

**UNIVERSIDADE PAULISTA – UNIP**

**PROGRAMA DE MESTRADO EM ENGENHARIA DE  
PRODUÇÃO**

**Documentos textuais jurídicos: um estudo de  
ferramentas de TI para sua recuperação**

EDIVALDO ANTONIO SARTOR

Dissertação apresentada ao  
Programa de Pós-Graduação em  
Engenharia de Produção da  
Universidade Paulista, para  
obtenção do título de Mestre.

São Paulo

2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**UNIVERSIDADE PAULISTA – UNIP**

**PROGRAMA DE MESTRADO EM ENGENHARIA DE  
PRODUÇÃO**

**Documentos textuais jurídicos: um estudo de  
ferramentas de TI para sua recuperação**

EDIVALDO ANTONIO SARTOR

Orientador: Prof. Dr. Marcelo Schneck de  
Paula Pessôa

Área de Concentração: Gestão da  
Informação

Linha de Pesquisa: Produção de  
Software

Projeto de Pesquisa: Aplicação de  
Sistemas Integrados de gestão

Dissertação apresentada ao Programa  
de Pós-Graduação em Engenharia de  
Produção da Universidade Paulista, para  
obtenção do título de Mestre.

São Paulo

2007

## **Dedicatória**

Dedico este trabalho a Deus, pela força, saúde, capacidade, paz, oportunidades e por várias outras bênçãos em minha vida.

Aos meus Pais que, mesmo sem terem tido muitas oportunidades de aprendizado, transmitiram-me, desde cedo, o valor do conhecimento.

À minha esposa, por seu amor incondicional.

Aos amigos, pelo incentivo e apoio nos momentos difíceis.

## **Agradecimentos**

A Deus, acima de tudo, pela paz e pela oportunidade de ter chegado até aqui.

À minha esposa, por sua paciência, dedicação e apoio em todos os momentos.

Aos meus pais e irmãs, que sempre estiveram ao meu lado e pelas inesquecíveis lições de caráter e integridade.

Ao professor e amigo Dr. Marcelo Pessôa, por acreditar e incentivar o desenvolvimento deste trabalho e por seus ensinamentos preciosos.

Aos professores, Dr. Fernando J. B. Laurindo e Dr. Antonio Roberto Albuquerque, por colaborarem com sugestões para aprimorar este trabalho.

A todos os professores do Programa de Pós-Graduação em Engenharia de Produção da UNIP, pelos ensinamentos.

A todos os amigos da STI/TJSP pela paciência, em especial ao “Mazola”, parceiro a quem coube a maior parcela das dúvidas.

A todos os amigos e familiares que, diretamente ou indiretamente, colaboraram para o desenvolvimento deste estudo.

## **Resumo**

Sartor, E. A. **Documentos Textuais Jurídicos: Um Estudo de Ferramentas de TI para sua Recuperação.** Dissertação (Mestrado em Engenharia de Produção) – Instituto de Ciências Exatas e Tecnológicas, Universidade Paulista, 2007.

O aumento expressivo da produção de informações jurídicas reflete, de um lado, a crescente complexidade da sociedade e, de outro, a necessidade de recuperá-las de forma eficiente, para que gerem grande benefício tanto para o funcionamento do sistema jurídico quanto para a sociedade.

As informações jurídicas, de modo geral, se traduzem em documentos textuais, que possuem características diferenciadas quando do seu armazenamento e recuperação. Considerando a TI como fator importante no processo de recuperação dessas informações e assumindo que novas tecnologias surgem a cada dia abrindo novos leques de utilização, este trabalho objetiva, por meio do amparo de ferramentas de TI, discutir formas de organizar e principalmente recuperar documentos textuais jurídicos, garantindo sua confiabilidade e disponibilidade, para servirem de referência a futuros casos.

A validade do estudo se dá através da análise de três unidades principais: pessoas, processos e tecnologia, tratadas por outros autores, na forma de análise técnica de ferramentas e estudo de caso com usuários envolvidos nos documentos textuais no Tribunal de Justiça de São Paulo (TJSP).

Palavras-chave: Recuperação de documentos, Tecnologia da Informação, Documentos Jurídicos, Documentos Textuais e Ferramentas de TI.

## **Abstract**

The expressive increase of the production of juridical information contemplates, on a side, to growing complexity of the society and of other, the need to recover them in an efficient way, so that this way, they generate great so much benefit for the operation of the juridical system as for the society.

The juridical information, in general, are translated in textual documents, that possess differentiated characteristics when of your storage and recovery. Considering TI with important factor in the process of recovery of these information and assuming that new technologies appear every day opening new use fans, this work is delimited the, aided by tools of TI, to discuss forms of organizing and mainly to recover juridical textual documents, guaranteeing your reliability and readiness, so that they serve from reference to futures cases.

The validity of the study feels through the tripod people's analysis, processes and technology, treated by other authors, in the form of technical analysis of tools and case study with users involved with the textual documents in the Tribunal of Justice of São Paulo (TJSP).

Word-key: Recovery of documents, Technology of the Information, Juridical Documents, Textual Documents and Tools of TI.

## Sumário

1. Introdução.....	12
1.1 Relevância .....	13
1.2 Formulação do problema.....	14
1.3 Objetivos .....	14
1.4 Organização do trabalho .....	15
1.4.1 Metodologia da pesquisa .....	15
1.4.2 Estrutura deste documento .....	16
2. Descrição de domínio .....	18
2.1 Considerações iniciais .....	18
2.2 Direito.....	18
2.2.1 Origem do Direito.....	19
2.2.2 Os sistemas jurídicos .....	21
2.2.3 Como está organizado o sistema jurídico brasileiro .....	22
2.3 Definição de termos jurídicos .....	24
2.3.1 Legislação .....	24
2.3.2 Doutrina .....	24
2.3.3 Sentença.....	25
2.3.4 Apelação / Recurso .....	25
2.3.5 Acórdão .....	26
2.3.6 Jurisprudência .....	26
2.3.7 Súmula vinculante .....	29
2.4 Informação e Direito .....	32
2.4.1 Sistemas de Informação.....	32
2.4.2 Características dos documentos jurídicos textuais .....	32
2.4.3 Pesquisa de documentos textuais .....	33
3. Recuperação da informação .....	35
3.1 Linguagens de indexação e busca .....	36
3.2 Modelos para recuperação de informação.....	37
3.3 Consultas jurídicas .....	38
4. Técnicas para recuperação de documentos .....	40
4.1 Considerações iniciais .....	40
4.2 Mecanismos de busca .....	40
4.3 Data mining .....	42
4.4 Técnicas de Inteligência Artificial (IA) .....	44
4.4.1 Algoritmos Genéticos .....	46
4.4.2 Programação evolutiva .....	48
4.4.3 Sistemas Difusos ou Lógica Fuzzy.....	49
4.4.4 Sistemas Baseados em Conhecimento.....	51
4.4.5 Programação Genética .....	53
4.4.6 Redes neurais .....	55
4.4.7 Raciocínio baseado em casos – RBC .....	56



4.5 Resumo das técnicas apresentadas para uso na recuperação de documentos textuais jurídicos .....	62
4.5.1 RBC na área jurídica .....	64
4.5.2 RBC e outras técnicas de IA.....	65
5. Protocolo do Estudo de Caso .....	66
5.1 Considerações iniciais .....	66
5.2 Local do estudo .....	67
5.3 Tipo de estudo .....	68
5.4 Sujeitos da pesquisa .....	70
5.5 Montagem da pesquisa .....	70
5.6 Coleta de dados .....	71
6. Tecnologia: Análise das ferramentas .....	73
6.1 JurisConsulta® .....	73
6.2 AlphaThemis® .....	78
6.3 Idol K2® .....	81
6.4 Comparação das ferramentas .....	83
7. Pessoas: Resultado do caso .....	86
7.1 Caracterização dos entrevistados .....	86
7.2 Necessidade do entrevistado quanto à utilização de documentos textuais em suas atividades .....	87
7.3 Formas atuais de desenvolvimento do trabalho relacionado a documentos textuais .....	88
7.4 Características importantes de um sistema de consulta de documentos textuais .....	92
8. Processos: Resultado do caso .....	96
9. Análise crítica da pesquisa .....	98
10. Considerações finais.....	100
Referências bibliográficas.....	102
Anexos.....	110
Anexo 1 .....	110

## Lista de Figuras

Figura 2.1 Organograma da Jurisprudência.....	28
Figura 2.2 Diagrama de alto nível da utilização/ criação das súmulas.....	30
Figura 3.1 Evolução da consulta jurisprudencial.....	39
Figura 4.1 Mineração de dados como um campo multidisciplinar.....	44
Figura 4.2 Escolha do paradigma Fuzzy ou RNA.....	50
Figura 4.3 Ciclo criar-testar-modificar.....	54
Figura 4.4 Ciclo do RBC.....	61
Figura 5.1 Sistemas de Informação.....	66
Figura 6.1 Interface do sistema JurisConsulta.....	75
Figura 6.2 Processo de recuperação baseado em casos do sistema JurisConsulta.....	76
Figura 6.3 Definição das expressões indicativas no sistema JurisConsulta.....	77
Figura 6.4 Interface AlphaThemis, ressaltando o uso de pesos dinâmicos.....	79
Figura 6.5 Interface demonstrativa da ferramenta Idol K2, desenvolvida na fase de análise.....	83

## Lista de Quadros

Quadro 4.1 Resumo das técnicas apresentadas.....	62
Quadro 6.1 Análise técnica comparativa dos <i>softwares</i> pesquisados.....	84
Quadro 7.1 Necessidade dos documentos textuais.....	87
Quadro 7.2 Ferramentas utilizadas no trabalho de pesquisa.....	88
Quadro 7.3 Campos mais utilizados nas ferramentas de busca.....	89
Quadro 7.4 Números de palavras utilizadas na pesquisa.....	90
Quadro 7.5 Opinião sobre a produtividade das ferramentas utilizadas.....	91
Quadro 7.6 Opinião dos entrevistados a respeito da usabilidade das ferramentas.....	92
Quadro 7.7 Comentários das implementações de demonstração.....	94

## Lista de Abreviações

AG: Algoritmo Genético

Art.: Artigo

Caput.: Refere-se à primeira parte, ou à parte mais alta de um artigo de lei.

CEC: *Congress on Evolutionary Computation*. Congresso de Computação Evolutiva

CF: Constituição Federal

CS: Conselho de Segurança (relacionado à ONU)

IA: Inteligência Artificial

IEEE: *Institute of Electrical and Electronics Engineers*. Instituto de Engenharia Elétrica e Eletrônica.

KDD: *Knowledge Discovery in Database*. Descoberta de Conhecimento em Banco de Dados

ONU: Organização das Nações Unidas

RBC: Raciocínio Baseado em Caso

RNA: Rede Neural Artificial

SBC: Sistema Baseado em Conhecimento

SE: Sistema Especialista

SRH/TJSP Secretaria de Recursos Humanos do Tribunal de Justiça de São Paulo

STF: Superior Tribunal Federal

STI/TJSP: Secretaria de Tecnologia da Informação do Tribunal de Justiça de São Paulo

STJ: Superior Tribunal de Justiça

TI: Tecnologia da Informação

TJSP: Tribunal de Justiça de São Paulo

TRF: Tribunal Regional Federal

WEB: Rede mundial de computadores

## **1. Introdução**

O Direito é uma ciência que visa tratar dos problemas inerentes ao convívio em sociedade. Para tal, é efetivado através da Legislação, Doutrina e Jurisprudência, elementos considerados pilares do Direito (GUSMÃO, 1996). No entanto, esses elementos geram grande quantidade de informações, documentos e, por conseqüência, dados, que precisam ser tratados, armazenados, recuperados e manipulados, para que possam servir de referência futura.

Considerando a informática como fator importante no processo de recuperação da informação, percebe-se que a cada momento surgem novas tecnologias, e cada uma delas abre um grande e também novo leque de utilização. Como a informação vem assumindo um papel de real importância na sociedade atual, transformando-se em ferramenta imprescindível na solução de problemas em diversas áreas do conhecimento, este trabalho visa discorrer sobre os métodos a serem utilizados para que a informação tenha a confiabilidade e disponibilidade necessárias para sua correta e ágil utilização no campo do Direito. Em uma conotação mais prática, buscar-se-á discutir formas para organizar e, principalmente, recuperar documentos textuais, que são largamente utilizados pelos juristas, exemplificados em jurisprudências, súmulas e acórdãos, entre outros, fazendo uso do cenário atual do Tribunal de Justiça de São Paulo.

Objetiva-se ainda, através do apoio da TI (Tecnologia da Informação), apresentar alternativas de ferramentas que apresentem características importantes e adequadas para a recuperação de informações textuais jurídicas.

## **1.1 Relevância**

A relevância dos documentos textuais no Direito já foi abordada em Bueno (1999) apud Mattos (2002), que cita que as jurisprudências são de suma importância para os juristas, auxiliando-os com uma valiosa fonte de interpretação da lei. Ainda em Mattos (2002), verifica-se que a principal deficiência dos sistemas existentes nesse segmento é seu embasamento em bancos de dados relacionais, fazendo com que o sistema de buscas convencionalmente implementado, via palavras-chave, dificulte o encontro de casos realmente similares ao caso em avaliação.

A relevância do estudo mostra-se ainda nos grandes volumes de documentos textuais existentes e naqueles que ainda serão gerados no órgão objeto deste estudo – TJSP. O TJSP conta atualmente com bases de dados referentes a documentos textuais de ementas, súmulas e acórdãos arquivados a partir do ano de 1998, com três milhões de documentos multipaginados, perfazendo aproximadamente 15 milhões de páginas arquivadas em diversos formatos eletrônicos, além da disponibilização de 350 mil ementas em portal de Internet.

Percebe-se, pelos números envolvidos, além da leitura dos autores citados, a importância do correto armazenamento dos documentos textuais como informação e sua conseqüente recuperação dentro do cenário jurídico.

## **1.2 Formulação do problema**

Conforme já citado, o Direito é efetivado através da Legislação, Doutrina e Jurisprudência, elementos considerados pilares dessa ciência (GUSMÃO, 1996). Todos esses pilares são constituídos basicamente de documentos textuais que necessitam ser armazenados e obviamente recuperados para servirem como referência aos juristas. Esses documentos apresentam particularidades que devem ser levadas em conta no momento de sua pesquisa e armazenamento.

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 (MATTOS, 2002). Corroborando a afirmação anterior, o estudo de caso proposto no TJSP mostra esse mesmo cenário, porém com algumas particularidades pertinentes ao órgão como: grande volume de documentos, diversas bases de dados independentes, diversos tipos de arquivos, falta de padronização na geração desses tipos de documentos, entre outros.

Diante do apresentado, o problema encontrado, e que será objeto do estudo de caso do TJSP, é a dificuldade de acesso às informações textuais que servem de embasamento para decisões dos juristas: jurisprudências, legislação e doutrina.

## **1.3 Objetivos**

Tendo em vista a questão levantada no Item 1.2, o objetivo do trabalho é validar, através de estudo de caso, a real importância dos documentos textuais no contexto jurídico, além de demonstrar como seu correto armazenamento somado a

ferramentas apropriadas de recuperação contribui para melhores resultados nos procedimentos judiciais no TJSP. Em outras palavras, buscar-se-á analisar como são tratados os pilares dos sistemas de informação, já explorados por outros autores (LAUDON e LAUDON, 2004), quando do trabalho com documentos textuais no órgão em questão.

Como objetivo secundário, será feita a análise de ferramentas para a recuperação de documentos textuais abordando-se características importantes aos usuários finais.

Como contribuição adicional, embora não desenvolvida a fundo, o trabalho abordará aspectos relacionados ao processo de desenvolvimento de sistemas para fins jurídicos, discutindo os que são desejáveis a essa finalidade.

Embora não seja um objetivo isolado e sim um meio para se chegar aos que se pretende, este trabalho contribui com um estudo resumido a respeito de técnicas de Inteligência Artificial.

## **1.4 Organização do trabalho**

### **1.4.1 Metodologia da pesquisa**

Este trabalho faz um estudo de caso da aplicabilidade e aceitação dos sistemas jurídicos para recuperação de informações textuais abrangendo os conceitos específicos a esses sistemas, além de abordar a importância dos documentos textuais no órgão objeto de estudo. As técnicas pesquisadas na literatura serão analisadas, comparadas e validadas através de estudo de caso,



visando sua adequação às características e particularidades inerentes ao domínio estudado.

Na metodologia de trabalho adotada, podem-se identificar as seguintes etapas:

1. Identificação do problema genérico e específico ao órgão objeto do estudo de caso – TJSP.
2. Identificação dos elementos pertinentes ao universo jurídico através de pesquisa bibliográfica.
3. Revisão bibliográfica dos métodos para recuperação de informações, para obter-se um panorama das diferentes técnicas que podem ser utilizadas para tal finalidade.
4. Análise dos métodos identificados e verificação de adequação diante do domínio estudado.
5. Obtenção dos resultados através da análise das diferentes técnicas, proporcionando a resposta ao problema inicial.
6. Análise do caso na utilização de ferramentas específicas de recuperação de informação dentro do domínio jurídico.

#### **1.4.2 Estrutura deste documento**

No Capítulo 2, *Descrição de domínio*, procura-se definir o domínio jurídico e suas particularidades relacionadas a documentos textuais, necessários para compreensão do problema proposto. Nesse capítulo, dá-se a definição dos termos e conceitos utilizados no trabalho.

No Capítulo 3, *Recuperação da informação*, busca-se demonstrar a importância da recuperação das informações bem como as formas utilizadas para isso. Nesse capítulo, citam-se características da recuperação da informação voltadas para consultas jurídicas.

No Capítulo 4, *Técnicas para recuperação de documentos*, é feita a definição de algumas técnicas utilizadas para recuperação de documentos textuais, como Inteligência Artificial e suas variadas áreas de atuação. Busca-se também descrever outras técnicas que podem ser aplicadas no domínio jurídico.

No Capítulo 5, *Protocolo do estudo de caso*, é feita a caracterização do ambiente estudado, assim como os sujeitos nele envolvidos. Aborda-se ainda, a definição do tipo de pesquisa a ser desenvolvida, com os respectivos critérios adotados além da análise e interpretação dos dados coletados. São ainda expostas algumas aplicações desenvolvidas, utilizando-se as técnicas mencionadas no capítulo anterior.

No Capítulo 6, *Análise das ferramentas*, é feita a análise do funcionamento e das funcionalidades das ferramentas pertinentes ao domínio estudado.

No Capítulo 7, *Resultado do caso*, todos os dados analisados são tabulados no intuito da validação das premissas do estudo.

No Capítulo 8, *Considerações finais*, são feitas as conclusões do trabalho desenvolvido além de se propor possibilidades de trabalhos futuros.

## **2. Descrição de domínio**

### **2.1 Considerações iniciais**

Pretende-se aqui definir o domínio jurídico, com definição de termos e conceitos que serão utilizados no desenvolvimento deste estudo. Propõe-se ainda demonstrar a definição e as particularidades dos documentos textuais no âmbito do direito, como armazenamento e recuperação apropriada. Por domínio jurídico entende-se o conjunto de elementos que caracterizam determinadas aplicações, estabelecendo fronteiras com outros tipos de sistemas.

### **2.2 Direito**

Conforme Gusmão (1996), de acordo com a etimologia, o termo *direito* vem do latim *directum*, do verbo *dirigere* (dirigir – orientar – endireitar). Aquilo que é *reto, direito, conforme a razão*. Já no sentido didático, o Direito é uma ciência que estuda as regras impositivas, sob os pontos de vista legal e doutrinário, do relacionamento social. Dessa forma, pode-se concluir que o Direito é um conjunto de regras específicas com o intuito de disciplinar e tratar dos problemas e conflitos comuns da vida em sociedade, fruto de valores morais e éticos da conduta humana.

Nessa mesma conotação, Rosa (1970) afirma que o direito é fato social. Ele se manifesta como uma das realidades observáveis na sociedade. É o instrumento institucionalizado de maior importância para o controle social. Desde o início das sociedades organizadas manifestou-se o fenômeno jurídico, como um

sistema de normas de conduta que corresponde a uma coação exercida pela sociedade, segundo certos princípios aprovados e obedientes e formas predeterminadas; e Nóbrega (1962) declara que: a religião, a moral, a educação, a arte e o direito são processos adaptativos, que visam a modelar, corrigir, melhorar a psique do homem, para melhor integração dele à vida social. E conclui ainda: o direito não tem sede nem na natureza, nem no mundo dos valores, mas participa a um só tempo de um e de outro: da natureza, porque tem base na vida humana, nas relações sociais, e dos valores, pela significação que imprime a essas relações, orientando-as para a satisfação dos interesses comuns. É, portanto, natureza valorada, modelada pelo valor e valor objetivado através de dados naturais. O que significa que o direito é fato cultural e se situa no mundo da cultura.

Parafraseando Rover (1997), vivemos em uma era com mais tecnologia, mais riqueza, maior número de informações e, conseqüentemente, maior número de direitos e, com isso, novos direitos para tratar toda essas relações.

### **2.2.1 Origem do Direito**

SECCO (1988) afirma que o ser humano é, essencialmente, um animal gregário. Significa dizer que não só é próprio da sua natureza, como também inerente às suas condicionantes de sobrevivência o inter-relacionamento com os semelhantes. De fato, o ser humano, sendo dotado de sentimentos e de razão, precisa comunicar-se, trocar experiências, produzir bens e desfrutar do produto do trabalho próprio e alheio, sendo impossível gerar sozinho tudo o que necessita para viver. Portanto, é próprio dele viver integrado em sociedade.

Diante da afirmação de Secco (1998), conclui-se que é próprio do ser humano viver em sociedade e, assim, integrado a dois mundos: o mundo natural e o mundo cultural. O mundo natural refere-se ao mundo da natureza, constituído pelos elementos que são inerentes a ela. Porém, o grande diferencial do ser humano vai se dar justamente pela criação do mundo cultural. Tal mundo foi criado por meio do exercício de sua inteligência e do seu trabalho, com o objetivo de modificar e buscar adaptar o mundo natural às suas necessidades e comodidades (MAÇOLI, 2005).

"A primeira coisa que o homem faz em sociedade é produzir para viver. Para produzir, usa instrumentos, aplica conhecimentos e experiências que decidem o tipo de relações que unirão e organizarão os indivíduos associados no trabalho."  
(LIMA, 1972)

O mundo cultural caracteriza-se então pelas realizações, criações, produções e tudo aquilo que o homem possa retirar da natureza para melhor satisfazer seu modo de vida. Dessa forma, como as necessidades humanas se revelaram sempre crescentes, exigia-se um relacionamento cada vez mais intenso com outros indivíduos, o que justifica a citação de Secco (1998), do ser humano como ser coletivo, que não sobrevive sozinho. Com o crescimento de relacionamentos e negócios feitos inter-indivíduos, começam a surgir os primeiros problemas de convivência, desigualdades e conflitos. Com o passar do tempo e o aumento da complexidade das relações, ficou claro para o ser humano que o mundo cultural necessitava se ampliar e, com isso, criou-se o mundo social. Dentro do mundo social, inicia-se a criação de normas e regras a serem

obedecidas por todos, surgem assim as primeiras “leis”, que atendiam as divergências e problemas encontrados nos convívios sociais (MAÇOLI, 2005).

O Direito ou mesmo as leis são elementos de base para os relacionamentos humanos, uma vez que os seres humanos divergem de pensamento e ponto de vista, e esses instrumentos constituem mecanismos reguladores do convívio social. Além disso, percebe-se que as leis se modificam através dos tempos, pois os relacionamentos humanos também se transformam.

### **2.2.2 Os sistemas jurídicos**

Atualmente, os dois principais sistemas legais, no mundo, são os de direito civil e de direito comum. A Europa Continental, a América Latina, a maior parte da África e muitos países da Europa Central e da Ásia fazem parte do sistema de direito civil; os Estados Unidos, além da Inglaterra e outros países que, no passado, faziam parte do Império Britânico, pertencem ao sistema de direito comum.

Segundo relata Messite (1999), o sistema de direito civil tem suas origens no antigo direito romano, que foi atualizado no século VI da era cristã pelo imperador Justiniano, e adaptado posteriormente por juristas franceses e alemães.

O sistema de direito comum começou a se desenvolver na Inglaterra há quase mil anos. Quando o parlamento da Grã-Bretanha foi fundado, seus juízes reais já tinham começado a basear suas decisões no direito costumeiro, "comum" no reino. Atualmente, existem características associadas ao sistema de direito comum que são encontradas no sistema de direito civil, podendo-se citar o júri. O júri é um grupo de cidadãos, normalmente composto de 12 pessoas, que são

convocados ao acaso para determinar os fatos em um processo. Quando é realizado um julgamento pelo tribunal do júri, o juiz orienta a respeito das leis, mas cabe ao júri decidir.

### **2.2.3 Como está organizado o sistema jurídico brasileiro**

O sistema jurídico brasileiro tem como base o chamado sistema continental de origem Romano-Germânica adotado na Europa, principalmente na França, Itália e Alemanha (MESSITE, 1999).

As normas que compõem o ordenamento jurídico encontram-se dispostas segundo uma hierarquia. A Constituição da República ocupa o ponto mais alto, tendo as demais normas como subordinadas. Estar juridicamente subordinada implica que determinada norma prevaleça sobre a inferior em qualquer caso em que com ela conflite.

A organização do Poder Judiciário, mediante atribuição de funções e competências aos seus órgãos, é estabelecida por meio da Constituição Federal, Constituições Estaduais, Leis de Organização Judiciária de cada Estado-membro, Lei Orgânica da Magistratura e Regimentos Internos dos Tribunais.

De acordo com o art. 92/98 da CONSTITUIÇÃO FEDERAL, os órgãos do Poder Judiciário são:

- STF – Supremo Tribunal Federal, órgão de cúpula do Poder Judiciário, que tem por função essencial guardar a Constituição Federal, incumbindo-lhe, portanto, a jurisdição constitucional.

- STJ – Superior Tribunal de Justiça tem por objetivo essencial preservar a autoridade e a correta aplicação da lei federal e uniformizar o seu entendimento.
- TRF – Tribunais Regionais Federais e Juízos Federais, compostos por Juízes Federais – órgãos de primeiro grau, e pelos Tribunais Regionais Federais, tendo sido a área de jurisdição de cada um deles fixada pela divisão do País em regiões, atualmente em número de cinco.
- Tribunais e Juízes do Trabalho, aos quais compete conciliar e julgar os dissídios individuais e coletivos entre trabalhadores e empregadores, e outras controvérsias decorrentes de relação de trabalho.
- Tribunais e Juízes Eleitorais, cabem-lhes processar e julgar as causas relativas à organização e ao exercício de direitos políticos, essencialmente os de votar e de ser votado, são compostos dos seguintes órgãos: Tribunal Superior Eleitoral; Tribunais Regionais Eleitorais; Juízes Eleitorais e Juntas Eleitorais.
- Tribunais e Juízes Militares, compostos pelo Superior Tribunal Militar; Tribunais e Juízes Militares instituídos por lei, competindo-lhes processar e julgar os crimes militares definidos em lei.
- Tribunais e Juízes dos Estados e do Distrito Federal e Territórios: compete-lhes organizar a sua justiça, observados os princípios estabelecidos pela Constituição Federal.

Por fim, no âmbito da Justiça estadual e do Distrito Federal, foram instituídos por meio da Lei nº. 9.099/95, os juizados especiais cíveis e criminais. São órgãos especiais do Poder Judiciário que têm por competência julgar causas



cíveis de menor complexidade, assim consideradas em razão do valor da causa (até quarenta salários mínimos) ou da matéria, e causas criminais de infrações de menor potencial ofensivo.

### **2.3 Definição de termos jurídicos**

Para fins de entendimento de alguns tópicos deste trabalho faz-se necessário conceituar alguns termos utilizados no Direito, para que possam sustentar as bases desta proposta:

#### **2.3.1 Legislação**

Conforme Cintra (1991), a legislação estabelece as normas que, segundo a consciência dominante, devem reger as mais variadas relações, dizendo o que é lícito e o que é ilícito, atribuindo direitos, poderes, faculdades, obrigações; são normas de caráter genérico e abstrato, ditadas a princípio sem distinção particular a nenhuma pessoa e a nenhuma situação concreta. Tem-se então na legislação, a forma documentada da lei.

#### **2.3.2 Doutrina**

A doutrina (jurídica) é também uma das fontes subsidiárias do Direito. Consiste em uma forma expositiva e esclarecedora do Direito, feita pelo jurista a quem cabe o estudo aprofundado da ciência. A Doutrina é, conseqüentemente, o Direito resultante de estudos voltados para a sua sistematização, esclarecimento, adequação e até mesmo inovação. (SECCO, 1998)

A doutrina é o resultado de experiências adquiridas pelos juristas, mestres de jurisprudência e dos juízes. A doutrina tem como função a formação das leis, acompanhando a dinâmica da vida social e, por consequência do direito, apontando as necessidades e oportunidades de reformas jurídicas.

### **2.3.3 Sentença**

Segundo Atienza (1979): "Sentença é uma decisão individualizada, aplicável a um caso concreto", ou seja, quando um juiz toma uma decisão dentro de um processo, a qual irá definir o andamento desse processo. Essa decisão é baseada na lei e interpretada pelo respectivo juiz que promulgou tal sentença.

### **2.3.4 Apelação / Recurso**

Segundo Atienza (1979):

Recurso interposto de juiz inferior para superior. Mantém, ainda, a apelação, a mesma significação quando um dos recursos de que se pode utilizar a pessoa prejudicada pela sentença, a fim de que, subindo a ação à superior instância, e conhecendo esta de seu mérito, pronuncie uma nova sentença, confirmando ou modificando o que se proferiu na jurisdição de grau inferior.

Ainda de acordo com Atienza, pode-se concluir que uma apelação é feita quando uma das partes envolvidas em um processo não concorda com a sentença proferida pelo juiz da primeira instância ou se julga prejudicada, dessa forma é aplicada a apelação da sentença para que seja reavaliada pela instância superior.

### 2.3.5 Acórdão

Conforme Guimarães (1994):

Para a delimitação conceitual do Acórdão no Direito Brasileiro, é mister recorrer ao Direito Positivo, especificamente, à definição estabelecida pelo art. 163 do Código de Processo Civil: Recebe a denominação de acórdão o julgamento proferido pelos tribunais. Dois elementos podem ser depreendidos, de pronto, dessa definição legal:

Sua natureza – julgamento;

Sua fonte geradora – tribunais.

Diante da definição de Guimarães, pode-se concluir que o acórdão é a sentença proferida pelo tribunal superior ou de segundo grau, no qual os desembargadores podem dar o provimento ao recurso, negar o provimento ou até mesmo dar-lhe provimento parcial, ou, em outras palavras, a sentença proferida pelo juiz da primeira instância pode, em uma instância superior, ser modificada, mantida ou modificada parcialmente.

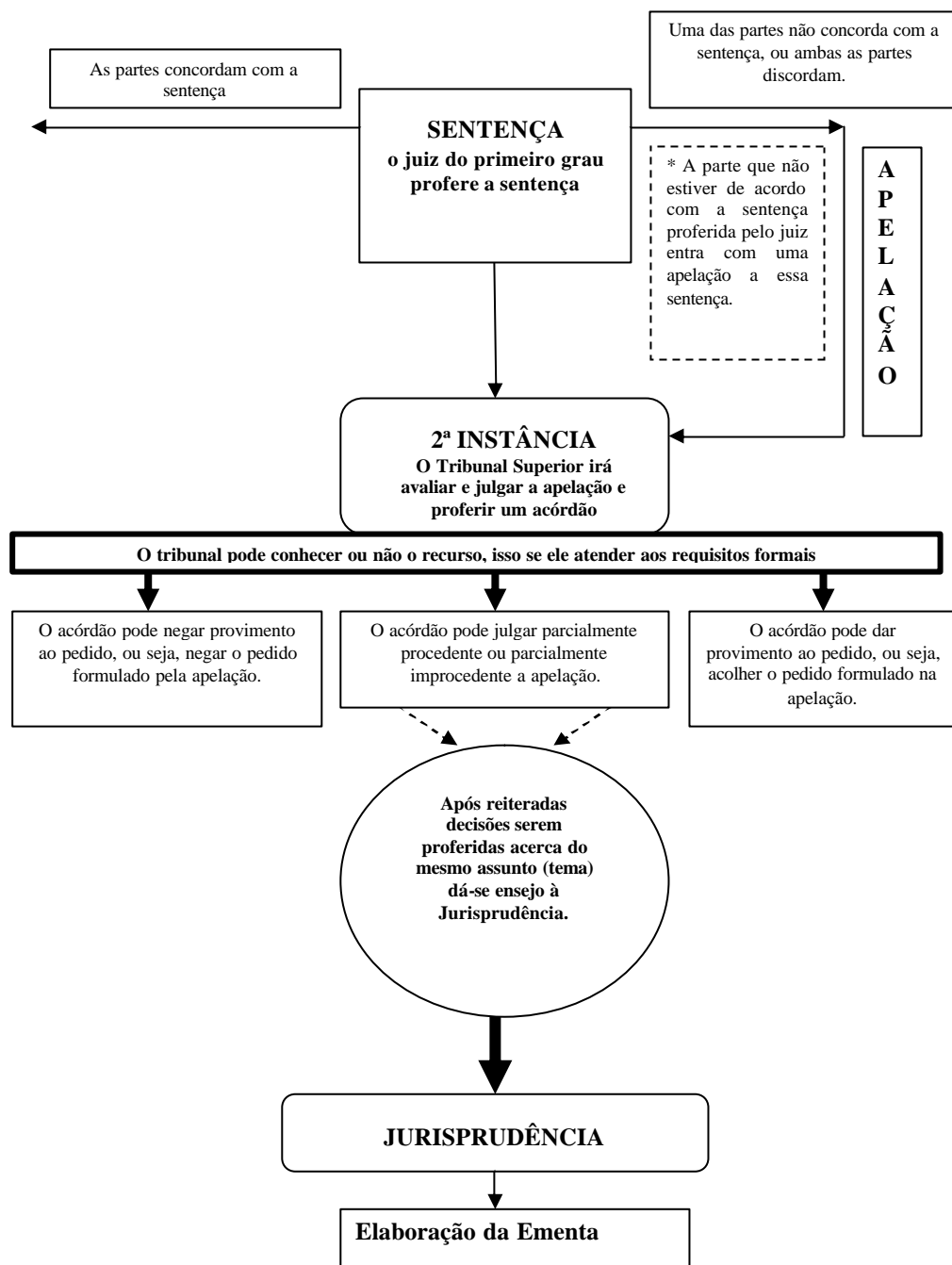
### 2.3.6 Jurisprudência

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. Modernamente, Jurisprudência é aplicada também no sentido de Ciência do Direito.

De acordo com Atienza (1979), a palavra *Jurisprudência* pode possuir significados diferentes: Pode indicar a Ciência do Direito, em sentido estrito,

também denominada Dogmática, Jurídica ou Jurisprudência; pode se referir ao conjunto de sentenças dos tribunais, em sentido amplo, e abranger tanto a Jurisprudência uniforme como a contraditória. E, em sentido estrito, Jurisprudência é apenas um conjunto de sentenças uniformes, assim, falamos em firmar a Jurisprudência ou contrariar a Jurisprudência.

A Jurisprudência consiste no modo pelo qual os tribunais se orientam na solução das diferentes questões. Expressa-se por meio das sentenças e acórdãos nas decisões. Essas decisões, quando tomadas em determinado sentido, passam a ser invocadas como precedentes a serem seguidos, muitas vezes, com força obrigatória. Maçoli (2005) elaborou um organograma, mostrado na Figura 2.1, a seguir, que ilustra o surgimento das jurisprudências e posteriores ementas.



**Figura 2.1 Organograma da jurisprudência**

Fonte: Adaptado de Maçoli, 2005.

Em outras palavras, quando não há uma lei específica determinando como proceder, a Jurisprudência serve como embasamento para o juiz estabelecer sua sentença baseada em casos similares já julgados.

A Jurisprudência dos tribunais pode ser unificada em um documento, como ocorre com a Súmula das Decisões do Supremo Tribunal Federal.

### **2.3.7 Súmula vinculante**

A palavra *súmula* é originária do latim *summula* que significa sumário, restrito.

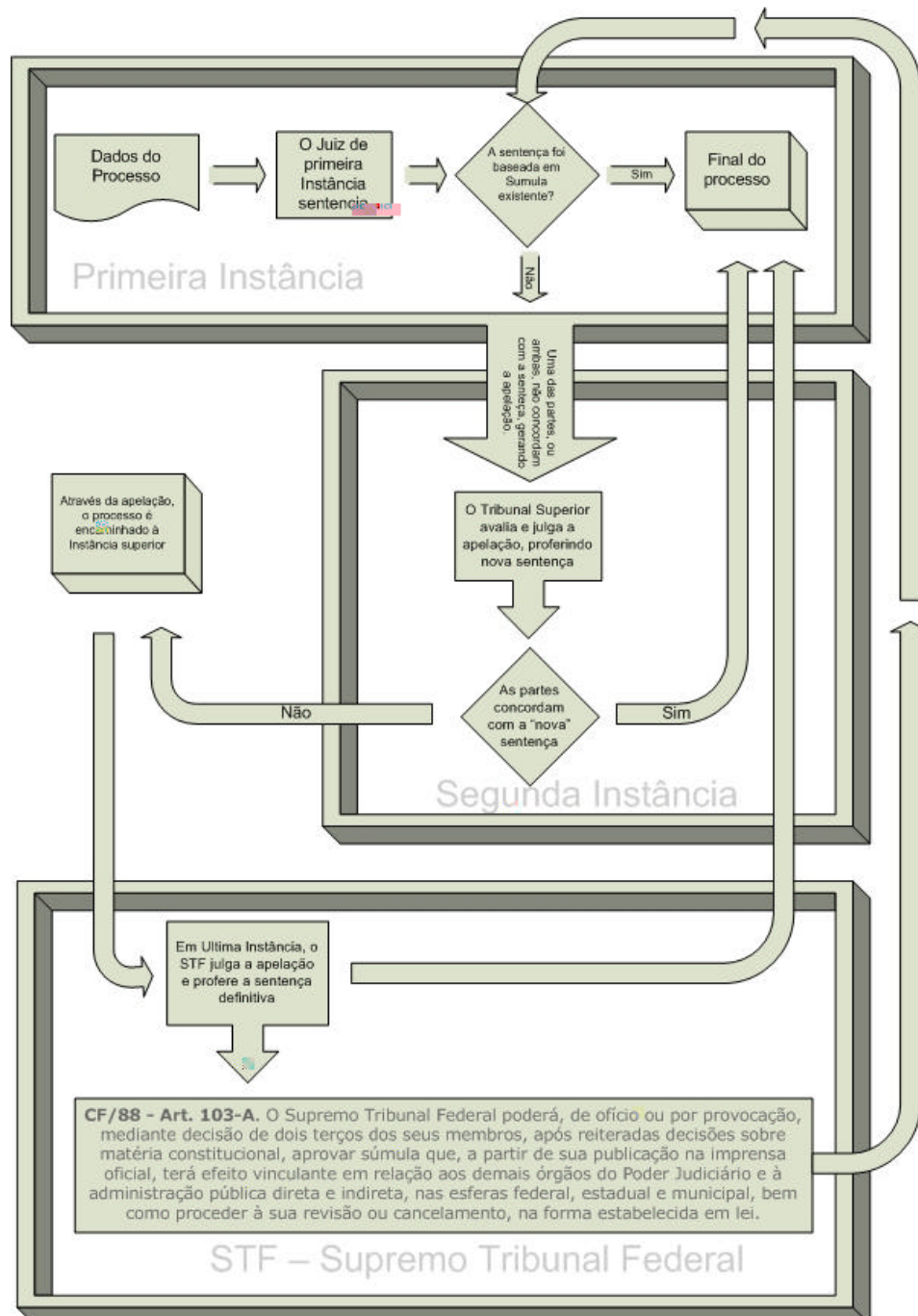
No direito brasileiro, significados e conceitos diferenciados têm sido atribuídos à palavra *súmula*, podendo, simplesmente, significar o resumo de um julgado enunciado pelo órgão julgador, ou a síntese da orientação jurisprudencial de um tribunal que é editada em numeração seqüencial (GUIMARÃES, 1999).

O efeito de vinculação, ou a súmula vinculante, foi um dos pontos discutidos pela reforma do Judiciário, aprovada em 17 de novembro de 2004, pelo Senado Federal e comentados por Soibelman (2005):

A Emenda Constitucional nº. 45, de 08.12.04, acrescentou à Constituição Federal CF/88 o art. 103-A, que institui a súmula de efeito vinculante, dispondo que o STF poderá, de ofício ou por provocação, mediante decisão de dois terços dos seus membros, após reiteradas decisões sobre matéria constitucional, aprovar súmula que, a partir de sua publicação na imprensa oficial, terá efeito vinculante em relação aos demais órgãos do Poder Judiciário e à administração pública direta e indireta, nas esferas federal, estadual e municipal, bem como proceder à sua revisão ou cancelamento, na forma estabelecida em lei.

Diante do novo cenário, pode-se dizer que a súmula vinculante traz a obrigatoriedade da utilização de certas jurisprudências/ ementas, que tinham até então caráter apenas orientativo. A Figura 2.2, a seguir, mostra um diagrama de

alto nível com as etapas pelas quais um processo pode passar, podendo ter sua decisão baseada em uma súmula existente ou mesmo originar uma nova, após tramite e aprovação pelo STF.



**Figura 2.2 Diagrama de alto nível da utilização / criação das súmulas**

Muitos são os argumentos referentes à utilização da súmula no direito brasileiro, porém cabe salientar que o foco deste trabalho não é a discussão sobre a validade jurídica ou a defesa de posições em relação às súmulas, e sim procurar estabelecer formas de auxiliar os processos de armazenamento e recuperação delas, utilizando-se para tanto, de técnicas apoiadas pela TI.

Um dos principais argumentos em prol da instituição das súmulas vinculantes no Brasil é a lentidão da Justiça. Este fato pode ser comprovado pelo enorme volume de processos que se acumulam em todas as instâncias dos tribunais (MORAES, 2002).

A vinculação, em tese, traz um nível de automação aos julgamentos, impedindo, por exemplo, que recursos sejam protocolados apenas com o intuito de retardar o procedimento judiciário. Tal efeito pode ser observado na Figura 2.2, que demonstra que decisões que tenham sido embasadas em súmulas já existentes, findam-se sem a possibilidade de que requerimentos de novos julgamentos sejam protocolados em instâncias superiores.

Moraes (2002) afirma ainda que, além da preocupação com a morosidade processual, a vinculação traz também a sensação de “segurança jurídica”, ou seja, o princípio da igualdade de todos perante a lei (art.5º, “caput”, CF/88).

Acredita-se também que a súmula vinculante torne-se um elemento de estabilidade que contribuirá para atrair investidores estrangeiros, que fogem de nosso país não só pela instabilidade econômica, mas também pela falta de previsibilidade jurídica (GOMES, 1997).



## **2.4 Informação e Direito**

A informação, cada vez mais, torna-se recurso fundamental na sociedade atual, atuando como ferramenta na solução de inúmeros problemas em diversas áreas do conhecimento. Neste tópico, conceituam-se sistemas de informação, conceito importante ao desenvolvimento do estudo, além de mostrar a importância da informação, especificamente para o domínio jurídico, mostrando as particularidades inerentes à sua recuperação.

### **2.4.1 Sistemas de Informação**

Segundo Laudon e Laudon (1999): “Um sistema de informação (S.I.) pode ser definido como um conjunto de componentes inter-relacionados trabalhando juntos para coletar, recuperar, processar, armazenar e distribuir informação com a finalidade de facilitar o planejamento, o controle, a coordenação, a análise e o processo decisório em empresas e outras organizações”.

Dentro do contexto deste estudo, os sistemas de informação definidos por Laudon e Laudon, (1999) terão a tarefa de recuperar informações textuais com a finalidade de facilitar o processo decisório dos usuários envolvidos.

### **2.4.2 Características dos documentos jurídicos textuais**

Apesar de textuais, os documentos jurídicos podem ser classificados como semi-estruturados, ou seja, apresentam suas informações de maneira que se possam identificar certas características.

Dados semi-estruturados trazem uma representação estrutural heterogênea, não sendo nem completamente não-estruturados nem estritamente tipados (MELLO et al. 2003).

Como exemplo dessa semi-estruturação pode-se citar a jurisprudência que, além do conteúdo textual, normalmente apresenta: relator, juiz, data do processo, local, entre outros (MATTOS, 2002), ou ainda o acórdão que, ademais do conteúdo textual, traz: fato, instituto jurídico, entendimento e o argumento (GUIMARÃES, 1994).

### **2.4.3 Pesquisa de documentos textuais**

Estudos recentes afirmam que 80% da informação de uma companhia estão contidos em documentos textuais. (TAN, 1999, apud LOH, 2001). No âmbito do Direito, nota-se também um grande volume de informações contidas em documentos textuais: jurisprudências, súmulas, ementas, sentenças, entre outros.

Encontrar informações presentes em textos pode se tornar uma tarefa árdua em razão do grande volume a ser pesquisado. A dificuldade vem habitualmente da heterogeneidade dos formatos e bases que contêm esses documentos, além dos sistemas utilizados para essas buscas que, de acordo com o método utilizado, podem produzir como resultado uma grande quantidade de referências, ou não retornar fontes que teriam utilidade.

Nem todas as palavras num texto não igualmente importantes para representá-lo semanticamente. Geralmente, substantivos (ou grupos de

substantivos) são mais representativos do conteúdo de um documento (FIGUEIREDO et al. 2003).

A alta heterogeneidade de formatos eletrônicos e bases distintas de armazenamento tornam complexa a atividade de pesquisa, uma vez que não existe um esquema uniforme a partir do qual uma consulta possa ser formulada. As consultas são, em geral, realizadas através de busca por palavras-chave, ficando a análise de relevância da informação basicamente por conta do usuário.

Uma melhor elaboração desses documentos em sua origem, somada às tecnologias de informação disponíveis atualmente, permitiria um tratamento muito mais efetivo e inteligente dessas grandes bases de dados.

Algumas empresas comercializam, atualmente, soluções voltadas à pesquisa de documentos jurídicos, inclusive na forma textual: Sistema *OLIMPO* – Tecnologia da Informação Jurídica para o conselho de segurança da ONU (HOESCHL e BARCIA, 2001); *ALPHATHEMIS* – Sistema para recuperação de súmulas baseado em Inteligência Artificial (COSTA et al., 2003), *IDOL K2* – Verity Division, entre outras. Nas soluções citadas, existem variações quanto ao mecanismo de trabalho, fazendo-se uso inclusive de técnicas de Inteligência Artificial como o RBC; que já foram objeto de análise de outros autores como Rissland (1995) apud Bueno (1999), "Raciocínio Baseado em Casos (RBC) é uma poderosa tecnologia da Inteligência Artificial para o desenvolvimento de sistemas inteligentes para a área jurídica".

### **3. Recuperação da informação**

Conforme Rowley (2002), os sistemas de recuperação da informação quase chegaram a ser sinônimos de computadores, mas não podemos nos esquecer de que esses sistemas em forma de papéis, fichas e anotações são os principais precursores dos computadores atuais, mesmo porque tais sistemas manuais também cumpriam seu papel de recuperar as informações.

Ainda de acordo com Rowley (2002), todos os sistemas de recuperação da informação podem ser compreendidos e estudados em três etapas:

- **Indexação:** é o processo de atribuir termos ou códigos de indexação a um registro ou documento. Termos ou códigos esses que serão úteis, posteriormente, na recuperação do documento ou registro. Tal indexação pode ser feita por meio manual, ou automático, feita por uma máquina ou computador, que, no entanto, somente pode selecionar termos de indexação de acordo com um conjunto de instruções. A seleção dependerá da ocorrência das palavras.
- **Armazenamento:** os sistemas de recuperação da informação utilizam o próprio computador para armazenar tanto os arquivos de documentos quanto os arquivos de índices, bem como a manutenção de base de dados.
- **Recuperação:** o processo de recuperação depende das etapas de indexação e armazenamento, as quais determinam, em grande medida, a melhor estratégia possível para as buscas feitas num sistema de recuperação da informação. O usuário e as consultas feitas ao sistema,

em geral, não sofrem alterações de um sistema para outro. As necessidades do usuário não se alteram devido à existência de determinados sistemas, ainda que possam se tornar mais elaboradas à medida que ele adquira mais experiência com o uso desses recursos.

É conveniente compreender que tanto o armazenamento das informações como seu tratamento e, conseqüentemente, sua recuperação sofreram mudanças, alterações, e até mesmo influência da tecnologia, que vem colaborando para o desenvolvimento e aprimoramento das ciências da informação.

Dessa forma, podemos descrever o processo de recuperação de informações em duas formas distintas:

- A primeira forma de recuperação da informação organizada por meio de sistemas manuais que consistem na utilização do catálogo de fichas, fichários e fichário decimal, todos eles executados e tratados manualmente pelo usuário.
- A segunda forma de recuperação da informação nos traz aos dias atuais, e aos recursos evidenciados pela trajetória histórica que a informática vem apresentando nos últimos anos e nos denominados sistemas automáticos de recuperação. (MAÇOLI, 2005)

### **3.1 Linguagens de indexação e busca**

Conforme Rowley (2002), define-se uma linguagem de indexação como sendo os termos ou códigos que podem ser usados como pontos de acesso num

índice. O propósito da elaboração de índices é possibilitar a recuperação das informações contidas no documento a ser indexado.

Em indexação trabalha-se com os conceitos expressos nos documentos pelos autores. Apesar de indexar palavras, estas têm que ser entendidas como rótulos que podem ser expressos de várias formas (LAAN et al., 2004).

Ainda citando Rowley (2002), compreende-se uma linguagem de busca como sendo os termos usados por um consulente ao especificar uma necessidade de busca. Se os termos ou código forem atribuídos por um indexador quando a base de dados é criada, então a linguagem de indexação estará sendo usada no momento de uso. Ainda que a linguagem de indexação possa ser diferente da linguagem de busca, é evidente que, para a recuperação tenha êxito, as duas devem estar intimamente relacionadas.

### **3.2 Modelos para recuperação de informação**

De acordo com Figueiredo et al. (2003), os modelos clássicos de recuperação de informação apresentam estratégias de busca de documentos relevantes para uma consulta. Tanto a consulta feita pelo usuário, quanto os documentos que compõem a coleção a ser pesquisada são representados pelos seus termos. Os modelos clássicos são: booleano, vetorial e probabilístico.

Ainda conforme Figueiredo et al. (2003), o modelo clássico mais utilizado em consultas a banco de dados é o booleano. Esse autor o descreve como um modelo clássico que considera uma consulta como uma expressão booleana convencional, que liga seus termos através de conectivos lógicos AND, OR e NOT. No modelo booleano, um documento é considerado relevante ou não

relevante a uma consulta, não existe resultado parcial e não há informação que permita a ordenação do resultado da consulta. Esse modelo é muito mais utilizado para recuperação de dados do que para recuperação de informação e torna-se interessante apenas para quem entende de álgebra booleana, o que normalmente não é o caso do usuário.

### **3.3 Consultas jurídicas**

Inicialmente, os operadores do Direito faziam suas consultas em livros. Eram necessárias muitas horas 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 devia ser constantemente atualizado. O custo para manter acomodações amplas, ou ampliá-las, passou a ser muito oneroso, dificultando a guarda do material de apoio (ATHENIENSE, 1996, apud MATTOS, 2002).

Com o advento da informática, surgiram sistemas de banco de dados que auxiliaram e, conseqüentemente, facilitaram as pesquisas 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 considerarem-no muito técnico.



**Figura 3.1 Evolução da consulta jurisprudencial**

Fonte: Adaptado de Mattos, 2002.

A figura 3.1 ilustra a evolução da consulta jurisprudencial comentada por Mattos (2002) partindo-se das consultas em livros e evoluindo até os sistemas de banco de dados e utilização da internet.



## **4. Técnicas para recuperação de documentos**

### **4.1 Considerações iniciais**

Busca-se aqui, através de levantamento bibliográfico, a caracterização de ferramentas para recuperação de informações no domínio jurídico, além de técnicas baseadas em Inteligência Artificial (IA) com a mesma finalidade.

Ao final do capítulo, mostra-se a aplicabilidade de uma técnica de IA, o raciocínio baseado em casos, dentro da área jurídica.

### **4.2 Mecanismos de busca**

Segundo a enciclopédia digital *Webopedia*, os mecanismos de busca (*search engines*) são: “Programas que pesquisam em documentos por palavras-chave especificadas e recuperam uma lista de documentos onde as palavras-chave foram encontradas”.

Embora ‘mecanismos de busca’ sejam uma classe geral de programas, o termo é frequentemente usado para descrever especificamente sistemas como *Google* e *AltaVista*, que permitem a usuários pesquisar por documentos na *World Wide Web* .

Segundo Alencar (2001), um mecanismo de busca trabalha enviando um robô (*spider*) para buscar o maior volume de documentos possível. Outro programa, chamado indexador, lê esses documentos e cria um índice baseado nas palavras contidas em cada um. Cada mecanismo usa um algoritmo próprio para

criar seu índice, de tal modo que, em condições ideais, só resultados significativos sejam recuperados.

Nas principais ferramentas, duas técnicas de recuperação são utilizadas no processo de busca: a indexação de termos ou palavras que, geralmente, é *full-text*, ou seja, todas as palavras dos documentos tornam-se disponíveis no índice, e o usuário recebe como resposta os textos que contêm as palavras fornecidas como entrada; e a catalogação de documentos, que ocorre quando alguma pessoa define o assunto explanado; o usuário, então, precisa escolher um entre os assuntos já predefinidos, para então receber os documentos relativos àquele tema (LOH et al., 2000).

Em algumas ferramentas, há uma linguagem própria para consulta, que utiliza conectivos e símbolos lógicos para eliminar documentos com determinados termos ou para recuperar somente aqueles que contenham obrigatoriamente certos termos.

Apesar da utilidade dessas ferramentas, alguns problemas podem ocorrer.

Boa parte dos usuários que utiliza os recursos de localização é inexperiente tanto no assunto que procura quanto no uso da ferramenta em si. Portanto, têm dificuldades em definir o contexto da informação de que necessitam utilizando palavras e conectivos (LOH et al., 2000).

Um estudo apresentado por Iivonen (1995) apud Loh et al. (2000) demonstra que aqueles que conhecem o funcionamento interno da ferramenta e possuem mais experiência com a linguagem de consulta (que é também específica da ferramenta) têm mais facilidade de encontrar informações úteis.

Além disso, as ferramentas que utilizam a técnica de indexação retornam grandes volumes de documentos sem a certeza de que a informação desejada se encontra em um deles. Isso acontece porque a técnica de indexação é baseada unicamente na presença de termos. Assim, podem ser retornados documentos que contêm as palavras fornecidas, mas que se referem a outro contexto, devido à possibilidade de as palavras terem vários significados. Outro problema é que poderão deixar de ser recuperados documentos relevantes para o assunto escolhido, justamente porque não possuem os termos fornecidos (LOH et al., 2000).

### **4.3 Data mining**

"Mineração de dados é a exploração e a análise, por meio automático ou semi-automático, de grandes quantidades de dados, a fim de descobrir padrões e regras significativos" (BERRY e LINOFF, 1997).

Os resultados obtidos com a mineração de dados podem ser usados no gerenciamento de informações, sendo indispensáveis na tomada de decisão, entre outras aplicações.

A mineração de dados pode ser aplicada de duas formas: como um processo de verificação e como um processo de descoberta (GROTH, 1998). No processo de verificação, o usuário sugere uma hipótese acerca da relação entre os dados e tenta prová-la aplicando técnicas como análises estatística e multidimensional sobre um banco de dados contendo informações passadas. No processo de descoberta, não é feita nenhuma suposição antecipada. Esse processo

usa técnicas, tais como descoberta de regras de associação, árvores de decisão, algoritmos genéticos e redes neurais.

O grande volume de dados disponíveis nas organizações cresce a cada dia e desafia a capacidade de armazenamento, seleção e uso dessas informações.

Os recursos da Tecnologia da Informação, principalmente no que diz respeito à capacidade do *hardware* e *software*, tornam possível efetuar em algumas horas atividades que, tradicionalmente, levariam meses.

A mineração de dados pode ser considerada como uma parte do processo de Descoberta de Conhecimento em Banco de Dados (*KDD – Knowledge Discovery in Databases*).

Segundo Goebel e Gruenwald (1999), o termo KDD é usado para representar o processo de tornar dados de baixo nível em conhecimento de alto nível, enquanto mineração de dados pode ser definida como a extração de padrões ou modelos de dados observados.

A mineração de dados combina métodos e ferramentas das seguintes áreas: aprendizagem de máquina, estatística, banco de dados, sistemas especialistas e visualização de dados, conforme Figura 4.1 (CRATOCHVIL, 1999).



**Figura 4.1 Mineração de dados como um campo multidisciplinar**

Fonte: Cratochvil, 1999.

Outra definição muito utilizada para *Data Mining* foi elaborada por Usama Fayyad (FAYYAD et al., 1996): "... o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e, ultimamente, compreensíveis".

#### **4.4 Técnicas de Inteligência Artificial (IA)**

Dentre os vários conceitos de IA existentes, optou-se, neste trabalho, por aqueles que tratam os sistemas de IA como aqueles que "pensam como seres humanos"; como Durkin, (1994) apud Modro, (2000), "Inteligência Artificial é o campo de estudo na ciência da computação que persegue o objetivo de fazer um computador raciocinar de maneira semelhante aos humanos", ou ainda Arariboia (1988), "a Inteligência Artificial é um campo que usa técnicas de programação que procuram, por intermédio de máquinas, resolver problemas do mesmo modo que um ser humano os resolveria", e Nikolopoulos (1997):

Inteligência Artificial é um campo de estudos multidisciplinar, originado da computação, da engenharia, da psicologia, da matemática e da cibernética, cujo principal objetivo é construir sistemas que apresentem comportamento inteligente e desempenhem tarefas com um grau de competência equivalente ou superior ao grau com que um especialista humano as desempenharia.

A Inteligência Artificial, como o próprio nome sugere, permite ao computador "pensar", simplificando a maneira como os programas são construídos, diminuindo assim o número de alterações e correções. A IA tende a imitar o processo de aprendizado humano, no qual novas informações podem ser incorporadas sem intervir em todos os outros fatos previamente armazenados. Nesse contexto, cabe citar O'Brien (2001) que afirma que a Inteligência Artificial está retornando à corrente principal da tecnologia nas organizações. Concebida mais para fomentar as capacidades humanas do que para substituí-las.

De acordo com Moscove (2002), existem vários campos de estudo dentro da IA com o propósito de dotar a máquina de capacidade de raciocínio e aprendizado como robótica, visão e reconhecimento de voz, processamento de linguagem natural e sistemas de solução de problemas, como sistemas especialistas em redes neurais e sistemas de raciocínio baseados em caso.

Ganascia (1993) apud Valle (2003) considera como os principais modelos de IA: os algoritmos genéticos, a programação evolutiva, os sistemas difusos, os sistemas baseados em conhecimento, a programação genética, as redes neurais e o raciocínio baseado em casos.

#### **4.4.1 Algoritmos Genéticos**

Algoritmos Genéticos, AG's, é um modelo para o aprendizado da máquina, inspirado no livro *Origem das espécies, através da seleção natural*, escrito pelo naturalista inglês Charles Darwin (DARWIN, 1859), criador da teoria evolucionista, segundo a qual somente os mais aptos sobrevivem. Foram criados por John Holland (1975), que utilizou o conceito de cromossomos como estrutura básica do indivíduo, representado computacionalmente por *strings* binários (0s e 1s). Koza (1992) observou que outras estruturas de dados poderiam ser evoluídas, como as árvores. Nasceu assim a Programação Genética.

De acordo com Minku et al. (2002), são muitas as utilizações dos AGs, todas relacionadas de uma forma ou de outra a uma análise multidimensional, com a qual se busca conseguir uma solução global. Geralmente, o material genético é utilizado para codificar os valores dos vários parâmetros que definem o espaço de resultados admissíveis, procurando-se encontrar o valor desses parâmetros (valor dos cromossomos), que solucionam um certo problema de otimização.

Os Algoritmos Genéticos vêm sendo utilizados em várias áreas de pesquisa e em situações do mundo real:

- Na evolução musical – Apresentado em 1999 na CEC99 – IEEE – *International Conference on Evolutionary Computation* – um ambiente interativo, utilizando Algoritmos Genéticos, para a avaliação de músicas (seqüências de acordes) tocadas em arquivos MIDI. O método emprega o formalismo difuso e é apresentado como uma otimização baseada em fatores relevantes à audição de músicas. No caso, os indivíduos da população foram definidos em

grupos de quatro vozes (soprano, contralto, tenor e baixo) ou coros. Cada um é avaliado segundo três critérios: melodia, harmonia e oitavas. A composição desses três critérios define a aptidão (*fitness*) estabelecida pela função de seleção, que retorna o melhor indivíduo ou melhor coro. Um ciclo genético é operacionalizado, criando novos indivíduos a partir dos anteriores e procurando sempre pelo melhor. Quando um novo grupo é selecionado, ele é tocado em MIDI. A duração do ciclo genético determina o ritmo da evolução. O sistema criado foi denominado *Vox Populi*. (FUKUSHIMA, 1999 apud FIALHO, 1999)

- Melhoria em telecomunicações – Segundo Blanchard (1994) apud Fialho (1999), o último WCCI'94 – *World Congress on Computational Intelligence* – ocorrido em Orlando, na Flórida, mostrou uma série de possíveis soluções, bastante promissoras, para situações reais utilizando Algoritmos Genéticos. Blanchard mostrou o caso da *US West*, uma companhia regional de telecomunicações do estado do Colorado, Estados Unidos, que vem usando um sistema baseado em AGs que possibilita projetar, em duas horas, redes ópticas especializadas, trabalho que levaria **seis meses** utilizando especialistas humanos. O sistema produz resultados ainda 10% (dez por cento) melhores que os realizados pelo homem.
- Outras aplicações – A AIS (Barcelona, Espanha) utilizou um sistema apoiado em AGs e Sistemas Especialistas (SEs) para programar os Jogos Paraolímpicos de 1992, já que nas Olimpíadas os atletas são



organizados em duas grandes classes, masculino e feminino, e os competidores paraolímpicos são divididos em mais de 100 (cem) subclasses, segundo certas deficiências.

- Um sistema em construção na *New Mexico State University* descreve imagens faciais de criminosos a partir de testemunhas do crime, utilizando AGs. O sistema tem se mostrado mais efetivo na produção de retratos falados de criminosos do que qualquer outra técnica de obtenção de informação de imagens.
- Syswerda e Palmucci (1991) apud Fialho (1999) relataram a execução de um otimizador para uma aplicação prática de programação de recursos no laboratório SITS – *System Integration Test Station Laboratory* – da Marinha Americana, para o desenvolvimento do jato F-14.

Minku et al. (2002) esclarece ainda que os AG's são apropriados para problemas de otimização complexos, que envolvem muitas variáveis e um espaço de soluções de dimensão elevada.

#### **4.4.2 Programação evolutiva**

A Programação Evolutiva é o campo da IA concebido por Lawrence J. Fogel. Assemelha-se aos Algoritmos Genéticos, sendo dada maior ênfase na relação comportamental entre os parentes e seus descendentes. As soluções para os problemas são obtidas por meio de tentativas e transmitidas para a nova população (simulada em programas). O método se caracteriza pela orientação a

um objetivo. Em outras palavras, este método seleciona a melhor saída para um objetivo predeterminado (*Top-Down*).

Nesse processo, o algoritmo seleciona os "filhos" (que foram gerados através de reprodução e, sobretudo repetitivas mutações) com melhores performances e os compara com seus "pais". Se o "filho" for superior ao "pai", ele o substitui, senão o "filho" é eliminado e o processo se repete.

Posteriormente, o processo "filho", que foi superior ao "pai", é comparado com o melhor indivíduo da população e se obtiver uma performance melhor ou igual a este "melhor indivíduo", uma nova geração forte é criada, e suas informações genéticas são armazenadas para que distorções ambientais inesperadas não o modifiquem.

Todas essas performances avaliadas significam, obviamente, adequação ao ambiente em que se insere a população e é assim que os algoritmos da Programação Evolutiva conseguem se adaptar a uma situação, ou escolher a melhor saída para uma problemática.

#### **4.4.3 Sistemas Difusos ou Lógica Fuzzy**

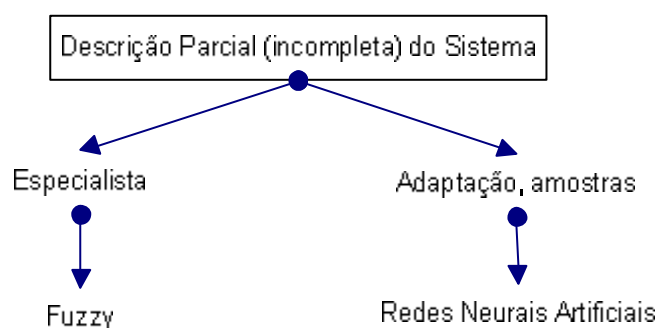
Os Sistemas Difusos ou Lógica Fuzzy foram estruturados por Lofti Zadeh, da Universidade da Califórnia, no ano de 1965. É uma metodologia que serve para representar, manipular e modelar informações incertas.

A característica especial da Lógica Fuzzy é representar uma forma inovadora no manuseio de informações imprecisas, de modo distinto da teoria de probabilidade. A Lógica Fuzzy provê um método de traduzir expressões verbais, imprecisas e qualitativas, comuns na comunicação humana em valores numéricos

(MARÇAL e SUSIN, 2005). O sucesso, atualmente reconhecido, de sistemas de modelagem e controle baseados em Lógica Fuzzy, em aplicações industriais, tem comprovado sua utilização como mais de uma ferramenta (ou tecnologia) para a resolução de problemas de engenharia de controle industrial, manufatura, manutenção, comunicação homem-máquina e em sistemas de tomadas de decisão (MARÇAL e SUSIN, 2005).

A Lógica Fuzzy permite criar sistemas especialistas utilizando variáveis lingüísticas para criar uma base de regras. Expressões lingüísticas são típicas da natureza humana de tomar decisões. Por exemplo: "Se estiver quente vou ligar o ar condicionado no máximo". Quente e máximo não significam um valor particular de temperatura e potência, mas podem assumir uma faixa considerável de valores. Pessoas diferentes também podem ter diferentes aceções para o mesmo conceito lingüístico (BAUCHSPIESS, 2004).

De acordo com a disponibilidade de um especialista ou de amostras de um sistema, o paradigma Fuzzy ou RNA é mais indicado, conforme ilustrado por Bauchspiess (2004), pelo que demonstra a Figura 4.2:



**Figura 4.2 Escolha do paradigma Fuzzy ou RNA**

Fonte: Adaptado de Bauchspiess, 2004.

Em aplicações de engenharia, trabalha-se com números (temperatura, pressão, força etc.), as variáveis são contínuas. Por outro lado, a inferência Fuzzy utiliza variáveis lingüísticas.

#### **4.4.4 Sistemas Baseados em Conhecimento**

Os Sistemas Baseados em Conhecimento implementam comportamentos inteligentes de especialistas humanos. Desse modo, são classificados também como sistemas especialistas (SE).

Os sistemas baseados no conhecimento foram alvos de várias pesquisas em Inteligência Artificial realizadas com sucesso. Esses sistemas são baseados em um modelo explícito de conhecimento destinado a solucionar problemas (PY, 2003).

Sistemas baseados em conhecimento (SBC) ou *knowledge-based systems* são sistemas que aplicam mecanismos automatizados de raciocínio para a representação e inferência de conhecimento. Esses sistemas costumam ser simplesmente identificados como de "Inteligência Artificial aplicada" e representam uma abrangente classe de aplicações da qual todas as demais seriam aproximadamente subclasses (PY, 2003).

Sistemas especialistas são sistemas desenvolvidos para conter em si o conhecimento de um ou mais especialistas, ou seja, são sistemas projetados para solucionar problemas e realizar tarefas simulando a tomada de decisão de especialistas. Segundo Hubert Dreyfus: "O nosso sistema especializado médico 'MYCIN', por exemplo, é superior aos especialistas da Escola de Medicina da Universidade de Stanford" (PESSIS-PASTERNAK, 1993, apud LIMA, 2005).

Dentre os programas que utilizam a tecnologia dos SE, podem-se citar:  
(LIMA, 2005)

- DENDRAL – Foi o primeiro sistema especialista a surgir (1971), criado por Edward Feigenbaum na Universidade Stanford. Serve para detectar a estrutura molecular de um componente orgânico, sendo de grande ajuda na química e na medicina.
- MYCIN (SHORTLIFFE, 1976) – É um programa que tem na sua base de conhecimento dados inseridos por diversos especialistas da área da saúde. Serve para realizar consultas, solicitar exames e diagnosticar doenças como a meningite e infecções do sangue.
- PROSPECTOR (HART et al., 1978) – É um programa que contém na sua base de conhecimento informações sobre mapas geológicos e levantamento de terrenos. Serve para auxiliar na localização de veios de minerais.

Segundo Machado e Borges (2005), algumas vantagens podem ser elencadas na utilização dos SE, entre elas: redução de falhas humanas, estabilidade, flexibilidade, aumento de desempenho e invulnerabilidade a questões psicológicas, estresse e fatores externos. Ainda segundo esses autores, algumas dificuldades podem ser apresentadas na utilização dos SE como: Ausência de metacconhecimentos, dificuldade no processo de aquisição de conhecimento e falta de conhecimentos genéricos.

#### **4.4.5 Programação Genética**

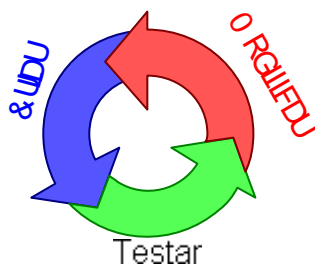
A Programação Genética é um campo de estudo da IA voltado para a construção de programas que visam imitar o processo natural da genética.

O paradigma da Programação Genética foi desenvolvido por John Koza (KOZA, 1992) com base nos trabalhos de John Holland em Algoritmos Genéticos (HOLLAND, 1975). Atualmente, representa uma área muito promissora de pesquisa em Inteligência Artificial devido à sua simplicidade e robustez. Seu uso tem sido estendido a problemas de diversas áreas do conhecimento, por exemplo: biotecnologia, engenharia elétrica, análises financeiras, processamento de imagens, reconhecimento de padrões, mineração de dados, linguagem natural, entre muitas outras (WILLIS et al., 1997).

A Programação Genética é a evolução de um conjunto de programas com o objetivo de aprendizagem por indução (BANZHAF et al., 1998). A idéia é ensinar computadores a se programar, isto é, a partir de especificações de comportamento, o computador deve ser capaz de induzir um programa que as satisfaça (KOZA, 1992). A cada programa é associado um valor de mérito (*fitness*) representando o quanto ele é capaz de resolver o problema.

Basicamente, a Programação Genética mantém uma população de programas de computador, usa métodos de seleção baseados na capacidade de adaptação de cada programa (escolha dos “melhores”), aplica operadores genéticos para modificá-los e convergir para uma solução. O objetivo é encontrar uma solução no espaço de todos os programas possíveis (candidatos) usando apenas um valor de *fitness* como auxílio no processo de busca (GATHERCOLE, 1998).

O mecanismo de busca da Programação Genética pode ser descrito como um ciclo "criar-testar-modificar", representado na Figura 4.3, muito similar à forma com que os humanos desenvolvem seus programas. Inicialmente, programas são criados com base no conhecimento sobre o domínio do problema. Em seguida, são testados para verificar sua funcionalidade. Se os resultados não forem satisfatórios, modificações são feitas para melhorá-los. Esse ciclo é repetido até que uma solução satisfatória seja encontrada ou um determinado critério seja satisfeito (YU, 1999).



**Figura 4.3 Ciclo criar-testar-modificar**

Fonte: Adaptado de Yu, 1999.

A especificação de comportamento é feita normalmente através de um conjunto de valores de entrada-saída, denominados *fitness cases*, representando o conjunto de aprendizagem ou treinamento (*training set*). Com base nesse conjunto, a Programação Genética procura obter um programa que, de acordo com O'Reilly (1995), cumpra o seguinte:

- Produza, de forma não trivial, as saídas corretas para cada entrada fornecida. Isto implica que o programa não deve mapear as entradas e saídas através de alguma forma de tabela de conversão. Portanto, o programa deverá aprender necessariamente alguma forma de algoritmo;

- Calcule as saídas de tal forma que, se as entradas forem representativamente escolhidas, o programa será capaz de produzir saídas corretas para entradas não cobertas inicialmente.

A especificação do domínio do problema é feita simplesmente pela definição dos conjuntos de funções e terminais (KOZA, 1992).

#### **4.4.6 Redes neurais**

As redes neurais possuem várias denominações, dentre elas redes neuronais, modelo conexionista, neurocomputação, modelo de processamento paralelo distribuído, sistemas neuromórficos e computadores biológicos. As redes neurais artificiais são definidas como redes massivamente paralelas e interconectadas, de elementos simples, com organização hierárquica. Esses elementos devem interagir com objetivos do mundo real, da mesma maneira que o sistema nervoso biológico (KOHONEN, 1972).

A tentativa de simular a rede neural do cérebro deu origem à chamada Rede Neural Artificial. Os trabalhos sobre redes neurais se iniciaram, na década de 1940, com a publicação de um artigo que estabeleceu uma analogia entre o processo de comunicação das células nervosas vivas e o processo de comunicação por transmissão elétrica, e propuseram a criação de neurônios formais.

As RNA têm como principal objetivo simular a capacidade de aprendizado e a capacidade de generalização do cérebro humano, podendo executar tarefas que os programas convencionais não conseguiam realizar, pois eles não têm a capacidade de aprendizado. A primeira rede capaz de imitar o cérebro humano utilizando processadores paralelos (em vez de um único processador), surgiu em



1959 com a estruturação da Rede ADALINE (*Adaptative Linear Element*). Mais tarde essa rede foi reestruturada e denominada MADALINE (*Many ADALINE*).

Atualmente, existem alguns modelos de redes neurais estruturados para as mais diversas aplicações, como: análise de assinaturas, análise de características demográficas para marketing, análise do grau de satisfação de um cliente, controle de processos industriais, monitoramento para manutenção de motores, previsão da bolsa de valores e cotação de moedas.

Krose (1993) apud Almeida (2000) define como principais características de uma rede neural artificial: arquitetura, capacidade de aprendizado e habilidade funcional, e cita vantagens como: inferência de múltiplas variáveis, grande tolerância a falhas, modelamento direto do problema e paralelismo inerente. Ainda de acordo com Krose, 1993 (apud ALMEIDA, 2000), uma das dificuldades das redes neurais artificiais é que trabalham com um número reduzido (centenas) de neurônios artificiais, enquanto as redes neurais biológicas trabalham com milhões de neurônios.

#### **4.4.7 Raciocínio baseado em casos – RBC**

"Eu prevejo que nos próximos 20 anos, RBC será a mais importante aplicação de IA" (MINSKY, 1991, apud WANGENHEIM, 2003).

Segundo Watson e Marir (1994) apud Abel (1996), os estudos de Raciocínio Baseado em Casos, doravante identificado por RBC, tiveram seu marco inicial no trabalho de Schank e Abelson (1977), que propôs que nosso conhecimento geral sobre as situações é registrado na forma de *scripts*. Os *scripts* descrevem seqüências de passos ou etapas que nos permitem antecipar como os

acontecimentos devem se suceder e realizar inferências a partir dessa expectativa. Os *scripts* são propostos como uma estrutura para a memória conceitual descrever informação sobre eventos estereotipados, como ir a um restaurante ou visitar um médico. Os experimentos mostraram, no entanto, que *scripts* não podiam ser considerados um modelo completo de representação da memória, uma vez que uma pessoa pode "lembrar-se" de *scripts* que são compostos por pedaços de diferentes seqüências de eventos, ou seja, lembrar-se de algo que na verdade foi criação de sua mente. Outros mecanismos atuam na formação da memória e são explicados por teorias da Filosofia e Psicologia de solução de problemas, formação de conceitos e aprendizado experimental. Embora as raízes filosóficas da teoria de RBC possam ser encontradas em diversos trabalhos de vários pesquisadores no campo da Psicologia ou Ciência da Computação, foram os trabalhos do grupo de Schank, na Universidade de Yale, no início dos anos 1980, que produziram o modelo do RBC e as primeiras aplicações baseadas nesse modelo. O grupo de Yale desenvolveu diversos sistemas de RBC, como o MEDIATOR, (SIMPSON, 1985), o CHEF, (HAMMOND, 1986) e o PERSUADER (SYCARA, 1987), entre outros.

Em 1986, um trabalho alternativo foi desenvolvido na Universidade do Texas, utilizando recursos de classificação heurística e aprendizado de máquina, para unificar em um modelo o conhecimento genérico do domínio e o conhecimento específico de casos. O resultado foi o sistema PROTOS com seu modelo de memória de casos (PORTER e BAREISS, 1986).

O desenvolvimento desses trabalhos e conceitos evoluiu para inúmeras aplicações de sistemas baseados em casos, em vários domínios como Direito,

Medicina e Engenharia. As aplicações, normalmente, buscam resolver problemas de classificação, projeto, diagnóstico ou planejamento, especialmente em domínios nos quais, naturalmente, o especialista utiliza casos anteriores como base para a solução do problema. Dessas aplicações, o Direito tem mostrado os resultados mais notáveis até o momento (ABEL, 1996).

O Raciocínio Baseado em Casos é o campo de estudo da IA que utiliza uma grande biblioteca de casos para consulta e resolução de problemas. Os problemas atuais são resolvidos através da recuperação e consulta de casos já solucionados e da conseqüente adaptação das soluções encontradas. Como exemplo, pode-se citar o Sistema CASEY, que faz o diagnóstico em pacientes cardíacos baseado na consulta de arquivos de pacientes com o mesmo diagnóstico (GRESSE, 2002, apud VALLE, 2003).

O RBC é um tipo de raciocínio que busca soluções para um determinado problema mediante a análise comparativa da realidade anterior e uma outra semelhante apresentada (BUENO, 1999).

É prática comum do ser humano basear-se em situações passadas e experiências semelhantes para tomar decisões e resolver problemas, além de aprender e adquirir conhecimento com cada nova situação.

Raciocínio Baseado em Casos é um enfoque para a solução de problemas e para aprendizado baseado em experiências passadas. Resolve problemas ao recuperar e adaptar experiências passadas, chamadas casos, armazenadas em uma base de casos (WANGENHEIM, 2003).

Uma diferença muito importante entre sistemas de RBC e sistemas que utilizam outras técnicas de IA é sua capacidade de aprendizado. Utilizando

critérios de similaridade ou analogia, os problemas são resolvidos com base em situações que aconteceram e, após sua resolução, são armazenados e constituem base para futuras soluções. (CARVALHO et al., 2005). Em uma conotação mais prática do funcionamento do RBC, podemos utilizar um exemplo cotidiano: Um médico recebe em seu consultório um paciente com os seguintes sintomas: febre alta, congestão nasal e dores pelo corpo. Utilizando-se de conhecimentos prévios, adquiridos na faculdade e/ou em sua vida profissional até o momento, o médico compara os sintomas e faz o diagnóstico, por exemplo, de gripe, baseado nas similaridades encontradas. Até esse ponto, não existe a certeza do diagnóstico, porém em caso negativo, este será um novo caso armazenado pelo médico para uso em outro caso futuro. Este exemplo mostra que, pelo fato de trabalhar com similaridades, a solução de um problema é possível, mesmo que não seja exatamente igual a um já ocorrido e solucionado.

De acordo com Wangenheim (2003), para se criar um sistema de RBC, são necessários alguns elementos básicos:

- Representação do Conhecimento
- Medida de Similaridade
- Modelo de Adaptação
- Modelo de Aprendizado

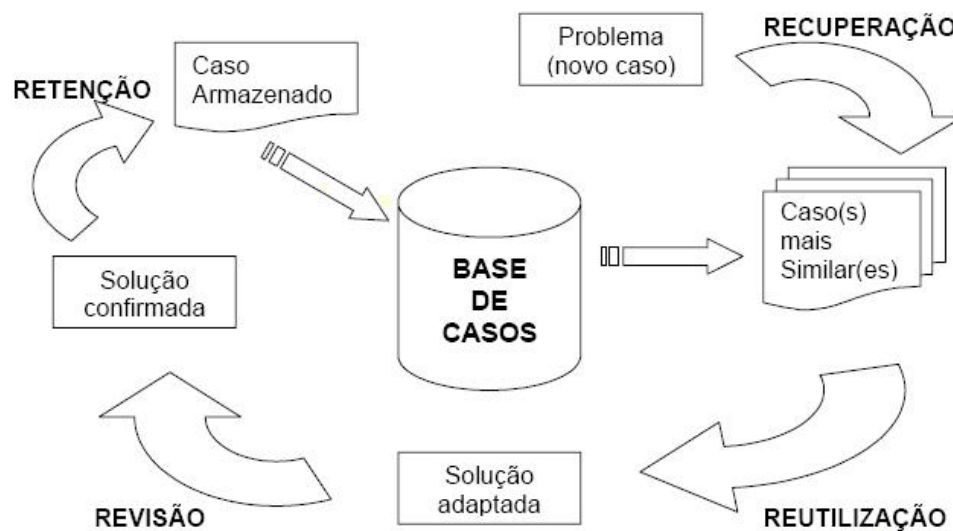
Em um sistema de RBC, o conhecimento é representado por casos que descrevem situações, as experiências passadas. De maneira simples, um caso descreve o que aconteceu, o problema e como foi resolvido. Assim, a estrutura básica de um caso compõe-se de: descrição do problema e descrição da solução, ou das soluções do problema. O sistema de RBC deve levantar, então, o grau de

similaridade dos casos, ou seja, estabelecer o nível de semelhança entre a situação atual e alguma já armazenada num banco de casos. A medida de similaridade indica a utilidade de um caso para a resolução de um problema, ou seja, o quanto ele é útil para ser utilizado.

A medida de similaridade a ser utilizada por um sistema de RBC vai depender do domínio a ser tratado, porém baseia-se na premissa de que "problemas similares possuem soluções semelhantes" (WANGENHEIM, 2003).

Difícilmente, uma situação ou problema é exatamente igual a uma situação anterior, por isso, por mais similar que um caso seja a outro, este pode não resolver completamente o problema atual apresentado. Surge assim a necessidade de uma adaptação da solução para que satisfaça adequadamente os requisitos da situação atual.

No sistema de RBC, sempre que um caso utilizado é adaptado à situação atual, resolvendo um problema, este é acrescentado na base de casos, como um novo instrumento para consultas futuras. A aprendizagem faz com que o sistema evolua continuamente, aumentando sua base de conhecimento e, dessa maneira, tornando-o cada vez mais eficiente, o que pode ser ilustrado na Figura 4.4 (WANGENHEIM, 2003, apud CARVALHO et al., 2005):



**Figura 4.4 Ciclo do RBC**

Fonte: Wangenheim, 2003, apud Carvalho et al., 2005.

A Figura 4.4 ilustra o que foi descrito anteriormente, ou seja, a partir de um problema ou novo caso, faz-se a recuperação dos casos mais similares provindos da base de casos. Depois de feita a seleção do caso que mais se encaixa à nova situação, este é analisado para a reutilização como solução do problema proposto. Como, de modo geral, o caso não é exatamente igual ao problema proposto, surge a necessidade de adaptação da solução. Após a validação da nova solução, esta é armazenada como um novo caso, aumentando assim o leque para próximas consultas.

Salienta-se ainda que um sistema de RBC pode e, certamente, deve armazenar também experiências negativas, falhas, para que não sejam cometidas novamente, aumentando assim a precisão de novas consultas. O armazenamento das falhas ou o grau da adaptação para a solução do novo caso pode ser baseado em intervenção humana no sistema.

## **4.5 Resumo das técnicas apresentadas para uso na recuperação de documentos textuais jurídicos**

<b><i>Search Engines</i></b>	
Programas que pesquisam em documentos por palavras-chave especificadas e recuperam uma lista de documentos nos quais as palavras-chave foram encontradas.	Possui sua utilização em vários sites de Tribunais para fins de pesquisas; Como a busca é baseada em palavras-chave e operadores lógicos, acabam gerando dificuldade na utilização e um grande numero de informações, o que dificulta o foco da pesquisa.
<b><i>Data Mining</i></b>	
Mineração de dados é a exploração e a análise, por meio automático ou semi-automático, de grandes quantidades de dados, a fim de descobrir padrões e regras significativos.	A técnica se baseia normalmente em análises estatísticas e descoberta de padrões; Por essas razões, acaba por não apresentar características interessantes dentro do domínio estudado.
<b><i>Algoritmo Genético</i></b>	
Modelo para o aprendizado da máquina. Utiliza o conceito de cromossomos como estrutura básica do indivíduo, representado computacionalmente por <i>strings</i> binários (0s e 1s)	Podem ser utilizados para geração e adaptação de casos, com busca na evolução de soluções anteriores
<b><i>Programação Evolutiva</i></b>	
Assemelha-se aos algoritmos genéticos, sendo dada maior ênfase na relação comportamental entre os parentes e seus descendentes. As soluções para os problemas são obtidas por meio de tentativas e transmitidas para a nova população (simulada em programas).	Semelhante ao algoritmo genético

**Quadro 4.1** Resumo das técnicas apresentadas

(continuação)

<b>Lógica Fuzzy</b>	
<p>E uma metodologia que serve para representar, manipular e modelar informações incertas.</p> <p>A característica especial da lógica Fuzzy é representar uma forma inovadora no manuseio de informações imprecisas, de forma distinta da teoria de probabilidade.</p>	<p>Pode ser utilizada para verificar o grau de certeza de um determinado caso buscado, baseado em regras de consulta, fornecendo soluções aproximadas, porém aceitáveis.</p>
<b>Sistemas Baseados em Conhecimento</b>	
<p>Os Sistemas Baseados em Conhecimento implementam comportamentos inteligentes de especialistas humanos.</p> <p>Sistemas baseados em conhecimento são sistemas que aplicam mecanismos automatizados de raciocínio para a representação e inferência de conhecimento.</p>	<p>Pode ser utilizado para implementar facilitadores e autonomia aos sistemas jurídicos, como assistentes dos usuários, nas formas de agentes de procura, agentes de apresentação ou ainda interfaces tutoriais.</p>
<b>Programação Genética</b>	
<p>Voltado para a construção de programas que visam imitar o processo natural da genética.</p> <p>A idéia é ensinar computadores a se programar, isto é, a partir de especificações de comportamento, o computador deve ser capaz de induzir um programa que as satisfaça.</p>	<p>Por se tratar de “<i>auto-programação</i>”, não se aplica no contexto de desenvolvimento aqui proposto.</p>

**Quadro 4.1 Resumo das técnicas apresentadas**



(continuação)

<b>Redes Neurais</b>	
As redes neurais artificiais são definidas como redes massivamente paralelas e interconectadas, de elementos simples, com organização hierárquica. Esses elementos devem interagir com objetivos do mundo real, da mesma maneira que o sistema nervoso biológico.	Podem ser treinadas para análise de características semelhantes em uma base de casos, com o intuito de melhor adaptação e ou estatística.
<b>RBC</b>	
O Raciocínio Baseado em Casos é o campo de estudo da IA que utiliza uma grande biblioteca de casos para consulta e resolução de problemas. Os problemas atuais são resolvidos através da recuperação e consulta de casos já solucionados e da conseqüente adaptação das soluções encontradas.	Bastante utilizada, com vários exemplos de utilização como o <i>Judge</i> já citado, <i>Ontojuris</i> (consulta de jurisprudências de alguns Tribunais), <i>AlphaThemis</i> (sistema para pesquisa contextual em documentos como jurisprudências e súmulas), <i>Olimpo</i> (busca textual em resoluções da ONU), entre outros.

#### **Quadro 4.1 Resumo das técnicas apresentadas**

##### **4.5.1 RBC na área jurídica**

Os sistemas baseados em RBC estão presentes também na área jurídica e já foram explorados por outros autores como Rissland (1995), "Raciocínio Baseado em Casos (RBC) é uma poderosa tecnologia da Inteligência Artificial para o desenvolvimento de sistemas inteligentes para a área jurídica"; e Bueno (1999), que o considerou como uma técnica muito adequada ao domínio jurídico, pois utiliza o mesmo tipo de raciocínio utilizado pelos juristas na solução de um problema.

Dentre os sistemas já criados no Brasil nessa área, podem-se citar: JurisConsulta, Olimpo e AlphaThemis, desenvolvidos por pesquisadores do Instituto Jurídico de Inteligência e Sistemas (IJURIS), da Universidade Federal de Santa Catarina.

#### **4.5.2 RBC e outras técnicas de IA**

Até este ponto foram abordadas, de forma isolada, técnicas de IA associadas a aplicativos específicos, inclusive com o RBC surgindo como uma forte alternativa para utilização no domínio do direito. Entretanto, outras técnicas como banco de dados relacional, *search engines*, entre outros, podem e devem conviver com o RBC, seguindo-se tendências atuais de construção de sistemas híbridos inteligentes, com o intuito de maior aproximação ao conhecimento do especialista, conforme já citado nos quadros comparativos dos métodos estudados.

Mattos (2002) comenta este tipo de interação quando diz que 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 podem ser treinadas a partir de bases de casos. Agentes inteligentes podem implementar autonomia a sistemas de RBC, como assistentes dos usuários, que fundamentam sua ajuda em interações anteriores.

## 5. Protocolo do Estudo de Caso

### 5.1 Considerações iniciais

Neste capítulo é feita a caracterização do ambiente estudado, assim como os sujeitos nele envolvidos.

Considerando que Laudon e Laudon (2004) estabelecem que os sistemas de informação resultam da integração da tríade: pessoas, processos e tecnologia, (Figura 5.1), esse protocolo foi elaborado de forma que se possa analisar a tecnologia, verificar sua viabilidade de uso com um grupo de pessoas (estudo de caso) e, finalmente, fazer uma avaliação da sua viabilidade dentro do contexto estudado.



**Figura 5.1 Sistemas de informação**

Fonte: Adaptado de Laudon e Laudon, 2004.

Cabe salientar que este estudo visa a solução do problema levantado no TJSP, com suas especificidades e não representa um comparativo total entre ferramentas existentes. Por conseqüência, as ferramentas analisadas no Capítulo 6 foram escolhidas por apresentarem particularidades interessantes ao problema

proposto e ao local de estudo, descrito no Item 5.2, não significando, porém que sejam as únicas existentes para esta finalidade.

## **5.2 Local do estudo**

A pesquisa aqui tratada foi desenvolvida no Tribunal de Justiça de São Paulo, TJSP, maior poder judiciário da América Latina em volume de processos, que perfazem, atualmente, cerca de 15 milhões, em andamento.

O órgão é composto por 309 foros distribuídos por todo o Estado de São Paulo.

Em seu quadro, o TJSP conta, efetivamente, com 45.157 funcionários, sendo 1.996 juízes de Direito distribuídos entre capital e interior.<sup>1</sup>

Para cada caso julgado diariamente no órgão, é gerada uma sentença, que é um documento textual. Além das sentenças geradas todos os meses, que não estão disponíveis para consulta por falta de ferramentas tecnológicas, o órgão conta atualmente com bases de dados referentes a documentos textuais de ementas, súmulas e acórdãos arquivados a partir do ano de 1998, com 3 milhões de documentos multipaginados, que perfazem aproximadamente 15 milhões de páginas arquivadas em diversos formatos eletrônicos. Destas, por limitações dos sistemas atuais de buscas, apenas 350 mil ementas são disponibilizadas no portal do órgão.<sup>2</sup>

Dentro do mesmo levantamento realizado pela Secretaria de Tecnologia da Informação do órgão, constatou-se que o crescimento mensal dos acórdãos é de 45

---

<sup>1</sup> Fonte: Censo SRH/TJSP – dez/06.

<sup>2</sup> Fonte: Levantamento STI/TJSP – nov/06.

mil documentos, que somarão, ao término de dois anos, 22 milhões de páginas armazenadas.

De acordo com as informações obtidas no *site* oficial do órgão,<sup>3</sup> o acervo bibliográfico da biblioteca é composto de 70 mil títulos, bases de dados de legislação nas esferas federal, estadual e municipal, bem como Normas e Atos Administrativos de interesse do Poder Judiciário.

Existe ainda, por parte da presidência do TJSP, um interesse na disponibilização do teor das sentenças de 1º grau para consulta através da *Web*, o que gera uma demanda que chega à cifra de 15 milhões de processos em andamento.

Por todos os números apresentados, principalmente no tocante a problemas relacionados a documentos textuais, o TJSP foi escolhido como objeto de estudo.

A análise da estrutura atual de armazenamento e recuperação de documentos textuais despertou a visão de se criar facilitadores para o desenvolvimento desses trabalhos.

Portanto, a avaliação de formas de recuperação efetiva, sua aceitabilidade e produção fazem-se objeto desta pesquisa, para que se possa identificar pontos de relevância bem como possibilidades de melhorias.

### **5.3 Tipo de estudo**

Yin (2001) ressalta dentre as estratégias de pesquisa a serem consideradas: a análise de arquivos, a pesquisa histórica, o experimento e o estudo de caso, esta última sendo a que está mais próxima da situação a ser analisada. O mesmo autor

descreve que um estudo de caso é uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre fenômeno e contexto não estão claramente definidos.

O projeto começa com a definição do tipo de questão a ser feita, podendo envolver aquelas do tipo “quem”, “o que”, “onde”, “como” e “por quê”. Dentro do tipo de pesquisa que é o estudo de caso, seria preferível a utilização dos dois últimos tipos, sendo que desses, principalmente ao “como”, será dada preferência no trabalho de pesquisa a ser realizado (CARMO, 2003).

Por se tratar de “uma fonte essencial de evidências para o estudo de caso” (YIN, 2001) e por se adequar ao tipo de levantamento necessário, o caso será estudado através de entrevistas estruturadas.

Segundo Contandriopoulos (1999), o estudo de caso é uma estratégia que o pesquisador adota em sua pesquisa, sobre uma ou um pequeno número de unidades a serem analisadas. Na pesquisa aqui delineada foram utilizados casos múltiplos, com o objetivo de melhor explorar o assunto proposto.

O propósito desta pesquisa é identificar “quem” são os interessados na recuperação eficiente de documentos jurídicos textuais, “o que” deve ser recuperado dentro do universo proposto, “como” os serviços de pesquisa impactam na produção do judiciário paulista e, por consequência, qual será o ganho conseguido ao se adotar uma ferramenta adequada e com bons índices de aceitabilidade por parte dos usuários.

---

<sup>3</sup> Disponível em: [www.tj.sp.gov.br](http://www.tj.sp.gov.br)

## 5.4 Sujeitos da pesquisa

Farão parte desta pesquisa, os Exmos. Srs. Desembargadores e Juízes do órgão estudado, além de funcionários e responsáveis pelas áreas de biblioteca e TI. Os profissionais diretamente interessados por esta solução são os juízes e desembargadores, do lado da geração e uso das informações; e bibliotecários e gestores de TI na divulgação e suporte das informações.

Salienta-se que a pesquisa não se delimita a números e percentuais e sim à forma que as atividades são realizadas.

## 5.5 Montagem da pesquisa

Para levantamento dos casos seguiram-se quatro passos:

- **Preparação** – formação do conjunto de conteúdos e estrutura sobre o escopo da pesquisa e o contexto onde se aplica.
- **Conceituação de um pré-modelo** – execução de testes em um caso, com possíveis implementações de melhorias.
- **Aplicação da pesquisa.** E
- **Análise de resultados.**

Ainda com o intuito de melhor demonstrar o domínio proposto, o roteiro da entrevista será dividido em quatro grandes blocos:

- **Caracterização do entrevistado** – enquadramento do entrevistado dentro do domínio proposto.

- **Necessidade do entrevistado quanto à utilização de documentos textuais em suas atividades** – levantamento da necessidade dos documentos textuais nas atividades diárias do entrevistado.
- **Formas atuais de desenvolvimento do trabalho relacionado a documentos textuais** – levantamento das formas de trabalho atual do entrevistado visando descrever os facilitadores e os problemas elencados com o uso das ferramentas atuais.
- **Características importantes inerentes a um sistema de consulta de documentos textuais** – levantamento das necessidades do entrevistado para um sistema de pesquisa para documentos textuais eficiente.

## **5.6 Coleta de dados**

Esta etapa da pesquisa consiste em reunir os dados para análise. De acordo com Yin (2001), a evidência para estudos de caso pode ser originária de seis fontes, sendo: documentos, registros arquivais, entrevistas, observação direta, observação participante e artefatos físicos.

A coleta dos dados foi feita pelo próprio investigador, por meio de entrevistas semi-estruturadas, conforme Anexo 1, havendo possibilidade, se permitido pelo entrevistado, de uso de gravadores para enriquecimento do processo, sendo os resultados descritos no Capítulo 7.



**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo - 5 Protocolo do Estudo de Caso**

As entrevistas foram previamente agendadas, tendo em vista a não disponibilidade integral dos sujeitos entrevistados, evitando interrupções em suas atividades rotineiras.

## **6. Tecnologia: Análise das ferramentas**

Com o fim de contribuir para a solução do problema proposto no Item 1.2, foram analisados alguns sistemas disponíveis no mercado que recuperam documentos textuais e podem ser utilizados no domínio jurídico.

Embora existam outras ferramentas com o intuito de indexar informações para fins de pesquisa e busca de informações, apenas as aqui referenciadas apresentaram foco nas necessidades estabelecidas nesta pesquisa.

As seguintes ferramentas foram analisadas por se enquadrarem no cenário: JurisConsulta, AlphaThemis e Idol K2.

O trabalho de análise foi desenvolvido a partir de uma tese de doutorado (BUENO, 2005), contato com os fornecedores das soluções e análise de versões de demonstração.

Nas ferramentas citadas, foram analisadas características que se enquadram ao domínio proposto como: permitir recuperação de documentos textuais através da coincidência de palavras e campos específicos, permitir pesquisa por texto livre, permitir a busca em diversos formatos de arquivos disponibilizados, entre outros.

### **6.1 JurisConsulta<sup>®4</sup>**

O sistema JurisConsulta teve por objetivo modelar e aplicar a teoria do RBC à área do Direito Criminal, visando a recuperação de jurisprudências com

---

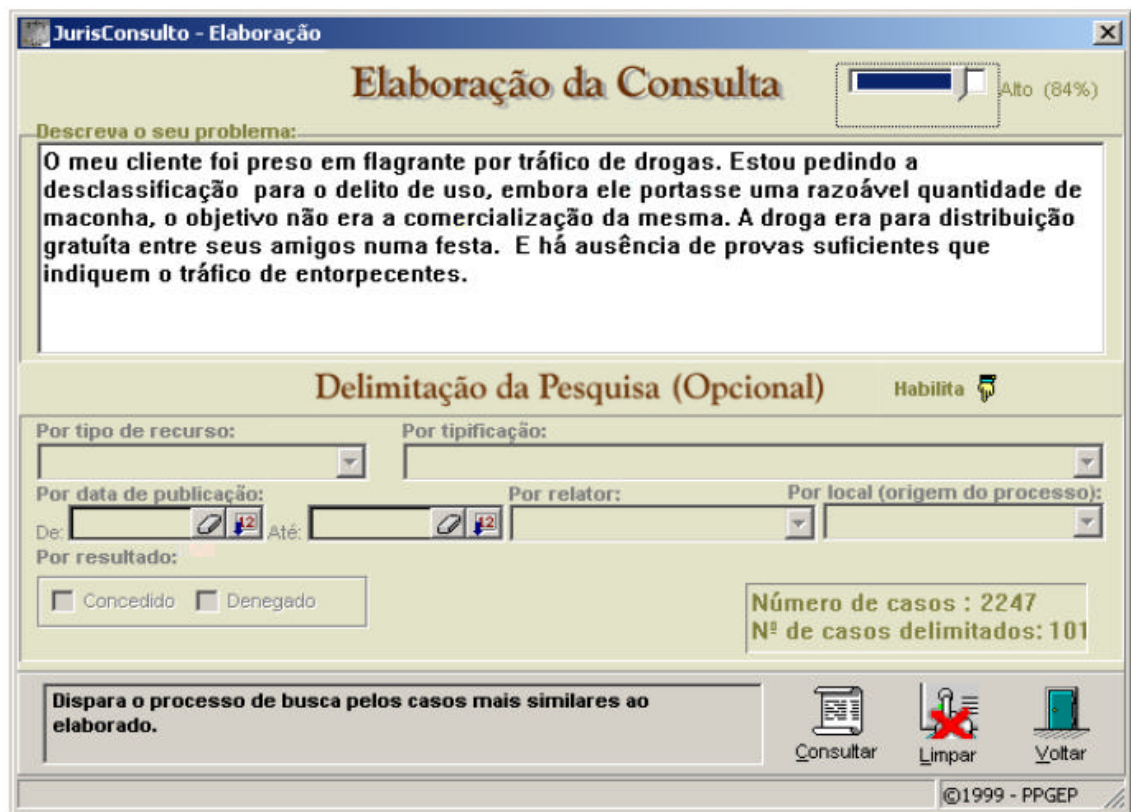
<sup>4</sup> Todos os direitos reservados da marca e uso da aplicação para o IJURIS – Instituto de Governo Eletrônico, Inteligência e Sistemas. Copyright© 2002

características semelhantes e, além disso, permitindo a inclusão de novos documentos, automaticamente, na base de conhecimento (BUENO, 1999).

O JurisConsulta é um sistema de recuperação de jurisprudência, desenvolvido com técnicas de Inteligência Artificial, que permite dentre outras coisas, facilitar o acesso do profissional do Direito e também do cidadão a informações armazenadas em banco de dados disponíveis na Internet (Figura 6.1)

Em Bueno (2005), a autora descreve detalhadamente o funcionamento desse sistema:

O JurisConsulta utiliza como técnica de Inteligência Artificial o Raciocínio Baseado em Casos (RBC). O RBC usa experiências anteriores e semelhantes para a solução de um problema novo. É baseado no princípio de analogia, assumindo que problemas semelhantes têm soluções semelhantes (AMONDT, 1994, apud BUENO 2005). Por essa razão, o RBC é uma técnica muito adequada ao domínio jurídico, pois emprega o mesmo tipo de raciocínio utilizado pelos juristas na solução de um problema.

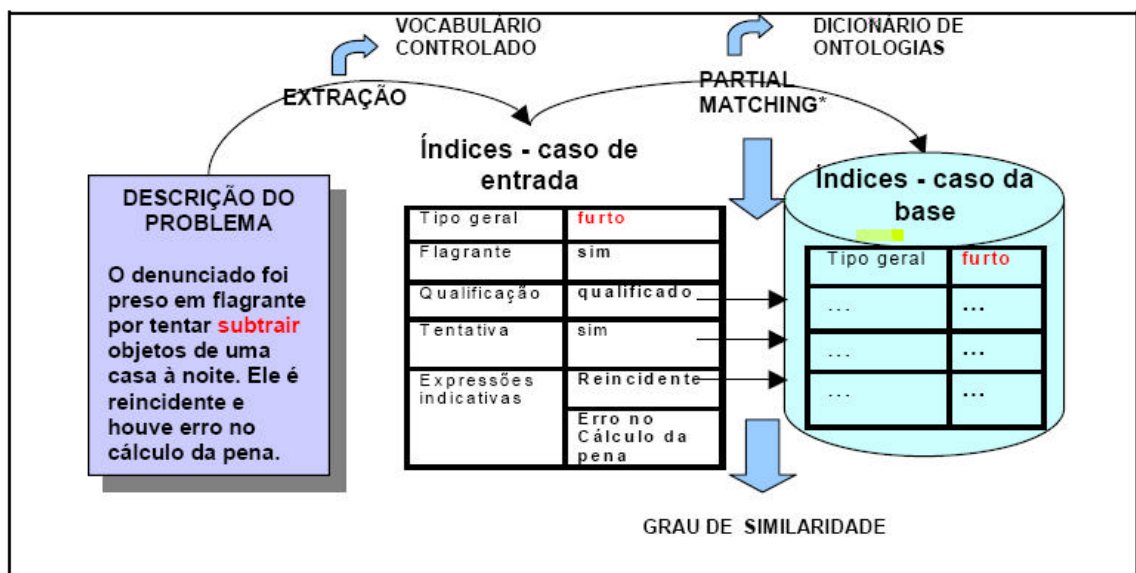


**Figura 6.1 Interface do sistema JurisConsulta**

Fonte: Bueno, 2005.

Para possibilitar o processo de conhecimento textual não estruturado, ‘um Vocabulário Controlado e um Dicionário de Termos Normativos’ – baseados na terminologia jurídica usual e na teoria jurídica – são integrados na recuperação e no processo de extração do conhecimento. Assim, o usuário pode descrever uma questão em linguagem natural, e o ajuste da situação é iniciado pela comparação deste 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, ele será ordenado de forma decrescente, de acordo com os valores obtidos com a métrica de similaridade; e os casos mais úteis, ou seja,

aqueles 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.

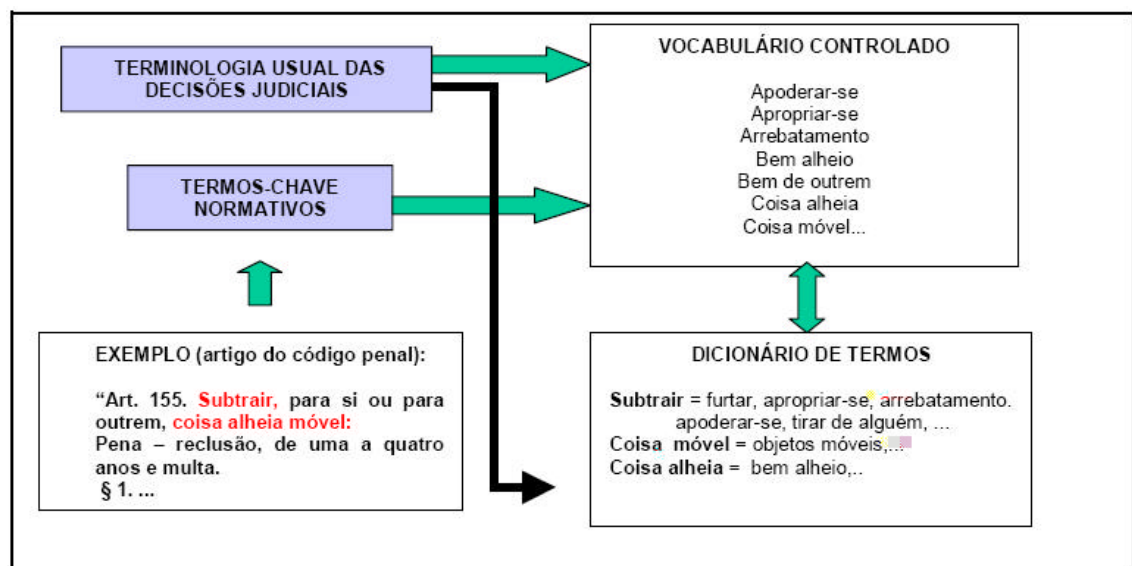


**Figura 6.2 Processo de recuperação baseado em casos do sistema**

**JurisConsulta**

Fonte: Bueno, 2005.

Bueno (2005) complementa ainda que a representação e recuperação de casos no sistema JurisConsulta considera que um termo normativo pode ter várias interpretações nas decisões jurídicas, e são os juristas que indicam quais são os termos relevantes para serem empregados na construção de argumentos persuasivos na solução de problemas jurídicos. Desse modo, para apoiar o processo de recuperação e a extração automática dos índices relativos ao conteúdo jurídico dos textos da jurisprudência criminal, um Vocabulário Controlado e um Dicionário de Termos Normativos foram desenvolvidos (Figura 6.3).



**Figura 6.3** Definição das expressões indicativas no sistema JurisConsulta

Fonte: Bueno, 2005.

O processo de recuperação do sistema JurisConsulta é baseado no cálculo da similaridade realizado através do *matching* parcial entre o caso fornecido inicialmente em linguagem natural e os casos na base de casos, usando uma medida de similaridade.

Os valores da similaridade local e global são calculados somente para os índices 'Tipificação' e 'Expressões Indicativas'. Os outros índices (número do Acórdão, data da publicação etc.) não entram no cálculo da similaridade, são considerados excludentes.

Com base no ajuste da situação fornecida, os casos relevantes são recuperados de uma base de casos. Para identificar os mais úteis em relação ao problema inicial fornecido, o ajuste da situação é 'comparado' parcialmente com cada caso da base. A similaridade de cada índice do problema inicial com cada caso da base é determinada pela 'Medida da Similaridade Local' (BUENO, 2005).

## **6.2 AlphaThemis<sup>®5</sup>**

O AlphaThemis consiste em um sistema inteligente para a extração do conhecimento contido nas súmulas dos tribunais nacionais, configurando uma potente ferramenta de mineração de textos em 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 (BUENO et al., 2005).

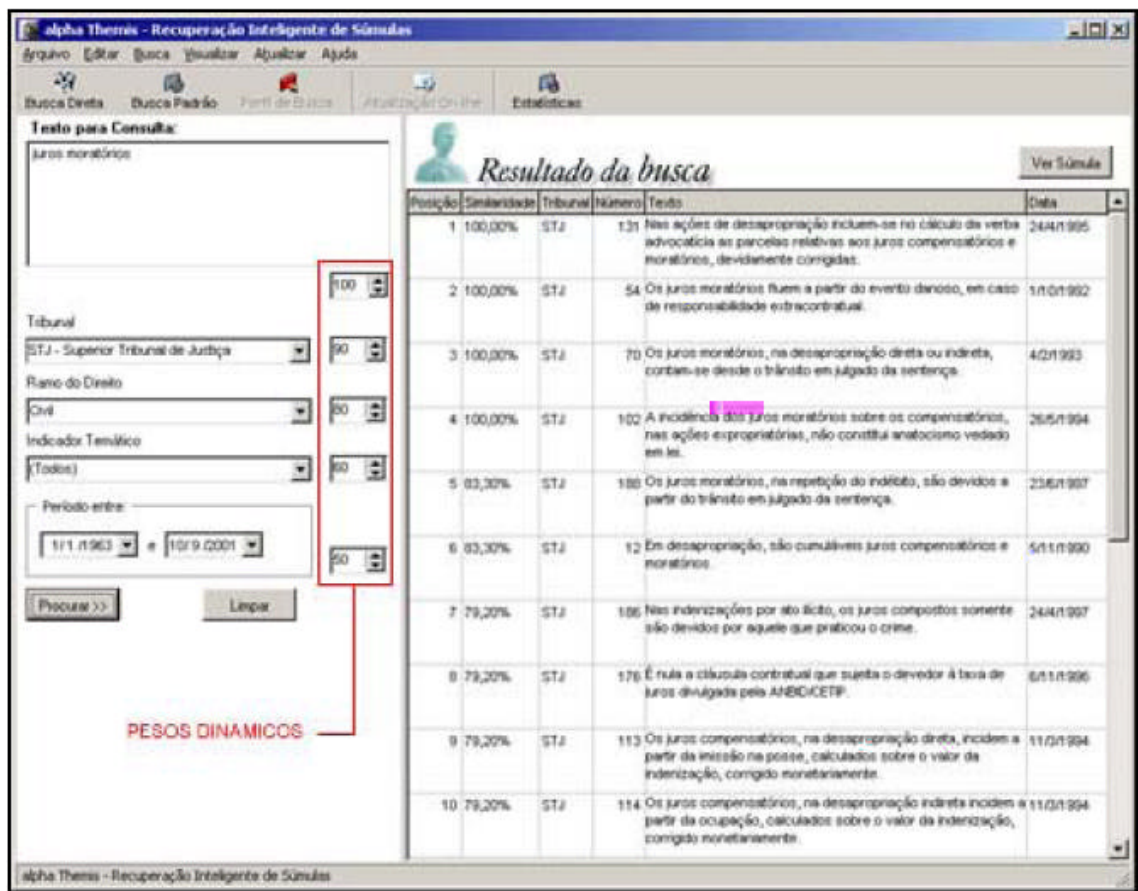
O uso de pesos dinâmicos nos filtros de busca torna o AlphaThemis ainda mais flexível na formulação da pesquisa. Trata-se da possibilidade de alteração dinâmica da relevância dos atributos dentro do contexto pesquisado, calibrando-se, individualmente, os pesos de cada um deles, dentro de uma escala de 0 a 100, conforme demonstrado em destaque na Figura 6.4.

O resultado da pesquisa é apresentado ao usuário em uma interface de saída que mostra os resultados hierarquizados de acordo com o percentual de similaridade constatado entre a consulta e os enunciados das súmulas. O sistema AlphaThemis apresenta em sua interface de saída, além do texto da súmula recuperada, todos os mecanismos que o usuário acionou na realização da consulta, bem como os valores respectivamente atribuídos (BUENO, 2005).

---

<sup>5</sup> Todos os direitos reservados da marca e uso da aplicação para o IJURIS – Instituto de Governo Eletrônico, Inteligência e Sistemas. Copyright© 2002

**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo 6 – Tecnologia: Análise das ferramentas**



**Figura 6.4 Interface AlphaThemis, ressaltando o uso de pesos dinâmicos**

Fonte: Bueno, 2005.

Semelhante ao JurisConsulta e ao Olimpo, como técnica de extração de informações, o AlphaThemis identifica os 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 (BUENO, 2005).

Em Bueno (2005), é explicada ainda a forma detalhada de operação desse sistema.



Inicialmente, a operação do sistema AlphaThemis foi fundamentada na extração de dez atributos provenientes dos enunciados das súmulas, com os seguintes valores a eles atribuíveis:

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

Para o cálculo da métrica de similaridade, o primeiro nível leva em consideração os atributos fixos, contidos nos campos de '1' a '6', que equivalem a 50% da métrica. Somados, perfazem o valor máximo de 5,0. Exemplo: de seis (6) atributos escolhidos, quatro (4) foram encontrados (1, 2, 4 e 5). A pontuação total, neste caso, é de 3,0, ou 30% da similaridade total. Os valores do primeiro nível são fixos, ou seja, 5,0 e sempre será igual a 50%,  $3,5 = 35\%$ , etc.

O segundo nível leva em consideração as palavras informadas na entrada, operando sobre os campos '7' a '9', que equivalem aos outros 50% da métrica

total. Cada palavra equivale a 1,0. Exemplo: se sete (7) palavras foram informadas,  $7 = 50\%$ . Na busca, duas (2) não foram encontradas; uma (1) foi encontrada no campo nove (9) (recebe o valor 0,5), duas (2) foram encontradas no campo oito (8) (recebem o valor 0,7 cada uma delas), e as demais foram encontradas no campo sete (7) (recebe o valor 1,0 cada uma delas). O valor total é  $0,5 + 1,4 + 2,0 = 3,9$  que corresponde a 27,85% da similaridade total (100%), tendo em vista que  $7 = 50\%$ . O valor final da similaridade é primeiro + segundo nível, o que, no exemplo citado, corresponderia a  $30 + 27,5 = 57,85\%$ .

### **6.3 Idol K2<sup>®6</sup>**

Por se tratar de uma solução totalmente comercial, não foi possível uma análise mais aprofundada na forma de funcionamento interno desse sistema, o que é descrito a seguir são particularidades comerciais do produto disponíveis no *site* da empresa que o produz ([www.autonomy.com](http://www.autonomy.com)), e observações feitas a partir de instalação piloto de uma versão de demonstração.

A solução corporativa Idol K2 (*Knowledge to*) compreende as funcionalidades de pesquisa, classificação, recomendação e redes sociais. A arquitetura da solução Idol atende os requisitos técnicos de segurança, escalabilidade, balanceamento de carga, possibilita a fácil e rápida recuperação das informações, classificação automática dos documentos e informações relevantes, bem como o compartilhamento dessas informações entre os departamentos e especialistas.

---

<sup>6</sup> Todos os direitos reservados da marca e uso da aplicação para o Autonomy Corporation plc Copyright © 2006

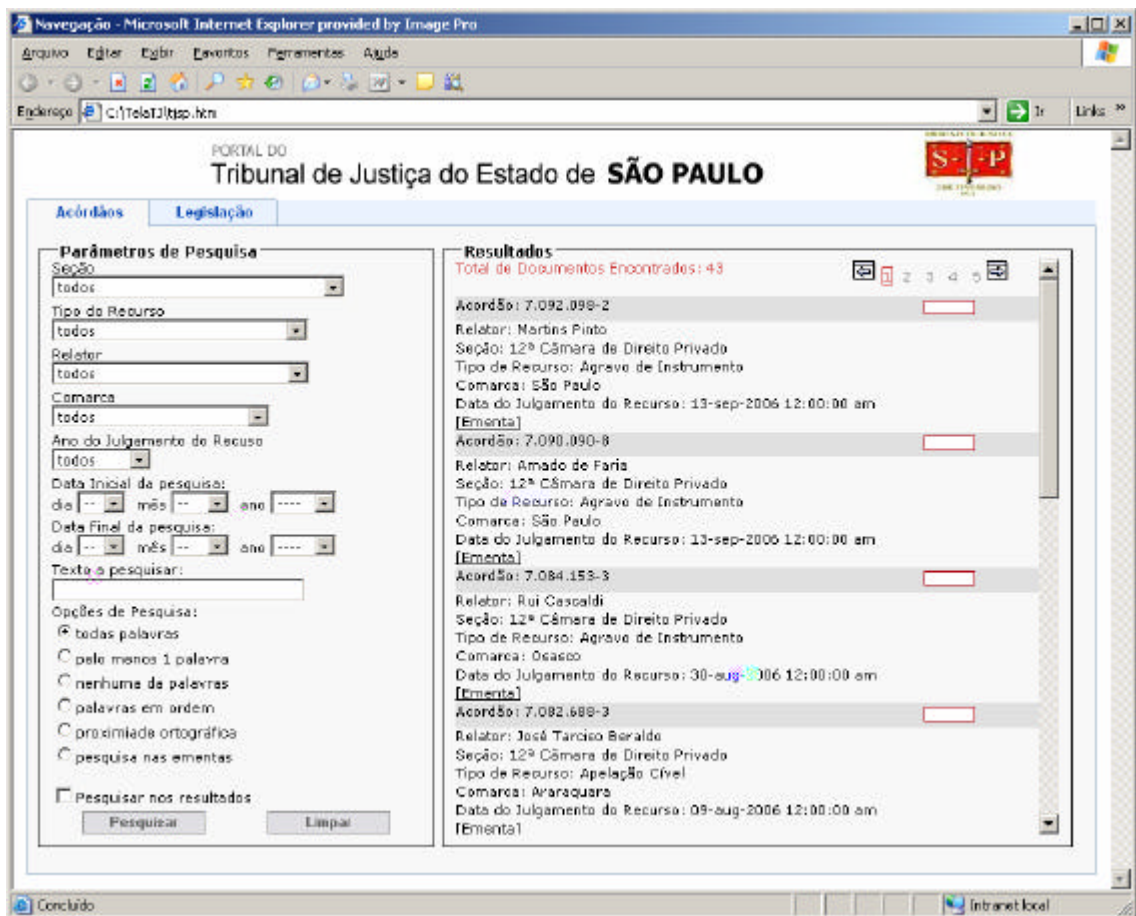
A solução propõe disponibilizar a informação de forma concisa, independente de onde ela esteja armazenada e do formato, permitindo ainda a recuperação e o compartilhamento das informações em vários idiomas.

A solução apresenta ainda algumas funcionalidades que se mostram interessantes ao domínio estudado como:

- Indexação de formatos eletrônicos comuns, como documentos do MicrosoftOffice, arquivos em formato pdf, html, entre outros.
- Busca *full text* por palavra-chave.
- Exclusão de *noise words*.
- Tolerância a erros ortográficos.
- Sugestão de grafia na pesquisa.
- Definir distância entre duas palavras pesquisadas (*near*).
- *Hightighting* de palavra encontrada no texto.
- Visualizador html para principais tipos de documento.
- Busca paramétrica com *full text*.
- Elaboração de resumos automáticos.
- Federação de busca para outros portais.

Para análise da solução, foi disponibilizada uma versão teste do produto, que teve sua interface de consulta customizada para o projeto em questão, sendo baseada na interface *Web* do STJ (Superior Tribunal de Justiça), conforme mostra a Figura 6.5.

**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo 6 – Tecnologia: Análise das ferramentas**



**Figura 6.5 Interface demonstrativa da ferramenta Idol K2, desenvolvida na fase de análise**

Houve ainda no decorrer do desenvolvimento deste estudo a análise de outras ferramentas, não citadas aqui por não apresentarem características mínimas necessárias para a solução do problema proposto.

## **6.4 Comparação das ferramentas**

O trabalho desenvolvido foi dividido em duas etapas, sendo uma delas a análise técnica das ferramentas, com o intuito de verificar as funcionalidades

pertinentes ao domínio proposto, e a outra, um estudo de caso com potenciais usuários.

Da análise técnica das ferramentas obtiveram-se algumas características importantes, demonstradas no Quadro 6.1.

Embora na literatura alguns autores (TONINI, 2003) recomendem que na seleção de *software* seja importante a identificação inicial de requisitos mínimos que o software deva atender, neste trabalho optou-se por fazer o inverso, ou seja, analisar as características das ferramentas existentes para se ter uma idéia do tipo de funcionalidades que elas apresentam.

<b>Funcionalidade/Software</b>	<b>JurisConsulta</b>	<b>AlphaThemis</b>	<b>Idol K2</b>
Permite indexação e busca em diversos formatos eletrônicos.	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Permite busca <i>full text</i> por palavra-chave;	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Apresenta tolerância a erros ortográficos (digitação).	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
<i>Highlighting</i> de palavra encontrada no texto.	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Permite o uso de diversos campos específicos para pesquisa	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Permite o uso de pesos dinâmicos para cada campo da pesquisa.	<i>Não</i>	<i>Sim</i>	<i>Não</i>
Disponibiliza visualizador padrão para os principais tipos de documento.	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Elaboração de resumos automáticos	<i>Não</i>	<i>Não</i>	<i>Sim</i>

**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo 6 – Tecnologia: Análise das ferramentas**

Permite busca por dicionário de sinônimos.	<i>Sim</i>	<i>Sim</i>	<i>Sim</i>
Permite busca por contexto da palavra (uso de ontologias).	<i>Sim</i>	<i>Sim</i>	<i>Não</i>

**Quadro 6.1 Análise técnica comparativa dos softwares pesquisados**

Observe-se que por se tratar de ferramentas muito específicas, todas se equiparam na maioria dos itens analisados. Essa semelhança é total se observadas as duas primeiras ferramentas citadas, JurisConsulta e AlphaThemis, pois são de um mesmo fornecedor e incorporam conceitos de Inteligência Artificial.

Outro item pertinente a ser observado é que a terceira ferramenta, Idol K2, em uma estimativa de custo de implementação, ficou aproximadamente 20% mais onerosa em relação às outras duas em um mesmo cenário de implantação.

## **7. Pessoas: Resultado do caso**

Este capítulo apresenta o resultado do caso proposto cujo protocolo se encontra no Capítulo 5, focando o modo como as pessoas se utilizam das ferramentas para recuperação de documentos textuais.

Embora Yin (2001) relate que ao se adotar o estudo de caso como método de pesquisa seja interessante a realização de estudo de casos múltiplos, o trabalho realizado nesta pesquisa abordou um único caso, o do Tribunal de Justiça de São Paulo que, por ser o maior tribunal da América Latina, foi considerado suficiente para a finalidade proposta. Por outro lado, tendo em vista que os juízes e desembargadores possuem autonomia no desenvolvimento de suas atividades, tomou-se o cuidado de se entrevistar individualmente as pessoas envolvidas.

De acordo com Marconi e Lakatus, (2005), uma vez manipulados os dados e obtidos os resultados, o passo seguinte é sua análise e interpretação

As entrevistas, baseadas no roteiro constante do Anexo 1 foram realizadas com oito profissionais, sendo quatro juízes de Direito, dois desembargadores, um coordenador da área de TI e um funcionário da biblioteca responsável pela divulgação de informações no órgão.

### **7.1 Caracterização dos entrevistados**

Conforme previamente citado, os entrevistados somam oito profissionais, sendo:

- quatro juízes de direito, doravante designados como J1, J2, J3 e J4;
- dois desembargadores, doravante designados como D1 e D2;

- um coordenador da área de TI, doravante designado como TI1, e
- um funcionário da biblioteca, doravante designado como B1.

## **7.2 Necessidade do entrevistado quanto à utilização de documentos textuais em suas atividades**

O Quadro 7.1 ilustra a necessidade dos documentos textuais para os entrevistados. Salienta-se que, nesse momento, antes das respostas, os entrevistados eram esclarecidos sobre a definição de documentos textuais jurídicos utilizados no presente trabalho.

	<b>Indispensável</b>	<b>Importante</b>	<b>Indiferente</b>	<b>Sem importância</b>
<b>J1</b>	X			
<b>J2</b>	X			
<b>J3</b>	X			
<b>J4</b>	X			
<b>D1</b>	X			
<b>D2</b>	X			
<b>TI1</b>		X		
<b>B1</b>	X			

**Quadro 7.1 Necessidade dos documentos textuais**

Observe-se que é compreensível o fato de o profissional de TI considerar os documentos textuais como importantes e não indispensáveis, pois documentos textuais não fazem parte de suas atividades rotineiras como o fazem para todos os outros entrevistados.



### **7.3 Formas atuais de desenvolvimento do trabalho relacionado a documentos textuais**

O Quadro 7.2, a seguir, demonstra quais são as ferramentas utilizadas atualmente para os trabalhos de pesquisa dos entrevistados. Salienta-se que as ferramentas foram colocadas na mesma ordem em que foram citadas.

<b>J1</b>	<b>Sítes de outros tribunais, principalmente, STJ; Publicações de editoras especializadas; Livros particulares</b>
<b>J2</b>	Sítes de outros tribunais; Publicações de editoras especializadas (em CD's)
<b>J3</b>	Site do STJ; Publicações de editoras especializadas (em CD's); Livros; Biblioteca do TJSP
<b>J4</b>	Publicações de editoras especializadas (em CD's); Sítes de outros tribunais
<b>D1</b>	Serviços da Biblioteca; Site do STJ; Publicações de editoras especializadas (CD's)
<b>D2</b>	Sítes de outros tribunais; Publicações de editoras especializadas (CD's e DVD's); Biblioteca do TJSP
<b>TI1</b>	Sítes de busca
<b>B1</b>	Sistemas da Intranet do TJSP; Sítes de outros tribunais; Publicações de editoras especializadas (CD's e DVD's); Livros e Revistas especializadas

**Quadro 7.2 Ferramentas utilizadas no trabalho de pesquisa**

O Quadro 7.3, a seguir, evidencia quais são, dentro dessas ferramentas, os campos mais utilizados, habitualmente, para se obter os documentos desejados. Salienta-se que de modo semelhante ao utilizado para os sistemas, os campos foram colocados na mesma ordem em que foram citados.

<b>J1</b>	<b>Assunto; Período dos documentos; Texto livre</b>
<b>J2</b>	Texto livre; Desembargador;
<b>J3</b>	Período; Assunto; Texto livre
<b>J4</b>	Assunto; Texto livre
<b>D1</b>	Desembargador/Ministro; Assunto;
<b>D2</b>	Numero do documento; assunto;
<b>TI1</b>	Texto livre;
<b>B1</b>	Texto livre; Período; Desembargador;

**Quadro 7.3 Campos mais utilizados nas ferramentas de busca**

Importante notar que todos os entrevistados que citaram “Assunto” como campo de pesquisa, o utilizam na forma de coincidência de palavras no texto encontrado. Note-se ainda que alguns entrevistados fizeram uso de um *site*/sistema em execução no momento da entrevista para se certificarem dos campos utilizados.

Ainda em referência ao trabalho dos entrevistados relacionado às pesquisas eletrônicas, o Quadro 7.4 mostra o número médio de palavras, quando existem, utilizadas para a pesquisa de texto livre, além de identificar se isto se deve a algum fator limitante da ferramenta em questão. Mais uma vez, os campos foram colocados na mesma ordem em que foram citados.

	<b>Número de palavras</b>	<b>Limitação da ferramenta</b>
<b>J1</b>	Duas ou três, mas varia bastante de acordo com a dificuldade encontrada na busca.	Em algumas, sim.
<b>J2</b>	Depende da busca, mas, normalmente, utilizo o máximo permitido.	Sim
<b>J3</b>	Depende do caso	Não sei
<b>J4</b>	Normalmente começo com uma e vou acrescentando para melhor filtro.	Às vezes, sim.
<b>D1</b>	Uma ou duas	Não sei
<b>D2</b>	Depende da busca, mas, normalmente, uma.	Não sei
<b>TI1</b>	Depende da dificuldade da busca	Sim
<b>B1</b>	Depende da dificuldade da busca	Sim

**Quadro 7.4 Número de palavras utilizadas na pesquisa**

Observe-se que, novamente, alguns entrevistados se utilizaram de um *site*/sistema em execução no momento da entrevista para se certificarem das respostas.

Note-se também que os entrevistados J3, D1 e D2 desconhecem as limitações das ferramentas na busca por texto livre, fato justificado nos campos mais utilizados (Quadro 7.3), respectivamente: Período, Assunto e Número do documento.

Diante do questionamento, o entrevistado B1 comentou que, muitas vezes, possui o texto da jurisprudência na íntegra, mas não consegue inseri-lo nas ferramentas para localização, por exemplo, do número da publicação.

Outro item levantado, com resultados expostos no Quadro 7.5, foi a produtividade das ferramentas utilizadas, fazendo-se a comparação entre o volume de informação gerada pelas ferramentas de busca *versus* o aproveitamento útil de

cada informação. Quando essa relação era considerada como improdutiva o entrevistado era indagado sobre o motivo.

	<b>Retorno produtivo</b>	<b>Motivo</b>
<b>J1</b>	Sim	
<b>J2</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno.
<b>J3</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno; Erro de leitura nas mídias fornecidas (CD's)
<b>J4</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno.
<b>D1</b>	Sim	
<b>D2</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno; dificuldade de encontrar a mídia específica (CD's e DVD's).
<b>TI1</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno.
<b>B1</b>	Não	Perda de tempo com a filtragem de vários documentos irrelevantes à pesquisa ou nenhum retorno.

**Quadro 7.5 Opinião sobre a produtividade das ferramentas utilizadas**

Foi analisada ainda, conforme demonstrado no Quadro 7.6, a opinião dos entrevistados sobre a usabilidade das ferramentas previamente citadas.

Observe-se que nesse momento, era desvinculado o conceito de usabilidade do conceito de produtividade, havendo ainda em praticamente todas as

entrevistas, conversas sobre os conceitos de usabilidade, uso de operadores lógicos, visual das interfaces, formas de acesso, entre outros.

<b>J1</b>	Simples utilização – Dentro das funcionalidades utilizadas
<b>J2</b>	Problemas com operadores lógicos
<b>J3</b>	Não encontra dificuldade de uso – Em pesquisas básicas
<b>J4</b>	Dificuldade de filtragem na utilização de diversos campos concomitantemente
<b>D1</b>	Simples utilização
<b>D2</b>	Normal
<b>TI1</b>	De modo geral, intuitivas
<b>B1</b>	Dificuldade de filtragem na utilização de diversos campos concomitantemente

**Quadro 7.6 Opinião dos entrevistados a respeito da usabilidade das ferramentas**

Observe-se novamente que o item analisado referia-se à usabilidade das ferramentas e não à sua produtividade. Alguns dados eram esperados, como a familiarização do profissional de TI aos *softwares* e ferramentas de modo geral, pois são inerentes à sua função.

#### **7.4 Características importantes de um sistema de consulta de documentos textuais**

Neste item, os entrevistados são questionados sobre particularidades que, se incorporadas às ferramentas, contribuirão para o processo de recuperação de documentos textuais. Nesse momento da entrevista, eram explicadas e exemplificadas aos entrevistados algumas formas possíveis de busca pelos documentos como percentual de certeza, operadores lógicos, sugestão de correção

ortográfica automática, uso de dicionário jurídico, uso de contexto jurídico das palavras (ontologias), entre outros.

Em sua totalidade, os entrevistados, mesmo em momentos separados, concordaram que a ferramenta ideal é a que mais se aproxima da forma de escrita e entendimento humano. Em outras palavras, gostariam de ter uma ferramenta em que fosse possível a inserção de uma frase, trecho ou texto integral, sem a preocupação de filtros, operadores, campos específicos, entre outros; e que o retorno dessa ferramenta fosse exatamente o documento desejado.

Nesse contexto de expectativa dos entrevistados, a forma de recuperação que mais se aproximou do chamado “sistema ideal”, foi a baseada em RBC – Raciocínio Baseado em Casos.

Explorando ainda as características importantes inerentes a um sistema de consulta de documentos textuais, os entrevistados foram convidados a fazer uma breve análise das implementações de demonstração das ferramentas. Os comentários estão descritos no Quadro 7.7, a seguir.

**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo 7 – Pessoas: Resultado do caso**

	<b>AlphaThemis</b>	<b>JurisConsulta</b>	<b>Idol K2</b>
<b>J1</b>	Mescla clareza com detalhamento.	Intuitiva e de uso simples.	Muita informação na tela.
<b>J2</b>	Campo de pesos interessantes.	Bom campo de pesquisa e barra de semelhança interessante.	Boa, mas com muitas opções de pesquisa. Confunde um pouco.
<b>J3</b>	Não houve tempo disponível.	Não houve tempo disponível.	Não houve tempo disponível.
<b>J4</b>	Ferramentas semelhantes e destaque para os campos de pesos.	Ferramentas semelhantes e destaque para percentual de semelhança além do foco na pesquisa por texto livre.	Ferramentas semelhantes e destaque para os campos de similaridade. Muitas opções de pesquisa, dificultando um pouco.
<b>D1</b>	Muito semelhantes, necessitaria de mais dados reais para comparação.	Muito semelhantes, necessitaria de mais dados reais para comparação.	Muito semelhantes, necessitaria de mais dados reais para comparação.
<b>D2</b>	Não houve tempo disponível.	Não houve tempo disponível.	Não houve tempo disponível.
<b>TI1</b>	Relativamente simples e destaque para a possibilidade de ajustar pesos.	A mais intuitiva para pesquisas diretas.	Bastante completa e destaque para o item de similaridade nos resultados, mas apresenta interface um pouco poluída.
<b>B1</b>	O campo de pesquisa por texto livre chama a atenção pelo destaque.	Bastante claro.	O campo de pesquisa por texto livre ficou sem foco no meio dos outros. Destaque para o campo de similaridade dos resultados.

**Quadro 7.7 Comentários das implementações de demonstração**

**Documentos Textuais Jurídicos: Um estudo de Ferramentas de TI para sua Recuperação**  
**Capítulo 7 – Pessoas: Resultado do caso**

Importante observar que os comentários foram relativos à poluição/clareza



## **8. Processos: Resultado do caso**

Este capítulo, assim como o anterior, apresenta o resultado do caso

Corroborando a disponibilidade para mudar alguns processos com o intuito de melhor recuperação, os entrevistados J1, J2, J3, J4, D1 e D2 afirmaram que um dos problemas encontrados no uso dos documentos textuais ocorre na sua geração e não apenas com sua recuperação. Faltam, segundo eles, padronização de procedimentos, termos e textos desses documentos. Os entrevistados declararam ainda que, a longo prazo, deverá existir algum tipo de padronização, pois já existe por parte do STJ – Superior Tribunal de Justiça, iniciativas nesse sentido e já há estudos referentes ao tema.

## **9. Análise crítica da pesquisa**

O estudo analisou os três fatores diretos dos quais descendem os sistemas de informação. Na faceta de tecnologia, buscou-se verificar características importantes nas ferramentas disponíveis para recuperação de informações textuais; na faceta pessoas, objetivou-se averiguar como as pessoas envolvidas com essas ferramentas as utilizam, bem como apontar suas características importantes; e na última faceta, processos, procurou-se verificar os processos atuais de geração dos documentos textuais, ressaltando que os processos futuros não foram alvo de análise devido à não implementação, até o momento, das ferramentas citadas.

Ao analisar criticamente a pesquisa realizada, alguns pontos foram identificados como merecedores de destaque.

Observe-se, inicialmente, que as pessoas entrevistadas foram muito solícitas, colaborando bastante para o desenvolvimento da pesquisa e as informações obtidas foram consideradas suficientes para atender o objetivo proposto.

Salienta-

intuitivas e de simples utilização, mas se mostram improdutivas, se comparadas às quantidades de documentos recuperados *versus* documentos úteis recuperados.

Nas ferramentas utilizadas, existem características mais importantes e utilizadas que outras, exemplificado na busca por texto livre, citado pela maioria dos entrevistados, independente da ordem, como campo de pesquisa.

As melhores ferramentas, na opinião dos entrevistados, são aquelas que mais se aproximam da linguagem humana, mostradas no decorrer da pesquisa, por técnicas de Inteligência Artificial.

Por fim, analisando-se os resultados, acredita-se que a pesquisa corrobore a afirmação de Laudon e Laudon (1999), quando dizem que “... A tecnologia mais avançada de computação é essencialmente sem valor, a menos que as empresas possam fazer uso da tecnologia e a menos que os indivíduos se sintam à vontade ao usá-la”.

## **10. Considerações finais**

A pesquisa demonstrou, no contexto do estudo realizado, como os pilares dos sistemas de informação (LAUDON e LAUDON, 2004) são importantes e se integram. Independente da qualidade das ferramentas analisadas compete às pessoas envolvidas a melhoria dos processos de criação dos documentos textuais, criando mecanismos de padronização para eles, viabilizando, desta forma, seu futuro processo de recuperação.

Notou-se neste contexto que, apesar de ainda não ser realidade, o órgão objeto de estudo, Tribunal de Justiça de São Paulo, caminha no sentido de unificação e padronização de suas informações textuais.

Esta pesquisa se propôs a analisar o problema de recuperação de documentos textuais e, para tal, foram identificadas algumas ferramentas que cumprem essa finalidade. Elas foram analisadas e, para verificar a viabilidade de sua implantação, foi realizado um estudo de caso no Tribunal de Justiça de São Paulo, no qual se verificou ser bastante viável a utilização de ferramentas baseadas em Inteligência Artificial, mais especificamente o RBC – Raciocínio Baseado em Casos.

O estudo de RBC evidenciou que a escolha da ferramenta apropriada facilita bastante a recuperação de informações textuais jurídicas, minimizando o tempo gasto nas novas decisões dos Srs. Magistrados, contribuindo, assim, para a melhoria da produtividade dos serviços prestados. Com isso, o trabalho atinge seus objetivos de estudar o cenário e elencar técnicas e ferramentas apropriadas para o problema levantado.

Outros pontos foram identificados ao longo desta trajetória e não foram pesquisados por não pertencerem ao foco abrangido pelos objetivos propostos, todavia, são sugeridos como continuidade desta pesquisa: Analisar a viabilidade do uso das ferramentas pesquisadas em outros tribunais, verificando, da mesma forma, como são tratados os documentos textuais nesses órgãos, e verificar a viabilidade da utilização das ferramentas pesquisadas em outros domínios de aplicação.

## Referências bibliográficas

ABEL, Mara. *Um estudo sobre Raciocínio Baseado em Casos*. Trabalho apresentado no Programa de Pós-Graduação em Ciência da Computação. Porto Alegre: UFRGS, jan. 1996.

ALENCAR, Maria S. M. *Mecanismos de busca na Web: Uma análise da metodologia de estudos comparados*. Tese de Mestrado em Ciência da Informação. Rio de Janeiro: Universidade Federal do Rio de Janeiro; Escola de Comunicação, 2001.

ALMEIDA, Dayse. C. Súmula vinculante. *Boletim Jurídico*, ano III, n. 150, ISSN 1807-9008. jul. 2005. Disponível em: <<http://www.boletimjuridico.com.br/doutrina/texto.asp?id=704>>. Acesso em: 01 nov. 2005.

ALMEIDA, Maria A. F. *Apostila de pós-graduação em Ciência da Computação*. Florianópolis: UFSC; Departamento de Informática e de Estatística – INE, 2000.

ARARIBOIA, G. *Inteligência artificial*. Rio de Janeiro: LTC, 1988.

ATIENZA, Cecília Andreotti. *Documentação jurídica: Introdução à análise e indexação de atos legais*. Rio de Janeiro: Achiamé, 1979.

AUTONOMY Corporation plc, Disponível em [http://www.autonomy.com/content/Products/IDOL\\_K2/index.en.html](http://www.autonomy.com/content/Products/IDOL_K2/index.en.html): Acesso em 13 dez. 2005

BANZHAF, W; NORDIN, P.; KELLER, R. E.; FRANCONI, F. D. *Genetic programming: An introduction*. ISBN 155860510X. Oxford, UK: Academic Press, 1998. 512 p.

BAUCHSPIESS, A. *Apostila do curso de engenharia elétrica: Introdução aos sistemas inteligentes – Aplicações em engenharia de redes neurais artificiais, Lógica Fuzzy e Sistemas Neuro-Fuzzy*. Brasília: Universidade de Brasília, março 2004. Disponível em: <<http://www.ene.unb.br/adolfo/ISI>> Acesso em: 03 jun. 2006.

BERRY, M. J. A.; LINOFF, G. *Data mining techniques*. New York: John Wiley & Sons, 1997.

BUENO, T. *Retrieval in jurisprudencial text bases using juridical terminology*. International Conference in Intelligence Artificial and Law – ICAIL, 1999.

BUENO, T. *Uso da teoria jurídica para recuperação de acórdãos criminais em sistema baseado em casos*. Dissertação de Mestrado. Programa de Pós-Graduação em Engenharia da Produção. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 1999.

BUENO, T. *Engenharia da mente: Uma metodologia de representação do conhecimento para construção de ontologias em sistemas baseados em conhecimento*. Tese de Doutorado. Programa de Pós-Graduação em Engenharia da Produção – PPGEP/UFSC. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2005.

CARMO, J. B. M. de S. *Planejamento estratégico na Eletronorte no período de 1996 a 2002 – implementação e resultados*. Dissertação de Mestrado – Centro Tecnológico-UFSC. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2003.

CARVALHO, Alexey; ALMEIDA, Mauricio A. A.; GUIMARÃES, Rogério C.; BARBOSA, Adilson A. *Inteligência artificial aplicada ao processo de tomada de decisões*. Bauru: XII SIMPEP, nov. 2005.

CINTRA, Antonio Carlos de Araújo et al. *Teoria geral do processo*. São Paulo: *Revista dos Tribunais*, 8. ed. 1991.

CONSTITUIÇÃO do Estado de São Paulo. Decreto-lei Complementar Estadual n.º 3 de 27/08/1969 e alterações posteriores. Disponível em: <[http://www.camara.gov.br/internet/interacao/constituicoes/costituicao\\_sp.pdf](http://www.camara.gov.br/internet/interacao/constituicoes/costituicao_sp.pdf)>. Acesso em: jul. 2006.

CONSTITUIÇÃO da República Federativa do Brasil. Título IV, Capítulo III, Seção I, artigo 92, 1998. Disponível em: <[http://legis.senado.gov.br/con1988/CON1988\\_04.03.1996/CON1988.pdf](http://legis.senado.gov.br/con1988/CON1988_04.03.1996/CON1988.pdf)> Acesso em: 10 out. 2005.

CONSTITUIÇÃO da República Federativa do Brasil. Capítulo I, artigo 5, 1998. Disponível em: <[http://legis.senado.gov.br/con1988/CON1988\\_04.03.1996/CON1988.pdf](http://legis.senado.gov.br/con1988/CON1988_04.03.1996/CON1988.pdf)> Acesso em: 10 out. 2005.

CONTANDRIOPOULOS, A. P. *Saber preparar uma pesquisa*. São Paulo: Atlas, 1999

CRATOCHVIL, A. *Data mining techniques in supporting decision making*. Master thesis. Universiteit Leiden, Leiden 1999.



COSTA, Felipe Correa da; HOESCHL, Hugo Cesar; BUENO, Tania Cristina D'Agostini; RIBEIRO, Erica Bezerra Quiroz. *A representação do conhecimento para criação de uma central inteligente para padronização dos procedimentos de licenciamento ambiental – CIPPLA. II CEBERÉTICA – Simpósio Internacional de Propriedade Intelectual, Informação e Ética*. Santa Catarina: 2003

DARWIN, C. *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. London: Murray, 1859.

DIAS, Maria M. *Um modelo de formalização do processo de desenvolvimento de sistemas de descoberta de conhecimento em banco de dados*. Tese de Doutorado em Engenharia de Produção. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2001.

FAYYAD, Usama; PIATETSKI, S.; Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. In: *Communications of the ACM*, p.27-34, Nov. 1996.

FIALHO, Francisco A. P. *Introdução ao estudo da consciência*. Pós-graduação em Engenharia de Produção. UFSC, 1999. Disponível em <<http://www.geocities.com/Athens/Sparta/1350/ia/introduc.html>> Acesso em jun. 2006.

FIGUEIREDO, Cristhiane Xavier; FRANCISCANI, Juliana de Fátima; CARDOSO, Olinda Nogueira Paes. *Recuperação de informações e bibliotecas digitais*. Semana de Ciência da Computação, Universidade Federal de Lavras – UFLA-MG, 2003.

GATHERCOLE, C. *An investigation of supervised learning in genetic programming*. PhD thesis. University of Edinburgh, Scotland, UK 1998.

GIL, Antônio C. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 1991.

GOEBEL, M.; GRUENWALD, L. A survey of data mining and knowledge discovery software tools. *SIGKDD Explorations*, June 1999.

GOMES, Luiz Flávio. Súmula vinculante e independência judicial. São Paulo: *Revista dos Tribunais*, n. 739/11, 1997.

GROTH, R. *Data mining: A hands-on approach for business professionals*. Saddle River, N.J: Prentice Hall, 1998.

GUIMARÃES, Deocleciano Torrieri. *Dicionário técnico jurídico*. 2. ed. São Paulo: Rideel, 1999.

GUIMARÃES, José Augusto Chaves. *Análise documentária em jurisprudência: uma metodologia de indexação de acórdãos para instrumentos terminológicos*. Tese de Doutorado apresentada ao Departamento de Biblioteconomia e Documentação da Escola de Comunicações e Artes da Universidade de São Paulo. Departamento de Biblioteconomia. São Paulo: Universidade de São Paulo – USP, 1994.

GUSMÃO, Paulo Dourado de. *História do direito*. Introdução ao estudo do direito. 19. ed. Rio de Janeiro: Forense, 1996.

HAMMOND, K. J. *CHEF: a model of case based planning*. AAAI-86. Proceedings. AAAI. Cambridge, MA: MIT Press, 1986.

HARRISON, Thomas H. *Intranet data warehouse*. São Paulo: Berkeley Brasil, 1998.

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

HOLLAND, J. H. *Adaptations in natural and artificial systems*. 2. ed. Cambridge, MA: MIT Press, 1975.

KOHONEN, T. *Correlation matrix memories*. IEEE Transactions on Computers. v. 21, 1972.

KOZA, J. R. *Genetic programming: On the programming of computers by means of natura selection*. Cambridge, MA: MIT Press, 1992.

LAAN, Regina Helena van der; FERREIRA, Gloria Isabel Sattamini; BONOTTO, Martha E. K. Kling; NEVES, Iara Conceição Bitencourt; GASPERIN, Inês M.. Avaliação de descritores relativos às ciências da informação: Relato de pesquisa. Porto Alegre: *Em questão*, v. 10, n.2, dez. 2004.

LAUDON, Kenneth C.; LAUDON, Jane P. *Sistemas de informação gerenciais* 5. ed. São Paulo: Pearson, 2004.

LAUDON, Kenneth C.; LAUDON, Jane Price. *Sistemas de informação* Rio de Janeiro: Livros Técnicos e Científicos, 1999.

LIMA, Cynthia M. *Linguagens para inteligência artificial*. Universidade Federal do Maranhão – UFM, 2005. Disponível em: <[http://cynthia\\_m\\_lima.sites.uol.com.br/ia2.htm](http://cynthia_m_lima.sites.uol.com.br/ia2.htm)> Acesso em: dez. 2005.

LIMA, Hermes. *Introdução à ciência do direito*. 22. ed.. Rio de Janeiro: Freitas Bastos, 1972.

LOH, S.; WIVES, L. K.; OLIVEIRA, J. P. M. Concept-based knowledge discovery in texts extracted from the web. *ACM SIGKDD*, July/2000, v. 2, n. 1.

LOH, Stanley. *Abordagem baseada em conceitos para descoberta de conhecimento em textos*. Tese de Doutorado. Instituto de Informática. Universidade Federal do Rio Grande do Sul.. Porto Alegre: Universidade Federal do Rio Grande do Sul – UFRGS, 2001.

MACHADO, Marcos O.; BORGES, Robson M. *Sistemas especialistas – sistemas de apoio à tomada de decisão*. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2005.

MAÇOLI, Fábio. *Análise documentária das ementas cíveis. Uma experiência com acórdãos do Tribunal de Justiça do Estado de São Paulo*. Dissertação de mestrado. Pontifícia Universidade Católica de Campinas. Campinas: Pontifícia Universidade Católica de Campinas – PUCCAMP, 2005.

MARÇAL, Rui Francisco Martins; SUSIN, Altamiro A. *O emprego de inteligência artificial como ferramenta de apoio à tomada de decisão na manutenção industrial*. Trabalho apresentado no III Seminário Catarinense – ABRAMAN. Joinville-SC, 2005.

MARCONI, M. A.; LAKATOS. E. M. *Fundamentos de metodologia científica*. 6 ed. São Paulo: Atlas, 2005.

MATTOS, Eduardo da Silva. *RBC-Text – Modelo para tratamento de documentos textuais e validação no domínio jurídico*. Dissertação de mestrado. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2002.

MELLO, Ronaldo dos Santos; DORNELES, Carina Friedrich; KADE, Adrovane; BRAGANHOLO, Vanessa de Paula; HEUSER, Carlos Alberto. *Dados Semi-Estruturados*. Universidade Federal do Rio Grande do Sul – UFRGS/ Universidade Federal de Santa Catarina – UFSC/ Universidade de Passo Fundo – UPF, 2003.

MESSITE, Peter. Sistemas de direito comum X Sistemas de direito civil. *Revista Questões de democracia* (revista eletrônica da Agência de Informações dos Estados Unidos), v. 4, n. 2, set., 1999.

MINKU, F. L.; POZO, A. T. R.; VERGILIO, S. R. *Chameleon: uma ferramenta de programação genética orientada a gramáticas*. Depto. de Informática. Curitiba: Universidade Federal do Paraná – UFPR, 2002.

MOBASHER, B.; COOLEY, R.; SRIVASTAVA, J. Automatic personalization based on web usage mining. In: *Communications of the ACM*, v. 43, n. 8, p. 142-151, 2000.

MODRO, Nilson Ribeiro. *Sistema inteligente de monitoramento e gerenciamento financeiro para micro e pequenas empresas*. Dissertação de Mestrado em Engenharia de Produção. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2000.

MORAES, Alessandra. *Súmulas vinculantes: Evolução ou retrocesso?* (2002). Disponível em: <<http://www.alessandramoraes.com/artigos/artigo07.pdf>>. Acesso em: nov. 2005.

MOSCOVE, Stephen A.; SIMKIN, Mark G.; BAGRANOFF, Nancy A. *Sistemas de informações contábeis*. Tradução Geni G. Goldschmidt. São Paulo: Atlas, 2002.

NIKOLOPOULOS, C. *Expert systems*. New York: Marcel Dekker, 1997.

NÓBREGA, J. F. *Introdução ao direito*. 2. ed. rev. amp. Rio de Janeiro: José Konfino, 1962.

O'BRIEN, James A. *Sistemas de informação e as decisões gerenciais na era da internet*. Tradução Cid Knipel Moreira. São Paulo: Saraiva, 2001.

O'REILLY, U. *An analysis of genetic programming*. PhD thesis. Ottawa-Carleton Institute for Computer Science. Ottawa: Carleton University, 1995.

PORTER, B.; BAREISS, R. *Protos: An experiment in knowledge acquisition for heuristic classification tasks*. Proceedings of the First International Meeting on Advances in Learning (MAL). Les Arcs, FR, 1986. p.159-174.

PY, Mônica X. *Sistemas especialistas: Uma introdução*. Instituto de Informática. Porto Alegre: UFRGS, 2003.

REVISTA Consultor Jurídico, 7 de maio de 2005. Disponível em <<http://conjur.estadao.com.br/>> Acesso em: 01 nov. 2005.

RISSLAND, E. L.; DANIELS, J. J. *Using CBR to drive IR*. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence IJCAI-95. Canada, 1995.

ROSA, F. A de Miranda. *Sociologia do direito: o fenômeno jurídico como fato social*. Rio de Janeiro: Zahar, 1970.

ROVER, Aires José. A tecnologia como fator de democratização do direito. *Seqüência, estudos jurídicos e políticos*. Florianópolis: Universidade Federal de Santa Catarina – UFSC, n. 35, dez. 1997.

ROWLEY, Jennifer. *A biblioteca eletrônica*. Brasília: Briquet de Lemos, 2002.

SCHANK, R.C.; ABELSON, R. *Scripts, plans, goals, and understanding*. Erlbaum: Northvale, 1977.

SECCO, Orlando de Almeida. *Introdução ao estudo do direito*. 2. ed. rev. e atual. Rio de Janeiro: Freitas Bastos, 1988.

SIMPSON, R. L. A computer model of case-based reasoning in problem solving: an investigation in the domain of dispute mediation. *Technical Report*. GIT-ICS-85/18. Georgia Institute of Technology, School of Information and Computer Science, 1985.

SOIBELMAN, Félix. Súmula vinculante na Emenda Constitucional n. 45/2004. *Jus Navigandi*, Teresina, a. 9, n. 618, 18 mar. 2005. Disponível em: <<http://www1.jus.com.br/doutrina/texto.asp?id=6392>> Acesso em: 08 jun. 2005.

SYCARA, E.P. Resolving adversarial conflicts: an approach to integrating case-based and analytic methods. *Technical Report*. GIT-ICS-87/26. Georgia Institute of Technology, School of Information and Computer Science, 1987.

TONINI, Antonio Carlos. *Sistemas ERP no Brasil*. São Paulo: Atlas, 2003.

TRIBUNAL de Justiça de São Paulo, Disponível em <http://www.tj.sp.gov.br>: Acesso em: 03 jan. 2007

VALLE, A. M. F. *Um modelo para implementação de consciência em robôs móveis*. Dissertação de Mestrado em Engenharia de Produção. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2003.

WANGENHEIM, Cristiane G. *Raciocínio baseado em casos*. Barueri: Manole, 2003.

WILLIS, M. J.; HIDDEN, H. G.; MARENBACH, P.; MCKAY, B.; MONTAGE, G. A. *genetic programming: an introduction and survey of applications*. The Second International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications, 1997.

YIN, Robert K. *Estudo de caso*, 2. ed. Porto Alegre: Bookman, 2001.

YU, T. *Structure abstraction and genetic programming*. Proceedings of the 1999 Congress on Evolutionary Computation (CEC99). ISBN 0780355369. IEEE Press, 1999.

## **Anexos**

### **Anexo 1**

Roteiro para entrevista semi-estruturada.

O roteiro aqui proposto visa analisar ferramentas e sua aplicabilidade dentro do cenário levantado e fomentar algumas premissas defendidas durante a execução do trabalho como a real importância dos documentos textuais no âmbito jurídico.

Antes da efetiva aplicação deste roteiro de entrevista, caracterizam-se previamente os documentos textuais jurídicos, atrelando-os a documentos como jurisprudências, súmulas, doutrinas, decisões anteriores de outros magistrados e sentenças, entre outros.

1. Caracterização do entrevistado.
  - 1.1. Nome
  - 1.2. Cargo
  - 1.3. Setor de trabalho
2. Necessidade do entrevistado quanto à utilização de documentos textuais em suas atividades.
  - 2.1. Para execução de suas atividades profissionais, como você classificaria a necessidade dos documentos textuais (previamente caracterizados):  
Indispensáveis, Importantes, Indiferentes, ou Sem importância.
3. Formas atuais de desenvolvimento do trabalho relacionado a documentos textuais.
  - 3.1. Quais as ferramentas utilizadas atualmente para os trabalhos de pesquisa?

- 3.2. Qual o número de palavras usadas em média para a pesquisa destes documentos? Por conta de limitações da ferramenta?
- 3.3. Levando-se em conta o volume de informação gerada *versus* o aproveitamento de cada uma, você consideraria as ferramentas em questão como produtivas? Motivos?
- 3.4. Qual sua opinião a respeito da usabilidade destas ferramentas? (definir previamente o termo usabilidade).
4. Características importantes inerentes a um sistema de consulta de documentos textuais.
  - 4.1. Os sistemas utilizados atendem plenamente ao propósito da pesquisa ou existe alguma particularidade que, se incorporada, contribuiria para a recuperação dos documentos? (explicar e exemplificar alguns exemplos de busca)
  - 4.2. O entrevistado teria a disponibilidade para analisar algumas ferramentas em suas versões de demonstração e comentar algo a respeito?
5. Em se tratando do modo como os documentos textuais são gerados, o entrevistado estaria disposto a alterar alguns processos de geração de documentos textuais se isso gerasse mais qualidade na recuperação destes mesmos documentos no futuro?
6. Existe algum ponto que não foi abordado e que o entrevistado considera relevante?



# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)