, na verdade, é que a jurisprudência firmada, em sucessivas decisões, vale como verdadeira lei. Jurisprudência. Extensivamente assim se diz para designar o conjunto de decisões acerca de um mesmo assunto ou coleção de decisões de um tribunal“.

## 1.8. Relevância da Jurisprudência no Direito

Segundo Cintra et al., (1999), chamam-se fontes formais do Direito os meios de produção ou expressão da norma jurídica. Tais meios são as leis (em sentido amplo, abrangendo a Constituição e as leis em geral), os usos e costumes e o negócio jurídico.

Cintra et al., (1999) continua informando que é controversa a inclusão da jurisprudência entre as fontes de Direito: de um lado encontram-se aqueles que, partindo da idéia de que os juízes e tribunais apenas devem julgar de acordo com o Direito já expresso por outras fontes, dele não se podem afastar; de outro lado, os que entendem que os próprios juízes e tribunais, através de suas decisões, dão expressão às normas jurídicas até então não declaradas por qualquer das outras fontes.

Independente da visão que tenha prevalência, o RBC possui alternativas que satisfaça ambas. Pois, através da apresentação ou omissão da etapa de aprendizagem o jurista poderá armazenar ou não as suas decisões.

## 1.9. A Informática para o Direito

Este tópico apresenta os benefícios trazidos pela informática aos profissionais do Direito, os quais contribuíram de maneira considerável no sentido de dinamizar e minimizar o tempo gasto nas suas atribuições.

No artigo *A Importância da Informática para o Profissional de Direito*, Andrade (1998), descreve sobre como as novas tecnologias são importantes para os operadores do Direito. Ele aponta que uma das facilidades que a informática trouxe para os juristas foi o advento dos sistemas direcionados às pesquisas jurisprudenciais, e também o

fato que a internet consolidou todo um processo de informatização que facilitou o acesso do jurista à informação.

Saraiva (1994) comenta em seu artigo que, "A utilização dos computadores em larga escala na criação e aplicação do Direito é recente, não ultrapassando os últimos vinte anos. Todavia, já produziu mudanças significativas e indeléveis no *modus operandis* dos autores do Direito". Ele conclui afirmando que, "É, pois, tempo de rendermo-nos ao computador, não como quem capitula frente ao inimigo, nem o encarando como substituto do jurista - o Direito, ciência social complexa como é tratando de grandezas e realidades ontologicamente inexatas e mutáveis, jamais prescindirá do discernimento do homem para sua realização -, nem como eliminador do esforço pessoal, mas como ferramenta utilíssima e, hoje, indispensável à consecução da justiça que diuturnamente buscamos".

Baseado nestas informações consta-se que os profissionais do direito estão cientes das facilidades e benefícios trazidos pela informática. Existindo cada vez menos, rejeição com relação aos sistemas informatizados, no tocante ao temor pela substituição do homem pela máquina. Verifica-se, também, que o campo do Direito está evoluindo com relação à informatização dos seus documentos.

## **1.10. Consulta Jurisprudencial**

Este tópico apresenta uma síntese da evolução das consultas jurisprudenciais disponíveis, até o momento, aos operadores do Direito. A

Figura 2 ilustra tal evolução:

Inicialmente os operadores do Direito faziam suas consultas em livros. Eram necessários dias de trabalho nas prateleiras das bibliotecas em busca dos textos legais almejados. Nos escritórios havia a necessidade de reservar um espaço amplo para guardar o acervo bibliográfico, o qual sofria constante atualização. O custo para manter acomodações amplas, ou ampliá-las, passou a ser muito oneroso, dificultando a guarda do material de apoio (Atheniense, 1996).

Com o advento da informática, surgiram sistemas de banco de dados, que auxiliaram e conseqüentemente facilitaram as pesquisa dos juristas. Fazendo com que o tempo dispensado nas consultas ao material bibliográfico fosse reduzido consideravelmente. Entretanto, tais sistemas continuam tomando muito tempo dos profissionais, pois trazem um

número demasiado de informações. E suas opções de refinamento nem sempre são facilitadoras. Muitos profissionais não entendem seu funcionamento por acharem muito técnicas.

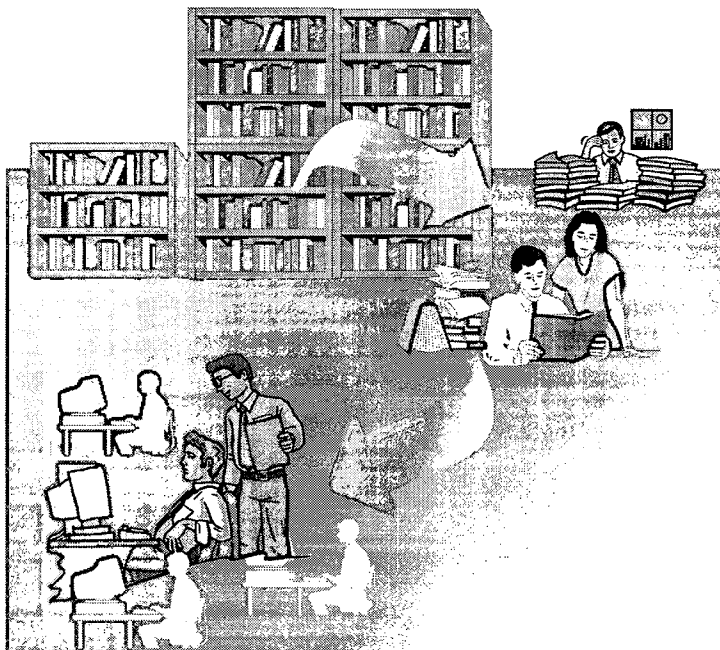


Figura 2 - Evolução da consulta jurisprudencial

### 1.10.1. Consultas em banco de dados

As consultas em banco de dados são realizadas através de palavras-chave. O grande problema da utilização pura de palavras-chave é a falta de flexibilidade, para que o usuário possa fazer uma consulta mais personalizada e mais perto da realidade do problema a ser solucionado. Os recursos oferecidos para que a consulta seja melhor elaborada, geralmente, são os operadores lógicos &, OR, NOT, XOR (veja Figura 3). Estes operadores, normalmente, não contribuem muito para uma consulta mais qualificada, pois os profissionais do Direito não os utilizam com muita frequência. Um dos principais motivos para esta não utilização, é o não entendimento de suas funções.

Segundo, Weber-Lee (2001), a consulta básica dos bancos de dados implica no uso de um conector, o que limita a recuperação somente de documentos que necessariamente contenham todas as palavras presentes na consulta. Este formato impede a recuperação de documento que possua qualquer valor alternativo para os índices. Conforme a visão dos especialistas, um documento que contenha o mesmo conjunto de expressões com

até duas ou três alterações ainda pode ser útil. Uma solução para este problema seria o uso de conectores tais como ou e XOR normalmente disponíveis. O uso do XOR também é exclusivo e o uso do conector ou acarreta um dilema de custo-benefício, na medida em que mais documentos podem ser recuperados, baixando o índice de precisão dos resultados.

Noortwijk (2000) aponta que os usuários do Direito precisam fazer certos malabarismos para elaborarem suas consultas, por conta da utilização dos operadores booleanos. Os juristas sabem o assunto que precisam pesquisar, porém, eles precisam especular quais palavras devem indicar para que a matéria mais adequada seja recuperada. Palavras comuns ou muito específicas devem ser evitadas. As comuns podem estar em quase todos documentos e as muito específicas podem estar em número muito reduzido. E, finalmente a saída apresenta muitos documentos, e na maioria das vezes não apresenta o grau de relevância entre eles.

**Pesquisa na Infobase:  
JURISPRUDÊNCIA DO TRT DA PRIMEIRA REGIÃO**

Entre com dados para a sua pesquisa: (Se for frase, colocar entre aspas. Usar nos demais casos os operadores &, OR, NOT, XOR. Consultar o botão Ajuda em caso de dúvida.)

ASSUNTO:

TIPO E Nº DO PROCESSO:  TURMA:

DATA DE JULGAMENTO:  (DD-MM-AA)

DATA DE PUBLICAÇÃO:  (DD-MM-AA)

JUIZ RELATOR:

PESQUISA LIVRE:

Pesquisar   Limpar   Ajuda   Assuntos   Voltar

Figura 3 - Pesquisa jurisprudencial (www.trtrio.gov.br)

Outro ponto negativo apresentado pelos sistemas de banco de dados, é justamente, o refinamento. Quando uma consulta retorna um número muito grande de registros, o usuário pode indicar mais algumas palavras-chave com intuito de refinar a sua pesquisa. Portanto, mesmo que a quantidade de registro de saída seja menor, não necessariamente isto significa um retorno com mais qualidade. Pois pode-se perder informações importantes após a indicação de novas palavras.

### 1.11. Inteligência Artificial no Direito e Recuperação Textual



A Inteligência Artificial está cada vez mais presente nas novas aplicações. Ferramentas utilizando técnicas da referida área (IA), como RBC, surgem para dinamizar as atividades dos operadores jurídicos. Baseados no paradigma do raciocínio analógico, estes sistemas buscam e apresentam experiências anteriores relevantes ao contexto atual. A utilização de uma interface amigável, a qual não exige um conhecimento técnico muito aprofundado, para que o usuário formule a sua pesquisa, facilita a sua operacionalização.

No artigo *A Discourse on Law and Artificial Intelligence*, Aikenhead (2000) aborda as técnicas de IA mais utilizadas no domínio do Direito. Ele menciona que os estudos no campo de IA e Direito vêm sendo realizados há longo tempo, e que atualmente estão surgindo resultados bastante significativos. As técnicas de IA que predominam são: Sistemas Especialistas e Raciocínio Baseado em Casos.

Aikenhead (2000) aponta que as duas técnicas, apesar de serem as mais eficientes para a área do Direito, ainda falham por não conseguirem contemplar na sua totalidade todas as características deste domínio. Porém, para a criação de sistemas inteligentes a partir das jurisprudências, ele indica a técnica de RBC como a mais apropriada, pelo fato de trabalhar com problemas que apresentam experiências anteriores.

Um ponto que vale salientar neste tópico é a formatação das jurisprudências. As jurisprudências são documentos semi-estruturados, ou seja, apresentam suas informações de maneira que possam ser identificadas certas características. Por exemplo: apresentam o relator, o juiz, a data do processo, o local, etc... Isto serve como um aporte para validar a técnica sendo empregada neste trabalho, pois segundo Lenz (2001) em muitas áreas as experiências estão disponíveis apenas em documentos textuais. E ele comenta que existe duas razões para estas informações requererem uma abordagem diferente da tradicional: Primeiramente, estes documentos são geralmente semi-estruturados, isto é, eles contêm componentes textuais, porém apresentam formatos específicos, tais como seções pré-definidas e informações estruturadas contidas no texto, como pares atributo-valor. Segundo, existe um vasto número destes documentos e uma vez que tal coleção é conhecida, ela geralmente cresce rapidamente.

Outra informação importante apresentada por Lenz (2001), é que um importante aspecto do RBC direcionado a documentos textuais, que o distingue de outras técnicas que trabalham com textos, tais como *recuperação da informação*, é habilidade para

considerar o conhecimento durante o processo de aquisição do conhecimento e assim capturar experiências de um domínio específico.

Em Hoeschl (2001), são apresentadas produções que integraram a conferência de IA e Direito no ano de 1999, realizada em Oslo, Noruega. Tal conferência, ICAIL, é considerada como o principal fórum de discussão do tema IA e Direito em nível mundial. A seguir, com base na obra de Hoeschl (2001), são destacadas algumas pesquisas que apresentam técnicas de IA, distintas das de RBC, sendo utilizadas no domínio jurídico, a saber:

- *Algoritmos Difuzos: AI Techniques for modeling legal negotiation.* Neste texto, Belluci e Zeleznikow (1999) *apud* Hoeschl (2001) 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*”.
- *Agentes Inteligentes e Sistemas Especialistas: Agents and norms: how to fill the gap?* Onde Conte, Falcone e Sartor (1999) *apud* Hoeschl (2001) 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; e 2. A da teoria de sistemas multiagentes (*multi-agents systems*, MAS) e aplicações computacionais relacionadas, especialmente nas áreas de 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 deôntica, cotejada com a teoria dos 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.

- *Agentes Inteligentes: Norms in artificial decision making*, texto no qual Boman (1999) *apud* Hoeschl (2001) 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.

No capítulo 1.18. , serão detalhadas as aplicações abaixo elencadas. Elas são direcionadas ao domínio do Direito e desenvolvidas segundo o paradigma da técnica de Raciocínio Baseado em Casos.

- JUDGE - 1986;
- HYPO - 1988;
- GREBE – 1991;
- MEDIATOR – 1985;
- PERSUADER – 1998;

## 1.12. Considerações Finais

Apresentou-se a área do Direito e como a informática vêm sido nela empregada. Os sistemas atuais têm como principal problema o excesso de itens nos resultados da buscas e a dificuldade de entendimento do uso de conectivos lógicos nas buscas.

A inteligência Artificial vem sendo aplicada ao Direito com destaque para Sistemas Especialistas e Raciocínio Baseado em Casos. Os primeiros fundamentam-se em princípios deducionistas, em áreas do Direito em que é possível modelar o conhecimento jurídico na forma de regras SE-ENTÃO. O problema desta abordagem está no escopo limitado de aplicações e na dificuldade de modelar situações do Direito em que a dedução não é genérica ou baseado em leis. RBC, por outro lado, aparece em sistemas baseados no conhecimento analógico, em áreas do Direito em que é possível modelar o conhecimento jurídico na forma de casos. O diferencial desta abordagem é que quando se modela o conhecimento embutido em uma situação/problema, pode-se generalizar tal modelagem para os demais problemas comuns.

## RACIOCÍNIO BASEADO EM CASOS (RBC)

Neste capítulo, aborda-se a técnica de RBC, sua definição, arquitetura, metodologia de desenvolvimento e aplicação em áreas semelhantes à proposta do trabalho.

### 1.13. Introdução

Raciocínio Baseado em Casos (RBC) é uma técnica de raciocínio da Inteligência Artificial (IA). O paradigma que fundamenta esta técnica é o de buscar uma solução para um caso de entrada através da comparação com casos similares passados. A maneira como RBC é utilizado caracteriza-se da seguinte forma: a partir de situação nova (problema de entrada) o sistema busca na base de casos por experiências similares, verifica a necessidade de adaptação e aplica o conhecimento da(s) experiências passadas para resolver uma nova situação (Weber-Lee, 1996).

Segundo Aamodt & Plaza (1999) RBC é a resolução de um novo problema através da lembrança de situações anteriores similares e da utilização de informações e conhecimentos daquela situação. Aamodt & Plaza (1999) continuam sua explanação com o seguinte exemplo: "Um médico - após examinar um paciente em seu consultório - recorda de um outro paciente que atendeu há duas semanas. Supondo que a lembrança foi causada por similaridades de sintomas importantes (e não a cor dos cabelos ou olhos), o médico usa o diagnóstico e o tratamento do paciente anterior para determinar a doença e o tratamento para o paciente atual".

Watson et al. (1994) utilizam o conceito de Riesbeck & Schank, onde um sistema de raciocínio baseado em casos resolve novos problemas adaptando soluções que foram usadas para resolver problemas antigos. Watson (1994) comenta, ainda, que o RBC na sua forma mais simples é baseado na observação de que quando o ser humano resolve problemas, ele geralmente baseia sua solução em uma utilizada para um problema similar anterior.

Segundo Kolodner (1991), em RBC, uma pessoa lembra de situações antigas similares com relação às atuais e as utiliza para ajudar na resolução dos problemas novos. RBC pode adaptar soluções antigas para encontrar novas necessidades, usar casos

antigos para explicar situações atuais, utilizar experiências anteriores para interpretar uma nova situação (similar ao que fazem os advogados) ou criar uma solução justa para um novo problema (similar ao que fazem os mediadores).

Raciocínio baseado em casos resolve novos problemas adaptando soluções anteriores bem sucedidas em problemas similares (Watson et al., 1994). Algumas características do RBC são:

- não requer um modelo do domínio explícito e, sendo assim, a tarefa de extração de conhecimento é realizada através de reunião de casos antigos;
- a implementação é reduzida para a identificação de características significantes que descrevem um caso, uma tarefa mais fácil do que a criação de um modelo explícito;
- a aplicação das técnicas de banco de dados em grandes volumes de informações pode ser gerenciada; e
- os sistemas de RBC podem aprender através da aquisição de novos conhecimentos na forma de casos, tornando a manutenção mais fácil. (Watson et al., 1994).

Sistemas especialistas ou baseados no conhecimento (KBS - Knowledge-based systems) são um dos sucessos históricos da Inteligência Artificial. Em uma pesquisa realizada no Departamento do Comércio e Indústria do Reino Unido foram encontrados mais de 2000 KBS em operações comerciais (a pesquisa excluiu os KBS dos laboratórios de pesquisa das universidades) (DTI, 1992 *apud* Watson et al., 1994).

Entretanto, em muitos setores, desenvolvedores destes sistemas têm encontrado alguns problemas:

- dificuldade no processo de extração do conhecimento;
- dificuldade no processo de implementação do KBS, requerendo conhecimento específico do domínio e muitas vezes muitos homens ano;
- os KBS implementados freqüentemente são lentos e incapazes de acessar ou gerenciar um grande volume de informações; e

- dificuldades na manutenção do KBS; (Bachant & McDermot, 1984; Coenen & Bench-Capon, 1992; Watson et al., 1992b *apud* Watson et al., 1994)

Soluções para estes problemas foram encontradas através de melhores técnicas de extração (Motta et al., 1989; Brooke & Jackson, 1991 *apud* Watson et al., 1994) e ferramentas melhores que as shells e ambientes de KBS, metodologias de desenvolvimento aperfeiçoadas (Diaper, 1989; Inder & Filby, 1991; Watson et al., 1992a; Wielinga et al., 1992 *apud* Watson et al., 1994), linguagens de modelagem do conhecimento e ontologias (Alexander et al., 1986; Steels, 1990; Chandraskaren 1986 & 1990; Wielinga et al., 1992 *apud* Watson et al., 1994), facilitando a integração entre KBS e base de dados em base de dados especialistas e base de dados dedutivas (Gallaire et al., 1981; Minker, 1988; Marir & Yip, 1992; Marir 1993 *apud* Watson et al., 1994), e técnicas e ferramentas para a manutenção de sistemas (Bench-Capon & Coenen, 1992, Watson et al., 1992b *apud* Watson et al., 1994).

Aamodt (1994) acrescenta que RBC é um paradigma de resolução de problemas que em muitos aspectos é fundamentalmente diferente de outras técnicas de IA. Em vez de contar somente com o conhecimento genérico de um domínio de problema, ou fazer associações junto com relacionamento genérico entre descritores de problemas e conclusões, RBC é capaz de utilizar o conhecimento específico de experiências anteriores, situações de problemas concretos (casos). A segunda diferença importante é que RBC também é uma abordagem incremental, mantém a aprendizagem, desde que uma nova experiência seja retida cada momento que um problema é resolvido, tornando-o imediatamente disponível para a resolução de problemas futuros (Aamodt et al., 1994).

Nos últimos anos, RBC vem evoluindo de uma específica e isolada área de pesquisa para um campo de vasto interesse. Várias atividades estão sendo exploradas, como é visto no crescente número de artigos, avaliação de produtos comerciais entre outros tipos de estudos realizados segundo este paradigma (Watson et al., 1994).

Para um melhor entendimento do funcionamento da técnica de RBC, são ressaltados os seguintes passos:

- ◆ identificação de um problema a ser resolvido (problema de entrada);

- ◆ definição das principais características que identificam este problema;
- ◆ busca e recuperação na memória de casos com características similares;
- ◆ seleção de um ou mais dentre os casos recuperados;
- ◆ revisão deste(s) caso(s) para determinar a necessidade de adaptação;
- ◆ reutilização do caso adaptado para resolver o problema de entrada;
- ◆ avaliação da solução do problema de entrada e inclusão do caso adaptado na memória de casos (aprendizagem).

#### **1.14. Histórico do Raciocínio Baseado em Casos**

Raciocínio Baseado em Casos é uma técnica direcionada aos campos de resolução de problemas e aprendizagem. Originária nos Estados Unidos, atualmente vem conquistando um grande espaço entre os pesquisadores dos outros continentes, também. Com destaque para o continente europeu, onde a Alemanha desponta-se como o país com mais trabalhos na área. Este crescimento é comprovado através do considerável aumento no número de *papers* e produção de produtos à luz de tal paradigma.



Na Tabela 2 e Tabela 3 um breve histórico do RBC nos Estados Unidos e Europa, respectivamente, demonstrado em (Bergmann, 1998):

### Histórico do RBC no Estados Unidos

Ano	Pesquisador(es)	Instituição	Pesquisa	Obra
1977	Roger Schank	Yale University	Cognitive Science	Scripts para representação do conhecimento (Schank, Abelson)
1983	Roger Schank	Yale University	Cognitive Science	Teoria da Memória Dinâmica, Pacotes de Organização de Memória. CYRUS: Primeiro sistema de RBC implementado (Kolodner)
1983-1988	Roger Schank	Yale University	Cognitive Science	Outros sistemas, Ex...: JUDGE, SWALE, CHEF
1986-89	Bruce Porter	Austin Texas	Concept Learning	Sistema PROTOS (Representação dos conceitos baseados em exemplares)
1990-92	Edwina Rissland	U. of Massachusetts	Cases in Law (desde 1983)	Sistema HYPO (Ashley) e CABARET (Skalak)
Desde 1990	Jaime Carbonell & Manuela Veloso	Carnegie Mellon U	Analogy	Prodigy/Analogy: Case-based Planning using analogy
Desde 1988				Vários DARPA e Workshops AAAI
<b>Observação geral:</b> Aumento do interesse em RBC no Estados Unidos (são criados vários grupos de pesquisadores)				

Tabela 2 - Histórico do RBC nos EUA

Ano	Pesquisador(es)	Instituição	Pesquisa	Obra
1988-1991	Michael M. Richter	U. Kaiserslautern, Germany	CBR for Expert Systems	Sistemas MOLTKE e PATDEX (Diagnóstico técnico)
Desde 1991	Michael M. Richter	U. Kaiserslautern, Germany	CBR for Expert Systems	Planejamento Baseado em casos: Sistemas Caplan/CbC, PARIS
Desde 1992	Michael M. Richter	U. Kaiserslautern, Germany	CBR for Expert Systems	Projetos Europeus: INRECA, INRECA-II
1990	Ramon Mantaras, Enric Plaza	IIIA Blanes, Spain	CBR and ML	Aprendizagem Baseada em Casos para diagnóstico médico
1991	Agnar Aamodt	U. Trondheim, Norway	CBR and Knowledge Acquisition	Sistema CREEK: Integração dos Casos e Conhecimento Genérico
Desde 1988	Mark Keane	Trinity College, Dublin	Cognitive Science	Teoria do Raciocínio Analógico
1991				Primeiro German CBR Workshop (AKCBBR, GWCBBR)
1993				Primeiro European CBR Workshop (EWCBBR)
1995				Primeiro International CBR Conference (ICCBR)
Desde 1991				Aumento do interesse em RBC na Europa (são criados vários grupos de pesquisadores)

Tabela 3 - Histórico do RBC na Europa

O trabalho de Schank e Abelson em 1977 é considerado como origem técnica de RBC. Eles propuseram que o conhecimento humano sobre situações é armazenado na forma de *scripts*, os quais permitem montar expectativas e gerar inferências. Os *scripts* foram apresentados como uma estrutura para os descritores de informações de uma memória conceitual sobre eventos estereótipos tais como, ir a um restaurante ou fazer uma consulta médica. Entretanto, as experiências com os descritores de informações mostraram que eles não representavam uma teoria para a representação da memória - as pessoas geralmente confundiam os eventos que tinham descritores similares (Watson et al., 1994).

Roger Schank continuou a explorar o papel da memória de situações anteriores (casos) e situações padrões ou pacotes de organização de memória (MOP - Memory Organization Packets) cumpriam na solução e aprendizagem dos problemas.

Outros pontos de origem dentro do campo de RBC vêm dos estudos do Raciocínio Analógico, realizados por Gentner em 1983, e da teoria da Formação de Conceito, Solução de Problemas e Aprendizagem através de Experiências dentro da Filosofia e Psicologia. (Wittgenstein, 1953; Tulving, 1972; Smith, 1981 *apud* Aamodt et al., 1994). Por exemplo, Wittgenstein observou que conceitos que fazem parte do mundo real, tais como, pássaro, cadeira, carro, etc, são polimórficos. Isto é, suas instâncias podem ser categorizadas por várias formas, e não é possível propor uma definição única em termos de um conjunto de características necessárias e suficientes, mas sim através de um conceito mais amplo definido pelo seu conjunto de instâncias ou casos (Aamodt et al., 1994).

Janet Kolodner, integrante do grupo de trabalho de Roger Schank, desenvolveu o primeiro sistema baseado na técnica de RBC em 1983. O sistema, denominado CYRUS, foi elaborado sob os modelos de memória dinâmica (MOPs), e possuía em sua base de casos, dados sobre as viagens e reuniões do Secretário de Estado Americano Cyrus Vance. O modelo de memória de casos do CYRUS foi utilizado posteriormente como base para vários outros sistemas, tais como MEDIATOR (Simpson, 1985 *apud* Watson et al., 1994), CHEF (Hammond, 1986 *apud* Watson et al., 1994), PERSUADER (Sycara, 1987 *apud* Watson et al., 1994), CASEY (Koton, 1989 *apud* Watson et al., 1994) e JULIA (Hinrichs, 1992 *apud* Watson et al., 1994).

O trabalho de Bruce Portes, na Universidade de Texas em Austin, também foi de grande importância para que a técnica de RBC se destacasse. Portes desenvolveu o PROTOS, o qual era direcionado para classificação de heurísticas e aprendizagem de

máquina. PROTOS uniu conhecimento genérico do domínio com conhecimento específico em um único modelo de memória de casos. Este trabalho resultou em outro sistema denominado GREBE, direcionado ao domínio legal (Branting, 91 *apud* Watson et al., 1994).

Os componentes principais do GREBE são um gerador de explicações, um gerador de memorando e uma base de conhecimentos de regras e casos. Quando o gerador de explicações recebe um novo caso e uma proposição sobre o caso, ele tenta construir uma explicação para a proposição através de encadeamento para trás (*back-chaining*), usando qualquer combinação de regras e casos anteriores. Se encontrar uma solução eficaz, o gerador de explicações gerará uma ou mais explicações da proposição. O gerador de memorando irá ordenar estas explicações pela avaliação heurística dos seus pesos. O GREBE, então, gerará uma saída em linguagem natural equivalente às explicações, utilizando uma biblioteca de modelos (*templates*) (Branting, Karl L, 1991).

Outro trabalho com contribuições significantes para o crescimento do RBC foi o sistema HYPO, desenvolvido por Edwina Rissland e seu grupo na universidade de Massachusetts. Com um grupo formado por vários cientistas da área do Direito, surgiu o interesse em trabalhar com o raciocínio precedente em julgamentos legais (Rissland, 1983 *apud* Aamodt et al., 1994). Os casos (precedentes) não foram utilizados para simplesmente produzirem uma resposta, mas sim interpretar uma situação na corte, e produzir e avaliar argumentos para as duas partes, apelante e réu (Ashley, 1990 *apud* Aamodt et al., 1994). Futuramente o HYPO combinou Raciocínio Baseado em Casos com Raciocínio Baseado em Regras, resultando no Sistema CABARET (Skalak, 1992 *apud* Aamodt et al., 1994).

As pesquisas no campo de RBC não ficaram restritas somente aos Estados Unidos. Porém, os primeiros resultados alcançados no continente europeu ocorreram a partir de 1988. Entre os primeiros trabalhos estava o sistema REFINER, desenvolvido pelo grupo de Derek Sleeman, na cidade de Aberdeen, Escócia. Eles estudaram o uso de casos para a aquisição de conhecimento (Sharma & Cleeman, 1988 *apud* Watson et al. 1994).

Na universidade de Kaiserslautern, Alemanha, Michael Richter e Klaus Althoff aplicaram RBC em diagnósticos complexos. Tais estudos resultaram no sistema PATDEX e mais tarde para ferramenta de RBC, S3-Case (Richter & Weiss, 1991 *apud* Watson et al. 1994). Agnar Aamodt, na universidade de Trondheim, Noruega, desenvolveu a parte da aprendizagem do RBC e a combinação dos casos e conhecimento genérico, resultando no Sistema CREEK. (Asmodt, 1989 & 1991 *apud* Watson et al. 1994).

No lado da ciência cognitiva, Mark Keane realizou um trabalho utilizando raciocínio analógico, no Trinity College, Dublin (Keane, 1988 *apud* Aamodt et al., 1994). E no grupo de Gerhard Strube, Universidade de Freiburg, o papel do conhecimento episódico no modelo cognitivo foi desenvolvido no projeto EVENTS.

### **1.15. Representação do caso**

Um caso é uma abstração de uma experiência descrita através de atributos devidamente valorados. Para descrever precisamente um caso, os atributos precisam detalhar não somente o conteúdo da experiência, mas também apresentar o contexto de ocorrência da experiência. A busca da resposta para o que seja um caso é bem mais complexa do que a delimitação do tipo de problema que pode ser utilizado num sistema de RBC. É possível tentar buscar o mesmo tipo de solução através de diferentes formulações do problema. A definição do que realmente seja um caso requer um estudo elaborado, pois poderá garantir o sucesso ou fracasso do sistema (Weber-Lee, 1996).

### **1.16. Estrutura do Raciocínio Baseado em Casos**

Neste tópico serão abordadas as principais etapas consideradas pelos pesquisadores da referida técnica. Alguns autores, como Aamodt e Plaza (1994), consideram um ciclo geral na definição das etapas envolvendo a técnica de RBC. O ciclo abordado é descrito por quatro processos, a saber: Recuperação, Reutilização, Revisão e Retenção, Figura 4.

A descrição inicial de um problema (topo da figura) define um novo caso. Este novo caso é usado para RECUPERAR um caso de uma coleção de casos antigos. O caso recuperado é combinado com o novo caso – através da reutilização – em um caso resolvido, isto é, uma solução proposta para o problema inicial. Através do processo de REVISÃO esta solução terá sua eficácia testada e será reparada se apresentar falha. Durante a RETENÇÃO, experiências úteis serão armazenadas para serem utilizadas futuramente, e com isto, a base de casos é atualizada através dos novos casos aprendidos ou pelas modificações de alguns casos existentes (Aamodt et al., 1994).

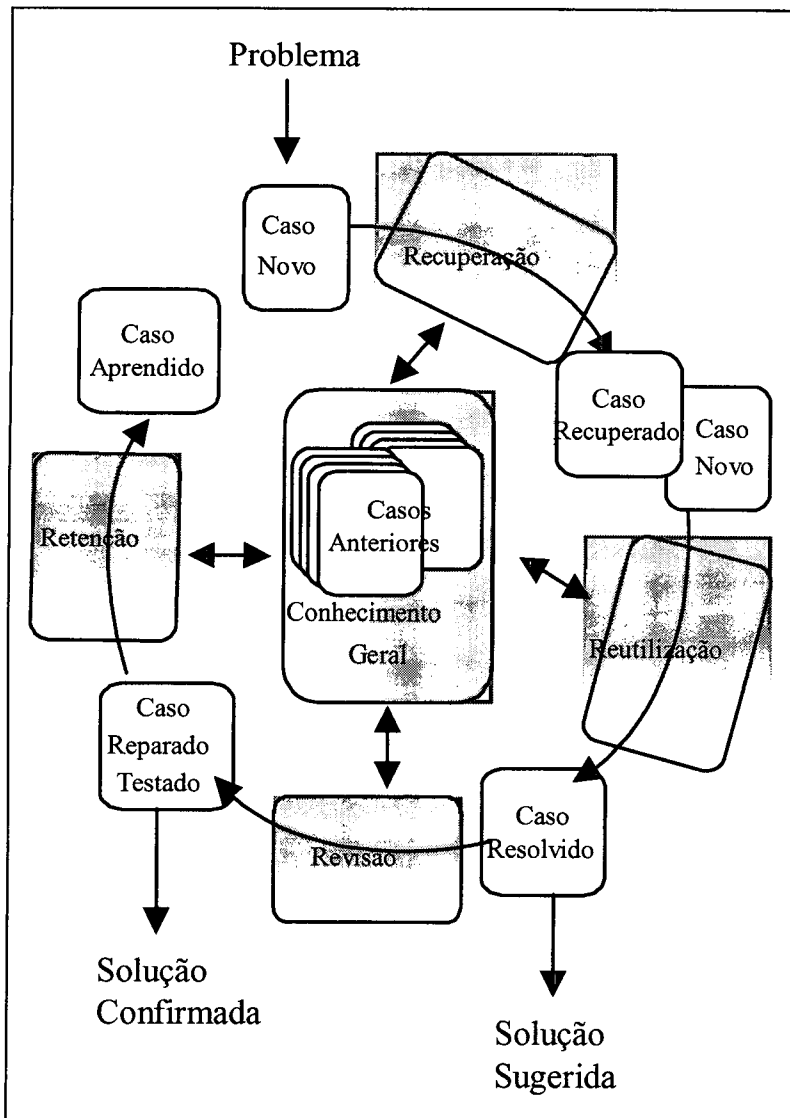


Figura 4 - Ciclo do RBC – (Aamodt et al., 1994)

### 1.16.1. Representação do Caso

A representação do conhecimento num sistema de RBC refere-se à determinação do que é um caso, como modelar a base de casos e como organizá-los. O conhecimento nos sistemas baseados em casos é essencialmente representado através da base de casos. São os casos que contém elementos para que a solução do problema proposto seja alcançada.

Esta seção é composta por:

- Modelagem dos Casos.
- Modelagem da Memória.
- Indexação.

## a) Modelagem dos casos

Na representação dos casos há dois componentes básicos: a descrição do problema e a descrição da solução. A descrição do problema é realizada através da atribuição de características que descrevem o problema de entrada. A descrição da solução consiste em determinar quais características descrevem a solução do caso, apontando a solução do problema de entrada e informando qual o resultado da aplicação desta solução ao problema de entrada. (Weber-lee, 1996).

## b) Modelagem da Memória

A base de casos consiste na coleção de casos que representam a base de conhecimento de um sistema de RBC. A memória compreende a base de casos e os mecanismos de acesso dessa base a outros módulos da arquitetura do sistema (Kolodner, 1993).

Segundo Slade (1991), o conhecimento permite que uma pessoa projete e resolva problemas. Conhecimento é um recurso, uma mercadoria. Memória é o repositório do conhecimento. A questão para os pesquisadores de IA é como representar o conhecimento no programa de computador e aplicá-lo na execução de tarefas específicas.

Apesar de ser um dos primeiros modelos de representação de conhecimento na IA, o modelo de memória com o uso de redes semânticas não é capaz de representar o conhecimento em sua totalidade, pois geralmente representa fatos estáticos sobre o mundo, tais como “Fido é um cachorro”, “um cachorro é um mamífero”. Em geral, este tipo de conhecimento não muda (Slade, 1991).

Como consequência dos esforços dos psicólogos também interessados na busca de um modelo de representação, Tulving (1972, 1983) propôs a teoria da memória episódica como um complemento para a memória semântica. Tulving descreveu memória semântica e episódica como dois sistemas de processamento de informações complementares. Memória semântica e episódica diferem no tipo de informação armazenada, referência autobiográfica versus referência cognitiva, condições de recuperação e consequências, a volatilidade das informações armazenadas e interdependência (Slade, 1991).

No intuito de encontrar um paradigma capaz de representar o conhecimento, os pesquisadores de IA continuaram sua busca, com os melhores resultados, em termos de raciocínio analógico, sendo alcançados pela equipe de Roger Schank (1972-1975). Suas pesquisas desencadearam o desenvolvimento do modelo de memória conceitual, dos scripts - análogos aos frames de Minsky - e, posteriormente, dos “MOPS” (*memory organization packets*), pacotes de organização de memória, e do modelo de Memória Dinâmica (Slade, 1991).

Dentro do escopo de RBC, ao tratarmos de modelagem de memória existem dois aspectos que precisamos enfocar separadamente. O primeiro trata-se do tipo de filosofia de representação que um dado sistema simula, que pode ser, entre outros, redes semânticas, memória episódica ou memória dinâmica. O segundo, a partir de um enfoque de implementação, busca a modelagem em termos da forma de estruturação da memória de casos. O estilo de organização dos casos na memória é escolhido de acordo com a natureza dos dados, com a forma de acesso aos casos e com outras etapas de desenvolvimento do sistema de RBC. Recuperação, adaptação e aprendizagem são as etapas importantes que devem ser projetadas com antecipação e consideradas na escolha do estilo de organização dos casos (Weber-Lee, 1996).

### c) Indexação

Os casos são representados a partir de um conjunto de características. Algumas destas características são usadas para orientar a recuperação, a isto se denomina *vocabulário de indexação*.

A indexação dos casos determina a similaridade, para fins de recuperação, ajudando na organização dos casos em função de uma ou mais características. Fazendo uma analogia com os índices de livros, é de extrema relevância que se determine nos casos os pontos mais importantes, para que sejam comparados com outros casos e recuperados os que apresentem um grau de similaridade suficiente para serem úteis na resolução do problema de entrada. A fim de proporcionar uma recuperação e reutilização dos casos de uma forma mais eficiente, é necessário que eles sejam propriamente indexados (Weber-Lee, 1996).

O vocabulário de índices de um caso está representado pelas características que são relevantes para a descrição dos casos e que orientarão a similaridade. A atribuição de índices é o processo de designação do vocabulário de índices (Kolodner, 1993).



Kolodner (1991) observa alguns pontos necessários para uma boa indexação, que segue:

- 1) os índices devem ser preditivos;
- 2) os índices devem ser abstratos o suficiente para fazer com que um caso seja útil em variedade de situações futuras;
- 3) os índices devem ser concretos o suficiente para serem reconhecíveis nos casos futuros; e
- 4) as predições a serem realizadas devem ser úteis.

O sucesso do vocabulário de índices será determinado pelo conhecimento do domínio apresentado pelo responsável pela atribuição de índices. Isto faz da atribuição de índices também um processo de representação de conhecimento.

Racine, 1996, considera o diagnóstico de um carro e seu provável reparo como exemplo de indexação. As características importantes extraídas deste domínio foram: a marca do carro, o modelo, o ano, o tipo de motor e a quilometragem. A Tabela 4 ilustra os casos através dos seus pares atributos/valores:

<b>CASE 1</b>		
<i>Marca:</i> Mazda	<i>Modelo:</i> 626	<i>Ano:</i> 1988
<i>Motor:</i> 20L EFI	<i>Quilometragem:</i> 20.114	
<i>Problema:</i> motor está apagando		
<i>Procedimento de validação:</i> injetor de combustível está obstruído		
<i>Reparo:</i> Limpar o injetor de combustível do Mazda 88		
<b>CASE 2</b>		
<i>Marca:</i> Toyota	<i>Modelo:</i> Camery	<i>Ano:</i> 1987
<i>Motor:</i> 2.8 L	<i>Quilometragem:</i> 108.121	

<b>Problema:</b> O marcador de combustível está desregulado.
<b>Procedimento de validação:</b> A bomba de combustível está quebrada.
<b>Reparo:</b> Repor a bomba de combustível do Toyota 87.

Tabela 4 - Exemplo de casos para diagnóstico e reparo automobilístico

### 1.16.2. Recuperação dos Casos

A etapa de recuperação, a partir do problema de entrada, parte da identificação das características deste problema. Após, faz-se uma busca na base de casos com o intuito de selecionar os casos mais similares, ou seja, casos que poderão ter aproveitamento na solução do problema que está sendo analisado (problema de entrada). A busca por casos é realizada através da avaliação e comparação da similaridade codificada por algoritmos ou funções que selecionam casos com determinada similaridade em relação ao problema de entrada (Weber-Lee, 1996).

Ainda trabalhando no conceito de recuperação, Ian Watson & Farhi Marir (1994) explicam que, a partir de uma descrição de um problema, um algoritmo de recuperação, usando os índices da memória de casos, deverá recuperar os casos mais similares ao problema inicial. O algoritmo de recuperação conta com os índices e a organização da memória para procurar pelos casos potencialmente mais úteis.

O resultado da recuperação é chegar ao conjunto de casos mais similares ao problema de entrada, apresentado-os em ordem de similaridade. Destes, um caso é selecionado ou alguns casos são combinados para compor a sugestão para o problema de entrada. As tarefas envolvidas na etapa de recuperação de casos são:

- Busca.
- Avaliação e Métrica da Similaridade.
- Seleção (Weber-Lee, 1996).

#### a) Busca

A busca ou recuperação dos casos mais similares é a etapa que avalia a similaridade de todos os casos da memória com o caso que representa o problema de entrada.

Segundo algum critério de avaliação desta similaridade e um determinado limiar, o resultado desta busca é um conjunto de casos mais similares (Weber-Lee, 1996).

## b) Avaliação e métrica da similaridade

Partindo da premissa que fundamenta a técnica de RBC, a similaridade é ponto crucial desta técnica. Pois é em virtude de haver um caso similar ao que está sendo avaliado, que o processo de raciocínio torna-se viável.

Na etapa de recuperação, a primeira questão a ser analisada é a similaridade entre o caso a ser solucionado (problema de entrada) e os casos candidatos. Somente dentro do domínio de uma determinada aplicação é que se saberá o que faz um caso ser similar ao outro. Um caso é similar ao outro quando são semelhantes as suas características que representam o conteúdo e o contexto da experiência. As características relevantes de um problema são aquelas que, combinadas entre si, são determinantes na escolha da solução para o seu problema. A métrica da similaridade é uma função que mede a similaridade entre dois casos, ela é utilizada para guiar a busca pelos casos mais similares que serão ordenados segundo a mesma (Weber-Lee, 1996).

Ian Watson & Farhi Marir (1994) apresentam alguns conhecidos métodos para recuperação de casos, a saber: *vizinho mais próximo*, *indução*, *indução guiada pelo conhecimento* entre outros. Eles podem ser usados sozinhos ou combinados em sistemas de recuperação híbridos.

## c) Seleção

Seleção é a última etapa da fase de recuperação, ela é mais específica do que a busca pelo conjunto de casos mais similares. A seleção pode ser implementada de várias formas, a mais óbvia é revisar os elementos que determinam a similaridade entre os casos, visando uma comparação mais apurada. A importância desta etapa se dá ao fato de que ela gera o resultado para solução do problema, ou seja, será a saída do sistema (Weber-Lee, 1996).

### 1.16.3. Adaptação

Após a recuperação do caso mais útil, o sistema de RBC poderá adaptar a solução armazenada no caso recuperado levando em consideração o caso de entrada. Na fase de adaptação procura-se por diferenças entre o caso de entrada e o recuperado, e então se aplica fórmulas ou regras satisfazendo o contexto em que a solução será apresentada (Watson et al., 1994).

A etapa de adaptação não pode ser implementada em todos os domínios. Uma das áreas que oferece restrições a aplicabilidade desta etapa é o domínio jurisprudencial. Pois para que uma jurisprudência seja válida há a necessidade que esta passe por vários processos legais. Sendo assim não se pode alterar/adaptar este tipo de documento constante em uma base de conhecimento.

### 1.16.4. Aprendizagem

Este é o processo de incorporar as informações úteis provindas do conhecimento existente nos novos problemas resolvidos. O aprendizado dos casos que apresentam sucessos ou falhas é disparado pelas tarefas de avaliação e adaptação de soluções. Este processo envolve a seleção das informações dos casos a serem retidos, qual a forma de armazená-las, como indexá-las para que participem do processo de *matching* e como integrar os novos casos na estrutura da memória (Aamodt et al., 1994).

## 1.17. Validação

Segundo Weber-Lee, 1996, a validação de um sistema de RBC pode ser feita adequando os métodos de validação de outros sistemas inteligentes, porém os sistemas baseados em RBC possuem um conjunto de técnicas de validação exclusivas. Uma das questões exclusivas dos sistemas de RBC, avalia se a tarefa para qual o sistema é proposto, quando executada por especialistas humanos, realmente baseia-se em casos.

Em termos gerais, estas são algumas das características a serem consideradas para a validação de um sistema de RBC:

- ◆ características técnicas: tais como estabilidade e operacionalidade;
- ◆ escolha do problema: se o problema é próprio para o tipo de raciocínio;

- ◆ características organizacionais: se o sistema é adequado à operação dentro de uma organização;
- ◆ características econômicas: retorno do investimento, aumento na qualidade de serviços;
- ◆ características estratégicas: se a memória de casos pode ser vista como um ativo;
- ◆ qualidade e eficiência com relação às principais etapas de um sistema baseado em casos: recuperação, adaptação, representação dos casos e aprendizagem;
- ◆ com relação à aprendizagem: além de avaliar a eficiência e qualidade, deve ser considerado se o aumento de robustez resultante da aprendizagem irá realmente beneficiar a qualidade do sistema ou diminuir sua velocidade, utilidade e eficiência;
- ◆ avalia-se o sistema comparativamente a especialistas humanos em dois momentos: num primeiro, apenas compara-se o sistema ao especialista; num segundo momento, o especialista utiliza o sistema como um colega ou um assistente técnico. Nesta etapa, calcula-se os percentuais em que o especialista utilizou as sugestões oferecidas pelo sistema.

### **1.18. Aplicações**

Três são os tópicos pertinentes às aplicações considerados relevantes: quais os tipos de aplicações são adequados; quais as ferramentas disponíveis para implementação; e quais são as importantes aplicações que podem orientar e ilustrar o potencial dos sistemas de RBC.

Apesar de RBC estar, ainda, num estágio de amadurecimento, existe uma gama significativa de aplicações que fazem uso de tal tecnologia. As aplicações não se limitam apenas ao âmbito acadêmico, observa-se também um grande número de implementações de cunho comercial. Os problemas mais comumente explorados pelos pesquisadores/desenvolvedores que estudam a técnica de RBC são: aquisição do conhecimento, raciocínio legal, explanação de anomalias, diagnósticos, arbitragem (jurídica), análise e elaboração de projetos, planejamento, reparos e adaptações, tutoriais.

Baseando-se nos levantamentos encontrados em Watson et al. (1994), apresentaremos algumas aplicações desenvolvidas à luz da técnica de RBC. Tais aplicações, Tabela 5, estarão divididas em acadêmicas e uma comercial.

### Acadêmicas

Ano	Nome	Autor(es)	Descrição
1988	REFINER	Sunil Sharma Sharma & Sileeman	É um sistema de aquisição de conhecimento que ajuda um especialista a refinar o seu conhecimento de uma forma mais natural do que através da extração de regras.
1986	JUDGE	Bain	Representa um modelo de base de casos de sentenças criminais. O programa inicia com um conjunto simples de estratégias para formação de sentenças e então inicia a recuperação na sua base de casos para o desenvolvimento de novas sentenças.
1988	HYPO	Ashley	Executa um RBC legal na área do Direito de patentes. Centrado nos casos de reclamação de violação de uma patente, tais como a revelação de um segredo de mercado, o HYPO usa a sua base de casos para gerar argumentos plausíveis para o processo e a defesa.
1991	GREBE	Branting	Usa o conhecimento na forma de generalizações e exemplos de categorias para determinar a classificação do novo caso.
1994	KICS	Yang & Robertson	Está sendo implementado no domínio da construção de regras.
1986	SWALE	Kass	Possui uma biblioteca de casos para explicar porque animais e pessoas morrem. Se ele receber um evento de anomalia, tal como a morte de um jovem e saudável cavalo, ele procura por explicações de tal morte em outros contextos, como na morte de um ator de rock causada por overdose ou no assassinato de um marido em virtude do seu seguro de vida. Então ele tenta adaptar tais explicações para encaixar na situação corrente.
1986	PROTOS	Porter & Bareiss	Foi desenvolvido no domínio da audiologia clínica.
1989	CASEY	Koton	É um sistema que faz o diagnóstico de parada cardíaca. Como entrada ele utiliza os sintomas do paciente e produz uma rede causal de possíveis estados internos que poderiam justificar tais sintomas.
1985	MEDIATOR	Simpsom	Trabalha no domínio da resolução de disputas judiciais. A partir de um conflito envolvendo várias partes, ele indica possíveis acordos.
1998	PERSUADER	Sycara	Objetiva resolver as disputas entre patrão e empregado. A entrada é uma descrição de uma disputa sobre benefícios salariais e a saída é um acordo adaptando contratos anteriores que foram utilizados em companhias similares.

Comercial

Ano	Nome	Autor(es)	Descrição
1994	BRITISH AIRWAYS – CASELINE	BRITISH AIRWAYS	<p>É a primeira geração de demonstradores de tecnologia usada pela British Airways (BA) para avaliar o potencial do RBC. CaseLine auxilia os engenheiros técnicos de suporte do Boeing 747-400 no diagnóstico das falhas das aeronaves e a repará-los entre a chegada e saída das mesmas. Quando uma falha em um Boeing 747-400 é detectada ou tem-se suspeita, os detalhes são transmitidos para a equipe de terra. A aeronave deve permanecer na pista por algumas horas ou o tempo em que os engenheiros necessitarem para identificar a causa da falha e restaurá-la. Isto é complicado, pois os defeitos, geralmente, são obscuros e apresentam causas complexas e inconsistentes. O atraso da decolagem irá atrapalhar a programação e trará prejuízo financeiro, por outro lado, se deixar a aeronave decolar com uma falha não resolvida poderá resultar em consequências catastróficas.</p> <p>Com o CaseLine os usuários entram com informações do diagnóstico e controlam a busca pelas sugestões de reparos disponíveis e recuperação das informações. O sistema possui cerca de 200 casos (antes de 1994) que descrevem instâncias de falhas anteriores e detalhes das ações para repará-las.</p>

Tabela 5 - Aplicações utilizando RBC



### 1.19. Ferramentas

Neste tópico serão apresentadas, baseado no artigo de Watson et al. (1994), algumas ferramentas direcionadas à modelagem dos conceitos que representam a técnica de RBC. Estas ferramentas contribuem para a disseminação da construção de aplicações sob a luz do paradigma de RBC, devido ao fato de abreviarem os esforços do engenheiro do conhecimento, uma vez que apresentam algumas etapas necessárias na construção de um RBC já implementadas.

CBR – *Express* – produzida pela Inferece Corporation, é especificamente direcionada para o mercado vertical, como o *Help Desk*. Alguns problemas similares no mercado de *Help Desk* são levantados e justificam a utilização de uma aplicação de RBC, a saber:

- treinar a equipe do *Help Desk* para que suas respostas sejam úteis,
- o treinamento toma tempo e dinheiro, e
- treinar uma equipe de *Help Desk* não é uma tarefa fácil, assim o troca dos profissionais é uma constante.

ReMind – produzido pela Cognitive System Inc., é uma ferramenta de RBC flexível que oferece uma grande variedade de métodos de recuperação de casos juntamente com conceitos interessantes tais como: modelos qualitativos e fórmula de adaptação visual. Ele não possui grandes características no tratamento de textos, porém apresenta algumas propriedades básicas do tratamento de linguagem natural via dicionários de termos que podem ser mapeados para hierarquia ordenada de símbolos. De qualquer maneira, em geral o usuários são mais forçados a selecionar do que descrever uma situação.

ReCall – é ferramenta de RBC pertencente à companhia francesa ISoft. Esta aplicação oferece uma combinação do vizinho mais próximo com a recuperação indutiva dos casos. Múltiplos índices hierárquicos são usados para propostas de organização e para recuperação eficiente de casos. ReCall fornece métodos diferentes para analisar automaticamente a base de casos fornecida para seleção dos índices, assim como a sua organização. De qualquer forma, desenvolvedores de aplicação podem impor a sua própria

organização. Procedimentos automáticos são baseados em técnicas indutivas. Os procedimentos automáticos consideram o domínio de conhecimento definido nos casos, ajudando os usuários a desenvolver aplicação interativamente.

## **1.20. RBC e Outras Técnicas de IA**

O ponto de argumentação mais relevante na utilização do paradigma baseado em casos ao realizar uma tarefa especialista é o fato de o que fortalece e caracteriza um especialista é a sua experiência. Se o conhecimento especialista fosse somente aquele que está nos livros, bastaria modelar as informações dos livros em regras e o sistema estaria pronto. Entretanto, num processo de extração de conhecimento especialista, observa-se que o especialista coloca sua experiência para tentar explicar como faz seu trabalho. Muitas vezes, sua especialidade é ilustrada através de exemplos, experiências passadas que justificam a abordagem escolhida para tratar determinado problema. Além disso, quando é possível estabelecer uma regra para determinada situação, esta vem sempre acompanhada de exceções.

Muitos são os artigos que apresentam os sistemas baseados em casos como uma alternativa aos sistemas baseados em regras. Esta forma é extremamente antídídática porque aproxima e confunde duas técnicas que não têm muito em comum, além do fato de serem ambas técnicas de IA e servirem para análise, diagnóstico e outros problemas comuns, embora com características bem distintas para os mesmos tipos de aplicação.

Como ocorre com todas abordagens novas propostas na Ciência e Tecnologia, a área de RBC tem buscado na comparação com outras técnicas a justificativa para sua atuação. Além de gerar resistências dos pesquisadores dedicados a outras áreas, a abordagem corporativa contradiz as tendências atuais de construção de sistemas híbridos inteligentes. RBC deve e pode conviver com outras técnicas de IA. Sistemas Especialistas podem contribuir significativamente na modelagem do conhecimento que permite adaptação de casos em determinado domínio. Sistemas Difusos, Reconhecimento de Padrões oferecem as funções de similaridade que se aproximam da cognição humana em contexto de incerteza. Algoritmos Genéticos podem ser empregados na geração/adaptação de casos, com busca na evolução de soluções anteriores. Redes Neurais pode ser treinadas a partir de bases de casos e o conhecimento adquirido elucidado por extração de regras, o que poderia contribuir no

ciclo do RBC original. Agentes Inteligentes podem implementar autonomia a sistemas de RBC, como assistentes dos usuários, que fundamentam sua ajuda em interações anteriores.

### 1.21. Resumo das Características do RBC

Esta seção apresenta as principais características oferecidas pela técnica de Raciocínio Baseado em Casos, baseado em (Weber-Lee, 1996), a saber:

- ◆ extração do conhecimento. basta ter um Banco de Dados. A difícil fase de extração do conhecimento especialista é facilitada nos sistemas de RBC, pois a representação do conhecimento pode ser feita com o preenchimento direto dos fatos que descrevem uma experiência. O paradigma proporciona um excelente mecanismo de aprendizagem, que pode ser utilizado para aquisição automática de conhecimento.

- ◆ representação do conhecimento. A tarefa de representação do conhecimento resume-se em escolher o tipo de estrutura da base de casos. Em aplicações onde a complexidade exigir, é necessária a criação de índices, entretanto muitas aplicações omitem este passo.

- ◆ reutilização do conhecimento. O conhecimento contido nos casos pode ser utilizado ou combinado e adaptado para gerar novas soluções além das originalmente presentes na memória.

- ◆ aprendizagem. A atualização do conhecimento pode ser feita automaticamente, na medida em que as experiências são utilizadas, assim o sistema pode crescer e incrementar sua robustez e eficiência.

- ◆ justificativas. As justificativas são sempre consistentes com as soluções por serem as próprias experiências, representando mais um aspecto de proximidade ao comportamento humano do paradigma. Além disso, as justificativas podem avisar sobre possíveis riscos que o uso de determinada abordagem pode implicar.

- ◆ raciocínio implícito. A incerteza implícita nas informações contidas nos casos é utilizada sem a necessidade de um tratamento específico desta incerteza.

## 1.22. Considerações Finais

RBC é a técnica de IA que está associada ao raciocínio analítico do ser humano. O paradigma que sustenta o RBC é a busca por experiências anteriores que possam ser utilizadas no presente com o intuito de ajudar na solução de um determinado problema.

A concepção do RBC deu-se nos Estados Unidos, tendo no trabalho de Schank e Abelson em 1977 o marco de sua existência. Atualmente sua pesquisa está em franco crescimento nos demais continentes, com destaque ao Europeu onde a Alemanha vem despontando.

Quando um domínio necessita recuperar informações anteriores para utilizar no presente, vê-se no RBC a técnica ideal para executar esta tarefa. Portanto qualquer área que apresente tal necessidade pode utilizar o RBC na confecção dos seus sistemas. Estas são algumas das áreas de aplicação do RBC: Diagnósticos Técnicos; Diagnósticos Médicos; Help Desk; Direito.

No ciclo que representa as etapas da técnica de RBC destacam-se: Recuperação; Reutilização; Revisão; Retenção.

A convivência da técnica de RBC com outras técnicas da área de IA é vista de maneira salutar. Atualmente existe uma tendência para a elaboração de aplicações compostas por mais de uma abordagem inteligente, onde as técnicas completar-se-ão atendendo as demandas de emulação de diferentes ramos da cognição humana, formando um sistema híbrido.

A utilização do RBC no domínio abordado na presente dissertação dá-se pela proximidade conceitual em que ambos estão amparados, ou seja, utilizam experiências anteriores para ajudar a resolver os problemas atuais.

## “RBC-TEXT” - MODELO PARA TRATAMENTO DE DOCUMENTOS TEXTUAIS E VALIDAÇÃO NO DOMÍNIO JURÍDICO

O presente capítulo estará comprometido na apresentação do modelo de “RBC-Text” concebido para atuar nos domínios que apresentem arquivos textos como fonte de informação e validação através do emprego no domínio jurídico.

### 1.23. Introdução

O modelo “RBC-Text”, Figura 5, é aplicado a domínios que apresentem documentos semi-estruturados e uma certa padronização na sua coleção. Esta estrutura permite a elaboração de métodos para a identificação dos atributos/índices que representarão um caso, bem como a elaboração de algoritmos para a valoração destes atributos/índices.

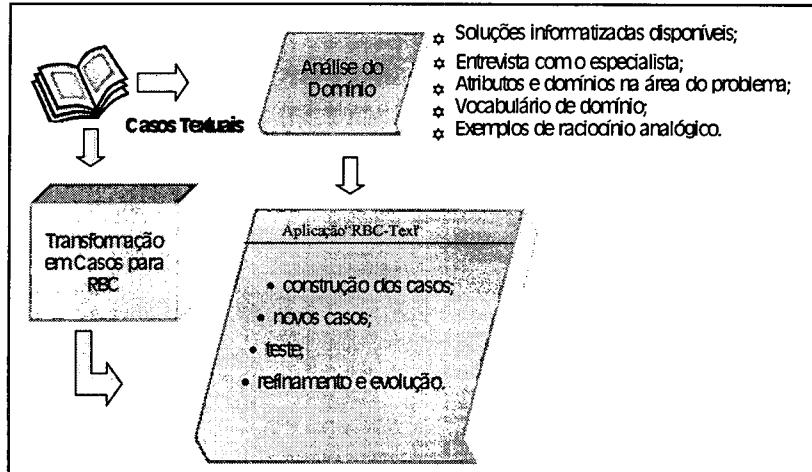


Figura 5 - Modelo "RBC-Text"

Para que uma solução seja construída utilizando o modelo “RBC-Text”, é de suma importância que o domínio sendo trabalhado seja baseado em documentos textuais. Na seqüência o domínio será analisado, objetivando verificar a sua conformidade com as características elencadas no modelo proposto. O sucesso no quesito *Análise do Domínio* viabilizará a continuação do processo, permitindo a execução da etapa *Aplicação “RBC-*

*Text*”, o qual transformará a coleção de documentos textuais em uma base de conhecimentos no formato de casos.

A aplicação do modelo a um determinado domínio está condicionada a presença das características, Figura 6, nele definidas, ou seja, é necessário que as informações constantes no domínio sendo analisado apresentem uma estrutura similar à proposta no modelo. Com o intuito de validar a modelagem de um determinado domínio a luz da proposta presente neste trabalho, apresenta-se uma metodologia de aplicação do modelo, Figura 6.

#### **1.24. Análise do Domínio**

A etapa de análise do domínio é determinante para o andamento do processo, pois é nesta etapa que definir-se-á se o domínio candidato a aplicação do modelo está apto para a sua utilização. É necessário observar se o problema a ser resolvido se enquadra nos conceitos da técnica de RBC, ou seja, se existe a necessidade de recuperar informações anteriores para que se aplique a sua experiência em situações presentes, e conseqüentemente se tal problema pode ser representado pelo raciocínio analógico. Um outro fator preponderante é o fato de o domínio ser representado por documentos textuais, haja visto, que o presente modelo é voltado para tal formato. Uma vez vencida esta etapa outros passos terão que ser contemplados até que se alcance a sua plenitude.

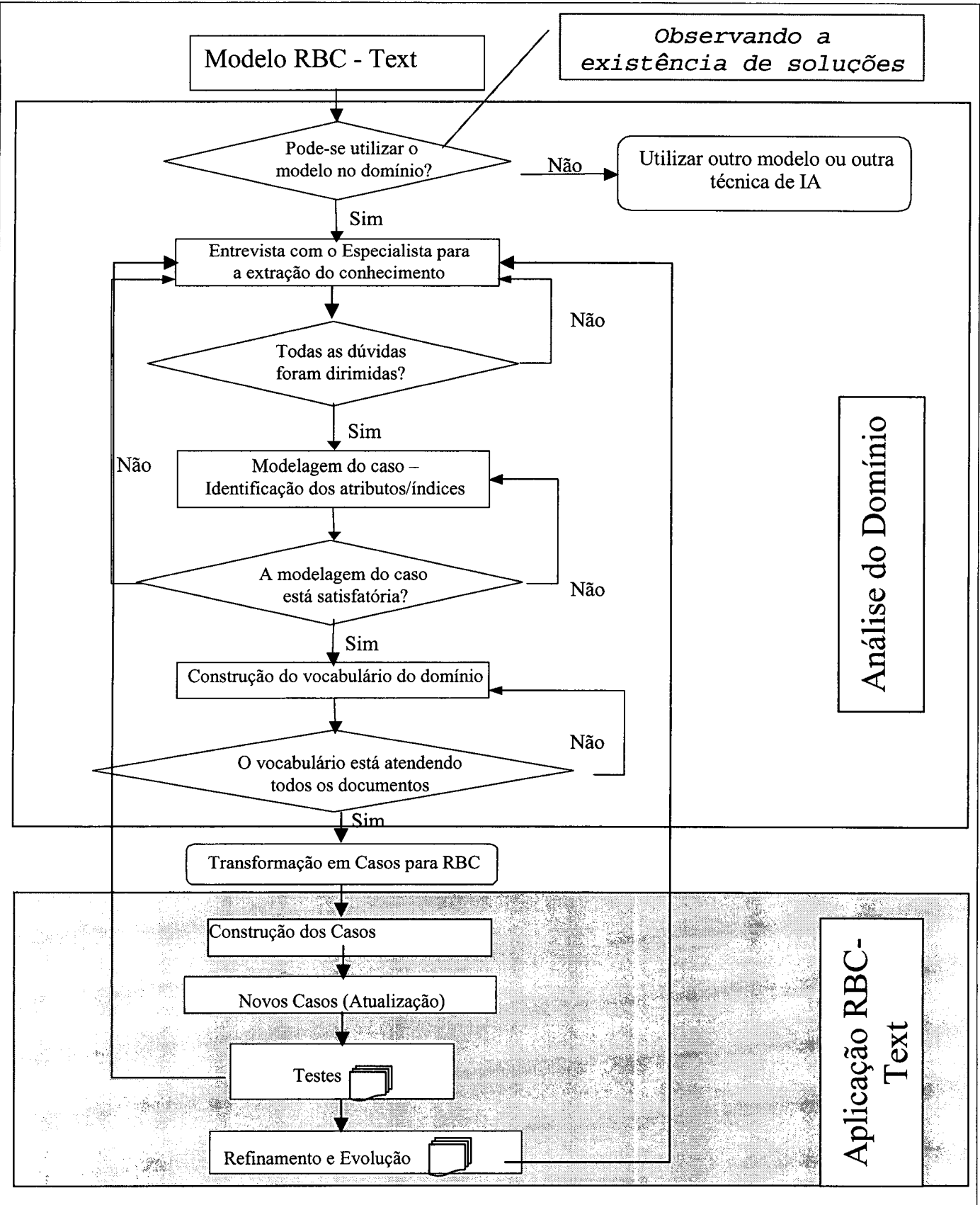


Figura 6 - Metodologia para aplicação do modelo

Nos próximos itens serão descritos os passos que complementam a análise do domínio, a saber: soluções informatizadas disponíveis, entrevistas com o especialista, atributos/índices da área do problema, construção do vocabulário controlado e exemplo de raciocínio analógico.

#### 1.24.1. Soluções informatizadas disponíveis

O primeiro passo dentro da análise do domínio é a realização de um levantamento de todos os aplicativos existentes para o domínio sendo analisado. De posse deste levantamento faz-se um estudo no softwares identificados, observando a sua eficácia. Pois, se já existirem soluções que satisfaçam a demanda solicitada e que apresentem uma relação de custo/benefício satisfatória, o desenvolvimento de um novo aplicativo não será mais necessário. Não é interessante construir um sistema inteligente, somente pelo anseio de querer empregar uma técnica de IA no seu desenvolvimento, isto é, se já existirem soluções prontas ou se uma abordagem tradicional resolver o problema, a opção por estas alternativas é o mais sensato.

#### 1.24.2. Entrevistas com o especialista

Para que um sistema seja elaborado a luz do paradigma de IA, é essencial a modelagem do conhecimento do especialista humano. Portanto, após identificar o problema como sendo factível utilizando técnicas de IA, é vital para o bom andamento do projeto, a realização das entrevistas com o especialista do domínio. Através das entrevistas, as quais fazem parte do processo de Engenharia do Conhecimento, é que o engenheiro identificará os elementos mais importantes do problema sendo estudado, os padrões existentes nos documentos, bem como as suas características específicas representando o mapeamento do conhecimento a ser extraído. Somente, quando os especialistas e engenheiros estiverem confiantes que todos os focos de extração de conhecimento foram identificados é que segue-se para o próximo passo.

#### 1.24.3. Atributos/Índices na Área do Problema



Após a série de entrevistas realizadas o engenheiro do conhecimento estará apto para eleger quais serão os índices e os atributos dos casos, isto é, fazer a modelagem do caso, identificando as características mais importantes no documento. Porém, quando da realização desta etapa for verificado que ainda existem dúvidas a serem dirimidas com o especialista do domínio, pode-se voltar a fazer mais algumas baterias de entrevistas até que os pontos duvidosos fiquem esclarecidos e o processo de escolha dos índices e atributos possam ser realizados com precisão.

#### 1.24.4. Vocabulário de Domínio

Com a solidificação do processo de modelagem do caso, o próximo passo é a construção do vocabulário do domínio ou a aquisição de um vocabulário já elaborado. O vocabulário do domínio é muito importante para o processo de indexação dos documentos, nele estão contidos os termos chaves para a boa representação do conhecimento. Existem várias técnicas para a sua construção, a saber: frequência de palavras, frequência inversa de palavras entre outras. Dependendo do domínio e do número de documentos a serem indexados a construção de um vocabulário totalmente manual pode ser viável.

#### 1.24.5. Exemplos de Raciocínio Analógico

Ainda, dentro da Análise do Domínio, pode-se observar a maneira como os especialistas raciocinam e ao comparar com o paradigma da técnica da RBC permite-se a confirmação da presença do raciocínio analógico nas duas pontas. Sempre que nos depararmos com problemas que necessitam se reportarem ao passado para resolverem os conflitos existentes no presente, permite-se o uso da abordagem de RBC neste contexto. A seguir serão indicados alguns exemplos de raciocínio analógico:

- Os juristas quando buscam pelas jurisprudências anteriores, com intuito de encontrarem experiências que possam lhes auxiliarem na resolução de um embate no qual está trabalhando.
- Os operadores de *call-center* quando procuram dar respostas aos seus clientes baseados em soluções já encontradas para problemas semelhantes.
- Os médicas quando fazem o diagnóstico de um paciente baseado nos mesmos sintomas encontrados em outro enfermo.

## 1.25. Transformação em Casos para RBC

Neste tópico será demonstrado a validação do modelo elaborado no domínio jurídico através da transformação dos textos em casos para o RBC. No campo do direito existem vários ramos com abordagens estritamente textuais, tais como: as súmulas, as jurisprudências e as resoluções do conselho de segurança da ONU (direito internacional) entre outros. No presente trabalho foi eleito o ramo das jurisprudências como sendo o objeto da aplicação do modelo, através do qual apresentamos os procedimentos utilizados na construção de um RBC Textual, como segue:

- **Análise do Domínio:** partindo da premissa que está associada a esta etapa, observamos o domínio das jurisprudências com o intuito de o validarmos como sendo apto à ser utilizado para a elaboração de um software utilizando a técnica de RBC. O primeiro ponto a favor do domínio candidato é o fato de as jurisprudências serem apresentadas em documentos textuais, uma vez que este é um pré-requisito para a aplicação do modelo. O outro fator que contribui para a sua escolha é que as jurisprudências são exemplo de raciocínio analógico, onde a resolução de um problema está fundamentada em resoluções anteriores de problemas similares. Tal conceito está em concordância com a abordagem na qual se enquadra a técnica de RBC. Portanto, podemos considerar que o passo de validação do domínio foi vencido, permitindo que possamos avançar nos pontos específicos da Análise do Domínio.
- **Entrevistas com o especialista:** nesta etapa foram extraídos dos especialistas as principais características que as jurisprudências apresentam. Vários encontros foram realizados até que chegou-se a um consenso com relação aos pontos que permitiam uma boa representação dos documentos sendo analisados. Vejamos as características que foram levantadas: *Data, Relator, Local, Tipo de Recurso, Resultado, Ementa, Expressões indicativas e Tipificação*.
- **Atributos/Índices na Área do Problema:** de posse das características que representam uma jurisprudência, o próximo passo foi a

identificação dos atributos e índices, os quais representaram um caso. A seleção de quais características eram mais importantes, portanto enquadrando-se na categoria de índice, e das menos relevantes, caracterizando-se como atributos, foi concluída após vários testes. Nestes testes foram levados em consideração a importância do atributo para o operador jurídico, bem como a sua contribuição para uma recuperação condizente com a esperada. Após vários ensaios os índices eleitos foram os seguintes: *Expressões indicativas e Tipificação*. E como atributos foram identificados os que seguem: *Data, Relator, Local, Tipo de Recurso, Resultado, Ementa*.

- Vocabulário de Domínio: nesta etapa foi utilizado um vocabulário já construído e utilizado pelos Tribunais. Tal vocabulário representa um conjunto de termos jurídicos que são usados normalmente nos Tribunais, que representam fatos enquadrados normativamente. Através desta lista (vocabulário) serão pesquisadas nas jurisprudências os termos nela constante.

## 1.26. Aplicação “RBC-Text”

Após vencida as fases de análise do domínio, onde examina-se o domínio com o intuito de verificar a necessidade do desenvolvimento de uma aplicação inteligente para a resolução do seu problema, e da transformação dos documentos deste domínio em casos, dá-se o início da construção do sistema, observando os seguintes tópicos: construção dos casos, novos casos (atualização), testes, refinamento e evolução.

### 1.26.1. Construção da Base de Casos

Neste tópico será apresentado o que é necessário para a boa representação de um caso e o exemplo de sua materialização em banco de dados relacionais, bem como a contribuição deste tipo de arquitetura na construção do sistema que faz uso da técnica de Raciocínio Baseado em Casos.

#### 4.4.1.1. Representação dos Casos

De acordo com a definição dos pesquisadores da técnica de Raciocínio Baseado em Casos, “um caso é uma abstração de uma experiência descrita através de atributos devidamente valorados”. Portanto, para representar a experiência na qual se está trabalhando é necessário que se faça uma ampla análise nos documentos envolvidos. O fator de grande relevância para que a representação seja feita com bastante precisão, é o envolvimento total dos especialistas do domínio em cada etapa da elaboração do trabalho. Através do conhecimento e da experiência dos especialistas consegue-se representar documentos textuais em pares de atributo-valor, viabilizando a elaboração da base de casos.

Uma vez tendo a base/banco de casos devidamente modelada, Figura 7, o domínio do problema torna-se compreensível para máquina, permitindo que o Engenheiro do Conhecimento elabore as tarefas a serem executadas computacionalmente.

#### 4.4.1.2. Banco de Dados Relacional

Após o processo de identificação das informações relevantes nos documentos, ou seja, os pares atributo/valor, é necessário uma estrutura para guardá-los. Neste trabalho foi selecionado o banco de dados relacional como o ideal para execução da tarefa de armazenamento.

Os bancos de dados relacionais foram apresentados pela primeira vez por Codd, pesquisador da IBM, em 1972. Codd partiu da noção matemática de relação. Numa visão intuitiva, pode-se entender uma relação como um conjunto de atributos associados à uma entidade do mundo real. Por exemplo, para descrever um “cadastro urbano” podemos utilizar a relação: <número\_lote, dono, endereço, area, IPTU>; ou representar através de uma tabela. Assim, por exemplo, a Tabela 6 apresenta a relação "cadastro urbano" com seus atributos e exemplos de valores para estes atributos. Cada conjunto de atributos é representado por uma linha ou registro (SPRING, 2002).

número_lote	Dono	endereço	area(m <sup>2</sup> )	IPTU(R\$)
195689	Guimarães, M.	Clóvis Bevilacqua, 768	900	350

Tabela 6 - Exemplo de Atributos para "Cadastro Urbano"

Uma grande vantagem da utilização de banco de dados para armazenar os casos é o fato de evitar a redundância das informações. Uma vez que cada conjunto de informações comuns serão armazenadas em tabelas próprias, possuindo um código único que representará cada registro. Na Figura 7, a tabela EN\_LOCAL apresenta os seguintes atributos: COD\_LOCAL e TXT\_LOCAL, onde um determinado local será sempre representado por um único COD\_LOCAL. Por exemplo, o atributo TXT\_LOCAL com valor de “Joinville” estará associado ao COD\_LOCAL com valor “5”, sendo assim, esta informação terá apenas uma ocorrência, evitando que uma tabela possua um dado repetido desnecessariamente.

#### 4.4.1.3. Busca

Outra característica explorada na visão de banco de dados relacional é a utilização dos operadores lógicos por meio da linguagem SQL. O efeito da aplicação de um desses operadores (ou, mais geralmente, alguma combinação válida desses operadores) a um banco de dados, será o de identificar algum subconjunto daquele banco. Esse subconjunto poderá, então, servir como base para uma variedade de operações, a saber: leitura; atualização; deleção; entre outros (DATE, 1988).

A aplicação de algoritmos de IA, mais precisamente, a métrica de similaridade em uma base que contenha um número elevado de registros, produz um resultado muito demorado. Tendo em vista de que não há necessidade de envolver todos os registros nesta métrica, pois alguns não possuem relevância para um determinado contexto em questão, seleciona-se, através da linguagem SQL, somente os registros de interesse, nos quais aplica-se a métrica de similaridade. Conseqüentemente o processo tornar-se-á mais rápido.

#### 1.26.2. Novos Casos

Sempre que novos documentos forem produzidos, a base de conhecimento precisa ser atualizada. Dependendo do domínio que se está trabalhando, o volume de novos documentos sendo gerados, tornará imprescindível que o processo de transformação dos documentos em casos seja feito de maneira automática.

A construção de algoritmos que extraiam as informações que representam o conhecimento especialista contido no textos é uma tarefa a ser executada pelo desenvolver em

conjunto com o engenheiro do conhecimento baseado nas informações coletadas com o especialista do domínio. Os algoritmos de atualização representam três formas de extração de informações: extração por localidade, extração por inferência e extração como auxílio de vocabulário controlado.

- Extração por localidade: este tipo de extração acontece quando existe uma padronização no local onde está situada um determinado tipo de informação ( ex.: a característica “*endereço*” aparece sempre na segunda linha do documento e após a palavra “ENDEREÇO: ”). A partir desta constatação elabora-se um algoritmo observando tal peculiaridade;
- Extração por inferência: esta situação é observada em documentos que apresentem expressões diferentes e representam um mesmo assunto, ou quando a informação desejada está localizada após a ocorrência de uma determinada expressão, ou ainda um misto destas situações, isto é, a existência de expressões diferentes que represente algo em comum e na seqüência desta estar o dado alvo (ex.: sempre que ocorrer a palavra “RESULTADO”, extrai-se a informação que segue até encontrar o primeiro ponto). Para a construção destes algoritmos, geralmente, são utilizadas regras.
- Extração com o auxílio de vocabulário controlado: este é o processo mais simples para a extração da informação desejada. Se dá através da verificação da existência de palavras/expressões constantes no vocabulário controlado nos documentos do domínio.

Ao finalizar estes algoritmos a aplicação permitirá a inserção de novos documentos, a medida que forem produzidos, sem a intervenção humana.

### 1.26.3. Teste

A fase de testes confirmará a validade das informações ou não, bem como se as informações sendo recuperadas são, realmente, as mais semelhantes. Nesta etapa, também, serão abordadas questões relacionadas a velocidade e segurança da solução.

Novamente, em conjunto, os desenvolvedores, os especialistas e engenheiros do conhecimento, gerarão um relatório apontando as falhas e suas respectivas melhorias a serem feitas e o grau de urgência para a sua execução.

#### 1.26.4. Refinamento e evolução

A última etapa do modelo é a fase de refinamento e evolução. Onde é recomendável o acompanhamento da utilização do sistema, com o intuito de detectar possíveis carências que não foram supridas nas fases anteriores, e também manter um monitoramento constante em novas abordagens que venha a surgir e que possa contribuir para uma evolução do sistema.

Novos relatórios são produzidos indicando as modificações a serem realizadas para as novas versões do software.

### **1.27. Considerações Finais**

A apresentação do modelo apontou as características essenciais para a modelagem de um determinado domínio, onde destaca-se a abordagem nos documentos textuais. Ficou evidente a necessidade de se vencer as etapas apontadas no modelo para que se tenha sucesso no desenvolvimento de uma determinada aplicação.

Com a validação do modelo no domínio jurídico (acórdãos), verificou-se a sua eficácia. Uma vez que todas as etapas indicadas no modelo encontraram suporte na aplicação real.

## **ESTRUTURA DO SISTEMA**

Neste capítulo será apresentado como está estruturado o sistema. Para isto apresentamos a modelagem do sistema, dicionário de dados, fluxo dos dados e também algoritmos utilizados no projeto.

### **1.28. Introdução**

O objetivo deste capítulo é ilustrar graficamente como está estruturada a base de casos do sistema, bem como os procedimentos que envolvem todas as etapas de desenvolvimento. Apresentaremos desde a modelagem geral do sistema, passando por suas minúcias, até os algoritmos que incorporam os processos envolvidos no projeto.

Haja vista, que o objeto desta dissertação é a apresentação de um modelo para o tratamento de documentos textuais a luz da técnica de RBC e sua validação no domínio jurisprudencial, a abordagem concernente à modelagem de banco de dados relacionais será superficial.

### **1.29. Modelagem em Banco de Dados Relacional**

O fato de os banco de dados relacionais estarem a bastante tempo no mercado e por possuírem uma grande aceitação e confiança, foi um dos pontos determinantes para o esforço dedicado na modelagem dos casos textuais na ótica relacional.

Verifica-se em Ribeiro (2002), que a modelagem de dados é a técnica que propõe-se a fazer uma investigação mais formal e rigorosa sobre os dados e seus significados, num determinado contexto. Investiga a organização e os relacionamentos dos dados no contexto da realidade onde estarão ocorrendo.

Portanto, para modelar casos de descrição textual em banco de dados relacional, primeiramente é necessário uma profunda análise nos documentos que compõem o



domínio sendo trabalhado, com o intuito de identificar os dados comuns e armazená-los em um agrupamentos distintos, dando-se início ao processo identificação das relações/tabelas.

Em BR-BUSINESS (2002), observa-se que o próximo passo é determinar as possíveis relações de um objeto com outro de forma a reproduzir situações reais, identificando os tipos de relacionamentos as chaves de cada tabela.

Ainda, conforme BR-BUSINESS (2002), verifica-se as três formas de se relacionar os objetos e, também, o conceito de chave da tabela, a saber:

- Um-para-um: Uma linha da tabela X relaciona-se com apenas uma linha da tabela Y e vice-versa..
- Um-para-vários: Para uma linha da tabela X existem várias linhas da tabela Y.
- Várias-para-várias: Para qualquer linha na tabela X, existem várias linhas na tabela Y e vice-versa.
- Chaves das colunas: Pode ser definida como a coluna de uma tabela da qual todas as outras colunas são dependentes, sendo que nenhuma outra linha da tabela pode ter o valor da(s) coluna(s) chave. Vale frisar que todas as colunas chave devem ser indexadas. O valor pode ser qualquer valor numérico ou combinação de caracteres que seja único como número do CIC, do RG, da conta corrente, etc...

Na Figura 7, podemos observar a modelagem dos acórdãos. A análise sobre um conjunto de acórdãos identificou sete grupos comuns de informações, os quais foram armazenados em tabelas diferentes. Este tipo de armazenagem apresenta facilidade para a identificação de acórdãos que contenha características específicas, de acordo com as relações existentes entre as tabelas. Por exemplo, quando da necessidade de se saber quantos acórdãos possuem resultado “positivo”, a única tabela a ser utilizada é a EN\_ATRIBUTO, a qual contém o atributo TXT\_RESULTADO.

Outra característica importante nesta abordagem é a possibilidade de se utilizar a linguagem SQL para manipulação de dados. Para obter uma resposta para a indagação referente a quantidade de acórdãos que possuem resultado “positivo” utiliza-se o seguinte comando: “SELECT COUNT(\*) AS TOTAL FROM EN\_ATRIBUTO WHERE TXT\_RESULTADO = ‘POSITIVO’ ”. Pode-se também identificar a quantidade de acórdãos

baseado em uma determinada localidade ou em um determinado relator ou ainda baseado em um determinado relator e em uma localidade específica.

Portanto para que a modelagem represente de maneira mais eficiente o domínio sendo trabalhado, a participação do especialista do domínio em conjunto com o engenheiro do conhecimento e analista de sistemas é de vital importância. Já que a boa representação das informações evitará redundâncias, implicando em maior velocidade no acesso as informações.

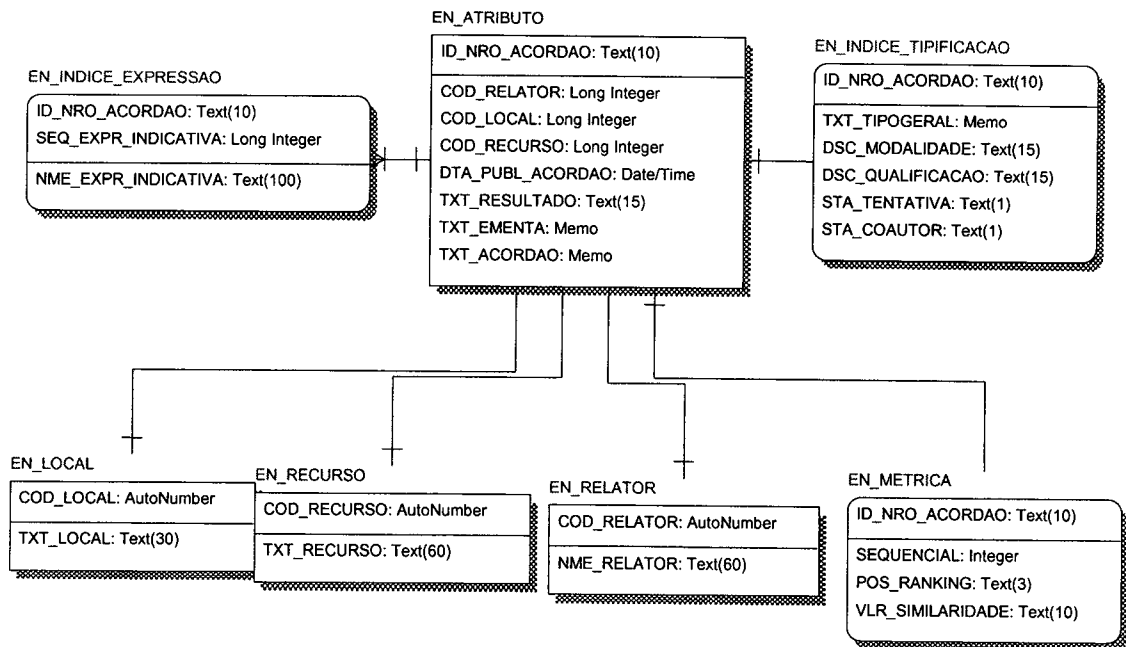


Figura 7 - Modelagem do JurisConsulta

### 1.29.1. Dicionário de Dados

Para que haja uma compreensão maior de todo o processo em torno da representação do conhecimento, ou seja, como o acórdão foi representado em pares de atributo-valor, apresenta-se um dicionário de dados ilustrando tais procedimentos.

A Figura 8, representa o ciclo envolvendo a pesquisa no JurisConsulta. Este ciclo se inicia com a requisição feita pelo profissional do Direito, onde ele escreve o seu problema e envia ao sistema. De posse do caso descrito o JurisConsulta procura em sua base de casos quais são as experiências mais similares ao problema em questão. Após analisar os acórdãos da base o programa apresenta os acórdãos mais similares para o usuário. Dentre os

casos eleitos, o profissional escolhe aquele que julgar ser mais útil para o problema sendo tratado.

O crescimento do banco de acórdãos é de responsabilidade do engenheiro do conhecimento, pois é ele que irá inserir novos casos. É também função do engenheiro gerenciar o banco, ou seja, verificar se as informações foram representadas corretamente.

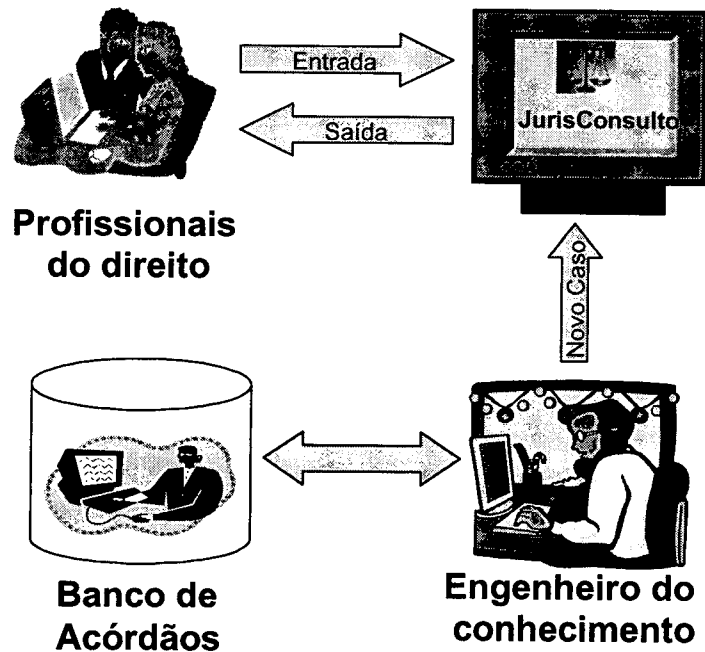


Figura 8 - Fluxo da

informação

<b>Entrada =</b>	
Descrição	: <i>texto</i>
TipoGeral	: <i>símbolo (texto)</i>
Modalidade	: <i>símbolo (texto)</i>
Qualificação	: <i>símbolo (texto)</i>
Tentativa	: <i>símbolo (texto)</i>
CoAutor	: <i>símbolo (texto)</i>
[ Recurso	: <i>símbolo (texto)</i>
Data início	: <i>date</i>
Data fim	: <i>date</i>
Relator	: <i>símbolo (texto)</i>
Local	: <i>símbolo (texto)</i>
Resultado	: <i>símbolo (texto)</i> ]

Figura 9 - Campos da entrada

O usuário tem a sua disposição os campos relacionados à entrada do sistema, Figura 9, para elaborar o caso que deseja resolver. Os campos possuem as seguintes funções:

- Descrição <sup>1</sup> - Campo do tipo texto, destinado ao usuário para que ele possa descrever o problema a ser utilizado no processo de recuperação por casos anteriores semelhantes a este.
- TipoGeral <sup>1</sup> - Campo do tipo símbolo (texto), permitindo que usuário indique qual o tipo geral do enquadramento do crime.
- Modalidade <sup>1</sup> - Campo do tipo símbolo (texto), proporcionando ao usuário a possibilidade de apontar a modalidade do enquadramento.
- Qualificação <sup>1</sup> - Campo do tipo símbolo (texto), oferecendo ao usuário a possibilidade de indicação da qualificação do enquadramento.
- Tentativa <sup>1</sup> - Campo do tipo símbolo (texto), permitindo que o usuário indique a existência ou não da tentativa no crime.
- CoAutor <sup>1</sup> - Campo do tipo símbolo (texto), no qual o usuário poderá apontar a existência ou não de co-autoria no crime.
- Recurso <sup>2</sup> - Campo do tipo símbolo (texto), oferecendo ao usuário a possibilidade de indicar o tipo de recurso.
- Data inicio <sup>2</sup> - Campo do tipo data, permitindo ao usuário selecionar uma data inicial.
- Data fim <sup>2</sup> - Campo do tipo data, permitindo ao usuário selecionar uma data final.
- Relator <sup>2</sup> - Campo do tipo símbolo (texto), por meio do qual o usuário poderá indicar um determinado relator.
- Local <sup>2</sup> - Campo do tipo símbolo (texto), onde o usuário poderá escolher uma determinada localidade.
- Resultado <sup>2</sup> - Campo do tipo símbolo (texto), proporcionando ao usuário a possibilidade de indicar um resultados desejado.

1: Campos que comporão a descrição do caso de entrada, sem a função de filtrar a base mediante a sua indicação.

2: Os campos utilizados como filtro estarão a disposição do usuário somente como uma maneira de reduzir o universo de busca, ou seja, terão a função de restringi-la em função dos valores eventualmente fornecidos.

<b>Saída_Principal =</b>	
[ Ranking	: <i>integer</i>
Similaridade	: <i>real</i>
Ementa	: <i>texto</i>
Data	: <i>date</i>
Relator	: <i>string</i>
Local	: <i>string</i>
Resultado	: <i>string</i> ] 10

Figura 10 - Campos da saída principal

O JurisConsulta utiliza os campos apresentados na Figura 10, para representar a saída principal. Estes campos possuem as seguintes funções:

- Ranking - Campo do tipo inteiro, apresenta a posição de cada caso após a realização do *matching*.
- Similaridade - Campo do tipo real, indica o valor obtido pelo caso após a realização da métrica de similaridade.
- Ementa - Campo do tipo texto, representa a ementa do acórdão.
- Data - Campo do tipo date, representa a data do acórdão.
- Relator - Campo do tipo string, indica o relator do acórdão.
- Local - Campo do tipo string, indica o local do acórdão.
- Resultado - Campo do tipo string, indica o resultado do acórdão.

Como resultado final o sistema utilizar o campo da Figura 11, para representar o caso escolhido pelo usuário.

- Acórdão - Campo do tipo texto, representa o acórdão.

<b>Saída_Acordao =</b>	
Acórdão	: <i>texto</i>

Figura 11 - Campos da saída do acórdão

<b>Caso_Base =</b>	
idCaso	: <i>integer</i>
Ranking	: <i>integer</i>
Similaridade	: <i>real</i>
[Expressões indicativas	: <i>lista string</i> ] 1..N
TipoGeral	: <i>string</i>
Modalidade	: <i>string</i>
Qualificação	: <i>string</i>
Tentativa	: <i>string</i>
CoAutor	: <i>string</i>
Recurso	: <i>string</i>
Data	: <i>date</i>
Local	: <i>string</i>
Relator	: <i>string</i>
Resultado	: <i>string</i>

Figura 12 - Casos do caso na base

Os campos apresentados na Figura 12 refletem a representação de um acórdão na base de casos. Os campos podem ser assim conceituados:

- IdCaso-Campo do tipo inteiro, representa o código que identifica um caso na base.
- Ranking - Campo do tipo inteiro, apresenta a posição de cada caso.
- Similaridade - Campo do tipo real, indica o valor obtido pelo caso.
- Recurso - Campo do tipo string, indica o tipo de recurso.
- Expressões indicativas - Lista de valores do tipo string, representam as expressões relevantes em acórdão.
- TipoGeral - Campo do tipo string, indica o tipo geral do enquadramento do crime.
- Modalidade - Campo do tipo string, indica a modalidade do enquadramento.
- Qualificação - Campo do tipo string, indica a qualificação do enquadramento.
- Tentativa - Campo do tipo string, indica a existência ou não da tentativa no crime.
- CoAutor- Campo do tipo string, indica a existência ou não da Co-autoria no crime.
- Data - Campo do tipo date, representa a data do acórdão.
- Local - Campo do tipo string, indica o local do acórdão.
- Relator - Campo do tipo string, indica o relator do acórdão.
- Resultado - Campo do tipo string, indica o resultado do acórdão

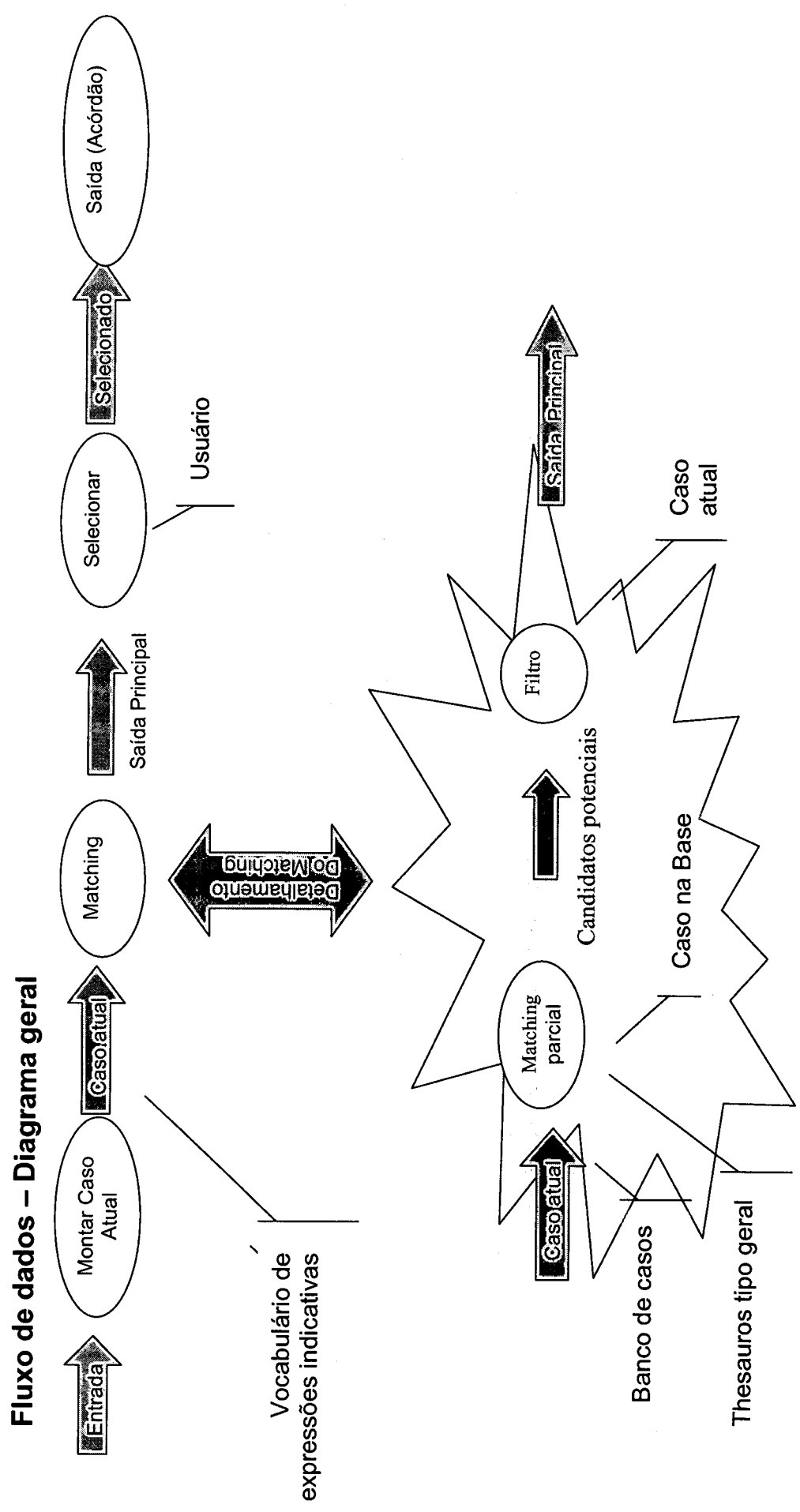


Figura 13 - Fluxo dos dados

### 1.30. Fluxo de Dados - Geral

Neste tópico é apresentada uma descrição do fluxo das informações, Figura 13, mostrando o tratamento recebido por elas durante o processo que envolve a recuperação dos casos mais semelhantes.

#### Caso de Entrada

Caso descrito pelo usuário, o qual será avaliado pelo sistema, com o objetivo de ser comparado com os outros casos da base de conhecimento, culminando na apresentação dos mais similares.

<b>Caso entrada =</b>	Descrição	: String (2000)
	TipoGeral	: <i>símbolo (texto)</i>
	Modalidade	: <i>símbolo (texto)</i>
	Qualificação	: <i>símbolo (texto)</i>
	Tentativa	: <i>símbolo (texto)</i>
	CoAutor	: <i>símbolo (texto)</i>
	[Recurso	: <i>símbolo (texto)</i>
	Data início	: <i>date</i>
	Data fim	: <i>date</i>
	Relator	: <i>símbolo (texto)</i>
	Local	: <i>símbolo (texto)</i>
	Resultado	: <i>símbolo</i>
		<i>[Concedido/Denegado] ]</i>

#### Montar Caso Atual

Representa o caso descrito pelo usuário, o qual faz o matching com todos os casos da base.

**Vocabulário Tipo Geral** = Lista de Símbolos : *string*



Possui uma lista de termos que serão acionados quando o usuário utilizar a restrição referentes ao atributo *Expr\_TipoGeral*.

**Vocabulário Expressões Indicativas = Lista de Símbolos : *string***

Fornecer uma lista de termos, os quais serão verificadas suas ocorrências na descrição do usuário.

**Caso Atual**

Representa, em forma de caso, para que o JurisConsulta possa fazer o *matching* com os casos da base, a informação inserida pelo usuário.

**Caso atual =**

Expressões indicativas	: lista de símbolo ( <i>texto</i> ) (not null)
TipoGeral	: símbolo ( <i>texto</i> )
Modalidade	: símbolo ( <i>texto</i> )
Qualificação	: símbolo ( <i>texto</i> )
Tentativa	: símbolo ( <i>texto</i> )
CoAutor	: símbolo ( <i>texto</i> )
[Recurso	: símbolo ( <i>texto</i> )
Data início	: <i>date</i>
Data fim	: <i>date</i>
Relator	: símbolo ( <i>texto</i> )
Local	: símbolo ( <i>texto</i> )
Resultado	: símbolo

[Concedido/Denegado] ]

**Matching Parcial**

Faz a comparação do caso atual com os casos armazenados, levando em consideração os valores contidos no *thesauros tipo geral*.

**Banco de casos =**

Registros (IdCaso	: Incremental
Ranking	: Integer
Similaridade	: Integer
[Expressões indicativas	: string (100)] 1..N
TipoGeral	: string (100)
Modalidade	: string (15)
Qualificação	: string (15)

Tentativa	: string (1)
CoAutor	: string (1)
Recurso	: string (100)
Data	: date
Relator	: string (100)
Local	: string (100)
Resultado	: string (30)
Ementa	: string (200)
TextoAcordao	: memo )

Indica os acórdãos no formato de casos, ou seja, utilizando um conjunto de pares atributo-valor na sua representação.

**Thesaurus tipo geral** = [ Termo : lista (texto) ] 3.

Define a similaridade lingüística dos termos, ou seja, contribui no matching fornecendo os sinônimos dos valores de *tipo geral* fornecidos pelo usuário.

### **Candidatos Potenciais**

Representam os casos potencialmente mais adequados ao problema de entrada, devidamente ordenados de acordo com o grau de similaridade.

**Candidatos potenciais** =

Registros = Banco de casos

### **Filtro**

Filtra os casos mais similares de acordo com as restrições fornecidas pelo usuário.

### **Saída Principal**

Apresenta os casos mais similares (ordem crescente) ao caso elaborado.

**Saída\_Principal** = [Ranking : integer  
 Similaridade : real  
 Ementa : string  
 Data : date  
 Relator : string  
 Local : string  
 Resultado : string ]10

### **Selecionar**

Serão apresentados os casos mais similares ao problema de entrada, dos quais o usuário seleciona o que melhor lhe convir segundo sua interpretação.

**Usuário** = Indivíduo : humano

### **Selecionado**

Caso escolhido pelo usuário como o mais adequada para auxiliar na tarefa em questão.

**Selecionado** = (Ranking : integer  
 Similaridade : real  
 Ementa : string  
 Data : date  
 Relator : string  
 Local : string  
 Resultado : string)

### **Saída**

Apresenta a saída do sistema, ou seja, fornece o acórdão na íntegra para que o usuário possa apreciá-lo e retirar as lições úteis para serem empregadas em um determinado problema.

**Saída** = Acórdão: memo (texto).

### 1.31. Fluxo de dados – Aquisição de novos casos

Nesta etapa trabalhou-se com os dados envolvidos na alimentação da base de casos. Com isto pretende-se representar os fluxos dos dados, Figura 14, a partir de um documento textual, acórdão, até sua inserção na base.

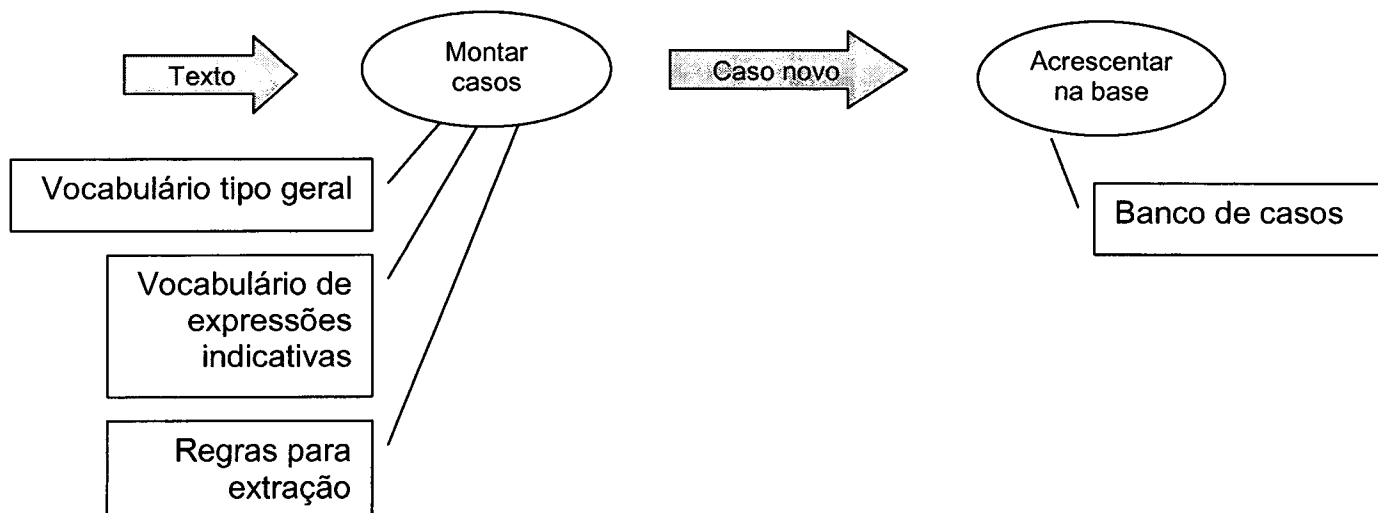


Figura 14 - Fluxo dos dados para a aquisição de novos casos

#### Texto

O processo de aquisição de novos casos iniciar-se-á com a leitura dos acórdãos (formato txt) que serão inseridos na base, para extrair os valores que os representarão na base de conhecimento.

#### Montar Casos

Nesta etapa novos casos serão inseridos na base. Esta inserção dar-se-á através da extração automática dos valores requisitados na modelagem dos casos, em função dos novos acórdãos apresentados ao sistema. Para a extração serão observadas algumas regras predefinidas. Regras estas que indicarão a maneira mais adequada e rápida para a devida valoração de cada atributo.

**Vocabulário Tipo Geral** = Define o tipo genérico da categoria do crime.

**Vocabulário Expressões Indicativas** = Determina as informações jurídicas importantes na petição e indica matéria jurídica de interesse.

**Vocabulário Tipo Geral** = Lista de Símbolos : *string*

**Vocabulário Expressões Indicativas** = Lista de Símbolos : *string*

**Regra para extração** = Para extrair os valores correspondentes aos atributos que representam um caso, foram formulados 3 procedimentos, que serão detalhadas no capítulo 1.35.1. Os procedimentos foram elaboradas de acordo com a disposição das informações no texto, a saber:

- **Por localização:** Nesta fase serão extraídos os valores que se encontram sempre na mesma posição nos acórdãos.

Ex.: Em todos os acórdãos o atributo *data publicação* encontra-se sempre na primeira linha e precedido da palavra DATA.

- **Por inferência:** Alguns atributos não são identificados de maneira explícita nos acórdãos, portanto, será necessário algumas inferências para a sua valoração.

Ex.: Tipificação.

- **Usando vocabulário controlado:** Através de uma lista predefinida de valores (vocabulário controlado), serão pesquisados nos acórdãos expressões existentes nesta lista.

Ex.: Expressões indicativas.

### **Caso Novo**

Finalizada as etapas de extrações de valores, o resultado será um novo caso a ser incluído na base de casos.

## Acrescentar na Base

Para concluir o processo de extração, acrescenta-se na base todos os novos casos. Os quais aumentarão a abrangência do conhecimento do sistema.

### 1.32. Algoritmos

Neste tópico serão apresentados os algoritmos desenvolvidos para a elaboração e construção para que o JurisConsulta aplicasse o modelo “RBC-Textual”.

#### Extração de novos casos

Ler arquivo texto

    Valorar atributos por localização

    Valorar atributos por inferência

    Valorar atributos usando vocabulário controlado

Fechar arquivo texto

#### Matching

Caso Entrada <= Descrição do usuário

    Enquanto não chegar ao final da base fazer

        Comparar Caso Entrada com Caso Base

        Aplicar métrica de similaridade

    Fim do Enquanto

    Guardar os valores da similaridade

    Ordenar Casos

    Filtrar base conforme as restrições valoradas e gerar base dos eleitos

    Apresentar os mais similares

    Se existir melhores sugerir

#### Filtrar

Fazer um SQL levando em consideração cada restrição fornecida pelo usuário.

#### Apresentar mais similares

Após filtro apresentar os mais similares.

#### Sugerir

Verificar se dentre os casos que ficaram de fora existe algum melhor

Se sim sugerir.

### **1.33. Considerações Finais**

Neste capítulo foi apresentado o mecanismo utilizado para o armazenamento e organização dos documentos na sua base de conhecimento. A utilização do banco de dados relacional apresentou uma performance satisfatória na aplicação do modelo. E, através da linguagem SQL conseguiu-se acessar, de maneira veloz e precisa, as informações desejadas.

Outros pontos abordados foram: os dicionários de dados e o fluxo de dados. Demonstrando qual o tipo de informação cada campo representava, bem como o fluxo que tais informações seguiam para que fossem coletados corretamente.

Finalmente, foram apresentados os algoritmos principais que nortearam a implementação do modelo.

## APLICAÇÃO

Este capítulo apresenta as fases envolvidas no desenvolvimento do sistema que tem como base o modelo proposto no trabalho.

### 1.34. Introdução

Nesta seção será descrita a implementação computacional de um sistema de RBC, utilizando o ambiente de programação Delphi o banco Access/SQL Server 6.5, respeitando as premissas teóricas vistas nos capítulos de Fundamentação Teórica.

A implementação computacional de um sistema de RBC para recuperação compreende um módulo de memória, um módulo de recuperação e um módulo de interface. A seguir serão demonstradas as etapas desenvolvidas na elaboração do protótipo que ilustra a aplicação do trabalho.

Neste protótipo, os casos originam-se a partir de textos (acórdãos) que descrevem situações legais. A Figura 15 mostra um acórdão em seu formato original de texto. A Figura 16 apresenta este mesmo caso (acórdão) conforme modelado na base de casos. Na seção seguinte serão discutidas as implementações utilizadas na transformação dos casos em formato de texto para o caso modelado.

DJ: 8.013 DATA: 17/05/90 PÁG: 08  
Apelação criminal n. 20.824, de Urubici.  
Relator: Des. Ernani Ribeiro.  
APELAÇÃO CRIMINAL. PEDIDO DE DESISTÊNCIA. HOMOLOGAÇÃO.  
Vistos, relatados e discutidos estes autos de apelação criminal n. 20.824, da comarca de Urubici, em que é apelante Antônio Domingos da Luz, sendo apelada a Justiça, por seu Promotor:  
ACORDAM, em Primeira Câmara Criminal, à unanimidade, homologar a desistência requerida.  
Custas de lei.  
Trata-se de pedido de desistência do recurso interposto por defensor em favor de ANTÔNIO DOMINGOS DA LUZ que na comarca de Urubici foi condenado à pena de 12 (doze) anos e 6 (seis) meses de reclusão, por infração ao art. 121, § 2o., IV do Código Repressivo. Presentes os pressupostos que autorizam o acolhimento da pretensão, homologa-se o pedido de desistência.  
Presidiu o julgamento, os Exmos Srs. Des. Ernani Ribeiro e participaram do mesmo, com votos vencedores, os Exmos Srs. Des. Marcio Batista e Wladimir d'Ivanenko.  
Florianópolis, 05 de abril de 1990.  
Nauro Collaço  
Presidente p/o acórdão  
Ernani Ribeiro  
Relator  
Nelson Ferraz  
Procurador de Justiça

Figura 15 - Acórdão



ID_NRO_ACORDAO	COD_RELATOR	COD_LOCAL	COD_RECURSO	DTA_PUBL_ACORDAO	TXT_RESULTADO
20824AC	6	22	1	17/5/1990	Concedido

ID_NRO_ACORDAO	SEQ_EXPR_INDICATIVA	RNE_EXPR_INDICATIVA
20824AC	1	Acórdão
20824AC	2	Câmara criminal
20824AC	3	Código
20824AC	4	Curador
20824AC	5	Custas
20824AC	6	Defensor
20824AC	7	homologa-se o pedido de desistência
20824AC	8	Marca
20824AC	9	presentes os pressupostos
20824AC	10	Pretensão
20824AC	11	Reclusão
20824AC	12	Reclusão
20824AC	13	Recurso
20824AC	14	Relator

Column: TXT_EMENTA	Column: TXT_ACORDAO
<p>APELAÇÃO CRIMINAL. PEDIDO DE DESISTÊNCIA. HOMOLOGAÇÃO.</p> <p>Acórdão: 20824AC - Relator: Ernani Ribeiro - Localização: Urubici</p> <p>Data: Resultado: Concedido</p>	<p>DJ: 8.013 DATA: 17/05/90 PÁG: 08</p> <p>Apeação criminal n. 20.824, de Urubici.</p> <p>Relator: Des. Ernani Ribeiro.</p> <p>APELAÇÃO CRIMINAL. PEDIDO DE DESISTÊNCIA. HOMOLOGAÇÃO. Vistos, relatados e discutidos estes autos de apelação criminal n. 20.824, de comarca de Urubici, em que é apelante Antônio Domingos da Luz, sendo apelada a Justiça, por seu Promotor.</p> <p>ACORDAM, em Primeira Câmara Criminal, à unanimidade, homologar a desistência requerida.</p> <p>Custas de lei.</p> <p>Título de pedido de desistência do recurso interposto por defensor em favor de ANTÔNIO DOMINGOS DA LUZ que na comarca de Urubici foi condenado à pena de 12 (doze) anos e 6 (seis) meses de reclusão, por infração ao art. 121, § 2o., IV do Código Repressivo. Presentes os</p>

EN_RELATOR	EN_LOCAL	EN_RECURSO
6	22	1

COD_RELATOR	NME_RELATOR
6	Ernani Ribeiro

COD_LOCAL	TXT_LOCAL
22	Urubici

COD_RECURSO	TXT_RECURSO
1	Apelação criminal

Figura 16 - Acórdão na base

### 1.35. Módulo de memória

Para que a técnica de RBC seja aplicada, primeiramente é necessário que exista um conjunto de atributos devidamente valorados (casos), formando a base de casos. Partindo desta premissa foi desenvolvido um módulo dentro do JurisConsulta que se chama Atualização da Base, Figura 17. Através deste módulo o sistema extrai dos textos (acórdãos) os dados necessários para representar o caso na base. Com isto tornou-se possível empregar a técnica de RBC para recuperar as experiências mais úteis.

Conforme ilustrado na Figura 8, a inserção de novos casos na base é de responsabilidade do engenheiro do conhecimento. Para realizar tal tarefa, vide Figura 17, basta o engenheiro selecionar os arquivos que deseja inserir o clicar no botão *Converter Arquivos*, após esta ação o sistema investigará os arquivos selecionados e extrairá as informações necessárias para a representação dos acórdãos na base. É o engenheiro, também, que irá dar autenticidade para cada valor extraído dos arquivos textos. Para isto, a interface do módulo de atualização permite que se faça alterações nos dados que formaram um novo caso. Uma vez que o engenheiro do conhecimento se certifique de que as informações estão corretas, o caso passa a estar efetivamente na base de casos, podendo participar do matching.

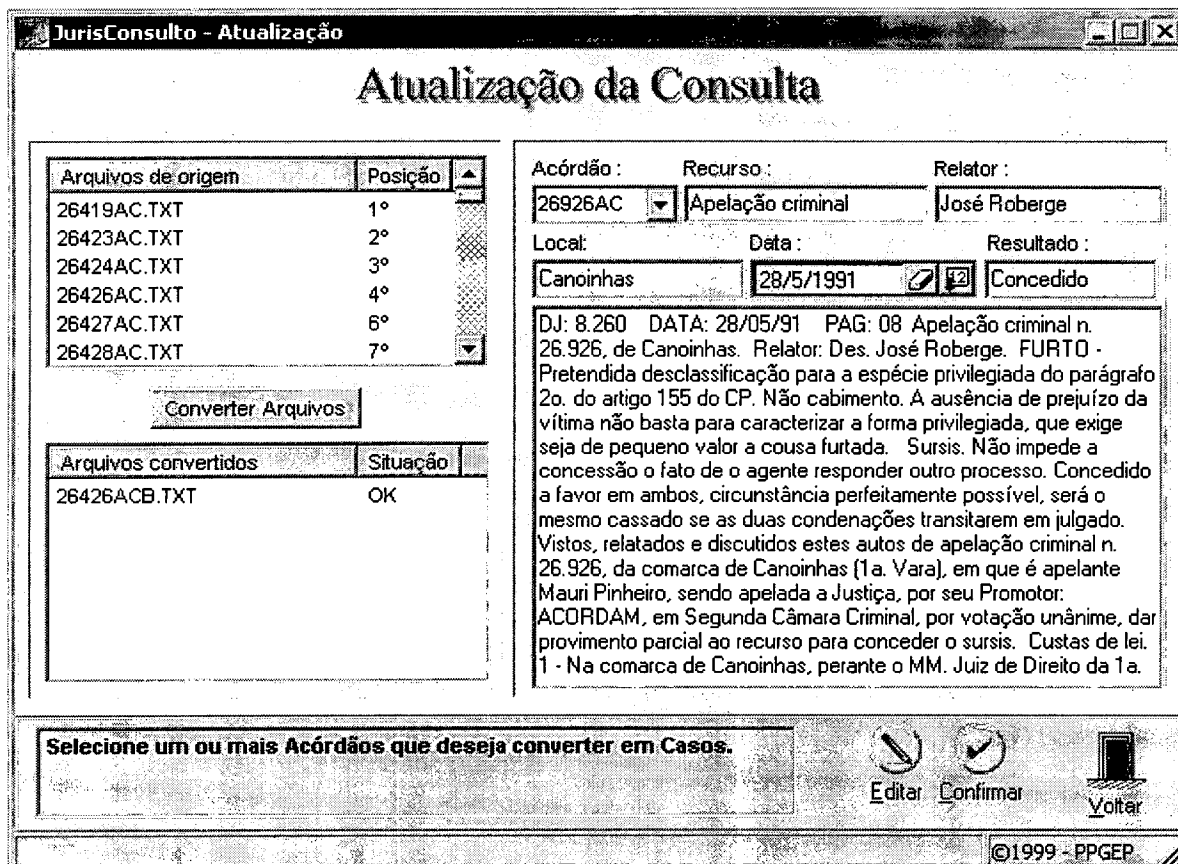


Figura 17 - Atualização da base

### 1.35.1. Extração dos dados

Nesta seção abordaremos as peculiaridades na valoração dos atributos, Tabela 7. O fato dos acórdãos manterem uma certa padronização, no que se refere à disposição das características que foram elencadas para representarem os casos, permitiu que estes dados fossem retirados diretamente dos acórdãos. Como já foi supracitado, a extração dos dados foi realizada de três maneiras:

- **Por localização;**
- **Por inferência;**
- **Usando vocabulário controlado.**

Atributos		Forma de extração
Número do acórdão		Por localização
Recurso		Por localização
Relator		Por localização
Local		Por localização
Data publicação		Por localização
Resultado		Por inferência
Ementa		Por localização
Acórdão		Por localização
Expressão indicativa		Usando vocabulário controlado
Tipificação	Tipo geral	Por inferência
	Modalidade	Por inferência
	Qualificação	Por inferência
	Tentativa	Por inferência
	Co-autoria	Por inferência

Tabela 7 - Atributos x Forma de extração

### Por localização

Valores que podem ser identificados no texto através da posição ou de palavras-chave. Por exemplo, os valores referentes aos atributos que representam o *tipo de recurso*, *número do acórdão* e *local* estão sempre localizados na segunda linha, Figura 18, do cabeçalho dos textos legais. Uma vez observando-se que em todos os acórdãos apresentam esta mesma características, converteu-se esta abstração em código no JurisConsulta. Procedimentos similares foram realizados para os outros atributos que, também, podem ser valorados de forma direta.

DJ: 8.444 DATA: 19/02/92 PAG: 09 <b>Apelação criminal n. 20.402, de Itajaí.</b> Relatora: Des. Thereza Tang. Crime de imprensa. Jornalista que entende estar acobertado pelo art. 27, inciso V, da Lei específica, argüindo, ainda, incompetência do juízo. Alegação inconsistente. Sentença mantida.
--

Figura 18 - Valoração dos atributos - por localização

### Por inferência

Alguns atributos não são identificados de maneira explícita nos acórdãos, portanto, é necessário fazer inferências para a sua valoração. Tais inferências traduzem as regras elaboradas através da avaliação de uma gama de acórdãos, onde se identificou características comuns para a valoração de determinados atributos. Por exemplo, o valor referente ao atributo *Resultado* é extraído da seguinte forma:

- Primeiramente encontra-se a linha que inicia com a palavra **ACORDAM**. Nesta linha verifica-se a existência de expressões, Tabela 8, que significam *Denegado* ou *Concedido*, após encontrar estas expressões valora-se os atributos com os símbolos correspondentes, ou seja, *Denegado* ou *Concedido*.

DENEGADO	CONCEDIDO
negar provimento	dar provimento
negando-se-lhes provimento	dar-lhe provimento
negar-lhe provimento	provimento parcial
negar-se provimento	extinção da punibilidade
negando-se provimento	extinta a punibilidade
desprover o recurso	prover o recurso
desprover o apelo	homologar a desistência
não conhecer do recurso	conceder
não conhecer do apelo	decretar a nulidade da sentença
não conhecer do pedido	anular o processo
prejudicado o recurso interposto	anular a sentença
de ofício	
prejudicado o recurso	
julgar deserta a apelação	
acolher a preliminar de	

Tabela 8 - Expressões para Denegado/Concedido

### Usando vocabulário controlado

Este tipo de extração ocorre quando se utiliza uma lista pré-definida de valores, vocabulário controlado<sup>4</sup>, para a valoração do atributo. Através desta lista serão pesquisadas nos acórdãos a existentes das expressões nela contida. Após a realização da correlação, os itens identificados serão acrescentados no caso correspondente. Este tipo de extração é utilizado para o atributo *Expressão Indicativa*.

Portanto, a alimentação da base de dados do JurisConsulta foi realizada através da codificação dos tipos de extrações supra-citados.

## **1.36. Módulo de recuperação**

O objetivo do sistema é recuperar os casos mais úteis para a solução de um problema jurídico atual. Após o problema ser descrito pelo usuário, o JurisConsulta inicia a comparação dele com os casos da base.

Para avaliar a similaridade, primeiramente se identificam as características do problema de entrada (pares atributos-valor correspondentes aos índices), para que se possa fazer a comparação com os casos candidatos da base de conhecimento. A obtenção do grau de similaridade entre os casos sendo comparados é dada por meio de uma medida de similaridade. Na seqüência, os casos são ordenados de maneira decrescente, considerando os valores resultantes da métrica de similaridade. Após realizada a ordenação, os casos mais úteis em relação a um determinado limiar, ou seja, com valores de similaridade mais altos, são apresentados ao usuário.

### **1.36.1. Indexação**

A partir da modelagem dos casos, Figura 7, é necessário definir o vocabulário de índices. A este processo dá-se o nome de *indexação*. O vocabulário de índices é composto pelos atributos que irão nortear a similaridade existente entre o problema de entrada e os casos da base, para que seja possível recuperá-los.

Com o intuito de representar os textos jurídicos de uma maneira acessível, o texto jurisprudencial é mapeado em uma representação estruturada de acordo com a utilizada, juntamente com o contexto das informações nele inseridas. Assim, um conjunto de índices é determinado para facilitar o processo de recuperação.

No contexto jurisprudencial, o principal problema é, justamente, a determinação dos índices e, conseqüentemente, os valores usados para a recuperação. O princípio da analogia sustenta a maneira de reutilização dos casos jurídicos; por esta razão, os atributos dos documentos textuais, usados como índices para a recuperação, têm que indicar a utilidade das informações do caso na situação presente. E assim, os índices são determinados do ponto de vista do especialista do domínio, em conformidade com a importância deles para o contexto de recuperação da informação jurídica. O resultado desta forma de representação é os índices que representam o conteúdo jurídico dos textos jurisprudenciais.

### Vocabulário de Índices

**Expressões indicativas:** Este índice determina as informações jurídicas relevantes e discriminantes num caso. Indica a matéria jurídica de interesse. É o principal atributo de indexação e recuperação dos casos, ou seja, é o que apresenta peso de maior valor, conseqüentemente é o que mais contribui na métrica de similaridade. Outra característica deste índice é sua multivaloração, ou seja, cada caso pode receber vários valores representando as *expressões indicativas*.

**Tipificação:** É o índice que representa a capitulação (enquadramento) do crime ou lei na qual algum réu é ou está incurso. Este índice é composto de um subconjunto de índices, conforme descrito a seguir e ilustrado na Tabela 9. Através da forma de extração *por inferência*, valora-se os atributos que compõem este índice, ou seja, utilizando o conhecimento do domínio modelado em regras. Os artigos que representam o enquadramento do delito no acórdão são precedidos por expressões padrões, Tabela 10, identificados pelos especialistas do domínio e terminam com a indicação da fonte, Tabela 11.

---

<sup>4</sup> No domínio do problema, o vocabulário é formado por um conjunto de termos jurídicos que são usados

ÍNDICE	VALOR
TIPIFICAÇÃO	“Art. 121, § 2., IV c/c art. 14, ambos do Código Repressivo”
Tipo geral	<b>Art. 121</b> – Homicídio
Modalidade do crime	§ 2 – Doloso
Qualificação do crime	<b>IV</b> – Qualificado
Tentativa	<b>Art. 14</b> – Sim . Quando este artigo não é citado, atribui-se o valor “Não” para este índice.
Co-autoria	<b>(Art. 29)</b> – sim; Quando este artigo não é citado, atribui-se o valor “Não” para este índice, como ocorre neste exemplo.

Tabela 9 - Valores do índice Tipificação

EXPRESSÕES PRECEDENTES
Previsto no artigo xx
Definido no artigo xx
Incurso no artigo xx
Infração ao artigo xx
Violação ao artigo xx

Tabela 10 - Algumas expressões precedentes

Fontes
Código Processual Penal
Código Penal
Código Penal Brasileiro
Código Repressivo
Lei antitóxicos

Tabela 11 - Algumas fontes

**Tipificação - Tipo geral:** Índice responsável pela designação do enquadramento legal genérico, que é o tipo de crime que deu origem ao processo judicial. Indica o tipo de crime: *homicídio, furto, estelionato*, etc...

---

normalmente nos Tribunais, que representam fatos enquadrados normativamente.

**Tipificação - Modalidade do crime:** Índice que aponta a modalidade do crime. Os crimes podem ser cometidos por uma pessoa quando a mesma tem a intenção clara e objetiva de cometê-lo, ou quando, por sua culpa, o resultado do crime foi atingido. Indica se o crime foi cometido na modalidade *culposa* ou *dolosa*. Sendo que a modalidade concernente ao dolo é punível de uma forma mais severa que a modalidade referente à culpa, no Direito penal.

**Tipificação - Qualificação do crime:** Índice que determina se um crime foi praticado em condições especiais (ex.: furto praticado à noite), ou se o agente possui uma característica especial que aumente ou diminua a pena (ex.: reincidência). Os valores deste índice indicam se o crime é *simples*, *qualificado* ou *privilegiado*. A qualificação também é usada para efeito de penalização de um crime.

**Tipificação - Tentativa:** Índice que indica se houve ou não tentativa de crime. Uma pessoa pode ser punida por um crime, mesmo quando o crime não atingir o resultado final. Por exemplo, errar o disparo de uma arma de fogo, demonstrando a clara intenção de matar a vítima. Determina se a ação criminosa atingiu o seu resultado ou se o criminoso não conseguiu atingir o seu objetivo, por motivos alheios à sua vontade.

**Tipificação - Co-autoria:** Índice que aponta se existiu ou não co-autoria no crime. O mesmo crime pode ser cometido por duas ou mais pessoas. Estas pessoas (agentes) serão punidas conforme a sua participação. Um crime cometido por duas ou mais pessoas é o chamado *concurso de agentes*.

### 1.36.2. Avaliação da similaridade

De acordo com as definições, é a avaliação da similaridade que possibilita a realização da comparação entre o caso de entrada e os casos da base, com o objetivo de encontrar as experiências passadas mais semelhantes e, conseqüentemente, mais úteis para o problema atual. Portanto para que o confronto entre os casos seja viável, é necessário que haja uma métrica de similaridade que represente numericamente o quanto um caso é semelhante a



outro. E, após encontrar os valores de semelhança, estes serão utilizados para, em ordem decrescente, apresentar os casos mais similares.

Somente os índices (*tipificação e expressões indicativas*) participam do cálculo da similaridade. Os demais atributos (*número do acórdão, data da publicação, etc.*) participam somente como delimitadores. Neste trabalho a avaliação da similaridade é baseada no método do *vizinho mais próximo*, Figura 19:

$$\frac{\sum_{i=1}^n w_i \times \text{sim}(f_i^I, f_i^R)}{\sum_{i=1}^n w_i}$$

Figura 19 - Vizinho mais próximo

Onde,

$w_i$  = peso do índice  $i$

$f_i^I$  = índice  $i^{\text{th}}$  do caso de entrada

$f_i^R$  = índice  $i^{\text{th}}$  do caso na base

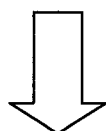
Com base na métrica de similaridade é necessário que cada índice do caso de entrada seja comparado com o seu correspondente na base de dados. Após a realização do confronto entre os índices, sob luz da métrica, os valores de semelhança de cada caso são encontrados. Observa-se, também, que, dependendo da importância de cada índice na recuperação dos casos, eles receberão pesos diferentes apontando o quanto um é mais relevante do que outro.

CASO DE ENTRADA				CASO NA BASE			
Tipificação	Tipo Genérico	Furto	I <sub>1</sub>	1	Tipificação	Tipo Genérico	furto
	Modalidade	Doloso	I <sub>2</sub>	1		Modalidade	doloso
	Caracterização	Qualificado	I <sub>3</sub>	1		Caracterização	qualificado
	Tentativa	Sim	I <sub>4</sub>	0		Tentativa	não
	Co-autoria	Não	I <sub>5</sub>	0		Co-autoria	sim
Expressões Indicativas		Flagrante	I <sub>6,1</sub>	1	Expressões Indicativas		flagrante
		reincidência	I <sub>6,2</sub>	0			menoridade
		erro no cálculo da pena	I <sub>6,3</sub>	1			erro no cálculo da pena

Tabela 12 - Caso de entrada x Caso na base

Peso das *expressões indicativas* = 5

Peso da *tipificação* = 1



$$\text{Sim} = \frac{1 \cdot W_{\text{tipoGeral}} + 1 \cdot W_{\text{mod}} + 1 \cdot W_{\text{car}} + 0 \cdot W_{\text{tent}} + 0 \cdot W_{\text{coAut}} + (2 \cdot W_{\text{ExprInd}})}{W_{\text{tipoGeral}} + W_{\text{mod}} + W_{\text{car}} + W_{\text{tent}} + W_{\text{coAut}} + (3 \cdot W_{\text{ExprInd}})}$$

$$\text{Sim} = \frac{1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + (2 \cdot 5)}{1 + 1 + 1 + 1 + 1 + (3 \cdot 5)} = \frac{13}{20} = 0.65$$

**Medida Similaridade ⇒ 65%**

Figura 20 - Medida de similaridade

A Tabela 12 ilustra o confronto entre um caso de entrada e um determinado acórdão modelado na base de casos. Os índices que apresentam valores iguais representam que sua similaridade local é igual a 1 (um) e os que apresentam valores diferentes indicam que sua similaridade local é igual a 0 (zero). Portanto, os valores referentes aos índices *tentativa*, *co-autoria* e um valor do índice *expressão indicativa* não participaram da métrica.

Na Figura 20 observa-se a contribuição de cada índice para se encontrar o valor de similaridade entre os casos. Constata-se, também, que o índice *expressões indicativas* é mais relevante e quanto mais valores de *expressões indicativas* coincidirem, maior será o valor de semelhança entre os casos.

### 1.37. Módulo de interface

Neste tópico apresentam-se a telas de navegação do sistema, devidamente ilustradas. Primeiramente abordaremos a etapa de *Elaboração da consulta*, Figura 21, seguida pela etapa de *Resultado* e, concluindo com a apresentação do acórdão na íntegra.

#### 1.37.1. Elaboração da consulta

É neste momento que o JurisConsulta captura a descrição do problema a ser trabalhado. Para que o usuário possa descrever o caso de entrada, o sistema disponibiliza na sua interface uma opção onde é permitida a digitação de um breve resumo do problema em questão.

Além da *descrição do caso*, o sistema ainda possui nesta tela mais duas situações que requerem a ação do usuário, são elas: *nível de restrição* e *delimitação da pesquisa*.

##### Descrição do caso

É através deste espaço que o usuário irá informar ao sistema, o problema. Problema este, que será o motivador para que o JurisConsulta pesquise na base, por casos similares. Na descrição, o profissional do Direito poderá contar um breve resumo do fato ocorrido, desta forma não precisará utilizar operadores lógicos na sua consulta. Através do resumo descrito o aplicativo irá inferir os valores para os índices que representarão o caso a ser pesquisado. Após concluída a descrição do caso o sistema já está apto para iniciar a busca.

Figura 21 - Elaboração da consulta

### Nível de restrição

Para que o JurisConsulta trabalhe com mais eficácia e rapidez, o usuário terá à sua disposição a opção de manipular a restrição dos casos, Figura 22, a serem pesquisados na base. O efeito da restrição, segundo Watson et al. (1999) é no sentido de disponibilizar, para a consulta, apenas os casos mais úteis com relação à entrada. Após ser escolhido um nível de restrição, o sistema faz uma pré-consulta, sem utilizar a métrica de similaridade, com intuito de selecionar os casos que apresentem informações relevantes. O nível escolhido influencia da seguinte forma: a partir da descrição, o sistema faz uma pré-consulta na base, selecionando somente os casos que apresentem pelo menos um número x de informações, dependendo do nível de restrição informado.

Os níveis de restrição possíveis são: *Baixo, Médio e Alto*.

- Baixo – com o nível de restrição baixo, o JurisConsulta irá pesquisar em mais casos na base, pois selecionará casos que apresentem um número menor de informações, com isto mais casos serão selecionados para participar da busca utilizando a métrica de similaridade.
- Médio – o nível médio irá pré-selecionar um nível ideal de casos para serem utilizados na pesquisa final. O tempo gasto pelo sistema, através deste tipo de restrição foi considerado o mais adequado.
- Alto - quanto mais restrito for o nível, menor será o tempo gasto na busca, pois a pré-consulta irá selecionar os casos que apresentem um número maior de informações. Sendo assim, menos casos serão selecionados após a pré-consulta.

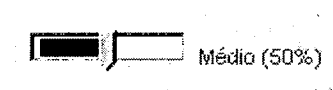


Figura 22 - Nível de restrição

### **Delimitação da consulta**

O objetivo da *delimitação da consulta* é fornecer ao usuário a possibilidade de ele poder restringir a gama de casos a participarem da métrica. Tal delimitação pode ser interessante quando o usuário deseja pesquisar somente nos acórdãos que possuam uma determinada característica específica. Por exemplo, se a intenção da consulta for atingir somente os casos ocorridos em um município definido, não é útil que o sistema faça a comparação em todos os casos da base. Portanto, basta que o usuário especifique o município ao qual a pesquisa será restrita para que o JurisConsulta reduza o espaço de busca somente a aos casos referentes a ele.

É importante ressaltar que o conjunto de campos que compõe a *delimitação da consulta* inclui apenas atributos, ou seja, pelo fato de não serem índices, não participam da métrica de similaridade.

### **Exemplo de uma *Elaboração da Consulta***

Baseado nas explanações concernentes à *Elaboração da Consulta*, apresenta-se um exemplo ilustrando o comportamento dos componentes participantes desta etapa.

Figura 23 - Caso de entrada

Primeiramente o usuário descreve um breve resumo do problema a ser consultado. Com já foi descrito, é através deste resumo que o sistema irá retirar os valores que representarão o caso a ser consultado.

Com a descrição do caso o JurisConsulta já está apto para fazer a busca na base de conhecimento. Porém, quando o usuário já tiver uma certa experiência com a utilização do sistema, ele pode julgar que não há necessidade de que a pesquisa seja realizada em todos os casos. Então, pode-se utilizar os recursos oferecidos na interface, com o intuito de diminuir o tempo de resposta.

Na Figura 23, além da descrição do caso, foi alterado o *nível de restrição*. Com isto a base de casos a ser pesquisada será mais restrita, ou seja, o JurisConsulta irá fazer a comparação do caso de entrada com menos casos na base. Pode observar ainda, na Figura 23, que o número de casos na base é 2247 e, após o aumento do nível de restrição o número de casos a serem pesquisados diminui para 101. Desta forma, comparando somente com os casos mais úteis, o tempo de resposta será bem menor.

Outra maneira de restringir o número de casos a ser pesquisado é através da utilização dos atributos que fazem parte da *delimitação da pesquisa*.

É importante salientar que este tipo de procedimento irá retirar alguns casos que poderiam ser comparados. Portanto, o usuário precisa já estar mais acostumado com o sistema para utilizar os recursos adicionais.

Uma vez que a *elaboração da consulta* esteja finalizada, pressiona-se o botão *Consultar*, para iniciar a recuperação, que representa o processo de raciocínio do sistema de RBC. É neste momento que o JurisConsulta irá fazer a comparação do caso de entrada com os casos da base, ativando a métrica de similaridade e verificando o percentual de semelhança entre eles.

### 1.37.2. Resultado da Consulta

Nesta etapa, após a finalização da consulta, o sistema apresenta os casos mais úteis encontrados na base de casos. Para que o usuário possa discernir com mais precisão e escolher dentre os casos resultantes qual é o mais útil com relação ao seu problema, na fase do *resultado da consulta* são apresentados alguns atributos que servem para elucidar melhor a importância de cada caso recuperado.

Os atributos apresentados são: *Ementa, IdArquivo, Relator, Local, Data, e Resultado*.

#### Exemplo de um Resultado da Consulta

Continuando o exemplo da *elaboração da consulta*, observa-se agora os casos apresentados no *resultado da consulta*, Figura 24, a partir da descrição realizada no

módulo anterior. Os casos são apresentados em ordem decrescente, observando os valores de similaridade obtidos na etapa de recuperação.

O *resultado da consulta* apresenta ao usuário os casos mais similares, para que ele possa escolher qual o mais relevante na utilização do seu problema em questão. O fato de um caso apresentar um valor maior de similaridade não implica que este será o mais útil para ser aplicado no problema sendo explorado. Portanto, o profissional do Direito observará os atributos que lhe são apresentados e resolverá qual resultado melhor se adequa ao seu problema.

Neste exemplo verificamos que os três primeiros casos, realmente, apresentam alguma semelhança com a descrição realizada na etapa de *elaboração da consulta*.

O usuário pode, ainda, não ficar satisfeito com nenhum dos resultados obtidos. Então, ele poderá retornar ao módulo da *elaboração da consulta*, e melhorar a descrição do seu caso.

Depois de eleger o melhor resultado, o usuário poderá pressionar o botão *Acórdão* para analisar o texto do acórdão escolhido na íntegra.

Posição	Similar.	Ementa
1º	76,90	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 da narcotraficância - Decisão mantida - Recurso ... Acórdão: 26016AC - Relator: Vladimir d'Ivanenko - Localização: São José - Data: 31/08/2090 - Resultado: Denegado
2º	69,20	TRÁFICO DE ENTORPECENTE. A jurisprudência e a doutrina predominantes não admitem a tentativa de tráfico de entorpecentes. Condenação baseada no conjunto probatório. Recurso provido. Acórdão: 27480AC - Relator: Odete Maria de Oliveira - Localização: Joinville - Data: 24/06/2092 - Resultado: Concedido
3º	61,50	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 PERÍCIA, MORMENTE SE MÍNIMA E INSIGNIFICANTE É ESSA ... Acórdão: 33123AC - Relator: Napoleão Xavier do Amarante - Localização: Curitiba - Data: 28/08/2095 - Resultado: Denegado

Acórdão    Ajuda    Elaborar

©1999 - PPGEP

Figura 24 - Resultado da consulta



### 1.37.3. Acórdão na integra

Para finalizar, nesta etapa o sistema apresenta, na integra, o acórdão, Figura 25, eleito pelo usuário. Com isto, o sistema completa o seu ciclo modelado sob o paradigma da técnica de RBC.

#### Exemplo de um Resultado da Consulta

Concluindo o exemplo iniciado com a *elaboração da consulta*, apresentamos o acórdão na integra, Figura 25, que representa o resultado que estava na segunda posição com 69,20% de similaridade, no módulo *resultado da consulta*.

Nesta fase o usuário pode, novamente, retroceder aos módulos anteriores, se não ficar satisfeito com a experiência contida no acórdão escolhido, e selecionar outros resultados ou ainda voltar ao módulo da *elaboração da consulta* melhorar a elaboração.

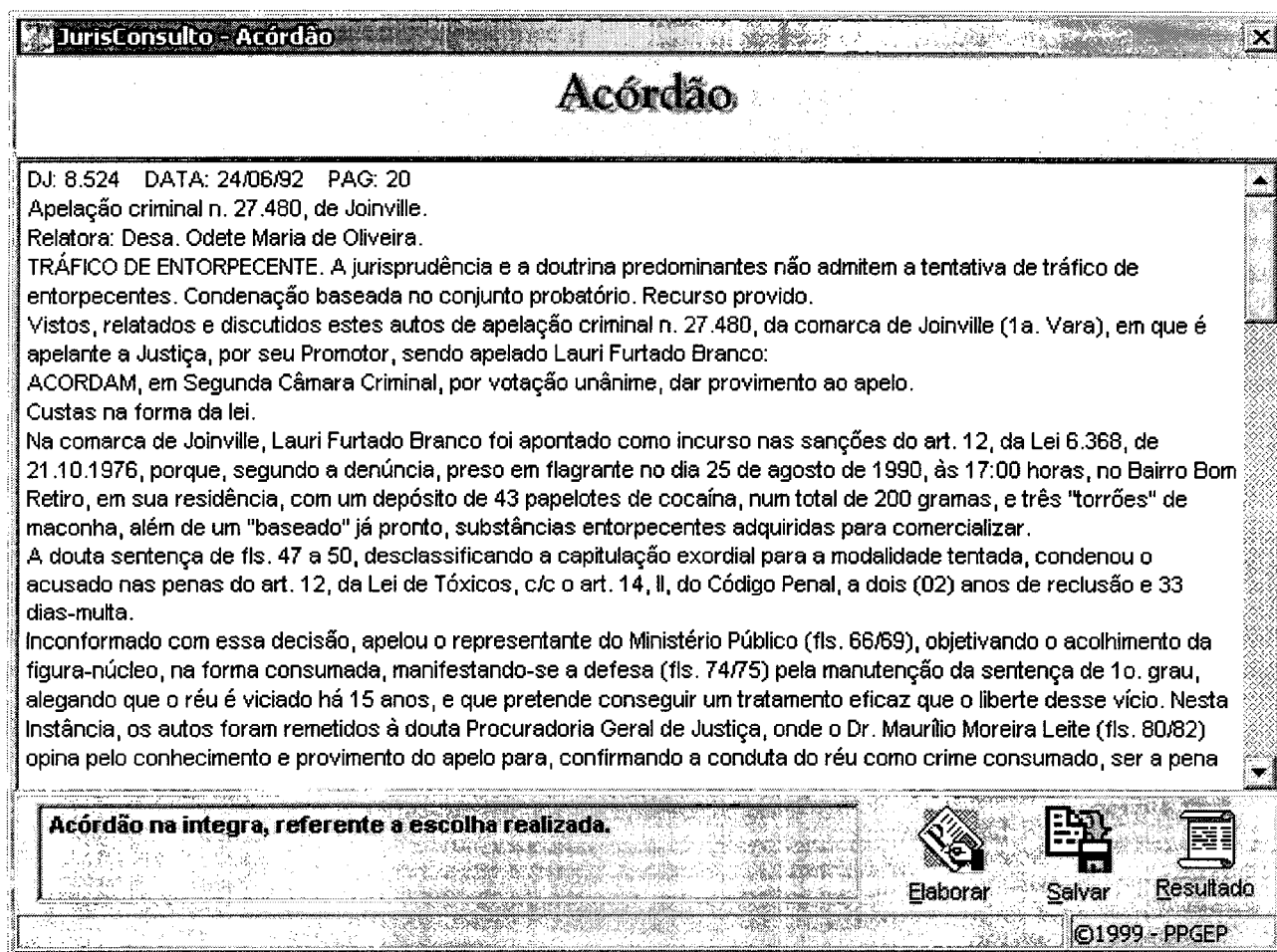


Figura 25 - Acórdão na integra

### **1.38. Considerações Finais**

Com o desenvolvimento da aplicação ficou comprovada a eficiência do modelo. Uma vez que todas as etapas previstas foram validadas através da modelagem dos documentos jurídicos (acórdãos) e, principalmente, pela materialização desta modelagem na aplicação apresentada neste capítulo.

## CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Alguns estudos apontam que, atualmente, mais de 90% das informações encontram-se em formato não-estruturado, ou seja, em documentos textuais. Entretanto, as ferramentas disponíveis para a recuperação destas informações não suprem as necessidades dos usuários por completo. Uma vez, que elas não são flexíveis, o suficiente, para compreender um pergunta em texto livre e retornar os assuntos mais relevantes baseados no contexto do questionamento.

A utilização de Inteligência Artificial para o tratamento de documentos textuais vem se destacando na comunidade científica, notadamente, para a recuperação de experiências anteriores que possam ser úteis em uma situação no presente, a técnica de Raciocínio Baseado em Casos é a mais recomendada.

Portanto, o presente trabalho apresentou um modelo (baseado na técnica de RBC) para trabalhar com documentos textuais, de maneira a estruturá-los a luz do conhecimento humano especialista. Permitindo que os aplicativos construídos tendo como base o modelo “RBC-Text” recuperem o conhecimento especialista mediante um pergunta em linguagem coloquial, ou seja, sem a utilização de palavras-chave e sim através de conceitos contextualizados.

O desenvolvimento do software JurisConsulta consolidou o modelo “RBC-Text” como sendo apto para a sua utilização em documentos textuais. Haja vista, que todas as fases elencadas no modelo foram executadas durante o processo de criação do aplicativo, apresentando um resultado eficiente.

No presente trabalho o modelo foi validado com o desenvolvimento de uma ferramenta de funcionamento local. Porém, com o advento e crescimento constante da “rede mundial de computadores”, os dados trafegam e chegam ao conhecimento do público de maneira instantânea. Informações que antigamente poderiam levar horas ou, até mesmo, dias até o seu destino, atualmente chegam em segundos. Esta evolução constante, aliando-se às tecnologias que permitem maior dinamismo e flexibilidade de interação com a Web (por exemplo, CGI, Java e ActiveX), têm levado muitos pesquisadores e desenvolvedores a utilizarem-na como meio de interação com as aplicações ( Raposo et al., 1999). Com a

massificação do uso da *Internet*, cresce a necessidade de aplicações que localize informações baseadas em conceitos, uma vez que perde-se muito tempo para a recuperação de conteúdos, realmente, úteis. Portanto, pretende-se evoluir os estudos concernentes ao tema abordado, com o intuito de agregar funções no modelo que permita o desenvolvimento de aplicações para a WEB.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. AAMODT, A; Plaza, E. *Case Based Reasoning: Foundational Issues, Methodological Variations, and Systems Approaches*. AICom - Artificial Intelligence Communications – Vol. 7, n. 1, pp. 39-59, 1994. [online] Disponível na Internet via WWW. URL:// <http://www.iiia.csic.es/People/enric/AICom.html>. 17 de março de 1999.
2. AIKENHEAD, M.; *A Discourse on Law and Artificial Intelligence*. [online] Disponível na Internet via WWW. URL:// <http://www.law.warwick.ac.uk/ljtj/5-1c.html>. 05 de agosto de 2000.
3. ANDRADE, P. G. S. *A Importância da Informática para o Profissional de Direito*. 1º Seminário Brasileiro de Divulgação Jurídica, 1998. [online] Disponível na Internet via WWW. URL://[http://infojur.ccj.ufsc.br/arquivos/artigos/A\\_importancia\\_da\\_informatica\\_para\\_o\\_profissional\\_de\\_direito.htm](http://infojur.ccj.ufsc.br/arquivos/artigos/A_importancia_da_informatica_para_o_profissional_de_direito.htm). 01 de agosto de 2000.
4. ATHENIENSE, A. R. *Advocacia e Informática*. Livros de Teses da XVI Conferência Nacional de Advogados, 1996. [online] Disponível na Internet via WWW. URL://[www.jus.com.br/doutrina/informa.html](http://www.jus.com.br/doutrina/informa.html). 20 de dezembro de 2000.
5. BERGMANN, R. *Introduction to Case-Based Reasoning*, 1998. [online] Disponível na Internet via WWW. URL: <http://www.cbr-web.org>. 25 de abril de 2000.
6. BRANTING, Karl L. *Reasoning with Portions of Precedents, Proceedings of the Third International Conference on Artificial Intelligence and Law*, Oxford, England, 25-28 de junho de 1991. [online] Disponível na Internet via WWW. URL: <http://pyramid.cs.uwyo.edu/~karl/pubs.html>. 23 de outubro de 2000.
7. BR-BUSINESS.; *Modelagem de dados*. [online] Disponível na Internet via WWW. URL: <http://www.br-business.com.br/bdados/modelagem.htm>. 20 de maio de 2002.
8. BUENO, T. C. Entrevista com o especialista em 22 de maio de 1999.
9. CINTRA, A. C. A.; GRINOVER, A. P.; DINAMARCO, C. R. *Teoria Geral do Processo*. 15ª Ed. São Paulo : Malheiros, 1999.
10. DATE, C. J.; *Banco de Dados – Tópicos Avançados*. 2ª Ed. Rio de Janeiro : Campus, 1988.
11. DEOCLECIANO, J.; GRISI, G.; BENÍCIO, J. C. et al. *A Informática e o Direito*. Opinião Jurídica, Paraíba, n. 21, jan/1999. <http://www.datavenia.inf.br/opiniao/infjur01.html>. 25 de maio de 2000.
12. HOESCHL, H. C. *Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU*. Tese de Doutorado, PPGE/UFSC. Florianópolis: 2001.

13. KOLODNER, J. *Improving Human Decision making through Case-Based Decision Aiding*. 1991. [online] Disponível na Internet via WWW. URL: <http://www.aaai.org/Library/Magazine/Vol12/12-02/vol12-02.html>. 10 de março de 2000.
14. KOLODNER, J. *Case-Based Reasoning* - Los Altos: Morgan Kaufmann Publishers, CA, 1993.
15. LENZ, M. *Defining Knowledge Layers for Textual Case-Based Reasoning*. 2001. [online] Disponível na Internet via WWW. URL: <http://citeseer.nj.nec.com/5305.html>. 22 de junho de 2001.
16. KOLODNER, J.; LEAKE, D. *A Tutorial Introduction to Case-Based Reasoning*. [online] Disponível na Internet via WWW. URL: <http://www.cc.gatech.edu/~jimmyd/summaries/kolodner1996-2.html>. 15 de junho de 2000.
17. LEAKE, D. B.; WILSON, D. C. *When Experience is Wrong: Examining CBR for Changing Tasks and Environments*. [online] Disponível na Internet via WWW. URL: <http://www.cs.indiana.edu/~leake/papers/p-99-04.pdf>. 25 de julho de 2000.
18. Mirra, A. L. V. *Inovações da jurisprudência em matéria ambiental*. Revista N.º 29 - Ano 15, 994 - p. 30-45. [online] Disponível na Internet via WWW. URL: [http://infojur.ccj.ufsc.br/arquivos/RevistasCCJ/Sequencia\\_numero29/Mirra-Inovacoes\\_da\\_jurisprudencia\\_em\\_materia\\_ambiental.html](http://infojur.ccj.ufsc.br/arquivos/RevistasCCJ/Sequencia_numero29/Mirra-Inovacoes_da_jurisprudencia_em_materia_ambiental.html). 28 de julho de 2000.
19. NOORTWIJK, K; MULDER, R. V. *The Similarity of Text Documents*. [online] Disponível na Internet via WWW. URL: [http://elj.warwick.ac.uk/jilt/artifint/97\\_2noor/default.htm](http://elj.warwick.ac.uk/jilt/artifint/97_2noor/default.htm). 8 de julho de 2000.
20. RACINE, K.; YANG, Q. *On the Consistency Management of Large Case Bases: the Case for Validation*, Burnaby, Canada, 1996. [online] Disponível na Internet via WWW. URL: <http://citeseer.nj.nec.com/racine96consistency.html>. 20 de janeiro de 2001.
21. RAPOSO, A. B.; MAGALHÃES, L. P.; RICARTE, I. L. M. *Interação na WEB*. Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação (SBC'99 - JAI), v. 2, Rio de Janeiro : EntreLugar, 1999. ISBN 85-87424-03-3.
22. RIBEIRO, M.; *Modelagem de Dados*. [online] Disponível na Internet via WWW. URL: <http://www.marcosribeiro.hpg.ig.com.br/modelagemdados.htm>. 21 de maio de 2002.
23. SARAIVA, W. C. *Aplicações da Informática à Atividade Jurídica*. Revista dos Tribunais – São Paulo, Vol. 710, Dezembro. 1994.
24. SILVA, P. *Vocabulário jurídico*. 15ª Ed. Rio de Janeiro : Forense, 1999.

25. SENGUPTA, A.; WILSON, D. C.; LEAKE, D. B. *Constructing and Transforming CBR Implementations: Techniques for Corporate Memory Management*. [online] Disponível na Internet via WWW. URL: <http://www.cs.indiana.edu/hyplan/leake/papers/p-99-07.pdf>. 15 de junho de 2000.
26. SYCARA, K. *Using Case-Based Reasoning for Plan Adaptation and Repair*. [online] Disponível na Internet via WWW. URL:// <http://online.loyno.edu/cisa494/papers/Sycara.html>. 03 de agosto de 2000.
27. SLADE, S. *Case-Based Reasoning: A Research Paradigm*. [online] Disponível na Internet via WWW. URL: <http://www.aaai.org/Library/Magazine/Vol12/12-01/vol12-01.html>. 12 de março de 2000.
28. SOUZA, F. A. C. *O Direito na Era Digital*. JUS – Direito e Informática. <http://planeta.terra.com.br/arte/arrabal/jus/>. 27 de julho de 2000.
29. SPRING; *Consulta a Banco de Dados*. [online] Disponível na Internet via WWW. URL: <http://www.dpi.inpe.br/spring/usuario/indicec.htm>. 23 de junho de 2002.
30. VASCONCELOS, A. V. R.; *Revista Panorama da Justiça* – São Paulo, n. 2, Outubro/Novembro. 1996.
31. VASCONCELOS, A. V. R.; *Revista Panorama da Justiça* – São Paulo, n. 3, Dezembro/Janeiro. 1996/1997.
32. WATSON, I.; MARIR, F., *Case Based Reasoning: A Review. The knowledge Engineering Review* – Londres, Vol. 9, n. 4, 1994.
33. WATSON, I.; GARDINGEN, D. *A Distributed Case-Based Reasoning Application for Engineering Sales Support*. 1999. [online] Disponível na Internet via WWW. URL: [http://www.cs.auckland.ac.nz/~ian/papers/cbr/watson\\_ijcai99.pdf](http://www.cs.auckland.ac.nz/~ian/papers/cbr/watson_ijcai99.pdf). 15 de junho de 2000.
34. WATSON, I.; *The Case for Case Based Reasoning*. [online] Disponível na Internet via WWW. URL: <http://146.87.176.38/ai-cbr/Papers/ita01.htm>. 12 de maio de 2000.
35. WEBER-LEE, R.; *Raciocínio Baseado em Casos*. [online] Disponível na Internet via WWW. URL: <http://www.eps.ufsc.br:80/~martins/fuzzy/cbr/intro.htm>. 23 de outubro de 1996.
36. WEBER-LEE, R.; *Pesquisa Jurisprudencial Inteligente*. [online] Disponível na Internet via WWW. URL: <http://www.eps.ufsc.br/teses98/rosina/index.html>. 18 de julho de 2001.

## ANEXO A

Artigo aprovado na II Encontro Nacional de Inteligência Artificial – ENIA99. 19-21 de julho de 1999, Rio de Janeiro, Brasil.

### Uso da Teoria Jurídica para Recuperação em Amplas Bases de Textos Jurídicos

**EDUARDO MATTOS, TÂNIA C. D. BUENO, CHRISTIANE VON WANGENHEIM,  
HUGO C. HOESCHL, RICARDO M. BARCIA**

Programa de Pós-Graduação em Engenharia de Produção da  
Universidade Federal de Santa Catarina, Brasil  
{ emattos, bueno, gresse, metajur, rbarcia }@eps.ufsc.br

**Resumo** No domínio do Direito, a jurisprudência tem um importante papel como fonte do direito e o seu conteúdo auxilia na interpretação da lei e sua aplicação na solução de um problema jurídico. Entretanto, estes documentos jurídicos são de difícil acesso devido ao grande número de decisões existentes e a forma que elas são armazenadas (livros, resenhas, banco de dados, etc.). Esta situação impele os profissionais jurídicos a despender mais tempo na busca por uma decisão jurídica adequada ao seu problema. A solução que apresentamos é o emprego de técnicas de IA (Inteligência Artificial) para minimizar o tempo de busca e melhorar a qualidade da informação recuperada. Utilizando a técnica de Raciocínio Baseado em Casos (RBC), elaboramos uma abordagem para recuperar documentos jurídicos semelhantes e adequados à solução de um caso jurídico. Isso é realizado com base no problema inicialmente descrito pelo usuário em linguagem natural. Para possibilitar este processo de conhecimento textual informal, um vocabulário controlado e um *dicionário* jurídico - baseado na terminologia jurídica usual e na terminologia normativa - são integrados na recuperação e no processo de extração do conhecimento. Além disso, esta definição possibilita a inclusão de novos documentos automaticamente na base de conhecimento. Este trabalho descreve um modelo de recuperação para amplas bases de textos jurídicos desenvolvido na área do direito criminal.

#### 1 Introdução

No sistema legal brasileiro, a jurisprudência é uma fonte formal do Direito Positivo [Reale, (1994)]. Quando um profissional do direito realiza uma pesquisa jurisprudencial, ele está buscando informações para reforçar o seu ponto de vista sobre a interpretação de uma norma jurídica e define argumentos persuasivos para fazer a analogia entre o seu problema atual e o anterior, já solucionado.

Muitos destes textos jurídicos estão disponíveis em bancos de dados, inclusive acessíveis na Internet (p. ex., Tribunais Federais, Tribunais Estaduais, ou Órgãos Públicos). No entanto, as buscas por informações jurídicas nesses sistemas requerem conhecimento jurídico e estão limitadas devido a problemas como a ambigüidade sintática e semântica, e também a incerteza existentes nos textos dos documentos [Lenz et al. (1998)]. Na prática, estes bancos de dados recuperam freqüentemente um grande número de informações irrelevantes, além de exigirem uma repetida reformulação da busca para alcançar um resultado satisfatório.

Nós utilizamos o Raciocínio Baseado em Casos (RBC) [Amondts-Plaza (1994)][Kolodner, (1993)], uma poderosa

tecnologia da Inteligência Artificial para o desenvolvimento da presente abordagem [Rissland—Daniels, (1995)]. O RBC usa experiências anteriores e semelhantes para a solução de um problema novo. O RBC é baseado no princípio de analogia, assumindo que problemas semelhantes tem soluções semelhantes. Por esta razão, o RBC é uma técnica muito adequada ao domínio jurídico, pois utiliza o mesmo tipo de raciocínio utilizado pelo juristas na solução de um problema [Bueno (1999)]. Para aplicar a técnica de RBC na recuperação de textos de jurisprudência, o domínio de aplicação (v.g., Acórdãos Criminais) precisa ser modelado de forma que características relevantes dos documentos sejam identificadas, assim como, a terminologia usada no domínio definida. Aqui, a teoria de argumentação jurídica [Warat, (1994)] ofereceu apoio teórico pela análise do significado do caso legal. Isso possibilita o processamento do conhecimento informal nas pesquisas por casos apropriados na base de casos em linguagem natural. A construção de um vocabulário controlado e de um dicionário jurídico permite o processamento dessa informação.

Com base nesta estrutura teórica, nós definimos a representação, o processo de recuperação e a extração



automática dos atributos dos novos documentos a serem incluídos na base de casos.

Nossa abordagem é apresentada na seção 2 e descreve a representação de casos e a sua extração de documentos textuais e a seção 3 descreve o processo de recuperação. A conclusão é apresentada na seção 4.

## 2 Apoio Inteligente Para Recuperação em Amplas Bases de Textos Jurídicos

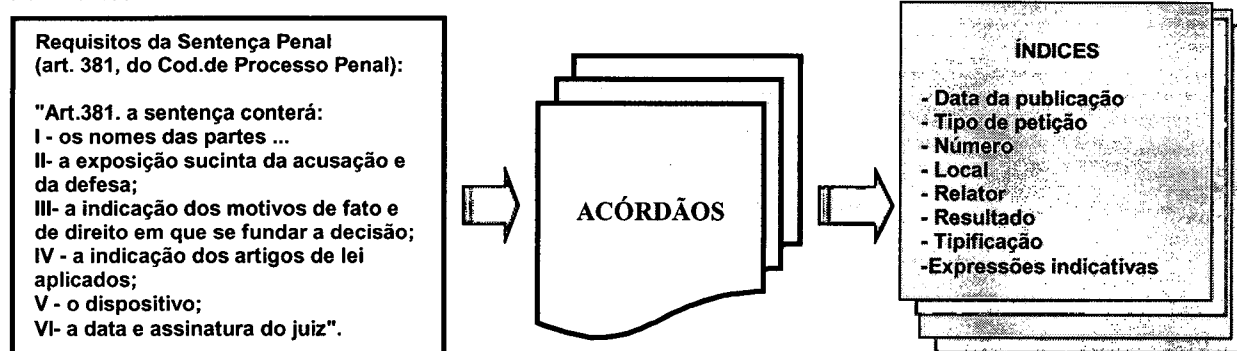


Figura 2. Representação de um Documento Jurisprudencial em Caso

Desta maneira, os casos jurídicos (jurisprudências) têm que ser representados e armazenados de um modo acessível, os casos pertinentes têm que ser recuperados e novos documentos jurídicos incluídos automaticamente na base de conhecimento. Nas seções seguintes são descritas as partes principais de nossa abordagem baseada em casos, que são:

- Definição da estrutura de representação de casos legais e identificação de informação relevante para a recuperação de jurisprudência apropriada.
- Extração automática da informação do texto legal na inclusão de novos casos jurídicos na base de casos.
- Processo de recuperação baseado na similaridade através do *matching* parcial entre o ajuste da situação fornecido inicialmente em linguagem natural, e os casos na base de casos, usando uma medida de similaridade.

### 2.1 Representação dos casos jurídicos

O conhecimento jurídico inserido numa jurisprudência é representado em forma de um caso que consiste no texto do documento original e um conjunto de índices em forma de pares atributo-valor (veja figura 2). Estes índices foram determinados em conformidade com os requisitos formais da sentença penal - exigência normativa do artigo 381, do Código de Processo Penal [Jesus, (1989)] e utilizados conforme a sua relevância para a recuperação [Bueno (1999)].

Pela análise dos textos jurídicos foi identificado um formato padrão. A jurisprudência é estruturada de acordo

No domínio jurídico, a meta da recuperação em textos é fornecer informações que sejam bons argumentos para justificar a aplicação da lei em um caso concreto. A informação recuperada tem que auxiliar o profissional do direito, entendendo e interpretando se a solução legal de um caso é aplicável na solução de um caso presente.

com o tipo de informação fornecida [Weber et al. (1997)], em 4 subestruturas básicas (veja tabela 1). Para reforçar esta forma de representação, o conhecimento do domínio é incluído em forma de um *vocabulário jurídico controlado* e um *dicionário de termos*. É este conhecimento de domínio que permite a recuperação dos documentos e o processo de extração automático, através da identificação de expressões indicativas e relevantes dos textos jurídicos em linguagem natural, juntamente com a modelagem explícita da semelhança destes termos jurídicos.

#### 2.1.1 Índices para a recuperação

Para representar os textos jurídicos de uma maneira acessível, o texto jurisprudencial é mapeado em uma representação estruturada de acordo com a utilizada e o contexto das informações inseridas nele. Assim, um conjunto de índices é determinado para facilitar o processo de recuperação (veja a tabela 1).

No contexto jurisprudencial, o principal problema é a determinação dos índices e os valores usados para a recuperação. O princípio da analogia sustenta a maneira de reutilização dos casos jurídicos; por esta razão, os atributos dos documentos textuais, usados como índices para a recuperação, têm que indicar a utilidade das informações do caso na situação presente. E assim, os índices são determinados do ponto de vista do especialista do domínio em conformidade com a importância deles para o contexto de recuperação da informação jurídica. O resultado desta forma de representação são índices que

representam o conteúdo jurídico dos textos jurisprudenciais.

No contexto da pesquisa jurisprudencial, o problema principal é a determinação dos índices e os valores deles usados para a recuperação. Como o uso de casos jurídicos está baseado no princípio de analogia, os atributos dos documentos de texto usados como índices para a recuperação têm que indicar a utilidade da solução do caso na situação presente. Na nossa abordagem, os índices

são determinados pelos requisitos formais [Jesus, (1989)] e retóricos da sentença penal [Warat, (1995)] em conformidade com a importância deles para o contexto de recuperação da informação jurídica. Nós usamos esta forma de divisão como base para definir o processo de extração do conhecimento através da localização do atributo (veja item 2.2).

Índices		Descrição	Importância para a recuperação
Número do Acórdão		Número fornecido pelo Tribunal.	É usado para identificar o recurso
Data da Publicação		Refere-se a data de publicação no Diário da Justiça.	Com essa informação, o usuário poderá determinar qual o período da pesquisa (na data de publicação há campo para a data da inicial e data final).
Local		Informa a cidade origem do processo.	Pode ser usado para definir a pesquisa por localidade.
Relator		Determina qual foi o juiz responsável pela elaboração da decisão.	Pode ser usado para recuperar os acórdãos proferidos por um determinado relator.
Tipo de recurso		Determina o tipo de recurso usado para rever uma decisão judicial	Serve para delimitar a pesquisa quanto ao tipo de procedimento legal usado para resolver uma questão jurídica.
Expressões jurídicas		Determina as informações jurídicas relevantes e discriminante num caso.	Indicam a matéria jurídica de interesse. É o principal atributo de indexação e recuperação dos casos
Resultado		Daqui se extrai o resultado do pedido feito por uma das partes do processo judicial.	Indica se os Juizes negaram ou aceitaram o pedido feito no recurso.
Tipificação	Tipo geral	define o enquadramento legal genérico, que é o tipo de crime que deu origem ao processo judicial.	indica o tipo de crime cometido: homicídio, furto, estelionato ...
	Modalidade do crime	os crimes podem ser cometido por uma pessoa quando o mesmo tem a intenção clara e objetiva de cometê-lo, ou quando, por sua culpa, o resultado do crime foi atingido.	indica se o crime foi cometido na modalidade Culposa, ou dolosa; O dolo é punível de uma forma mais severa que a culpa, no direito penal.
	Qualificação do crime	Determina se um crime foi praticado em condições especiais( ex.: furto praticado à noite), ou se o agente possui uma característica especial que aumente ou diminua a pena (ex.: reincidência).	Indica se o crime é simples, qualificado ou privilegiado. A qualificação também é usada para efeito de penalização de um crime.
	Tentativa	Uma pessoa pode ser punida por um crime, mesmo quando o crime não atingir o resultado final. por exemplo, errar o disparo de uma arma de fogo, demonstrando a clara intenção de matar a vítima.	determina se a ação criminoso atingiu o seu resultado ou se o criminoso não conseguiu atingir o seu objetivo, por motivos alheios a sua vontade, a pena referente ao crime será diminuída.
	Co-autoria	O mesmo crime pode ser cometido por duas ou mais pessoas. Estas pessoas (agentes) serão punidos conforme a sua participação no crime.	determina se o crime foi cometido por duas ou mais pessoas, é o chamado concurso de agentes

Tabela 1. Descrição dos índices e a sua importância para a recuperação

### 2.1.2 Conhecimento do Domínio Jurídico

No mundo jurídico, o profissional do direito, deve buscar argumentos persuasivos na solução de um caso concreto. Por esta razão, a jurisprudência é um lugar excelente para estudar o argumento jurídico [Bench-Capon (1995)]. Assim, para uma efetiva representação e recuperação destes casos jurídicos, é necessário considerar que um termo normativo pode ter várias interpretações nas decisões judiciais e todas indicam argumentos relevantes para a solução de problemas jurídicos. Do ponto de vista de uma estrita legalidade, um caso jurídico pode ser naturalmente incluído em uma classe normativa; no

entanto, aos olhos do julgador, essa inclusão pode parecer totalmente injusta. Normalmente, quando isso acontece, o juiz produz uma definição explicativa de caráter ideológico, apresentando como relevantes outras conotações para caracterização de algum termo-chave incluído na norma, e decide com isto uma mudança de linha de soluções. É por esta razão que somente uma lista com os termos-chave normativos não é suficiente para representar as informações juridicamente relevantes de um caso.

A Teoria da Argumentação Jurídica [Warat, (1994)] fornece a possibilidade de uma análise do significado da norma jurídica e seu significado individual de justiça.

Segundo essa teoria, nenhum discurso jurídico será considerado significativamente legítimo se contradizer as formas axiológicas predominantes. Os argumentos são persuasivos no direito, quando não contradizem à ideologia de seus destinatários, isto é, a ideologia dos Tribunais.

Assim, um vocabulário jurídico controlado pode ser determinado com termos usuais dos Tribunais (oriundos da aplicação da lei ao caso concreto) e dos próprios termos-chave normativos (veja figura 3).

Estrutura de um caso jurídico	Descrição	Índices
Estrutura 1	Aqui localizam-se as informações que são utilizadas como referências do acórdão. São elas que tornam a citação de uma jurisprudência válida Quando referenciada num processo judicial, ou numa doutrina.	- Número do Acórdão - Data da publicação - Localidade - Relator - Tipo de recurso
Estrutura 2	É a ementa do acórdão, onde esta o resumo dos principais argumentos que fundamentaram a decisão do recurso.	- Expressões indicativas
Estrutura 3	Informa o nome das partes do processo, a decisão do recurso e, ainda, se o resultado foi unânime ou não	- Resultado
Estrutura 4	É a principal estrutura de um acórdão, onde estão descritos os acontecimentos de fato e de direito relatados no processo, além da fundamentação da decisão.	- Tipificação ( <i>tipo geral; modalidade do crime; qualificação do crime; tentativa; co-autoria</i> ). - Expressões indicativas

Tabela 2. Estrutura de um caso jurídico

### 2.1.2.1 Vocabulário controlado

O vocabulário controlado define o valor dos índices usados para a recuperação de um acórdão, enfocando no domínio de aplicação específico (no caso, jurisprudência criminal). Este vocabulário controlado é constituído de termos jurídicos que são usados normalmente nos Tribunais que representam fatos enquadrados normativamente que, por sua vez, são definidos através dos termos-chave de uma norma. Nós elaboramos uma listagem com aproximadamente 3.200 itens para aplicações na área do direito criminal. Cada um desses itens constitui o que denominamos de *expressões indicativas*, atributo composto no mínimo por duas palavras e no máximo por três palavras ( por exemplo, nulidade de citação, ausência de autoria, etc.).

### 2.1.2.2 Dicionário de Termos Normativos

Toda vez que os critérios de relevância de um termo são alterados no seu uso contextual, ocorre uma redefinição. Para redefinir os juristas interferem diretamente sobre as significações dos termos, propondo novas regras designativas.

Tal mecanismo, salienta Warat [Warat, (1995)], facilita a adequação da jurisprudência às exigências reais ou supostas, de um dado momento histórico.

A construção e integração de um dicionário de termos normativos ao sistema tem por objetivo a adequação destas redefinições às expressões indicativas, ou seja, considerar casos com valores de índices semelhantes ao da situação presente durante o processo de recuperação.

O dicionário de termos normativos define a semelhança lingüística, semântica e axiológica de condições que permitem a determinação da semelhança local entre valores de um índice. O dicionário foi desenvolvido por profissionais do direito com base na experiência deles/delas e fundamentado pelo conhecimento de domínio geral no direito criminal.

### 2.2. Extração automática da informação textual

A recuperação de jurisprudência realizada em amplas bases textuais, requer uma automatização da extração das informações relevantes. Algumas informações são facilmente localizadas no texto, embora a identificação de outras informações necessita maior apoio teórico e conhecimento do domínio. Então, nós desenvolvemos técnicas diferenciadas para a extração de conhecimento de textos de jurisprudências:

Através da Localização: Os documentos jurisprudenciais são textos semi-estruturados. Pela análise de textos legais foi identificado um formato padrão (veja tabela 1). As jurisprudências são estruturadas de acordo com o tipo de informação fornecida em quatro subestruturas básicas que facilita a localização de informações relevantes. Além disso, parte das informações relevantes de um texto jurisprudencial são descritas explicitamente no texto numa posição pré-definida. Por exemplo, a data da publicação, sempre é descrita na primeira linha do acórdão precedida da palavra-chave "DATA". Assim, os valores são identificados pela localização deles no texto. Todas as informações juridicamente relevantes representadas nas estruturas 1 e 3 (veja tabela 2) podem ser extraídas por sua localização no texto. Isso inclui a identificação da subestrutura específica e das palavras-

chave pertinentes. Para cada um destes índices um algoritmo de extração específico foi desenvolvido com base neste conhecimento de domínio.

Através de Inferência: A determinação dos valores para o índice *Tipificação* requer a inferência de valores com base no Código Penal [Delmanto, (1991)], baseado nas informações do texto jurisprudencial. O índice *Tipificação* é composto de um subconjunto de índices: *tipo geral; modalidade do crime; qualificação do crime; tentativa; co-autoria*. Por exemplo, se a informação extraída do

texto for “**art. 121, § 2., IV c/c art. 14, ambos do Código Repressivo**”, os valores são automaticamente inferenciados pelo o sistema – pelo uso de conhecimento de domínio em forma de regras.

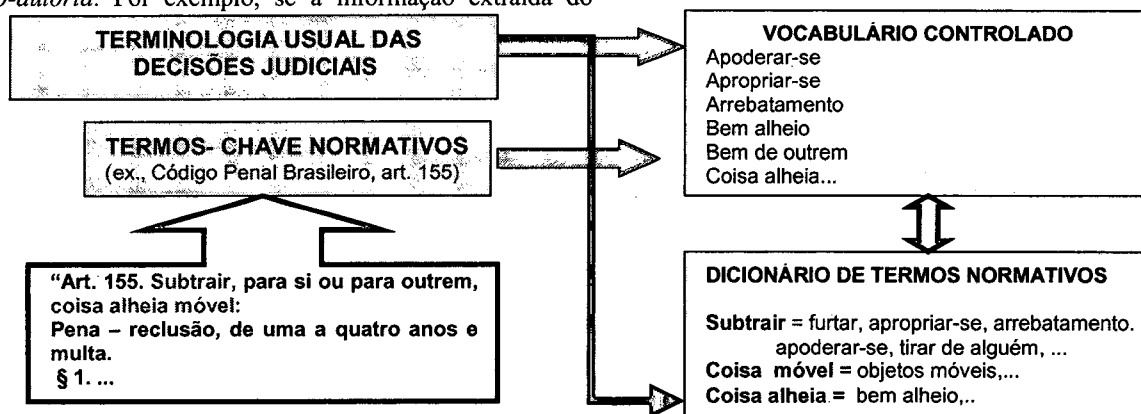


Figura 3. Construção do Vocabulário Controlado e do Dicionário de Termos normativos.

Através do uso de um Vocabulário Controlado: As informações relevantes num texto jurídico, que não podem ser identificadas unicamente pela sua localização no texto são extraídas pelo uso de um vocabulário controlado e um dicionário de termos jurídicos. Partes relevantes do Acórdão (veja tabela 2, estruturas 2 e 4) são avaliadas para localizar expressões do vocabulário controlado. Sinônimos das *expressões indicativas* são identificados pela interação com o dicionário de termos jurídicos (veja item 2.1.2.2.). A informação identificada e o índice *expressões indicativas* é instanciado com valores múltiplos. Por exemplo, no Acórdão exposto na Figura 4, três valores índices podem ser identificados: “*pedido de desistência do recurso*”, “*homologação do pedido*”, “*presentes os pressupostos*” (veja tabela 3).

OCASO 20824		
Índice	Valores	
Data da Publicação	17/05/90	
Localização	Ubirici	
Tipo de recurso	Apelação criminal	
Número do acórdão	20824	
Relator	ernani ribeiro	
Resultado	Concedido	
Tipificação	tipo geral	Homicídio
	modalidade do crime	Doloso
	Qualificação do crime	homicídio Qualificado
	Tentativa	Sim
	co-autoria	Não
Expressões indicativas	Pedido de desistência do recurso	
	Homologação do pedido de desistência	
	Presentes os pressupostos	

Tabela 3. Exemplo de um caso representado.

### 3 Recuperação de Casos Jurídicos

O objetivo do sistema é recuperar os casos mais úteis para a solução de um problema jurídico atual.

Este problema é descrito pelo usuário e o ajuste da situação é iniciado pela comparação dele com os casos na base de conhecimento.

O grau de similaridade entre os casos e o problema fornecido no ajuste da situação é determinado por uma medida de similaridade que nomeia um valor de semelhança a cada caso. Após cada caso da base ser valorado, eles serão ordenados de forma decrescente de acordo com os valores obtidos com a métrica de similaridade, e os casos mais úteis, ou seja, com valores de similaridade mais altos, serão apresentados como o resultado da consulta. O usuário poderá ainda, selecionar na íntegra o caso mais apropriado para solução do seu problema jurídico. O processo de recuperação é dividido em:

- Ajuste da Situação;
- Cálculo da Similaridade;
- Seleção do melhor caso

### 3.1 Ajuste da Situação

Um problema jurídico é descrito em linguagem natural para iniciar o processo de recuperação pelo usuário. O ajuste da situação é realizado para identificar os valores atuais de cada índice na nova situação legal. Baseado na descrição dada pelo usuário, as informações relevantes (*expressões indicativas*) são identificadas de uma forma automática através do uso do vocabulário controlado. Além disso, no ajuste da situação existem filtros que o usuário pode, como opção, usar na pesquisa (data da publicação, tipo de recurso, relator e local).

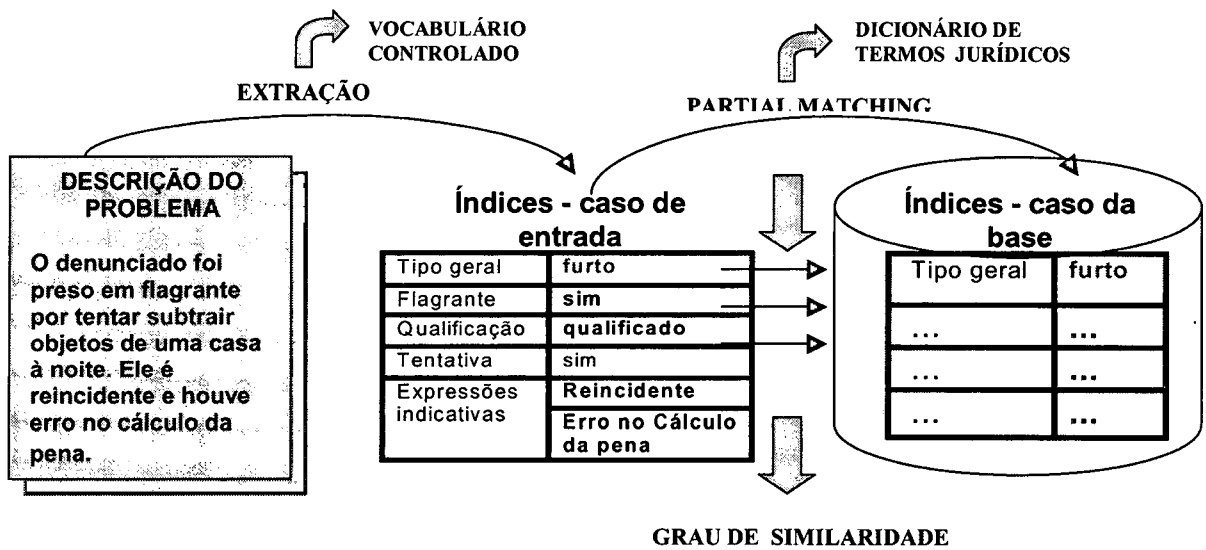


Figura 4. Processo de Recuperação.

### 3.2 Cálculo da Similaridade

Para identificar os casos mais úteis em relação ao problema inicial fornecido, o ajuste da situação é *comparado* parcialmente com cada caso na base de casos. A similaridade de cada índice do problema inicial com cada caso na base de casos é determinado pela *métrica da similaridade local* [Bueno et al. (1999)]. O grau total de similaridade de um caso é calculado pela soma dos valores de similaridade local multiplicada por um fator de pesagem relativo a importância de cada índice para a pesquisa (veja figura 5). O valor da similaridade local de atributos com valores múltiplos é calculado baseado no número total de valores normalizados.

#### 3.2.1 Medida da Similaridade Global

Com base na descrição da situação atual, casos relevantes são recuperados da base de conhecimento. De maneira a identificar os casos mais úteis em comparação com a situação atual, o ajuste da situação é parcialmente comparado com cada caso na base de conhecimento. Assim, o valor da similaridade global de um caso é calculado com base no *vizinho mais próximo* (veja figura 5)

$$Sim(S, C_i) = \sum_{j=1}^8 (f_j(I_{sj}, I_{cij}) W_j)$$

Onde,

S = Ajuste da Situação

$C_i$  = caso  $i$  da base de conhecimento  
 $f_j$  = índice  $j$   
 $I_{sj}$  = índice  $j$ th do ajuste da situação  
 $I_{cij}$  = índice  $j$ th do caso  $i$   
 $W_j$  = peso do índice  $j$

$$0 \quad I_s \neq I_{ci}$$

O índice expressões indicativas podem ser representados através de valores múltiplos ( no máximo 8 valores no protótipo atual). Para este índice, a similaridade local é calculada comparando cada valor fornecido pelo ajuste da situação com cada valor de cada caso, determinando o valor máximo de similaridade para cada valor fornecido. Então, a soma dos valores máximo da similaridade é calculado e normalizado pelo número de valores fornecidos no ajuste da situação. O valor da similaridade local do índice tipificação é calculado de acordo com os seus 5 valores (veja tabela 3).

**3.2.2 Medida da Similaridade Local**

A similaridade entre cada índice da situação atual com cada caso na base de casos é determinado pelas medidas locais de similaridade.

O valor da similaridade local para os índices: local, tipo de recurso, relator, resultado e tipificação é calculado através da seguinte fórmula:

$$f(I_s, I_{ci}) = 1 \quad I_s = I_{ci}$$

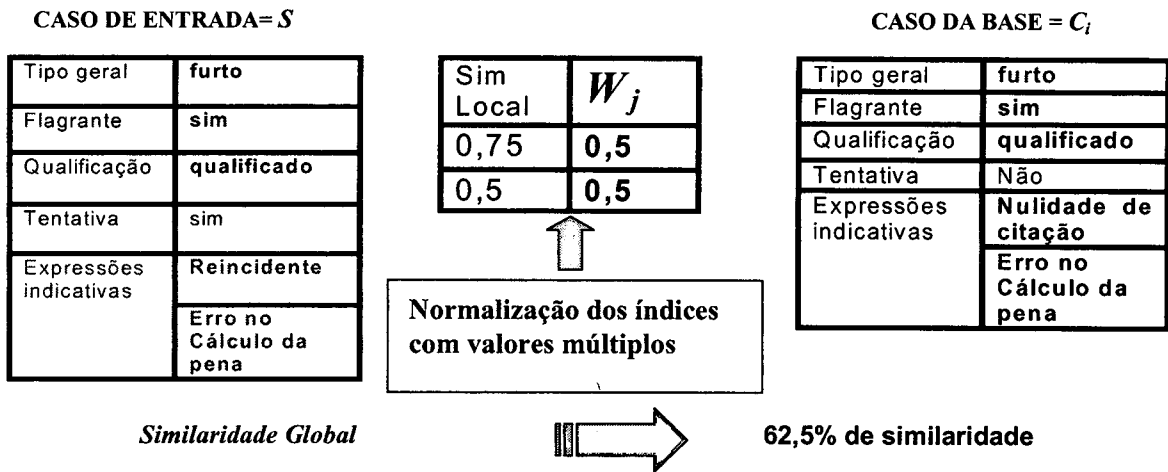


Figura 5. Cálculo da Similaridade.

O cálculo do valor local da similaridade é aprimorado pela integração de um dicionário jurídico (veja seção 2.1.2.2). Isto permite a consideração de valores similares mas não necessariamente iguais, por exemplo, se na descrição do problema é afirmado que o *tipo geral* do crime é assassinato e o *tipo geral* do caso da base é homicídio ( um sinônimo para assassinato), então pelo uso de um dicionário jurídico estes dois termos são considerados similares. Sem a integração de um dicionário jurídico a similaridade entre os dois casos não seria identificada e, possivelmente casos relevantes não seriam observados para serem recuperados.

O grau total da similaridade de um caso é calculado através da soma de todos os valores locais da similaridade e multiplicado por um fator de peso, como demonstrado acima.

**3.3 Seleção dos melhores casos**

Determinada a métrica de similaridade, os dez melhores casos são ordenados de acordo com o grau de similaridade. Os casos ordenados são apresentados ao usuário de modo que ele possa visualizar um resumo (Ementa) de todos os casos, com algumas informações sobre o documento jurídico ou, se o usuário quiser, o documento completo. Assim, o usuário poderá escolher entre os melhores casos, aquele documento que é o mais adequado.

**4 Conclusão**

A enorme quantidade de decisões produzidas pelos Tribunais, torna o domínio jurídico um campo fértil para o desenvolvimento de sistemas computacionais para a recuperação de dados. Neste contexto, o RBC oferece um potencial significativo para a recuperação inteligente de documentos jurisprudenciais. Seus principais benefícios são o enfoque no conhecimento em forma de episódios

individuais, em lugar de conhecimento de domínio genérico e a recuperação baseada na similaridade.

Recuperar textos é o objetivo de muitos trabalhos de pesquisa. No entanto, na área jurídica, poucos são os sistemas baseados em casos visando a recuperação deste tipo de informação. O Projeto Prudentia [Weber-Lee et al. (1997)][Weber et al. (1997)] é um exemplo de sistema para a recuperação da informação aplicado ao domínio jurídico. Existem também, sistemas de RBC aplicados para a recuperação da informação em diferentes áreas, como o Projeto FallQ [Lenz—Burkhard, (1997)].

Nós utilizamos estes trabalhos como parâmetros para o desenvolvimento da presente abordagem.

O próximo passo do presente trabalho é a validação da aplicação desta abordagem. Embora este trabalho seja um aprimoramento teórico do Projeto Prudentia, somente com a avaliação dos resultados da aplicação poderemos o grau de efetividade do sistema em relação ao Prudentia. A recuperação de jurisprudências feita com qualidade pode fornecer meios eficazes para que o julgador possa conciliar retoricamente os ditames de sua consciência com a expectativa das partes e dos órgãos judiciais superiores. Além disso, os profissionais do direito, por sua vez, poderão contar com mais um instrumento de orientação no sentido de selecionar os argumentos que colocarão nas mãos dos juízes com vista a suas pretensões.

## Agradecimentos

Os autores gostariam de agradecer as seguintes pessoas que contribuíram no desenvolvimento da presente pesquisa: Andre Bortolon, Cristina Santos, Luis Gustavo D'Agostini Bueno e Ricardo M. Barcia Filho.

## Referências bibliográficas

- A. Amond, E. Plaza, "Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches". *AI Communications*, 17(1), 1994.
- T. Bench-Capon, "Argument in Artificial Intelligence and Law". *Jurix*. 1995.
- T. Bueno. "Uso da Teoria Jurídica para Recuperação de Acórdãos Criminais em Sistema Baseado em Casos". Dissertação. Programa de Pós-Graduação em Engenharia da Produção da Universidade Federal de Santa Catarina, 1999.
- T. Bueno, "Retrieval in Jurisprudencial Text Bases Using Juridical Terminology". *International Conference in Intelligence Artificial and Law – ICAIL*, 1999.
- C. Delmanto, "Código Penal Comentado", 3.ed. ampliada e atualizada - Rio de Janeiro, Renovar, 1991.
- D. Jesus, "Código de Processo Penal anotado". 8. ed. atual. e aum. – São Paulo, Saraiva, 1989.
- J. Kolodner, "Case-Based Reasoning", Morgan Kaufmann, Los Altos, CA. 1993.
- M. Lenz, H.-D Burkhard, "CBR for Document Retrieval: The FallQ Project", *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*. D. Leake, E. Plaza (ed.), Springer Verlag, 1997.
- M. Lenz.; A. Hübner; M. Kunze, "Textual CBR". M. Lenz, B. Boutsh-Sporl, H.-D. Burkhard, S. Wess (eds.), *Case- Based Reasoning Technology*, Springer Verlag, 1998.
- M. Reale, "Fontes e Modelos de Direito". São Paulo:Saraiva, 1994.
- E. Rissland, J. Daniels, "Using CBR to drive IR". In *Proceedings of IJCAI-95, Canada*, 1995.
- L. A. Warat, "Interpretação da Lei: temas para uma reformulação", Porto Alegre: Sergio Antonio Fabris (ed.), 1994.
- L. A. Warat, "O Direito e sua Linguagem", 2ª ed., Sergio Antonio Fabris Editor, Porto Alegre, 1995
- R. Weber-Lee, R. Barcia, M. Costa, M.; I. Rodrigues Filho, H. Hoeschl, T. C. Bueno, A. Martins, R. Pacheco, "A Large Case-Based Reasoner for Legal Cases", *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*, D. Leake, E. Plaza (ed.), Springer Verlag, 1997.
- R. Weber, R Barcia, R. Pacheco, A. Martins, H. Hoeschl, T. Bueno, M. Costa, I. Rodrigues Filho, "Representing Cases From Texts In Case-Based Reasoning", III Congresso Internacional de Engenharia Industrial e XVII ENEGEP, Brasil, 1997.

## ANEXO B

Artigo aprovado na VII Conferência Internacional de Inteligência Artificial e Direito – ICAIL99. 14-17 de junho de 1999, Oslo, Noruega.

# JurisConsulta: Retrieval in Jurisprudencial Text Bases using Juridical Terminology

Tânia C. D'Agostini Bueno, Christiane Gresse von Wangenheim,  
Eduardo da Silva Mattos, Hugo Cesar Hoeschl, Ricardo M. Barcia  
Graduate Program - Production Engineering  
Federal University of Santa Catarina, Brazil  
{bueno, gresse, emattos, metajur, rbarcia}@eps.ufsc.br

### Abstract

In the legal domain, jurisprudence has an important role as a juridical source; its decisions support the application of the Law to a concrete case. The problem is that Brazilian Courts produce an enormous amount of decisions every year, turning these text sources larger every time and forcing juridical professionals to spend more time in the search for a relevant decision. Sophisticated AI techniques are needed to minimize searches time and improves the quality and appropriateness of the retrieved information. This paper describes a case-based approach for the intelligent retrieval of jurisprudencial texts. The approach enables the retrieval of adequate texts with characteristics similar to information supplied by the user in natural language. New documents are automatically included into the knowledge base by extracting relevant information. In order to enable the processing of informal textual knowledge in natural language, a controlled vocabulary and a juridical thesaurus based on common juridical terminology is integrated into the retrieval and extraction process. The approach is based on sentences to criminal proceedings in the domain of the Brazilian Right.

### 1. Introduction

In the Brazilian legal system, jurisprudence is a formal source of Law [9], and its importance resides in supplying an interpretation of the norm to a concrete case. Jurisprudence helps juridical professional to reinforce their point of view about the interpretation of a juridical norm. In this context, a juridical professional defines persuasive arguments through analogy between the current problem and a past one, already solved.

However, the number of jurisprudencial cases has grown largely, complicating the efficient access to this type of information. Despite the availability of those juridical texts in databases (such as, e.g., Informa Jurídico [6]), or in the Internet (e.g., Federal Tribunals [13] and State Tribunals [14], or public

organs [11]), the difficulty remains. The search in those systems requires juridical knowledge and is limited due to problems with syntactic and semantic ambiguity and existing uncertainties in the textual document [8]. In practice, these database and web-based technologies often retrieve a great number of irrelevant information and require a repeated reformulation of the search to achieve a satisfying result. Therefore, the development of efficient and effective systems for the retrieval of juridical information using techniques in the area of *Artificial Intelligence* (AI) is important for all juridical professionals and yields significant benefits for the solution of legal problems.

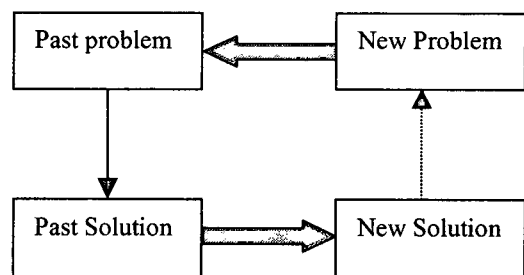


Figure 1. Process of analogy

For the operationalization of intelligent jurisprudencial research, we apply *Case-Based Reasoning* (CBR) [1][5]. CBR is a powerful Artificial Intelligence technology, which uses experiences from past similar problems as a basis for the solution of a new problem. Case-based reasoning is based on the principle of analogy, assuming that similar problems have similar solutions. In the juridical domain, professionals also apply analogy, when using past cases to justify arguments and interpret the law. For example, judges use this type of reasoning when interpreting a new case through its comparison with previous cases as a basis for decision.



In order to facilitate the search for relevant jurisprudence, we develop an intelligent system for the retrieval of legal texts. The approach of the system supports the representation and storage of textual jurisprudential knowledge and retrieves documents with similar characteristics, based on the initial information supplied by the user in natural language. New documents are included in the knowledge base by extracting automatically relevant information for the retrieval process. In order to enable the processing of informal knowledge in natural language, a controlled vocabulary and a juridical thesaurus based on the usual juridical terminology is integrated into the retrieval and extraction process. The representation of the cases and the required legal domain knowledge (e.g., thesaurus) have been developed based on the theory of juridical argumentation [15][16], which offered theoretical support for the analysis of the meaning of the legal cases. Our approach is based on sentences to criminal proceedings in the domain of the Brazilian Right originating from of the Tribunal of Justice of Santa Catarina<sup>[14]</sup>.

This article describes a case-based approach, JurisConsulta, for the retrieval of jurisprudential texts. In Section 2, an introduction to the application domain, jurisprudential research is given. Our approach is presented in Section 3, describing the representation of cases, their extraction from textual documents, and the retrieval process. A comparison of our approach with related research is given in Section 4 and conclusions are presented in Section 5.

## 2. Application Domain: Jurisprudential Research

Jurisprudence establishes a correlation between the Law and the fact and determines the most appropriate way of its application [12].

The theoreticians of the Common Law state, in an unanimous and exact way, that every juridical sentence creates Right. In contrast, the Brazilian Right, which adopts the continental system (dogmatic) does not consider jurisprudence in the same way. In our system, jurisprudence is not the primary source of law as in the common law and does not necessarily have to be considered for a juridical decision. However, jurisprudence provides support for the interpretation and application of the Law to concrete cases. The existence of a past decision similar to the present problem can reinforce the lawyer's argument and support the judge concerning the definition of the juridical decision. In this context, jurisprudence becomes important. For example, suppose that a lawyer's defendant was arrested in *flagrante delicto* and a mistake occurred with the penalizing condition. In this case, the lawyer aims at the reduction of the penalty, and concludes that *Habeas Corpus* is an adequate petition for the solution of the case. However, the Law does not specify that *Habeas Corpus* is applicable to this specific subject, as the Law is described in a generic form.

According to Bench-Capon [2], jurisprudence is an excellent form to study argument. *Juridical argumentation theory* [16] provides the possibility for the analysis of the meaning of juridical norms and their individual meaning of justice. The theory can help a judge to select arguments based on the Law, but considering different aspects (moral, political, social, personal, etc.). As a consequence, the Law is redefined in conformance with the Judge's interpretation, when the Judge

decides a case and determines the application of a specific Law for the particular case. Those different possible interpretations of the Law are the main reason for the existence of divergent decisions. Juridical professionals can use the redefinition of the normative terms<sup>5</sup> as described in the jurisprudence for the construction and reinforcement of arguments, which facilitates the application of the Law to situations not considered before. Therefore, jurisprudence also plays an important role in the Brazilian Right.

However, the search for appropriate past decisions can be quite difficult. For example, while searching in jurisprudential archives, the user may find only few cases applying to the present problem, or using database management systems frequently a large number of irrelevant cases is retrieved. In both cases it is necessary to repeat the search, focusing on other or refined aspects of the problem. This is even complicated, through the fact that the Juridical Institutions produces an enormous amount of decisions every year, which turns these text sources larger every time and forces the judicial professionals to spent more time searching for a relevant decision. Therefore, sophisticated and intelligent techniques are required to minimize search time and improve the quality and appropriateness of the retrieved information.

## 3. Intelligent Support for Jurisprudential Search

In the context of jurisprudential research, we focus on providing an approach that facilitates the retrieval of information to help juridical professionals concerning the elaboration of persuasive arguments in the solution of a new juridical case.

Therefore, legal cases have to be represented and stored in an accessible way and relevant cases have to be retrieved for a specific problem. New legal documents have to be automatically included into the knowledge base in order to be available for future decisions.

In the following sections, the main parts of our case-based approach are described:

- Representation structure of legal cases and identification of relevant information for the retrieval of appropriate jurisprudence.
- Automatic extraction of the information of the legal document for the inclusion of new legal cases into the knowledge base.
- Similarity-based retrieval process through partially matching the given assessment of the present situation and the cases in the knowledge base using a similarity measure.

### 3.1. Representation of Legal Cases

The jurisprudential knowledge is represented in form of cases, which consist of the original document text and a set of indexes in form of attribute-value pairs (see figure 2).

<sup>5</sup> Normative terms represent juridical facts [16].

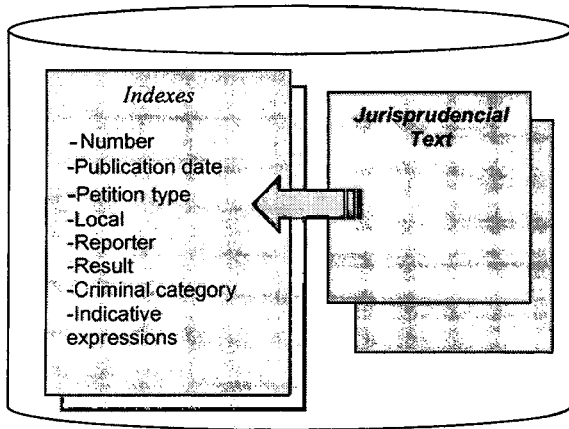


Figure 2. Representation of legal document.

The usage of indexes during the retrieval process allows an efficient access to the text documents by referring to their content. Besides, the jurisprudential text cases, general domain knowledge is included in form of a controlled juridical vocabulary and a juridical thesaurus. Definitions of the content

of norms are represented in form of rules. This domain knowledge supports the retrieval and extraction processes by identifying relevant terms in juridical documents given in natural language and explicitly modeling the similarity of juridical terms.

3.1.1. Retrieval Indexes

In order to represent the legal texts in an accessible way, the textual description of the legal case is mapped into a structured representation. This representation focuses on aspects of the legal documents relevant for the specific retrieval and legal area, e.g., reinforce argumentation in the sentences to criminal proceedings. Therefore, a set of indexes has been determined facilitating the retrieval process.

In the jurisprudential context, a major problem is the determination of those indexes and their values used for the retrieval. As the reuse of juridical cases is based on the principle of analogy, the attributes of the text documents used as indexes have to indicate the usefulness of the solution of the case in the present situation. Thus, the indexes are determined from the viewpoint of the domain expert in accordance with their importance for new situations.

Index		Description	Importance of Retrieval
Publication Date		Refers to the publication date in the Journal of the Justice.	Can be used to determine a period specific of interest (e.g., 11/11/1995 until 01/01/1998).
Petition type		States the resource type used to review a judicial decision originated from first degree Court.	Can be used to limit the search by the specific type of legal procedure used to solve a juridical subject
Document number		States the number supplied by the Court of Justice.	used to identify the petition
Local		States the city of origin of the process	Can be used to define a specific place of interest.
Reporter		States the responsible judge for the elaboration of the decision.	Can be used to retrieve the petitions uttered by a certain reporter.
Indicative expressions		Determines the important juridical information in the petition and indicates the juridical matter of interest.	Used to retrieve jurisprudence dealing with similar topics.
Result		Defines the result of the petition, indicating whether the Judges denied or accepted the request.	Used to determine the result of request.
Criminal Category	Generic type	Defines the generic type of criminal category, that is the crime type that created the lawsuit, e.g., homicide, rape, etc.	Used for discrimination between different types of crimes.
	Crime Modality	Defines the criminal intention (e.g. crime done on purpose or happened accidentally) of the person.	Used for discrimination between different types of crime modalities.
	Crime qualification	Determines the specific circumstances of the crime (e.g., theft practiced during night, or criminal record of the person).	The qualification is also used for effect of punishment of a crime.
	Attempt	Determines when a person did not get to reach her/his criminal objective, for strange reasons, the punishment regarding the crime will be decreased	Used for determine if a person was reached the goal of her/his criminal action.
Co-executor		Determines a number of persons involved in the crime. These persons can be punished according its participation in the crime.	Used to know the number of person involved in the crime.

Table 1. Description of the indexes and their importance for the retrieval.

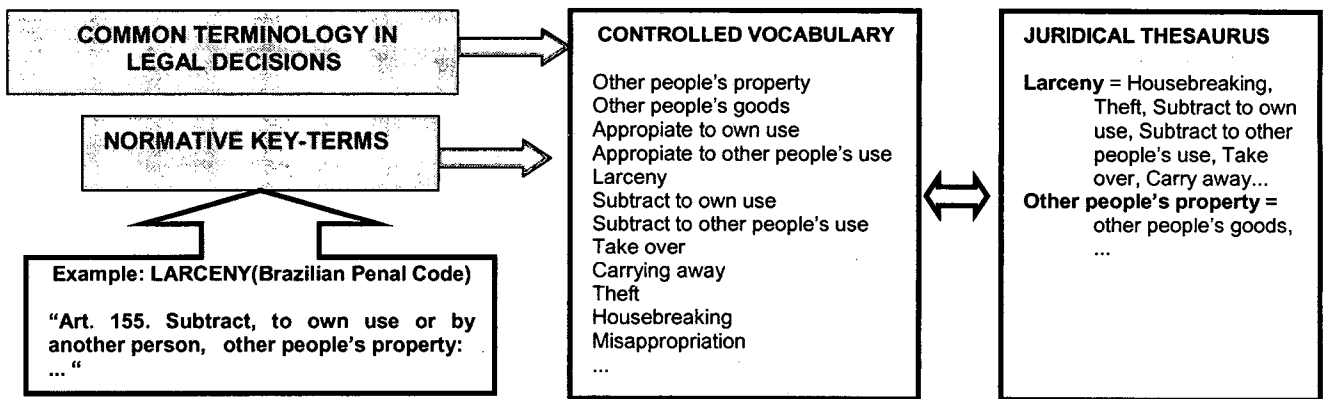


Figure 3. Controlled Vocabulary and Juridical Thesaurus.

As a result, the indexes *indicative expressions* and *criminal category* represent information related to the juridical content of the jurisprudential texts. In addition, information necessary to identify the jurisprudential document (*origin, number, date of publication, resource type* and *reporter*) is represented. This information can be used in order to limit the search, e.g., to a certain period of time.

### 3.1.2. Legal Domain Knowledge

To support the retrieval process and an automatic extraction of the indexes concerning the content of the jurisprudential texts, a controlled vocabulary and a juridical thesaurus is developed.

#### 3.1.2.1 Controlled Vocabulary

The vocabulary defines indicative expressions for a predictive indexation by restricting normative key-terms that represent juridical facts and common juridical terminology focusing on the specific application domain (see Figure 3). Here, an expression is considered as a potential index value, if indicates a juridical situation, based on the similarity wrt. normative terms. These index values also should not be very generic (for example, accident), or, on the other side, very specific (e.g., accident of traffic with fatal victims) [5].

Developing a juridical controlled vocabulary, we have to consider that a normative term can have several interpretations in juridical decisions, all indicating at strong arguments for the solution of juridical problems. Therefore, in the Brazilian Juridical System the construction of a list of expressions, based on the normative key-terms and the experience of the Tribunals, can have more juridical relevance than the extraction of knowledge done exclusively based on the information contained in the jurisprudential text itself. Consequently, a list of normative key-terms and common juridical terminology has been extract by analyzing and comparing codes, norms and jurisprudence [3], [4], [11], [13], [14]. We elaborated a list of approximately 3.200 items in the area of the Criminal Law, which represents juridical facts.

#### 3.1.2.2. Juridical Thesaurus

In order to enable the retrieval of cases with similar index values compared to the ones of the present situation, a juridical thesaurus is integrated into the system. The juridical thesaurus (see Figure 3) defines the linguistic similarity of terms, which allows the determination of the local similarity among values of an index.

The thesaurus is developed through juridical professionals based on their experience and guided through general domain knowledge on Criminal Right. In addition, the juridical thesaurus of the Brazilian Federal Senate was used to complement the definition of indicative expressions [11].

The juridical thesaurus allows to retrieve relevant cases, which are similar, although not equal to the given situation wrt. the individual values of the indexes. For example, if the present situation is characterized through the expression *presumed coercion*, cases in the knowledge base characterized through similar expressions, such as, e.g., *vis compulsiva* are also considered as relevant cases, even if the expression *presumed coercion* is not stated. This avoids the exclusion of a relevant and similar case, if index values are not equal.

### 3.2. Automatic Extraction of Textual Information

New jurisprudential texts are continuously produced through the Juridical Institutions. Thus, our approach enables the automatic inclusion of new documents into the knowledge base through the extraction of relevant information in order to make the new cases available.

Jurisprudential texts are semi-structured documents for which a standard format has been identified through their analysis. Legal texts can be structured wrt. the type of information represented [17]. Four basic substructures have been identified (see Table 2). This subdivision of the texts facilitates the

CASE STRUCTURE	DESCRIPTION	Index
Structure 1	Includes informations that are used as references to the petition.	-Publication Date -Petition Type -Document Number -Local -Reporters
Structure 2	Summarizes the main arguments on which the decision is based. This structure is the abstract of the juridical document.	- Indicative expressions
Structure 3	States the names of persons involved in the process and the result of the request.	- Result
Structure 4	Represents the main structure of a petition, which includes the facts and the arguments of the decision, including the arguments and related laws.	- Criminal category - Indicative expressions

Table 2. Structure of a jurisprudencial document.

extraction of relevant information according to its location in the substructures.

Some relevant information can be identified in the legal texts through their location. Although, the identification of other information demands theoretical support and domain knowledge. Therefore, we develop different techniques for the extraction of knowledge from jurisprudencial texts:

- by location
- by inference
- by usage of a controlled vocabulary

### 3.2.1. Extraction by Localization

Jurisprudencial texts include some relevant information, which is explicitly described in the texts at a predefined position. For example, the date of the publication, is always described the first line of the petition proceeded by the keyword "DATA". All information represented in structure 1 and 3 relevant for indexation (see Table 2) can be extracted through its location in the text. These indexes are:

- Publication date,
- Petition type,
- Number,
- Local,
- Reporter,

Based on expert knowledge, the positions of this information in the jurisprudencial texts have been specified. This includes the identification of the specific substructure and relevant keywords. For each index a specific extraction algorithm has been developed based on this domain knowledge.

### 3.2.2. Extraction by Inference

Information concerning some indexes, e.g., *Criminal Category*, is not explicitly stated in the legal documents, and can, therefore, not directly extracted from the texts. Those values

need to be inferred based on information extracted from the legal text.

The index *Criminal Category* is composed of a set of indexes (see Table 1): *general term; crime modality; crime qualification; attempt; co-responsibility*. In order to instantiated those values, information on the respective law used for the argumentation is extracted from the text document. The relevant information is identified in the text through certain keywords (e.g. *artigo, Art.*) followed by a numerical value. Then, based on the extracted information, the values for the index *Criminal Category* are inferred by rules representing general legal domain knowledge on the content of norms. Those rules map the legal categorization to a unique form of representation accordingly to the Brazilian Criminal Code (see Figure 4, structure 4). For example, if the information extracted is *art. 121, § 2., IV c/c art. 14, ambos do Código Repressivo*, the following values are inferred:

- general type: homicide
- crime modality: intentional
- crime qualification: qualified
- attempt: yes
- co-responsibility: no
- crime modality: homicide it will be framed as homicide, the same is made with the sub-indexes modality, qualification, attempt and co-responsibility.

Another index, *result*, is inferred, accordingly, by using general domain knowledge.

### 3.2.3. Extraction by Controlled Vocabulary

Information on the juridical argumentation in the petition is extracted by the usage of a controlled vocabulary. Relevant parts of the petition (see table 2, structure 2 and 4) are searched for expressions of the controlled vocabulary. Synonyms of relevant

expressions are identified through the interaction of the juridical thesaurus. The information identified is extracted and the index *indicative expressions* is instantiated with the identified values. For example, from the petition shown in Figure 4, three indicative expressions can be extracted: “*Desistence request*”, “*Ratification of request*” and “*Lawsuit conditions*” (see Table 3). As a consequence, the index *indicative expressions* can be instantiated with multiple values.

<b>Structure 1.</b>
DJ: 8.013 DATA: 17/05/90 PÁG: 08 Apelação criminal n. 20.824, de Urubici. Relator: Des. Emami Ribeiro.
<b>Structure 2.</b>
APELAÇÃO CRIMINAL. PEDIDO DE DESISTÊNCIA. HOMOLOGAÇÃO.
<b>Structure 3.</b>
Vistos, relatados e discutidos estes autos de apelação criminal n. 20.824, da comarca de Urubici, em que é apelante Antônio Domingos da Luz, sendo apelada a Justiça, por seu Promotor: ACORDAM, em Primeira Câmara Criminal, à unanimidade, homologar a desistência requerida. Custas de lei.
<b>Structure 4.</b>
Trata-se de pedido de desistência do recurso interposto por defensor em favor de ANTÔNIO DOMINGOS DA LUZ que na comarca de Urubici foi condenado à pena de 12 (doze) anos e 6 (seis) meses de reclusão, por infração ao art. 121, § 2º, IV c/c art. 14, ambos do Código Repressivo. Presentes os pressupostos que autorizam o acolhimento da pretensão, homologa-se o pedido de desistência. Presidiu o julgamento, os Exmos Srs. Des. Emami Ribeiro e participaram do mesmo, com votos vencedores, os Exmos Srs. Des. Marcio Batista e Wladimir d'Ivanenko. Florianópolis, 05 de abril de 1990. Nauro Collaço Presidente p/o acórdão Emami Ribeiro Relator Nelson Ferraz Procurador de Justiça

Figure 4. Localization of index values in the document.

CASO 20824		
Index	Values	
Document number	20824	
1. Publication Date	17/05/90	
2. Petition type	Criminal appeal	
3. Local	Ubirici	
4. Reporter	Ernani Ribeiro	
5. Result	accepted	
6. Criminal Category	General type	Homicide
	Crime modality	Intentional homicide
	Crime qualification	Qualified homicide
	Attempt	Yes
	Co-executor	No
7. Indicative Expressions		Desistence request
		Ratification of request
		Lawsuit conditions

Table 3. Example of a case representation

### 3.3. Retrieval of juridical cases

The objective of the system is to retrieve the most useful cases for the solution of a current juridical problem. The present problem is described by the user and the given situation assessment is compared with the cases in the knowledge base. As part of the comparison, the degree of similarity between each case of the knowledge base and the given situation assessment is determined through a similarity measure which assigns a similarity value to each case. The cases of the knowledge base are ranked by their similarity value and the most similar ones are presented to the user, who can explore them in detail and select the most appropriate for the solution of the present juridical problem. Thus, the retrieval process is refined into:

1. Situation assessment
2. Matching
3. Selection of the most useful case

#### 3.3.1. Situation Assessment

The present juridical problem is described as input for the retrieval process. The problem is described in natural language, p. ex. "The defendant was arrested for committed a homicide. After his judgment, the defendant was submitted for an exam of mental health, and it was verified that he was insane at the moment of the criminal action; the request is for annulment of the process". The user can further limit the search by period, place, reporter or/and resource type, e.g., retrieve only cases after 1995.

Based on this situation assessment provided by the user, relevant information is automatically extracted by using the controlled vocabulary and the juridical thesaurus. As a result, the present situation is described by attribute-value pairs representing the indexes used for the retrieval (see Figure 5).

In addition, the situation assessment is supported through the provision of terms related to the given description, if necessary. Those terms help the user to refine the present situation. Continuing the example above, the system could propose, e.g., the term exclusion of penalty. The proposed terms are identified based on an initial description of the problem stated by the user. Based on this description, relevant cases are identified in the knowledge base and their indexes are provided as possible terms for the refinement of the situation assessment.

#### 3.3.2. Matching

Based on the given situation assessment, relevant cases are retrieved from the knowledge base.

In order to identify the most useful cases in the present situation, the situation assessment is partially matched with each case in the knowledge base. Therefore, the global similarity value of a case is calculated based on nearest neighbor (see Figure 6):

$$Sim(S, C_i) = \sum_{j=1}^7 (f_j(I_{sj}, I_{cij}) * W_j)$$

Where,

$S$  = situation assessment  
 $C_i$  = case  $i$  of knowledge base  
 $I_{sj}$  =  $j_{th}$  index of situation assessment  
 $I_{cij}$  =  $j_{th}$  index of case  $i$

$W_j$  = weight of index  $j$   
 $f_j$  = local similarity measure for index  $j$

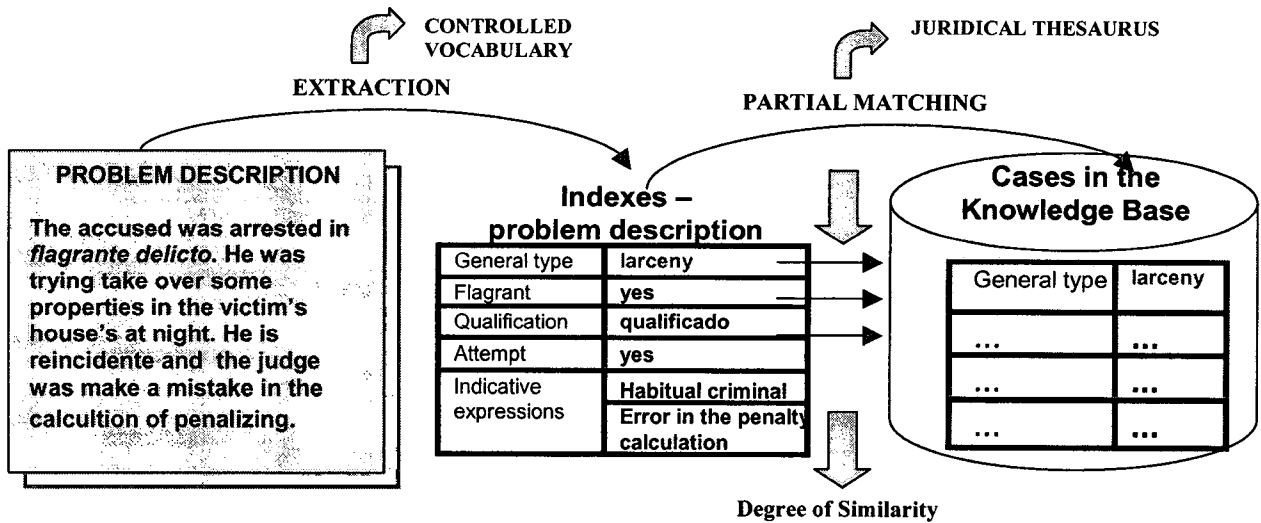


Figure 5. Retrieval Process.

The similarity of each index of the current situation to each case in the case base is determined by local similarity measures.

The local similarity value for the indexes: location, resource type, reporter, result, crime category is calculated through

$$f(I_s, I_{ci}) = \begin{cases} 1 & I_s = I_{ci} \\ 0 & I_s \neq I_{ci} \end{cases}$$

The index *indicative expression* can be represented through multiple values (maximal 8 values in the current prototype). For this index, the local similarity is calculated by comparing of each value supplied in the situation assessment with each value of the case, determining the maximum similarity value for each given value. Then, the sum of the maximum similarity values is calculated and normalized by the number of values supplied in the situation assessment. The local similarity value of the index crime category is calculated accordingly, including its five values (see Table 3).

The calculation of the local similarity value is further enhanced through the integration of the juridical thesaurus (see Section 3.1.2.2.). This allows the consideration of similar, but not necessarily equal values. For example, if in the given situation assessment the *general crime type* is described through "murder" and the *general crime type* stated in a case is "homicide" (a synonym for murder), then through the usage of the thesaurus these two terms are considered as similar. Without the integration of a thesaurus, this similarity between the two cases would not be identified and, possibly relevant case could be overseen and not retrieved.

The total degree of similarity of a case is calculated through the sum of all local similarity values multiplied by a weighting factor, as shown above.

### 3.3.3. Selection of the Most Useful Case

Based on the calculated similarity values, a total order between the cases in the knowledge base is introduced. The system presents the ten most similar cases to the user. In order to enable the evaluation of the retrieved cases through the user concerning his present situation, the system visualizes the *Abstract* of each case, which summarizes the jurisprudence. The user can further explore the cases in detail, by requesting the complete legal document. Based on the evaluation of the retrieved cases, the user can then select the most appropriate ones for the needs of his particular problem.

## 4. Discussion and Related Work

Modeling juridical texts in a comprehensible way has been the objective of many research works. Today, in practice exist a great variety of database management systems [6], and web-based information retrieval systems [11], [13], [14]. However, those systems often retrieve a large number of irrelevant cases, whereas important cases *similar* to the situation of interest are not found. The retrieval of cases is limited to a set of common keywords and strings, without considering similarity between terms and different relevance of indexes for the determination of relevant cases.

In this context, CBR offers great potential for the intelligent retrieval of jurisprudence documents. Its major benefits are its focus on knowledge in form of individual episodes, rather than general abstract domain knowledge, and similarity-based retrieval.

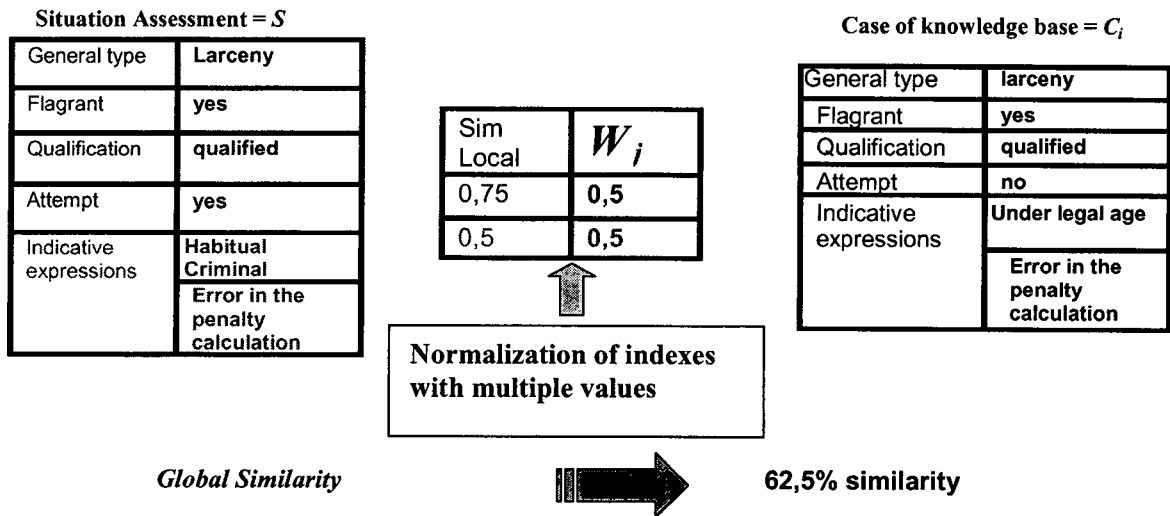


Figure 6. Calculation of similarity.

CBR has been applied in the system PRUDENTIA [17] for the retrieval of jurisprudential documents. In contrast to the system PRUDENTIA, JurisConsulto uses enhanced techniques for the representation and retrieval of juridical cases and the automatic extraction of information from legal texts. JurisConsulto applies besides the global similarity measure also the calculation of local similarity values for each index. This improved considerably the retrieval effectiveness and efficiency. Besides the knowledge represented in form of cases, as in the system PRUDENTIA, JurisConsulto integrates additional legal domain knowledge, such as a juridical thesaurus in accordance to the theoretical juridical framework.

CBR has also applied for document retrieval in other domains [10], such as e.g. the FallQ project in hotline support [7]. The objective of FallQ is to provide a tool that helps finding documents related to a given query to the hotline staff, such as answers in *Frequently Asked Questions*. Similar to our approach a controlled vocabulary and a thesaurus are used for the support of the information extraction and retrieval process. However, besides being applied in a different domain, the controlled vocabulary in our approach and the thesaurus were developed based on the theoretical juridical framework rather than based on the documents themselves as in the FallQ project.

**5. Conclusion**

In this paper, we present a case-based approach for the retrieval of text documents from jurisprudential text bases. A major

problem and issue of discussions in the juridical domain is the interpretation of the juridical language. Our approach is based on the theory of the juridical argument with defines important and discriminate information for the retrieval of relevant juridical documents. Based on the theoretical framework, a general legal domain knowledge has been modeled and represented, such as, e.g., a controlled vocabulary and a juridical thesaurus, to support the representation of juridical documents and to facilitate the interpretation of its content.

Our approach has been based on the domain of the Brazilian Criminal Law. Currently we are implementing a prototype including 3400 cases of Criminal Appeals originating from of the Court of Law of Santa Catarina. Further research work has to be performed in order to improve and broaden the application domain of the approach in different juridical areas based on the evaluation and feedback from its application in practice.

**Acknowledgements**

The authors would like to thank André Brotolon, Luiz Gustavo Bueno and Ricardo M. Barcia Filho, who contributed to our research.

## References

- [1] Amontdt, A., Plaza, E. Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches. *AI Communications*, 17(1), 1994.
- [2] Bench-Capon, Trevor. Argument in Artificial Intelligence and Law. *Jurix*. 1995.
- [3] Delmanto, C. *Código Penal comentado*. 3.ed. ampl. E atualizada.- Rio de Janeiro: Renovar. 1991.
- [4] Jesus, Damásio E. de. *Código de Processo Penal anotado*. 8. ed. ,atual. e aum. – São Paulo : Saraiva, 1989.
- [5] Kolodner, J. *Case-Based Reasoning*. Morgan Kaufmann, Los Altos, CA. 1993.
- [6] Informa Jurídico: jurisprudência e legislação. Pro-link software. CD-1. 1996
- [7] Lenz, M.; Burkhard, H.-D.. CBR for Document Retrieval: The FallQ Project. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*. D. Leake, E. Plaza (ed.)-Berlin: Springer Verlag, 1997.
- [8] Lenz M.; Hübner A.; Kunze M. Textual CBR. M. Lens, B. Boutsh-Sporl, H.-D. Burkhard, S. Wess (eds.). *Case- Based Reasoning Technology*. Springer Verlag. 1998.
- [9] Reale, Miguel. *Fontes e Modelos de Direito*. São Paulo:Saraiva, 1994.
- [10] Rissland, Edwina L.; Daniels, Jody. Using CBR to Drive IR. AAAI. Proceedings of IJCAI-95, Montreal, Canada, 1996.
- [11] Senado Federal .Available Online. [Http://www.senado.gov.br](http://www.senado.gov.br). [October, 20,1998].
- [12] Streak, L. L. Súmulas no Direito Brasileiro: Eficácia, Poder e Função. Porto Alegre: Livraria do Advogado,1995.
- [13] Supremo Tribunal Federal. Available Online. [Http://www.stf.gov.br](http://www.stf.gov.br). [November, 10,1998].
- [14] Tribunal de Justiça de Santa Catarina. Available Online. [Http://www.tj.sc.gov.br](http://www.tj.sc.gov.br). [November, 15,1998].
- [15] Warat, L. A. *O Direito e sua Linguagem*. 2ª versão. 2ª ed. Sergio Antonio Fabris Editor. Porto Alegre. 1995.
- [16] Warat, L.A. *Introdução Geral ao Direito. Interpretação da Lei.: temas para uma reformulação*. Porto Alegre: Sergio Antonio Fabris Editor, 1994.
- [17] Weber-Lee, R.; Barcia, R.; Costa, M.; Rodrigues Filho, I.; Hoeschl, H. C.; Bueno, T.; Martins, A., Pacheco, R. A Large Case-Based Reasoner for Legal Cases. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*. D. Leake, E. Plaza (ed.)-Berlin: Springer, 1997
- [18] Weber, R.; Barcia, R.; Pacheco, R.; Martins, A.; Hoeschl, H.; Bueno, T.; Costa, M.; Rodrigues Filho, I. Representing Cases From Texts In Case-Based Reasoning. *III Congresso Internacional de Engenharia Industrial e XVII ENEGEP*, Brasil, 1997.



## ANEXO C

Artigo aprovado para ser apresentado no I workshop IDEIA/Conferência IBERAMIA 2002. 12 de novembro de 2002, Sevilha, Espanha.

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

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

Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Campus Universitário, Trindade, Florianópolis, Santa Catarina, Brasil.  
CEP: 88040-900

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

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

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

**Submetido a:** IDEIA 2002

**Tópico:** Sistemas Baseados em Conhecimento

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

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

Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Campus Universitário, Trindade, Florianópolis, Santa Catarina, Brasil.  
CEP: 88040-900

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

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

## 1. INTRODUÇÃO

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

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

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

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

## 2. Embasamento Teórico

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

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

Tabela 13. Comparação entre Mineração de Dados e Mineração de Textos

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

Diante dessa conceituação verifica-se que a mineração de textos possui duas fases principais e seqüentes: a extração de informações e a própria mineração de dados. A primeira destina-se a extrair conceitos, estatísticas e palavras relevantes de um

conjunto textual para estruturá-los minimamente, preparando-os para a aplicação das técnicas de mineração de dados. Neste segundo momento aplicam-se as diretrizes e algoritmos de mineração de dados destinados a gerarem regras, classificações ou agrupamentos.[Figura. 1]

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

A extração de informações cumpre um importante papel no processo de mineração de textos, visto que ele permite o pré-processamento do corpo dos documentos textuais de maneira a facilitar o processo de mineração. Geralmente, o processo de extração de informações gera pares de atributos e valores que representam o conteúdo do documento. Normalmente, os atributos são um conjunto previamente definidos.[Tabela. 2]

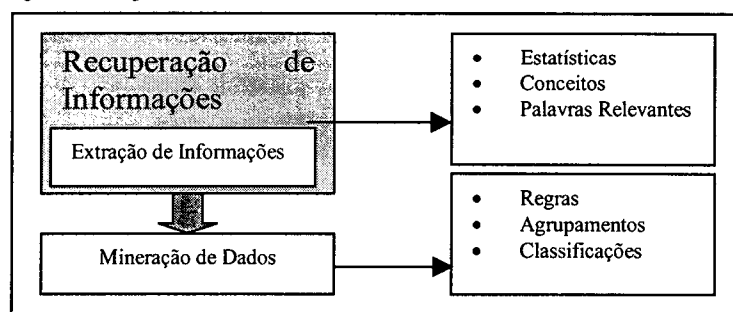


Figura. 1. Técnica de Mineração de Textos

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

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

1. A informação recuperada deve ter valor para o operador do conhecimento. A mineração de textos deve fornecer novas visões sobre os textos para os operadores do conhecimento. Isto permite a adição de valores à base de conhecimento;
2. O conteúdo deve ser baseado em texto. Para dados numéricos as tecnologias existentes estão melhor desenvolvidas;
3. O valor da mineração de textos é diretamente proporcional ao valor do dado que se está minerando. Quanto mais importante for o conhecimento contido na coleção de textos, mais valiosa será a mineração;

4. O conteúdo deve estar explicitamente declarado no texto, tais como textos técnicos e científicos;
5. A mineração de textos é mais valiosa quanto menos estruturado for o texto. Textos desorganizados, contendo acúmulo de informações são os mais indicados.

Tabela 14. Exemplificação da extração de conceitos em textos usando o Intelligent Miner da IBM.

Categoria	Palavra-chave_1	Palavra-chave_2	Palavra-chave_3	Maior Valor
Aplicações Intranet	de Título artigo	do	Aplicações Web	Avaliação de Rede R\$ 41.000,00
Política pessoal	de Licença Maternidade	do	Benefícios saúde	de Licença Paternidade R\$ 150,00
Informações sobre o Lotus Notes	Replicação Banco Dados	do	e-mail	de Aplicações colaboração R\$ 90,00
Informações para viajantes	Linhas ônibus	de	Opções tele-trabalho	de Estações Trem R\$ 2,50
Ergonomia do Escritório	do Descanso pulso	do	Curvatura coluna	da Reconhecimento de voz R\$ 208,00

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

## 2.1 RBC como algoritmo de recuperação

A técnica de Raciocínio Baseado em Casos (RBC) encontrada no campo da Inteligência Artificial (IA), notadamente a parte do estudo que trata da métrica de similaridade é aqui usada para a comparação de textos.

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

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

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

## 3. Estado da Arte

### 3.1 Mercado

O avanço das tecnologias disponíveis para o trabalho com informações superou a abordagem de dados armazenados para a junção destes em informações e, mais recentemente, com os processos de mineração de dados ressalta-se a descoberta do conhecimento nessas informações.

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

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

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

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

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

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

### 3.2 Ferramentas de mineração de textos

A análise das ferramentas presentes no mercado privilegiou três aplicações:

#### 3.2.1 TextAnalyst – Megaputer

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

- Aplicação *standalone*
- SDK de Componentes COM para fácil integração
- Funcionalidade do TextAnalyst:
- Purificação do Resultado (Rede Semântica);
- Navegação;
- Sumarização;
- Explicação em Tópicos;
- Agrupamento;
- Focalização dinâmica;
- Categorização (TextAnalyst COM).

#### 3.2.2 Semio

Características da suíte de informação:

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

#### 3.2.3 Intelligent Miner for Text - IBM.

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

O Intelligent Miner for Text inclui:

- Ferramentas de análise de texto;
- Um avançado mecanismo de busca;
- Ferramentas de acesso a Web.

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

Muitas das ferramentas no Intelligent Miner for Text podem ser vistas como metadados sobre os documentos. Metadados são os dados estruturados que podem ser

armazenados em uma base de dados e poderia ser usado objetivando a Mineração de Dados.

## 4. Aplicação

Pesquisas em inteligência artificial e inteligência jurídica conduzidas pela Web Intelligence Systems S.A. geraram uma aplicação inovadora denominada AlphaThemis<sup>®6</sup>.

Consiste em um sistema inteligente para a extração do conhecimento contido nas súmulas dos tribunais nacionais, configurando-se em uma potente ferramenta de mineração de textos destes documentos norteadores de decisões judiciais. A técnica responsável pela recuperação de súmulas empregada no aplicativo é derivada do desejo de se mesclar uma pesquisa estruturada em texto e atributos que pudessem ensejar a formação de um contexto (próprio para o sucesso da busca). A modelagem conferida ao software tem como idéia básica tratar o conhecimento contido nas súmulas e seus respectivos enunciados de forma tal que a interface do sistema permita ao usuário formular uma consulta na qual seja possível mesclar atributos fixos com texto livre.[Fig. 3]

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

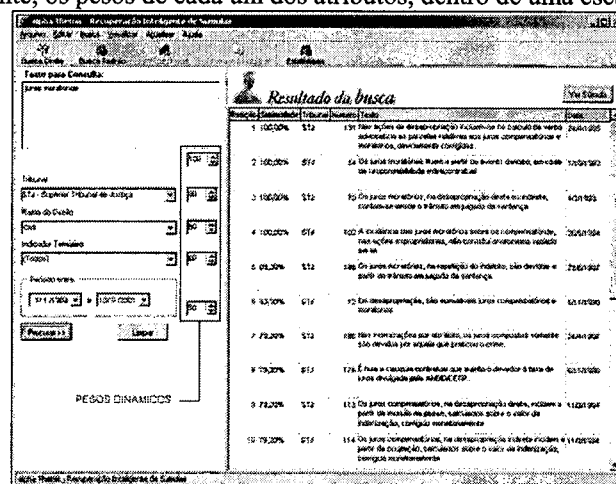


Figura. 2. Interface AlphaThemis<sup>®</sup>, mostrando a saída dos dados e ressaltando o uso de pesos dinâmicos

<sup>6</sup> Todos os direitos reservados da marca e uso da aplicação reservados para a Web Intelligence Systems S.A. Copyright 2001.



O uso de pesos dinâmicos nos filtros de busca torna o AlphaThemis<sup>®</sup> ainda mais flexível na formulação da pesquisa. Trata-se da possibilidade de alteração dinâmica da relevância dos atributos dentro do contexto pesquisado, calibrando-se, individualmente, os pesos de cada um dos atributos, dentro de uma escala de 0 a 100. O resultado da pesquisa é apresentado ao usuário em uma interface de saída que mostra os resultados hierarquizados de acordo com o percentual de similaridade constatado entre a consulta e os enunciados das súmulas. O sistema AlphaThemis<sup>®</sup> apresenta em sua interface de saída, além do texto da súmula recuperada, todos os mecanismos que o usuário acionou na realização da consulta, bem como os valores respectivamente atribuídos (a fim de que o usuário possa checar quais atributos e valores influenciaram na escolha do registro).

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

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

#### **4.1 AlphaThemis<sup>®</sup> como minerador de textos**

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

##### **4.1.1 Características do texto**

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

##### **4.1.2 Extração das informações**

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

##### **4.1.3 Conceitos**

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

Conforme demonstrado pela figura 4, a junção de várias palavras podem gerar conceitos diferenciados daqueles representados pelas palavras isoladamente.

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

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

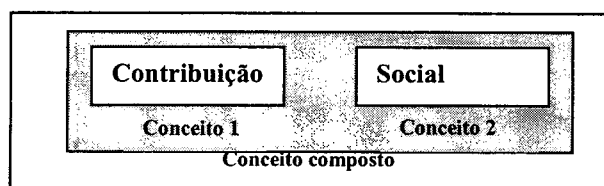


Figura. 3. Identificação de conceitos

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

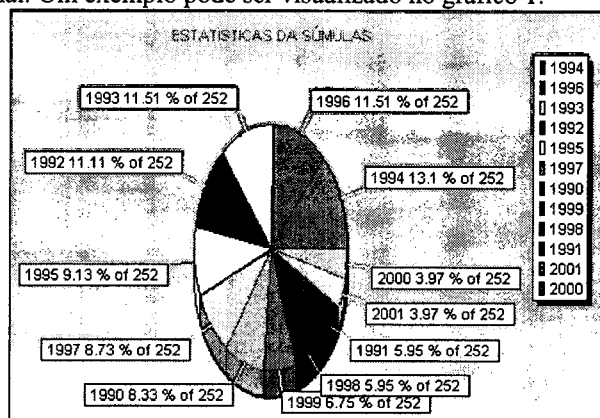


Gráfico 1. Resultado da produção anual de súmulas do Superior Tribunal de Justiça

## 5. Conclusão

Identifica-se claramente a mineração de textos como uma fronteira tecnológica, onde as aplicações estão em fase inicial e os conceitos não completamente definidos. No

mercado as empresas procuram consolidar excelências no fornecimento de produtos para esse fim e as indicações de crescimento e importância do segmento são claras. Novas tecnologias possuem a inovação como fundamento e o campo apresenta-se aberto para alternativas criativas que ampliem as possibilidades das ferramentas atuais. A incorporação da abordagem de Inteligência Artificial figura como norte estratégico nessa área, visto as características não estruturadas do conhecimento em foco.

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

Inserido no processo definido de mineração de textos o aplicativo descrito inicia seus trabalhos extraindo informações do texto das súmulas, sejam explícitas ou implícitas, por técnicas de localização ou inferência. Segue essa atividade de forma dinâmica a cada nova entrada na base ou a partir de demandas da consulta. Executa, conseqüentemente, a mineração de dados, extraindo estatísticas e agrupando conceitos.

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

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

## Referências

- [1]ANANYAN, Sergei. **Text Mining Applications and technologies**. Disponível em: <[http://www.megaputer.com/down/tm/Text\\_Mining.pps](http://www.megaputer.com/down/tm/Text_Mining.pps)>. Acesso em: 30/01/2002.
- [2]BUENO, Tania Cristina D'Agostini. **Recuperação da Informação Jurídica em Sistema Baseado em Casos**. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis.
- [3]BUENO, Tania Cristina D'Agostini; HOESCHL, Hugo César. **Representação e Extração do Conhecimento das Súmulas do STJ**. Florianópolis, 11 p. Trabalho não publicado. Acesso Restrito para Web Intelligence Systems S.A..
- [4]IBM. **IBM Intelligent Miner for Text**. In: IBM Business Intelligence Solutions CD. EUA: 1998. CD-ROM.
- [5]NAHM, Un Yong; MOONEY, Raymond J. **Text Mining with Information Extraction**. In: AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases. **Proceedings ...** Stanford: American Association for Artificial Intelligence, 2001.
- [6]SEMIO. **Semio Products and Solutions Overview**. Disponível em: <<http://www.semio.com/products/overview.html>>. Acesso em: 30/01/2002.
- [7]SEMIO CORPORATION. **Text Mining and the Knowledge Management Space**. Disponível em: <<http://www.dmreview.com/master.cfm?NavID=61&WhitePaperID=80>>. Acesso em: 30/01/2002.
- [8]TKACH, Daniel. **Text Mining Technology: Turning Information Into Knowledge. A White Paper from IBM**. In: IBM Business Intelligence Solutions CD. EUA: 1998. CD-ROM.
- [9]TEXT MINING RESEARCH GROUP AT THE UNIVERSITY OF WAIKATO. **Text Mining**. Disponível em: <<http://www.cs.waikato.ac.nz/~nzdl/textmining/>>. Acesso em: 30/01/2002.

- [10] Web Intelligence Systems S.A. **Documentos de registro da Tecnologia AlphaThemis**. INPI, 2001.
- [11] WOHL, Amy D. **Intelligent Text Mining Creates Business Intelligence**. In: IBM Business Intelligence Solutions CD. EUA: 1998. CD-ROM.