

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO PPGE/UFSC**

Pesquisa Jurisprudencial Inteligente

**Tese submetida à Universidade Federal de Santa Catarina
para obtenção do grau de Doutora em Engenharia**

ROSINA WEBER LEE

Orientador: Prof. Ricardo Miranda Barcia, Ph.D.

Florianópolis, 14 de maio 1998

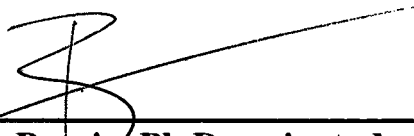
Pesquisa Jurisprudencial Inteligente

Rosina Weber Lee

Esta tese foi julgada adequada para obtenção do título de "Doutor em Engenharia", especialidade Engenharia de Produção e aprovada em sua forma final pelo Programa de Pós-graduação.

**Professor Ricardo Miranda Barcia, Ph.D.
Coordenador do Programa de Pós-Graduação em Engenharia de Produção**

Banca Examinadora:



Ricardo M. Barcia, Ph.D., orientador
Programa de Pós-Graduação em Eng. de Produção
Prof. Ricardo M. Barcia, Ph.D.
Coordenador

Karl Branting, Ph.D.

Ian Watson, Ph.D.

Alejandro Martins, Dr.

Fernando Alvaro Ostuni Gauthier, Dr.

Agradecimentos

Eu não estaria escrevendo estas palavras se não fosse por meu orientador, Professor Ricardo Barcia, e sua sutil condução para a direção certa. Obrigada por acreditar em mim.

Eu gostaria de agradecer aos três patrocinadores que me auxiliaram durante estes últimos anos permitindo-me total dedicação a esta pesquisa. Aos órgãos governamentais Capes e CNPq, especialmente na pessoa de Carlos Pittaluga cuja colaboração foi fundamental durante o período do programa *sandwich*. Ao Programa de Pós-Graduação em Engenharia de Produção (PPGEP) que, em função da audácia e visão de seu coordenador, Professor Ricardo Barcia, pode complementar recursos financeiros e manter recursos humanos tornando o PPGEP tão inigualável.

Jamais um empreendimento deste vulto poderia ser implementado individualmente. Todas as pessoas que contribuíram com este processo, sabendo ou não, merecem agradecimentos. Entretanto, há algumas que mencionarei.

A maior gratidão devo ao Professor Alejandro Martins por sua orientação, suas valiosas críticas e grande apoio durante o processo desta pesquisa, especialmente pela correção deste trabalho.

Também devo agradecimentos ao Professor Roberto Pacheco que muito contribuiu com meu aprendizado e evolução como pesquisadora e com sugestões e críticas construtivas ao trabalho.

Eu gostaria de agradecer aos membros da banca, especialmente aos examinadores externos por aceitarem o convite honrando-nos com sua presença.

Eu gostaria de agradecer ao Dr. Ian Watson por tão bem me receber na comunidade de RBC e por citar-me em seu livro.

Também fico agradecida ao Dr. Karl Branting cujas idéias tiveram forte influência sobre meu trabalho. Além disso, agradeço sua gentileza e consideração inclusive auxiliando-me com referências atualizadas da literatura de IA & Direito.

Desejo agradecer Dr. Fernando Gauthier por sua amizade e confiança com que sempre me presenteou desde os tempos do mestrado.

Agradeço ao meu co-orientador do programa *sandwich*, Dr. Suresh Khator, por suas valiosas sugestões, lições e cooperação de variadas formas durante este período.

Durante o período que estive na Universidade do Sul da Florida (USF) também pude contar com a ajuda de Dr. Paul Givens, responsável pelo departamento de Engenharia Industrial e Sistemas de Gerência (Dept. of Industrial & Mgmt. Systems), que ajudou a me sentir bem vinda nos EUA.

Eu sou imensamente grata a Eduardo Mattos por toda sua ajuda e todo o seu esforço ao programar o PRUDENTIA.

Eu gostaria de expressar minha gratidão para Tania Bueno e Hugo Hoeschel por inspirarem a idéia de utilizar RBC na pesquisa jurisprudencial e por incansavelmente dirimirem minhas dúvidas jurídicas e prontamente me atenderem sempre que preciso. Suas contribuições e sugestões foram de fundamental importância.

Ainda devo salientar que meus esforços no desenvolvimento desta pesquisa foram fortemente aprimorados pelas valiosas contribuições de Marcelo Eduardo de Oliveira, Antonio Carlos Marini Garcia, Cassiano Reis, Marcio C. da Costa, e Ilson Wilmar Rodrigues Filho.

Na minha jornada perseguindo a compreensão do domínio jurídico, devo ainda lembrar e agradecer a atenção, comentários e respostas proferidas pelos Professores Aires José Rover e Luiz Adolfo Olsen da Veiga que auxiliaram a aprofundar minha pesquisa. O nome do sistema, PRUDENTIA, devo ao Professor Olsen. - Obrigada!

Eu também gostaria de agradecer ao CIASC por disponibilizarem suas instalações para nossa equipe.

Eu sou muito grata a todos os outros colegas que, direta ou indiretamente, contribuíram com este trabalho. Quero agradecer à equipe do LIA, na pessoa do Neto, por todas suas sugestões e soluções.

Ainda desejo agradecer aos alunos e à talentosa equipe do PPGEP, especialmente à Rita, Lecir, Neiva, Rosangela e Airton, e a quem mais merecer que eu não tenha mencionado explicitamente.

Também agradeço a Marcelo Pezzi por ter me passado as cópias dos arquivos, incentivando e facilitando meu trabalho.

Finalmente, eu gostaria de agradecer a meus mais queridos amigos, especialmente a minhas irmãs e sobrinhas, por compreenderem minha ausência por tão longo período. Principalmente, eu gostaria de exaltar o mais especial de todos os agradecimentos a minha querida mãezinha, por toda sua devoção e paciência, todo o apoio e respaldos de toda a natureza, não somente durante este doutoramento, como em toda minha vida.

Dedico este trabalho à memória de João Pedro Osório, desejando que estudos científicos como este sobre reutilização de conhecimento, bem como outros na área médica, culminem com soluções que evitem outras tragédias como a que lhe sucedeu.

Resumo

Pesquisa Jurisprudencial Inteligente é uma proposta de um sistema inteligente para executar pesquisa jurisprudencial: a pesquisa executada por operadores do Direito na busca de decisões judiciais que possam ser úteis a uma nova situação jurídica por sua semelhança com a decisão passada. Quando a pesquisa jurisprudencial é praticada por humanos, emprega-se o raciocínio analógico ao comparar uma determinada situação atual com decisões passadas e apontar as afinidades entre elas. Raciocínio baseado em casos é uma técnica de Inteligência Artificial que modela aspectos de cognição humana para resolver problemas especialistas. Sistemas de RBC imitam o ato humano de recordar um episódio prévio para resolver um determinado problema devido à forte semelhança entre eles. No processo de recordar uma situação semelhante quando comparada a uma nova, sistemas de RBC simulam o raciocínio analógico. Assim, RBC é a tecnologia apropriada para lidar com o problema escolhido. PRUDENTIA é um sistema de raciocínio baseado em casos desenvolvido especificamente para a pesquisa jurisprudencial Brasileira. No Brasil, operadores do Direito têm duas fontes de pesquisa jurisprudencial: livros e sistemas de banco de dados. A procura em livros é demorada e imprecisa devido às limitações de memória dos humanos. Sistemas de banco de dados de textos disponíveis não garantem resultados eficientes em termos da utilidade dos documentos recuperados. Conseqüentemente, embora compensando limitações de memória dos humanos, esta ainda é uma opção pobre. Enfrentando esta carência de recursos, buscamos um sistema que melhore a eficiência da pesquisa em jurisprudências. Em vez de modelar a lei como um objeto real, optamos por modelar a Lei do modo como os profissionais interpretam fatos jurídicos. Os processos de aquisição de conhecimento tentam extrair de especialistas jurídicos o modo com que eles interpretam e vêem o mundo. Decisões judiciais são descritas através de textos em linguagem natural, exigindo grandes esforços de *engenharia do caso*. De uma perspectiva jurídica, um conjunto de documentos em forma de texto demanda um tratamento para sua compreensão; já, pelo enfoque da implementação de um sistema de RBC, exige-se um tratamento para modelar os documentos em uma representação estruturada. Desta forma, tornou-se necessário o desenvolvimento de meios automáticos para executar a engenharia do caso; essencialmente, referimo-nos à conversão automática de textos jurídicos em representações estruturadas para os casos, permitindo uma aplicação prática de um sistema de RBC no domínio jurídico. Exemplos no PRUDENTIA demonstram o poder de recuperação baseada em similaridade com relação a aplicações de banco de dados de texto, melhorando a qualidade (precisão e utilidade) dos documentos recuperados. Esta metodologia também pode ser estendida a outros domínios.

Conteúdos

| | |
|--|------|
| 1. Introdução | 1-11 |
| 1.1 Motivação | 1-13 |
| 1.2 Objetivos | 1-16 |
| 1.3 Fundamentos | 1-18 |
| 1.4 Importância..... | 1-21 |
| 1.5 Organização | 1-24 |
| 2. Inteligência Artificial e Direito | 2-25 |
| 2.1 Inteligência Artificial | 2-26 |
| 2.2 Engenharia do conhecimento | 2-27 |
| 2.2.1 Aquisição do Conhecimento | 2-27 |
| 2.2.2 Representação do Conhecimento | 2-29 |
| 2.2.2.1 MOPs | 2-29 |
| 2.2.2.2 Redes Semânticas | 2-30 |
| 2.2.2.3 <i>Frames</i> | 2-30 |
| 2.2.2.4 Conceitos, Objetos e Fatos | 2-31 |
| 2.2.2.5 Regras | 2-31 |
| 2.2.2.6 Representação Árvore | 2-31 |
| 2.2.2.7 Grafos Conceituais | 2-31 |
| 2.2.2.8 Representações Formulário | 2-32 |
| 2.3 Inteligência Artificial e Direito | 2-33 |
| 2.3.1 Conhecimento no Domínio do Direito..... | 2-39 |
| 2.3.2 Outros Esforços | 2-40 |
| 2.4 Conclusão | 2-43 |
| 3. Raciocínio Baseado em Casos | 3-44 |
| 3.1 Raciocínio Baseado em Casos..... | 3-45 |
| 3.2 Conhecimento Especialista em RBC..... | 3-49 |
| 3.3 Casos e Base de Casos | 3-51 |
| 3.4 Indexação | 3-53 |
| 3.5 Avaliação de Similaridade..... | 3-56 |
| 3.5.1 Métrica de Similaridade | 3-56 |
| 3.6 Ajuste da Situação | 3-58 |
| 3.7 Aprendizagem em RBC | 3-60 |
| 3.8 Conclusão | 3-61 |
| 4. Aquisição de Conhecimento de Textos | 4-63 |
| 4.1 Compreensão de Linguagem Natural..... | 4-64 |
| 4.1.1 Linguística Computacional..... | 4-64 |
| 4.2 Extração de Informação..... | 4-66 |
| 4.2.1.1 Extração de Informação baseada em conhecimento..... | 4-67 |
| 4.2.2 Template Mining..... | 4-67 |
| 4.2.3 Text windowing | 4-68 |
| 4.2.4 Documentos Auto-Explicativos..... | 4-68 |
| 4.2.5 Aquisição de Conhecimento de Textos | 4-68 |
| 4.2.6 Recuperação de Informação | 4-70 |
| 4.2.7 Recuperação de Textos..... | 4-73 |
| 4.2.7.1 Índices de Recuperação e Precisão | 4-74 |
| 4.2.7.2 Modelo Booleano ou <i>Exact Match</i> (Combinação Exata) | 4-75 |

| | |
|--|-------|
| 4.2.7.3 O Modelo de Vetor Espacial ou <i>Best Match</i> - Não Booleana | 4-76 |
| 4.2.7.4 Modelos Probabilísticos | 4-76 |
| 4.3 Conclusão | 4-77 |
| 5. Convertendo Textos em Casos: viabilização de uma aplicação prática .. | 5-78 |
| 5.1 Desenvolvimento da Metodologia | 5-79 |
| 5.1.1 Avaliação e Representação dos Casos | 5-79 |
| 5.1.2 Atributos e Índices | 5-81 |
| 5.1.2.1 Índices Jurídicos de Propósito Múltiplo | 5-83 |
| 5.1.3 Análise de Textos | 5-85 |
| 5.1.3.1 Associando Atributos às Subestruturas | 5-87 |
| 5.1.4 Conhecimento para a Extração de Informação | 5-87 |
| 5.1.5 Outros Métodos: Elaboração | 5-90 |
| 5.2 Implementação da Metodologia | 5-92 |
| 5.2.1 Extração de valores para a atribuição de índices | 5-92 |
| 5.2.1.1 Atributos Explícitos | 5-92 |
| 5.2.1.2 <i>Tipificação</i> | 5-93 |
| 5.2.1.3 <i>Fundamento</i> | 5-94 |
| 5.2.1.4 <i>Artigos Secundários</i> | 5-95 |
| 5.2.1.5 <i>Unanimidade</i> | 5-96 |
| 5.2.1.6 <i>Resultado</i> | 5-96 |
| 5.2.1.7 <i>Tema</i> | 5-98 |
| 5.3 Avaliação e Conclusão | 5-101 |
| 6. O Sistema | 6-105 |
| 6.1 PRUDENTIA | 6-106 |
| 6.2 Pesquisa Jurisprudencial e o Domínio do Direito | 6-107 |
| 6.2.1 O Escopo | 6-108 |
| 6.3 Arquitetura do Sistema | 6-109 |
| 6.3.1 Casos e a Base de Casos | 6-110 |
| 6.3.2 Atributos e Índices | 6-110 |
| 6.3.2.1 Índices Essenciais | 6-111 |
| 6.3.2.2 Índices Básicos | 6-111 |
| 6.3.3 Avaliação da Similaridade | 6-112 |
| 6.3.3.1 Avaliação de Similaridade dos Índices de Propósito Múltiplo | 6-113 |
| 6.3.3.2 Integrais Difusas | 6-115 |
| 6.3.4 Ajuste da Situação | 6-116 |
| 6.3.4.1 Metodologia | 6-119 |
| 6.4 PRUDENTIA: Estágios de Desenvolvimento | 6-121 |
| 6.5 Exemplos | 6-123 |
| 6.5.1 Exemplo "porte" | 6-123 |
| 6.5.1.1 Comparação | 6-128 |
| 6.5.2 Exemplo "abandono material" | 6-131 |
| 6.5.2.1 Comparação | 6-132 |
| 6.6 Verificação | 6-134 |
| 6.7 Conclusão | 6-137 |
| 7. Conclusões | 7-140 |
| 7.1 Desenvolvimentos Futuros | 7-146 |
| 8. Referências | 148 |
| 8.1 Bibliografia | 161 |

Lista de Tabelas

| | |
|--|--------------|
| Tabela 1 Implementações de Sistemas Especialistas no domínio do Direito. | 2-34 |
| Tabela 2 Sistemas de RBC no domínio do Direito. | 2-36 |
| Tabela 3 Resumo da comparação entre os projetos Flair e PRUDENTIA. | 2-42 |
| Tabela 4 Posição dos valores para os atributos na estrutura dos acórdãos. | 5-87 |
| Tabela 5 Resumo dos métodos empregados na mineração de valores para os atributos. | 5-98 |
| Tabela 6 Intervalo de valores de similaridade para <i>fundamento</i> . | 6-115 |

Lista de Figuras

| | |
|---|-------|
| Figura 1 O ciclo do RBC. | 3-46 |
| Figura 2 Estrutura retórica dos acórdãos. | 5-86 |
| Figura 3 Exemplo de subestruturas. | 5-88 |
| Figure 1 Figura 4 Partes de dois acórdãos onde o resultado é extraído. | 5-97 |
| Figure 2 Figura 5 Leitura do valor para resultado em dois acórdãos no Amzi! Prolog. | 5-98 |
| Figura 6 Interface introdutória do PRUDENTIA. | 6-106 |
| Figura 7 Arquitetura do sistema. | 6-109 |
| Figura 8 Atribuição do índice <i>tipificação</i> com auxílio da árvore de tipificações. | 6-118 |
| Figura 9 Usuário redige um resumo da nova situação. | 6-119 |
| Figura 10 Casos recuperados, exemplo "porte". | 6-124 |
| Figura 11 Resultados dos casos mais similares. | 6-125 |
| Figura 12 A representação tipo formulário do caso selecionado. | 6-126 |
| Figura 13 Trecho extraído do texto do acórdão. | 6-126 |
| Figura 14 Atributos dos casos recuperados. | 6-127 |
| Figura 15 Consulta para o exemplo "porte". | 6-128 |
| Figura 16 Resultados da consulta no banco de dados de texto. | 6-129 |
| Figura 17 Casos recuperados do exemplo "abandono material". | 6-132 |
| Figura 18 Caso alvo como parte da base de casos. | 6-134 |

Capítulo Um

1. Introdução

Pesquisa jurisprudencial inteligente é uma proposta para realizar pesquisa jurisprudencial com um sistema de Inteligência Artificial (IA). Pesquisa jurisprudencial é a busca empregada pelos operadores do Direito que procuram por decisões judiciais passadas que possam ser úteis a uma nova situação por sua semelhança. A tecnologia de inferência do sistema inteligente é Raciocínio Baseado em Casos (RBC), e PRUDENTIA é o sistema direcionado à jurisprudência brasileira que ilustra a proposta.

Raciocínio baseado em casos é uma técnica de IA que modela aspectos de cognição humana para resolver problemas especialistas. Sistemas de RBC imitam o ato humano de recordar um episódio prévio para resolver um determinado problema devido a forte semelhança entre eles. No processo de recordar uma situação semelhante quando comparado a uma nova, sistemas de RBC simulam o raciocínio analógico.

Pesquisa jurisprudencial é a procura por situações similares em uma fonte de decisões judiciais. Quando seres humanos pesquisam por jurisprudência, empregam o raciocínio analógico ao comparar uma nova situação com decisões passadas. Eles examinam as decisões comparando-

as e contrastando-as com a situação que motivou a pesquisa. A pesquisa jurisprudencial resulta em um conjunto de situações jurídicas semelhantes que podem substanciar diversas atividades judiciais.

As limitações de memória humana aliadas ao contínuo crescimento do volume de decisões armazenadas sugerem o desenvolvimento de um sistema computacional inteligente para apoiar os humanos na pesquisa jurisprudencial. A eficiência ao representar o raciocínio analógico faz de RBC a tecnologia indicada para orientar um sistema inteligente de pesquisa jurisprudencial.

PRUDENTIA— um sistema de RBC para pesquisa jurisprudencial — é um sistema interpretativo que recupera os casos mais úteis de uma fonte de decisões para auxiliar na pesquisa jurisprudencial. A meta definitiva deste projeto é proporcionar aos operadores do Direito uma fonte de pesquisas inteligente que possibilite um sistema judiciário rápido e eficiente. Neste RBC, os casos são as descrições de decisões judiciais que são escritas originalmente em textos em linguagem natural. A viabilidade deste sistema depende das necessidades de *engenharia de caso*¹ que envolve o desenvolvimento de um sistema de grande porte para abarcar todo o universo de casos jurídicos. Os casos jurídicos são escritos em linguagem natural, o que representa uma forma de difícil manipulação computacional, aumentando as necessidades de engenharia de conhecimento sobre os casos. Na medida que a codificação manual desta base de casos consiste em uma tarefa inviável, reunimos recursos e tecnologias na busca de meios automáticos de executar a engenharia dos casos. Esta tarefa, em essência, refere-se a uma forma de indexação automática dos casos.

¹ Esforços de engenharia de conhecimento empregados na modelagem dos casos.

1.1 Motivação

A pesquisa presente tem dois enfoques principais. O primeiro refere-se à viabilização de aplicações práticas de IA no domínio jurídico que vão além de projetos de pesquisa. O segundo é resolver o problema de pesquisa jurisprudencial que clama por uma ferramenta computacional confiável. Conseqüentemente, na procura de uma solução pragmática, precisamos inovar na forma de implementar IA no Direito já que os métodos tradicionais não têm sido capazes de cumprir os propósitos desejados e produzir soluções eficientes no domínio jurídico. Os relatos da literatura na área substanciam esta perspectiva que atribuímos à natureza incomum do Direito.

As pesquisas de aplicações de IA no domínio do Direito vêm seguindo os padrões de aplicações em outros domínios. Tradicionalmente, ao desenvolver sistemas de IA, adquire-se e representa-se o raciocínio empregado em determinada tarefa e aplica-se este raciocínio sobre uma representação do domínio em questão.

O Direito caracteriza-se por não se adequar às formas tradicionais de representação do conhecimento utilizadas na IA. O Direito, enquanto ciência e teoria, não se baseia em assertivas verdadeiras ou falsas; opostamente, o Direito prevê regras, normas e estatutos para regular, controlar e orientar a busca do valor verdade de uma afirmação. Podemos, então, vislumbrar a tarefa jurídica como uma prática de busca de valores verdade para os fatos. Assim, não há um modo plausível ou fácil de representar fatos verdadeiros ou falsos através de quaisquer formalismos quando seus valores verdade nem mesmo foram atribuídos ou se não há sequer um senso comum sobre estes valores.

Em vez de modelar o Direito como um objeto real, encaramos o Direito como um objetivo abstrato que é perseguido pelos operadores do Direito. O ensino do Direito trata-se da divulgação de um corpo de princípios da teoria jurídica, o que é sobrepujado por seu real conteúdo: a visão jurídica do mundo. Este é o objeto real que escolhemos representar, o modo com que os profissionais jurídicos interpretam fatos jurídicos. A principal fonte de conhecimento sobre esta visão de conhecimento jurídico reside na memória dos operadores do Direito. A aquisição de conhecimento busca extrair o modo com que os especialistas jurídicos interpretam e vêem o mundo.

No Brasil, os operadores do Direito têm duas fontes de pesquisa jurisprudencial: livros e sistemas de banco de dados. A procura em livros é executada por humanos que empregam raciocínio analógico na busca por decisões semelhantes. As desvantagens principais desta forma de pesquisa originam-se nas limitações da memória humana que tornam esta tarefa demorada e imprecisa.

Os sistemas disponíveis no mercado, em sua maioria, são bancos de dados de textos implementados com ementas de decisões judiciais. Blair & Maron (1985) indicam que sistemas de banco de dados de texto estão limitados a um índice de recuperação (*recall*) de 25%; indicando que somente 25% dos documentos relevantes presentes na base serão recuperados. Conseqüentemente, embora compensando as limitações da memória humana, esta opção ainda é insuficiente devido a sua baixa recuperação. Perante esta escassez de soluções, buscamos um sistema que melhore a eficiência da pesquisa jurisprudencial aumentando a qualidade e precisão da pesquisa. Assim, representando o raciocínio de analógico em um sistema de RBC, podemos fazer o devido uso do potencial de memória dos computadores para efetivamente colocá-lo a serviço dos seres humanos.

A contribuição mais importante do presente trabalho é oferecer um sistema eficiente que suporte a pesquisa jurisprudencial. Sempre que uma decisão judicial passada não for recuperada, uma injustiça pode prevalecer. Na medida em que a memória humana não é capaz de armazenar e avaliar várias situações passadas, é importante que a ciência desenvolva a tecnologia para compensar os humanos em tal limitação. Não seria justo privar os humanos de viajarem de avião ou de construir arranha-céus devido a suas limitações de memória. As tecnologias computacionais vêm ajudando os humanos a alcançar várias metas e superar muitas dificuldades. Esta pesquisa é uma tentativa para orientar a tecnologia hoje disponível na busca por uma sociedade mais justa.

1.2 Objetivos

A meta principal desta pesquisa é viabilizar o desenvolvimento de um sistema para pesquisa jurisprudencial que seja mais eficiente que os sistemas de banco de dados de texto disponíveis no mercado em termos da utilidade das decisões judiciais recuperadas.

Raciocínio baseado em casos é a tecnologia ideal para lidar com o problema escolhido: pesquisa jurisprudencial. Da perspectiva do domínio jurídico, temos uma vasta coleção de textos que representam problema específico que demanda tratamento. Na luz da IA, a engenharia de desenvolvimento de um sistema de RBC envolve a definição da fonte de conhecimento e a engenharia dos casos. Conseqüentemente, alcançamos o nível onde o objetivo secundário é converter documentos textuais em representações de casos em forma de formulários. Contudo, não devemos esquecer do aspecto global da implementação computacional que soma a este objetivo o caráter automático fazendo nosso objetivo mais específico a conversão automática de textos jurídicos em representação estruturada de casos (formulários).

Na busca para atingir os objetivos propostos, há metas correlatas como a delimitação do conteúdo dos casos, a definição de atributos que contemplem a representação do tipo formulário e os métodos de extração de valores para os atributos a partir dos textos jurídicos. A conclusão dos passos acima mencionados encerra um estágio intermediário na solução do problema escolhido.

Acima de tudo, objetivamos com o PRUDENTIA uma busca eficiente, que proporcione a recuperação de documentos que sejam úteis e relevantes na pesquisa. Pesquisa jurisprudencial vem sendo executada por humanos que

fazem uso de livros ou sistemas de banco de dados de texto. Assim, também é nosso objetivo demonstrar que empregando a tecnologia de RBC, podemos alcançar um desempenho melhor que os sistemas de banco de dados de texto existentes.

O resultado é um sistema que, a partir da entrada da nova situação que motivou a pesquisa, pode oferecer ao usuário todos os documentos que podem ser úteis em ordem de similaridade. O aumento da qualidade no índice de recuperação (*recall*) observa-se na medida que todos os documentos úteis são mostrados aos usuários. Assim, sempre que houver alguma característica relevante, este documento é apontado entre os mais similares. Desta forma, o usuário tem um resultado mais preciso sem a necessidade de selecionar somente os documentos relevantes entre os recuperados.

1.3 Fundamentos

Os objetivos da presente pesquisa estão orientadas a dois campos de conhecimento. Um deles refere-se a uma meta específica: de manipular textos jurídicos automaticamente. Desta forma, conduzimos nossa atenção para temas ligados à Leitura Automática de Textos, Compreensão de Texto, Compreensão de Linguagem Natural, Recuperação de Textos, Extração de Informação e Extração de Conhecimento de Textos. O outro campo de conhecimento refere-se às aplicações de RBC no Direito e conduz nossa atenção à teoria de RBC, sistemas jurídicos de RBC, IA no Direito, e em Engenharia do Conhecimento. Segue-se uma breve introdução.

A leitura automática de textos tem sido o foco de várias pesquisas como a de Cardie & Lehnert (1991), Lehnert & Sundheim (1991), Smeaton (1995), Riloff (1993), Waltz e Pollak (1985), Masand et al., e Ram (1991), estes das áreas de Compreensão de Linguagem Natural, Extração de Informação, Recuperação de Informação, como também RBC. A compreensão de textos é um dos problemas principais tratado na compreensão da linguagem natural de onde derivou-se o novo campo destinado à extração de informação de grandes volumes de texto.

Uma direção importante na pesquisa é o uso de técnicas da Extração de Informação que facilitam a extração automática de valores de textos para convertê-los em casos. Técnicas de compreensão de linguagem natural não se ajustam ao problema de procurar índices em domínios complexos. Smeaton, (1995) argumenta que, *"isto acontece porque técnicas de programação de linguagem natural derivam de aplicações como tradução automática e interfaces de linguagem natural que são simplesmente muito diferentes de recuperação de informação que tem tantos níveis de liberdade*

e imprecisão". Daniels e Rissland (1995) propuseram um sistema híbrido de RBC e Recuperação de Informação (IR) que usa os textos relacionados aos casos mais similares recuperados pelo módulo de RBC para gerar a consulta no sistema de IR. A consulta é gerada com os n termos mais importantes presentes no texto. Um dos fundamentos desta abordagem é a alegação de que textos não são tratáveis por métodos baseados em conhecimento, resultando em que não é explorada a utilidade da recuperação baseada em similaridade da tecnologia de RBC. Elas aumentaram a eficiência do índice de recuperação (*recall*); contudo, a consulta é gerada a partir dos resultados de um sistema de RBC com bases entre 25 e 45 casos. Salomon (Uyttendaele et al., 1996) também explora o processamento automático de textos jurídicos objetivando facilitar seu acesso. O sistema extrai informações dos textos para atribuir valores a nove atributos, que, segundo eles, podem determinar a relevância dos textos. O número fixo de nove atributos talvez limite a utilidade dos textos recuperados. Além disso, o uso de métodos estatísticos ao invés de abordagens baseadas em conhecimento, empobrece a qualidade do resultado.

A revisão da literatura aponta aplicações de sistemas inteligentes no domínio do Direito que utilizam complexos métodos de indexação, demandando, desta forma, grandes esforços de engenharia de caso. Alguns destes são HYPO (Ashley & Rissland, 1988a, 1988b), GREBE (Branting, 1991c), CABARET (Rissland & Skalak, 1991) e BankXX (Rissland et al., 1993 & 1996). As demasiadas necessidades de engenharia de caso limitam a implementação destes sistemas na forma de ferramentas comerciais realmente capazes de resolver problemas práticos.

A distinção entre a **teoria do domínio** e os **fatos da vida** tem sido praticada por muitos autores da Inteligência Artificial e Direito tais como Smith (1987, 1997a, 1997b), Branting (1991c), e Valente (1995). A tentativa de separar estes dois diferentes elementos nas experiências jurídicas não tem se demonstrado eficiente, na medida que todos estes

pesquisadores ainda procuram por abordagens que realmente possibilitem o desenvolvimento de sistemas inteligentes que efetivamente resolvam problemas práticos no universo do Direito. Desta forma, sugerimos a interpretação das experiências jurídicas de uma forma próxima à humana, que mantenha a teoria do domínio e os fatos da vida no mesmo prisma, ao invés de separar tais elementos.

Os operadores do Direito são os especialistas que interpretam e lidam com as experiências jurídicas. A especialidade dos advogados é, em essência, interpretar os textos jurídicos que descrevem experiências jurídicas. Assim, esta é a habilidade que acreditamos que deva ser representada num sistema inteligente para o domínio jurídico: a especialidade dos advogados ao interpretar as experiências jurídicas. Conseqüentemente, a estrutura dos casos deve seguir as diretrizes que compreendem o conhecimento obtido dos especialistas jurídicos e sua interpretação é o conhecimento que deve ser extraído.

1.4 Importância

A implementação de um sistema inteligente para apoiar a pesquisa jurisprudencial na forma como foi introduzida pode aumentar a velocidade desta tarefa por causa de três razões. Primeiro, uma recuperação mais confiável resultante da recuperação baseada em similaridade satisfaz o usuário evitando a necessidade de diversas execuções do programa até encontrar as decisões úteis. Segundo, um *ajuste de situação*² inteligente ajuda o usuário a informar a situação que originou a pesquisa evitando a necessidade de construção de consultas complexas. Terceiro, a interpretação especialista das experiências gera um resumo que evita a necessidade do usuário ler todo o texto para identificar sua utilidade. Estes três aspectos fundamentam o incremento de qualidade para a pesquisa jurisprudencial, resultantes tanto da redução de tempo destinado à pesquisa ou das diretrizes baseadas em conhecimento especialista, que conduzem a implementação do sistema sempre enfocando e resultando em melhor qualidade.

A demanda por uma ferramenta eficiente para auxiliar a pesquisa jurisprudencial tem dois ramos: o privado e o público. O mercado brasileiro vem indicando que esta ferramenta eficaz e eficiente tem um grande potencial de demanda na medida que novas versões de diversas empresas são freqüentemente lançadas no mercado. Entretanto, todas estas versões são implementações de bancos de dados de textos. Além disso, o aumento na qualidade sem qualquer desvantagem com relação aos sistemas existentes tende a fazer de seu desenvolvimento um investimento atraente.

² Etapa da engenharia de conhecimento de um sistema de RBC referente ao tratamento da entrada

A demanda governamental por uma ferramenta computacional para a pesquisa jurisprudencial é uma responsabilidade com a justiça. Todo o sistema judiciário pode beneficiar-se com uma ferramenta de tal qualidade e eficiência. A qualidade é melhorada na medida em que a pesquisa fica melhor sustentada, proporcionando maior consistência entre as decisões. A eficiência é aprimorada através de uma produção mais rápida que resulta em economia de tempo e recursos. Um sistema judiciário célere é o caminho para uma sociedade justa e honrada.

Um grande benefício do desenvolvimento de sistemas de RBC no domínio jurídico é fazer possível a reutilização do conhecimento embutido nas jurisprudências que são usadas como referências em novas decisões judiciais. Melhorando o acesso aos casos passados aumenta-se o horizonte de fundamentação das novas decisões e eleva-se a qualidade dos resultados do sistema judiciário. Demonstramos estas melhorias dentro do escopo específico de decisões do Tribunal de Justiça do Estado de Santa Catarina (TJSC).

Os tribunais de justiça produzem textos jurídicos descrevendo decisões sobre recursos interpostos em primeira instância. O fato destes textos serem altamente estereotipados é uma das razões que viabiliza nossa abordagem. Entretanto, o princípio fundamental é conhecimento – **conhecimento especialista**, isto é que realmente viabiliza a extração automática de informação. Nossa metodologia elucida como criar grandes sistemas de raciocínio baseado em casos que comportem grandes bases de casos reais, demonstrando que textos dentro de domínios específicos e limitados de conhecimento são, de fato, tratáveis por métodos baseados em conhecimento como os utilizados na tecnologia de RBC. A abordagem proposta pode ser generalizada para ser destinada a diferentes domínios fazendo do paradigma de RBC uma poderosa tecnologia para resolver problemas práticos com extensas bases de conhecimento.

Os benefícios desta metodologia são muitos. Qualquer organização que mantenha registros armazenados de forma digital pode fazer uso de um sistema de RBC para reutilizar o conhecimento contido nesses registros, independente de seu tamanho.

Assim, torna-se viável o uso de sistemas de RBC em problemas práticos do mundo real. A metodologia aqui proposta representa um meio de modelar bases de casos para alavancar conhecimento especialista existente, conforme já sugerido por Klahr (1996). Conseqüentemente, a importância dos sistemas de RBC e das aplicações da IA fica reiterada.

1.5 Organização

Esta tese trata essencialmente sobre a solução proporcionada pela metodologia para converter textos em casos viabilizando a aplicação de RBC na pesquisa jurisprudencial.

O capítulo 1 apresenta a introdução e o 2 oferece alguns fundamentos teóricos relacionados a IA e Direito. No capítulo 3, resumimos a teoria de RBC. A aquisição de conhecimento de textos é tratada no capítulo 4. A metodologia é descrita no capítulo 5 e o sistema viabilizado por ela é apresentado no capítulo 6, seguindo-se as conclusões no capítulo 7.

Capítulo Dois

Em Branting (1991c), *"There is a broad consensus that rule-based reasoning is insufficient in itself to model legal analysis (McCarty & Sridharan, 1982; Gardner, 1984; Burton, 1985)³"*.

Em J.C. Smith (1997b), *"Attempts to make machines think like lawyers by representing legal knowledge in the form of doctrinal rules have failed⁴"*.

2. Inteligência Artificial e Direito

O objetivo deste capítulo é introduzir alguns aspectos essenciais das tecnologias de IA que foram considerados, aplicados ou até mesmo renovados na presente pesquisa bem como as tecnologias de IA que já foram empregadas no domínio jurídico. Introduzimos alguns dos tópicos principais em IA e Engenharia do Conhecimento na luz de sua utilidade para o domínio jurídico. Como o domínio em enfoque é o Direito, introduzimos IA&Direito com algumas aplicações e particularidades do domínio jurídico. Por fim, apresentamos um método alternativo e comparamos com a pesquisa proposta.

³Há consenso de que raciocínio baseado em regras é insuficiente para medelar a análise jurídica.

⁴Tentativas de fazer as máquinas pensarem como advogados através da representação do conhecimento jurídico na forma de regras doutrinárias falhou.

2.1 Inteligência Artificial

Destaques importantes da história de IA vão de Turing a Minsky, de Holland a Schank. O primeiro estágio na história de IA inicia-se com McCulloch e Pitts (1943) e é conhecido como a era do "*Olhe, mamãe, sem as mãos!*" (Look, Ma, no hands era), (Russel & Norvig, 1995). Um segundo período pode ser caracterizado pela busca de Newell e Simon pelo *General Problem Solver* (GPS), um resolvidor de problemas genéricos (1961). Hoje, a IA envolve o desenvolvimento de sistemas de computadores inteligentes que tratam de problemas em *microworlds* (microdomínios). Definido por Minsky (Russel & Norvig, 1995), *microworlds* se referem a domínios limitados de aplicação. Esta era mais recente também é caracterizada pela busca pela técnica mais apropriada para resolver cada tipo de problema. Esta pesquisa é uma instância nessa busca. No andamento desta era, pesquisadores da IA confrontam-se com o problema de como desenvolver sistemas capazes de resolver problemas complexos compostos por distintas tarefas. Na medida em que deve haver uma técnica mais apropriada para resolver cada tarefa, culmina hoje a era dos agentes inteligentes, onde várias ferramentas de IA são combinadas em um único sistema inteligente que administra forças para resolver tarefas.

Inteligência Artificial (IA) é o campo de estudo destinado ao desenvolvimento de sistemas computacionais que simulam algum aspecto de cognição humana. São discernidas duas orientações a partir desta definição; a tecnológica, associada ao desenvolvimento de programas inteligentes; e o científico, que trata dos aspectos teóricos de cognição humana.

2.2 Engenharia do conhecimento

Um sistema inteligente é desenvolvido com conhecimento especialista sobre o domínio da aplicação e engenharia do conhecimento. A engenharia do conhecimento, por sua vez, contribui com o conhecimento especialista sobre extração de conhecimento, representação do conhecimento e linguagens de IA (Luger & Stubblefield, 1993).

Os aspectos da engenharia do conhecimento e IA que são considerados nas seções seguintes são principalmente contextualizados com Sistemas Especialistas e sistemas de RBC. Isto deve-se ao fato de outras tecnologias não terem produzido pesquisa significativa nem na única publicação periódica de IA&Direito⁵, nem nos congressos ICAIL (*International Conference on AI&Law*) ou em extensos artigos como o de Greinke (1994). Há apenas um artigo sugerindo o uso de redes neurais, Hunter (1996).

2.2.1 Aquisição do Conhecimento

Aquisição de conhecimento se refere à tarefa executada pelos engenheiros de conhecimento extraíndo conhecimento especialista para ser representado em sistemas inteligentes. A aquisição de conhecimento normalmente é comentada quando comparam-se sistemas de RBC com tecnologia de sistemas especialistas.

Os Sistemas Especialistas (SE) compreendem os sistemas baseados em conhecimento que executam tarefas especialistas (Durkin, 1994); eles podem ser classificados em sistemas baseados em regras, baseados em

⁵ Artificial Intelligence e Law. Editors: Carole Hafner, Giovanni Sartor, Anne Gardner, Marc Lauritsen, e Richard Susskind. Kluwer Academic Publishers.

frames e sistemas indutivos. Além de poder executar uma tarefa especialista dentro de um microdomínio, as necessidades de engenharia de conhecimento caracterizam os SE's. Mesmo em sistemas especialistas indutivos onde regras são derivadas por algoritmos indutivos, a essência do domínio ainda tem que ser representada por algum formalismo. Os SE's baseados em *frames* fazem uso de regras de produção dentro de métodos que conectam diferentes *frames* para representar o conhecimento sobre as interações entre as entidades. Sistemas especialistas baseados em regras demandam um processo de aquisição de conhecimento para extrair conhecimento genérico sobre o domínio, conhecimento procedural e heurístico para representá-los através de regras num motor de inferência. Assim, apesar das variações que podem assumir, SE's apresentam grandes necessidades de aquisição de conhecimento.

As necessidades de aquisição de conhecimento no desenvolvimento de sistemas de RBC são reduzidas em comparação às de SE's, e isto deve-se às já reduzidas necessidades de representação do conhecimento. Como apresentado na seção 3.2, o conhecimento especialista é representado em vários momentos do desenvolvimento de um sistema de RBC e este conhecimento varia em sua natureza. Assim, a extração de conhecimento é uma etapa necessária no desenvolvimento de sistemas de RBC que pode ser tanto extensa, como praticamente inexistente, isto variando de acordo com a tarefa do sistema e o domínio da aplicação. É possível desenvolver um sistema de RBC usando conhecimento embutido em um banco de dados. Neste caso, a única necessidade é de modelar a similaridade entre os casos. Conseqüentemente, a aquisição de conhecimento depende da necessidade de representação de conhecimento. Na próxima seção, discutimos estes assuntos com mais detalhes.

2.2.2 Representação do Conhecimento

Formalismos como redes semânticas, *scripts* e *frames* podem ser empregados para representar casos em sistemas de RBC e representar domínios em SE's. Regras representam conhecimento procedural em SE's e também são empregadas para representar conhecimento de adaptação em sistemas de RBC. Alguns formalismos são descritos a seguir.

2.2.2.1 MOPs

Pacotes de organização de memória (MOPs-*Memory Organization Packets*) são um conceito de representação de conhecimento desenvolvido por Schank (1982) para representar eventos estereotípicos. MOPs são uma evolução dos *scripts*⁶ - o conceito original. Os MOPs são organizados dentro de estruturas que reúnem eventos com características similares através de abstrações e hierarquias do tipo todo-parte. Com relação ao conteúdo, os MOPs são estruturas de conhecimento que representam experiências. Os MOPs representam eventos através de *cenos* que incluem situações e são representadas através de informação normativa e descritiva. *Cenas* são expectativas associadas às situações de uma experiência, conseqüentemente elas estão sujeitas a mudar com a experiência (Schank, Kass, & Riesbeck, 1994). Além disso, o conceito de um *script* não é compartilhado entre os indivíduos (Riesbeck & Schank, 1989), na medida que cada memória compreende um *script* sobre uma experiência a partir de seu próprio ponto de vista.

A entidade básica do modelo de Memória Dinâmica permite representar computacionalmente um modelo de organização de memória que compreende recordar, entender, experienciar e aprender.

⁶ See Kolodner, chapter 4 for a complete comparison of *scripts* e MOPs.

2.2.2.2 Redes Semânticas

Redes semânticas são grafos direcionados ligados por nós para representar objetos e conexões que representam a relação entre objetos (Quillian, 1968; Brachman, 1979; Dean et al., 1995). Redes semânticas são usadas para representar elementos de uma representação tal como uma classe, suas instâncias e suas características. Os arcos são direcionados e representam relações e atributos. Redes semânticas exigem tratamento de exceção (*exception handling*) quando um determinado atributo não deve ser herdado (Durkin, 1994).

2.2.2.3 Frames

Um *frame* é uma estrutura de dados que representa uma entidade através de suas características e potencialidades. De modo geral, as características são representadas através de pares atributo-valor e as potencialidades são representadas por métodos. Um *frame* abstrato (ou um *frame* de classe) não tem instâncias, assim seus atributos não são valorados, suas subclasses são ligadas a instâncias da entidade representada por esta classe. Conforme o autor, (Minsky, 1975),

*"A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed."*⁷

⁷ Um *frame* é uma estrutura de dados para representar uma situação estereotipada, como ir a um certo tipo de sala-de-estar ou ir a uma festa de criança. Anexado a cada *frame* estão vários tipos de informações. Parte destas informações referem-se a como usar este *frame*. Parte refere-se a sobre o que esperar por acontecer. Parte refere-se a o que fazer se tais expectativas não se confirmarem.

2.2.2.4 Conceitos, Objetos e Fatos

Um objeto é uma entidade básica que pode ser instanciada. Um conceito conta algo sobre o objeto. Um conceito pode ser representado por uma abstração de um objeto quando vários objetos podem ser agrupados sob o mesmo conceito; ou um conceito pode ser um atributo, quando conta algo exclusivamente sobre esse objeto ou se para fins de organização da análise como um todo não vale a pena representá-lo como uma abstração. Esta decisão repousa na análise global do conhecimento e enfoca especialmente o uso da representação. Quando um objeto é associado a um atributo valorado, isto é um fato. Um fato pode assumir valores verdadeiro ou falso (Durkin, 1994).

2.2.2.5 Regras

Regras são seqüências lógicas compostas por antecedentes (premissas) e conseqüentes (conclusões). Ambos antecedente e conclusão são fatos. O antecedente tenta verificar se o fato é verdadeiro ou falso; quando o fato que compõe o antecedente é verdadeiro, a conclusão é disparada. O antecedente pode ser composto de vários fatos conectados através de operadores do tipo E, OU, e NÃO. As conclusões normalmente modificam ou atribuem valores aos atributos de um objeto, chamam métodos ou disparam outras regras (Waterman, 1986; Durkin, 1994).

2.2.2.6 Representação Árvore

A árvore é um grafo direcionado que contém nós para representar regras e conexões que indicam as relações entre antecedentes e conseqüentes (Schneider et al., 1996).

2.2.2.7 Grafos Conceituais

Grafos conceituais são uma variedade de redes semânticas de onde estas herdam o poder de representar significados (Sowa, 1984). Uma definição

completa é dada por Cyre (1997), “A *conceptual graph* is a finite, connected, bipartitioned graph consisting of a set of labeled concept nodes, a set of labeled conceptual relation nodes, e a set of (directed) arcs linking concept e relation nodes”⁸. Um grafo conceitual é um formalismo que inclui as características necessárias para modelar a semântica da linguagem natural. Eles vêm sendo usados para representar significado na compreensão da linguagem natural (Luger & Stubblefield, 1993).

2.2.2.8 Representações Formulário

A representação formulário é composta por um conjunto de campos valorados. Este formato lembra registros de bancos de dados. Este formalismo é usado nos sistemas de RBC para representar casos em estruturas organizacionais planas (Kolodner, 1993).

Os formalismos enumerados acima são esforços para representar conhecimento em termos de teoria do domínio e fatos da vida em sistemas de IA. Dependendo da natureza do domínio e do próprio conhecimento, escolhe-se o formalismo apropriado. Este resumo sobre representação do conhecimento conclui os tópicos de deste capítulo. Os próximos tópicos relacionam-se à IA&Direito em termos de suas aplicações e da natureza do seu conhecimento.

⁸ Um grafo conceitual é um grafo finito, conectado e biparticionado que consiste de um conjunto de nós conceituais rotulados, um conjunto de nós relacionais conceituais rotulados e um conjunto de arcos (direcionados) ligando nós conceituais e nós relacionais.

2.3 Inteligência Artificial e Direito

A literatura da Inteligência Artificial descreve diversos esforços para modelar o raciocínio jurídico. Entretanto, a utilização da IA no domínio do Direito deve ser vista por uma perspectiva mais ampla, na medida que a meta final é desenvolver sistemas que auxiliem satisfatoriamente as atividades jurídicas.

Primeiramente, consideramos necessário uma espécie de classificação do campo IA&Direito. A primeira distinção faz-se a partir dos ramos científico e tecnológico. A agenda científica refere-se à pesquisa do raciocínio jurídico a partir de sua natureza cognitiva. Os aspectos cognitivos que embasam o raciocínio jurídico, na construção do argumento jurídico, sua contextualização e a resolução de conflitos normativos, são objeto de estudo de pesquisadores nesta agenda. Relevantes expoentes são Bench-Capon (1995), Prakken e Sartor (1995), Hage (1995) e JC Smith (1997b). Na condução da discussão para o escopo de interesse da presente, enfocamos o aspecto tecnológico associado às aplicações de IA&Direito. O âmago da agenda tecnológica refere-se à modelagem das diferentes formas de raciocínio praticadas nas atividades jurídicas a fim de desenvolvermos sistemas realmente capazes de resolver problemas práticos deste domínio.

Os primeiros sistemas inteligentes no domínio do Direito foram Sistemas Especialistas. Há referências desde 1975, conforme descrito na Tabela 1 Implementações de Sistemas Especialistas no domínio do Direito. Todavia, a grande maioria destas aplicações limitam-se a protótipos de pesquisa. J.C. Smith (1997b) alega, *“Attempts to make machines think like lawyers by representing legal knowledge in the form of doctrinal rules have failed”*⁹.

⁹ "As tentativas de fazer as máquinas pensarem como os advogados através da representação do

| | |
|-----------------------|---|
| Nome | tarefa/ferramenta de IA/comentários/base de conhecimento |
| DSCAS | Usa conhecimento jurídico para analisar reclamações de construção civil; Kruppenbacher, 1984; regras com encadeamento para frente; protótipo de pesquisa; plataforma: ROSIE. |
| JUDITH | um ambiente de IA para explorar o raciocínio jurídico para auxiliar advogados a raciocinarem sobre casos cíveis sugerindo premissas adicionais; Popp & Schlink, 1975; plataforma: FORTRAN; base de conhecimento contendo as premissas. |
| EVIDENT | Auxilia alunos de Direito a aprenderem sobre a admissibilidade de provas em normas federais; Liebowitz, 1988; baseado em regras, encadeamento para trás; plataforma EXSYS; base de conhecimento com regras sobre provas. |
| LDS | Determina o valor devido em reclamações de responsabilidade civil, através da análise da perda, da responsabilidade, das características e contexto; Waterman & Peterson, 1980, 1981; baseado em regras; desenvolvido na Re Corporation; plataforma: ROSIE. |
| LEGAL ANALYSIS SYSTEM | Relaciona um conjunto de fatos com a doutrina referenciando decisões judiciais; Meldman, 1975; redes semânticas; MIT, protótipo de pesquisa; base de conhecimento com conhecimento especialista jurídico, doutrinas e fatos. |
| LRS | Recupera informação sobre decisões judiciais e legislação no domínio de instrumentos negociáveis; Hafner, 1981; redes semânticas; protótipo de pesquisa; Universidade de Michigan; base de conhecimento: seis conceitos primitivos: partes, instrumento jurídico, obrigação, ação jurídica, conta e montante. |
| SAL | Estima quanto deve ser pago às vítimas expostas a minerais (<i>asbestos exposure</i>); Waterman, 1986; baseado em regras; plataforma: ROSIE; base de conhecimento: danos, obrigação do réu, vítima, responsabilidade da vítima. |
| SARA | Analisa decisões governadas por um tipo especial de normas (<i>discretionary norms</i>); Bing, 1980; baseado em <i>frames</i> ; protótipo de demonstração. |
| TAXADVISOR | Oferece ações a serem tomadas por clientes para estabelecer perfil financeiro, aquisição de seguros, aposentadorias, transferências de bens e testamentos; Michaelson, 1982; baseado em regras, encadeamento para trás; plataforma EMYCIN; protótipo de pesquisa; base de conhecimento: planos baseados em estratégias de advogados e livros. |
| TAXMAN | Auxilia na investigação do raciocínio jurídico e do argumento jurídico gerando uma metodologia de transformação para o reconhecimento de relações entre conceitos; McCarty & Sridharan, 1982 e Kedar-Cabelli, 1984; baseado em <i>frames</i> ; protótipo de pesquisa; plataforma AIMDS; base de conhecimento: casos sobre tributação, leis tributárias e princípios de transformação. |

Tabela 1 Implementações de Sistemas Especialistas no domínio do Direito.

Esta conclusão é confirmada pelo mesmo autor ao descrever a inexistência de sequer um sistema especialista na *1996 Legal Software Survey* (programas jurídicos do mercado canadense) que continha centenas de programas jurídicos. Branting (1991c) alega que há um consenso de que raciocínio baseado em regras é insuficiente para modelar a tarefa de análise jurídica, alegação esta que é sustentada ainda por McCarty &

conhecimento legal na forma de regras doutrinárias falharam."

Sridharan, (1982), Gardner, (1984), e Burton, (1985). Por outro lado, os sistemas de RBC vêm demonstrando (Tabela 2) ser esta uma técnica adequada para lidar com as idiossincrasias e peculiaridades do domínio jurídico. Os projetos de pesquisa¹⁰ que demonstram tal adequação empregam RBC para sentenciar, mediação e construção de argumentos. A tarefa de sentenciar é facilmente tratável em sistemas de RBC por ser esta uma tarefa essencialmente de adaptação; desde que tenhamos uma sentença anterior similar destinada a uma situação específica e conhecimento sobre adaptação, é fácil implementar esta tarefa automaticamente. A construção de argumentos é uma tarefa que combina elementos contrastantes de um dado assunto, o que pode ser determinado através de avaliação de similaridade. Mediação é uma tarefa que se executa com adaptação e as experiências de adaptação podem ser modeladas com a teoria de indexação de RBC, representando mais um forte exemplo da tratabilidade de questões jurídicas com RBC.

Uma das primeiras implementações de RBC no domínio do Direito foi JUDGE (Bain, 1984, 1986, 1989). Este sistema adapta casos passados para sentenciar crimes de delinquentes juvenis, inferindo as circunstâncias que caracterizam os crimes. Depois de gerar uma sentença, o JUDGE recupera outro episódio similar para avaliar a consistência da sentença obtida. O sistema MEDIATOR (Simpson 1985, Kolodner & Simpson, 1988) usa RBC na tarefa de planejamento para a mediação de partes envolvidas em uma disputa que têm metas e submetas a alcançar. O MEDIATOR foi o primeiro sistemas de RBC a usar casos *múltiplos* (Kolodner, 1993). O sistema PERSUADER (Sycara, 1987) também serve para mediar disputas através de planejamento, este sistema é aplicado ao domínio de negociações sindicais. O sistema adapta planos considerando metas e restrições.

¹⁰ Alguns projetos de pesquisas que demonstram o potencial das aplicações de RBC são HYPO (Ashley & Rissland, 1988a, 1988b, Ashley, 1990), JUDGE (Bain, 1984, 1986, 1989), e GREBE, (BRANTING, 1989 & 1991C).

Os esforços de desenvolver sistemas inteligentes para o domínio do Direito foram marcados com o HYPO – um sistema de RBC que constrói argumentos a partir de uma base de casos no domínio de segredos industriais (Ashley & Rissland, 1988a, 1988b, Ashley, 1990). Este programa utiliza dimensões para indexar e realizar análise de relevância de casos passados dinamicamente. Mais importante, o sistema demonstrou como tratar argumentos e lições presentes em casos jurídicos. Contudo, a necessária engenharia de casos representa um gargalo no desenvolvimento deste tipo de sistema para resolução de problemas práticos reais.

| Nome | resumo/referências |
|-----------|--|
| JUDGE | Adapta casos similares passados para sentenciar delinquentes juvenis. (Bain, 1984, 1986, 1989) |
| MEDIATOR | Utiliza tecnologia de RBC implementando a tarefa de planejamento onde as partes envolvidas nas disputas tem metas e submetas. (Simpson 1985, Kolodner & Simpson, 1988) |
| PERSUADER | Media disputas no domínio de negociações sindicais, adaptando planos e considerando metas e restrições na criação de argumentos. (Sycara, 1987a) |
| HYPO | Construção do argumento jurídico a partir de uma base de casos no domínio da legislação de segredos comerciais. (Ashley & Rissland, 1988a, 1988b, Ashley, 1990) |
| GREBE | Um sistema híbrido que integra casos e regras gerando explicações para fundamentar classificação na execução do raciocínio jurídico. (Branting, 1991c) |
| BANKXX | Criação de argumentos através de uma busca heurística baseada em casos numa rede altamente conectada contendo conhecimento jurídico na área de falências. (Rissland et al., 1993) |
| CABARET | Integra raciocínio baseado em regras e casos passados para fazer interpretação no domínio da legislação de imposto de renda. (Rissland & Skalak, 1991) |
| PRUDENTIA | Pesquisa por decisões judiciais passadas que possam ser úteis ensinando lições para uma nova situação executando uma Pesquisa Jurisprudencial Inteligente. (Weber, 1997; Weber et al., 1997a; Weber et al., 1997b; Weber et al., 1998) |

Tabela 2 Sistemas de RBC no domínio do Direito.

Em Smith et al. (1992) é proposto um procedimento para construir sistemas especialistas jurídicos conectados a um banco de dados jurídico no domínio das leis relacionadas a danos originados do uso da propriedade do réu ocorridos em outra propriedade. JC Smith argumenta que a estrutura de conhecimento jurídica é complicada pela existência de teorias conflitantes de jurisprudência além da total falta de consenso com relação

à lei. Como solução, Smith et al. (1992) propuseram uma metodologia de estruturação de conhecimento chamada *análise estrutural profunda* (Coval & Smith, 1986; Smith, 1976, 1984; Smith & Deedman, 1987). Esta metodologia busca encontrar os elementos conscientes e inconscientes que são levados em consideração na definição de conceitos como, por exemplo, uso não natural da terra. A análise tradicional compreende a definição da causa da ação em termos das condições necessárias e suficientes, tais como negligência, transtorno e travessia de animais, na definição de um conceito. O sistema especialista baseado em regras proposto pelos autores não consulta os casos associados antes de chegar a conclusões enquanto um sistema baseado em casos retorna um conselho. Uma estrutura de árvore é construída para refletir a estrutura da lei em detalhe conforme determinado pela análise estrutural profunda do domínio e representar a informação em termos dos atributos genéricos (estilo da causa, citação, jurisdição, nível do tribunal, data) e atributos específicos (natureza da substância causadora do dano, classificação do distrito onde o incidente ocorreu, atividade do réu).

GREBE (Branting, 1991c) é um sistema híbrido baseado em regras e baseado em casos para o raciocínio jurídico. Um importante tópico discutido pelo autor trata sobre a lacuna entre as leis na forma em que são usualmente representadas na teoria e a descrição de uma dada situação jurídica. A teoria do domínio especifica as leis e as situações em que estas se aplicam através de conceitos que devem ser inferidos de uma determinada situação para associar a situação à lei respectiva. Branting alega que o raciocínio jurídico é uma forma de classificação (1988) e emprega explicações para sustentar as classificações, separando as explicações dos fatos. Branting (1991b) também propõe o uso de elaboração de casos no tratamento de predicados de contexto aberto¹¹ na combinação de casos. Branting, (1991c) propõe a análise jurídica¹² automatizada

¹¹ Contexto aberto, tradução literal do termo em Inglês: *open-textured*.

¹² Em Branting (1991c), tarefa de inferir as conseqüências de um conjunto de fatos jurídicos

através da integração de casos e regras em explicações jurídicas. No GREBE, os casos complementam o raciocínio baseado em regras quando não há mais regras (Gardner, 1984). O sistema de RBC trata a incerteza que surge com as diferenças entre antecedentes de regras de contexto aberto e as novas situações em virtude das novas situações serem expressas no mesmo nível de abstração do que os casos precedentes. A incerteza na qual sistemas de RBC ficam expostos refere-se à diferença entre um caso passado e um novo.

Recentemente, Rissland, Skalak, e Friedman (1993) desenvolveram BANKXX, que gera argumentos realizando uma busca heurística baseada em casos através de uma rede altamente conectada de conhecimento jurídico no domínio de falências. Deste mesmo grupo, o sistema CABARET (CAse-BAsed REasoning Tool) integra o raciocínio com regras e com casos passados na tarefa de interpretação (Rissland & Skalak, 1991) no domínio de leis sobre imposto de renda. Ao examinarmos tantas aplicações, concluímos que a maioria da tecnologia de RBC não apenas parece mais adequada como também confirma e garante o desenvolvimento de sistemas inteligentes no domínio do Direito.

Daniels e Rissland (1995) construíram um sistema híbrido de RBC e Recuperação de Informação onde a tecnologia de RBC exerce o papel de aprimorar a consulta apresentada ao sistema de RI, aprimorando, em resultado, os resultados das consultas. Esta alternativa de implementação origina-se da alegação das autoras de que textos não são tratáveis por métodos baseados em conhecimento e, desta forma, sua proposta não se beneficia da recuperação baseada em casos, um dos pontos fortes de RBC.

Ao considerarmos as diferentes abordagens ao tratamento automático de textos jurídicos, Branting e Lester (1996) descreveram a tarefa de projeto de rascunhos de documentos com a reutilização de casos adaptados. Em sua abordagem, eles demonstram as estruturas retóricas e ilocucionárias

de documentos auto-explicativos (4.2.4), uma abordagem baseada em lingüística para identificar funções em extratos de textos.

O projeto SALOMON (Uyttendaele, 1996) visa facilitar o acesso de textos jurídicos através da geração automática de resumos de decisões judiciais em casos criminais. O resumo representa decisões judiciais através de nove atributos cujos valores são extraídos dos textos. A abordagem combina técnicas estatísticas com técnicas baseadas em conhecimento. Apesar de lidarem com indexação, eles não exploram a tecnologia de RBC.

A principal desvantagem da implementação de sistemas de RBC no domínio jurídico refere-se às necessidades de engenharia dos casos que vêm impedindo aplicações comerciais. Os casos devem ser indexados com atributos que representam tópicos relevantes e proporcionam utilidade no desempenho da tarefa do sistema que normalmente estão disponíveis em forma textual. A engenharia de casos bem sucedida engloba a modelagem do conhecimento jurídico, que discutimos na próxima seção.

2.3.1 Conhecimento no Domínio do Direito

O desenvolvimento de sistemas inteligentes em diferentes domínios requer a avaliação do conhecimento que este sistema compreenderá. Rapidamente identificam-se diferentes tipos de conhecimento no domínio do Direito. O conhecimento declarativo engloba objetos tais como as partes numa disputa, leis, agentes concretos e abstratos como imóveis, dinheiro, armas, liberdade, respeito, direito e medo. Os fatos são representados por ações realizadas pelas partes assim como por outros agentes. Os conceitos oriundam-se do conhecimento geral sobre o domínio, senso comum e princípios jurídicos. O conhecimento procedural é caracterizado pelas leis, jurisprudência, doutrinas, procedimentos e pelo argumento jurídico. O metaconhecimento está embutido no raciocínio dos advogados e refere-se a tópicos tais como a resolução de conflitos normativos que surgem quando leis conflitantes são aplicáveis numa mesma situação. O conhecimento

estrutural do Direito consiste no saber as combinações de leis possíveis e viáveis e os resultados das implicações de objetos, conceitos e suas inter-relações que possam ser obtidas nas leis, princípios e doutrinas. Todos os estilos de conhecimento presentes no Direito constituem o corpo de conhecimento que deve ser dominado pelo podemos chamar de um bom advogado: o especialista. Isto tudo resume-se no entendimento genérico e conciso de que as aplicações inteligentes do domínio do Direito são as capazes de representar o conhecimento especialista dos juristas. Este conhecimento especialista é o conhecimento jurídico que representa a fonte para resolver qualquer problema deste domínio. Especificamente, o raciocínio dos juristas refere-se a sua interpretação sobre os conceitos e entidades deste domínio e como empregar o metaconhecimento sobre procedimentos jurídicos, decisões judiciais e como construir argumentos.

Perante os obstáculos de modelar o domínio do Direito em sistemas inteligentes, uma das alternativas é lançar mão de técnicas estatísticas. O projeto Flair (1997) é um exemplo disto, que introduzimos na próxima seção.

2.3.2 Outros Esforços

Poucas implementações computacionais no domínio do Direito atuam sobre problemas práticos, a maioria contempla pequenas amostras que visam demonstrar ou introduzir idéias através de projetos de pesquisa. O grupo do projeto FLAIR¹³ vem trabalhando numa aplicação prática. Embora fundamentalmente baseando-se em métodos estatísticos, eles desenvolveram o Sistema de Recuperação de Informação Flexicon¹⁴.

O trabalho do grupo de JC Smith em IA e Direito baseia-se no desenvolvimento do projeto FLAIR, iniciado em 1985 com colaboração da

¹³ <http://www.flair.law.ubc.ca/flair>

¹⁴ <http://www.flair.law.ubc.ca/flexicon/index.html>

IBM, no valor de US \$ 3,2 milhões. O sistema FLEXICON é um mecanismo de busca estruturado e projetado segundo propriedades léxicas para combinar ordenamento de relevância com consultas com múltiplos termos. O modelo de representação de conhecimento estruturado projetado para o sistema serve tanto para um esquema de representação de conhecimento, aliado ao ordenamento estatístico, como para uma representação externa utilizada para resumir textos jurídicos visando uma rápida avaliação dos resultados da busca (Smith, 1997). Contudo, os requerimentos de manutenção e de recursos proporcionaram pouca aplicação prática levando-os a redirecionar a pesquisa no sentido da administração, representação e recuperação de informação a partir de bancos de dados, tornando-se este o âmago deste projeto.

Segue um resumo desta pesquisa conforme descrito em Smith (1997):

- três estágios de desenvolvimento: prova de conceito, 300 casos; protótipo com 1200 casos; 33.000 casos de decisões (em desenvolvimento);
- representa os documentos do banco de dados em termos de conjuntos de glossários organizados alfabeticamente, hierarquicamente e em termos de frequência;
- FlexNote: recupera documentos com diversos tipos de glossários que constituem o conteúdo com relação à frequência dos termos normalizados em relação ao total da base;
- os fatos em uma disputa jurídica são definidos por um conjunto de perguntas que identificam metas da situação, a ordem destas metas é identificada na base de dados e é responsável por retornar o conjunto de casos que definiram tais metas;
- aspectos secundários tais como a idade do caso, jurisdição e o nível do tribunal alimentam um algoritmo que atribui pesos a cada caso;

- o sistema calcula um resultado em termos percentuais e retorna os casos relevantes.
- o glossário hierárquico ainda não está pronto em virtude de:
 - a dificuldade de distinguir o que é conceitual do fatural;
 - como assegurar a classificação de um conceito jurídico;
 - algumas palavras têm significado tanto doutrinal quanto ordinário;
- os glossários de fatos ainda não estão reconhecendo todas as expressões, soluções estão em fase de implementação e teste.

Em função das afinidades e de algumas metas compartilhadas entre as pesquisas com FLEXICON e PRUDENTIA¹⁵, consideramos necessário mostrar uma revisão das diferenças entre as duas pesquisas.

| Flexicon vs. | PRUDENTIA |
|---|--|
| Teoria lingüista de Ferdinand de Saussure, significado e significantes. Baseado em estatística. Motivação científica. Objetiva construir uma solução ao advogado. | Indexação e representação baseada em casos. Teoria de RBC. Motivação tecnológica. Objetiva a Pesquisa Jurisprudencial Inteligente para todos os operadores do Direito. |
| Doutrina jurídica e conteúdo material são separados. | Ambos doutrina jurídica e conteúdo material são tratados igualmente e em conjunto. |
| Casos jurídicos sem elementos estruturais. Separa áreas de assunto. | Casos jurídicos com elementos estruturais definidos. Áreas distintas são identificadas por um índice. |
| Indexação estatística. | Mineração por modelos (4.2.2) baseada em conhecimento. |
| O modelo de representação de conhecimento estrutural serve para resumir os casos para a rápida avaliação dos resultados da pesquisa. Eles consideram mais vantajoso procurar métodos para gerar transformações criativas para resolver tarefas difíceis do que reproduzir ou simular a inteligência humana. | Usuários avaliam casos recuperados através de atributos que resumem seu conteúdo e o contexto. A motivação por trás da tecnologia do PRUDENTIA é fazer uso da inteligência humana para manipular conhecimento e criar sistemas que facilitem as tarefas humanas. |

Tabela 3 Resumo da comparação entre os projetos Flair e PRUDENTIA.

¹⁵ The system PRUDENTIA is fully described in chapter 6.

2.4 Conclusão

As aplicações de IA alternaram seus pontos de ênfase através de diferentes períodos. A escolha da ferramenta mais adequada para realizar cada tarefa deve ser combinada com o formalismo de representação do conteúdo do domínio que seja mais próprio. Engenheiros do conhecimento esforçam-se na busca de aplicações bem sucedidas que culminem na melhor combinação entre linguagem de IA, formalismo de representação e o próprio conhecimento.

As implementações de IA no domínio do Direito tiveram seu horizonte ampliado com o uso da tecnologia de RBC. As particularidades do conhecimento jurídico têm dificultado a obtenção de bons resultados práticos. Felizmente, as pesquisas mais recentes começam a apontar melhores direções.

A comparação de aplicações jurídicas com sistemas especialistas e (Tabela 1 e Tabela 2) a tecnologia de RBC indica a melhor adequabilidade apresentada pela última. O próximo capítulo apresenta um resumo da tecnologia de RBC.

Capítulo Três

De Riesbeck & Schank (1989), "*A case-based reasoner solves new problems by adapting solutions that were used to solve old problems*"¹⁶.

3. Raciocínio Baseado em Casos

A extensão e as implicações da definição de Raciocínio Baseado em Casos enquanto técnica de Inteligência Artificial (IA) sublinham o conteúdo deste capítulo. Descrevemos aspectos da implementação de um sistema de RBC dando ênfase às etapas utilizadas no presente trabalho, através das seguintes seções: Raciocínio Baseado em Casos 3.2, Conhecimento Especialista em RBC; 3.3, Casos e Base de Casos; 3.4, Indexação; 3.5, Avaliação de Similaridade; 3.5.1, Métrica de Similaridade; 3.6, Ajuste da Situação; 3.7, Aprendizagem em RBC.

¹⁶ Um sistema de raciocínio baseado em casos resolve novos problemas adaptando soluções que foram utilizadas para resolver problemas passados.

3.1 Raciocínio Baseado em Casos

Raciocínio Baseado em Casos (RBC) é uma técnica de Inteligência Artificial (IA) que reproduz aspectos da cognição humana para resolver problemas especialistas. Os sistemas de RBC simulam o ato humano de lembrar um episódio prévio para resolver um determinado problema em função da identificação de afinidades entre os mesmos. Definições e algumas referências gerais foram dadas por Riesbeck & Schank (1989), Kolodner (1993), Aamodt & Plaza (1994), Leake (1996) e Watson (1997).

O ato de lembrar um episódio anterior é simulado em um sistema de RBC por meio da comparação de um novo problema com um conjunto de casos do mesmo tipo. Este conjunto chama-se Base de Casos. A comparação é efetuada através da avaliação de similaridade entre o novo episódio com os contidos na base de casos. A avaliação de similaridade é executada ao nível dos índices. Os casos mais similares são recuperados. Uma fase de seleção determina qual o caso mais útil completando o ato de lembrar.

O raciocínio humano reproduzido no sistema de RBC é a analogia e há dois pressupostos que fundamentam o emprego deste raciocínio (Leake, 1996):

1. problemas similares têm soluções similares; e
2. os tipos de problemas se repetem.

Uma forma de classificar os sistemas de RBC é dividi-los em sistemas interpretativos e de solução de problema (Leake, 1996). A diferença entre eles pode ser explicada a partir das tarefas e metas de cada sistema. Os sistemas de RBC interpretativos usam casos passados como referência para classificar e caracterizar os novos casos; enquanto que nos sistemas de solução de problema, os casos são reutilizados para prover soluções

para o novo problema. Com relação às metas, os sistemas interpretativos procuram a interpretação do novo caso enquanto que resolver o novo problema é o objetivo dos sistemas de solução de problema. Esta diferenciação indica, por exemplo, que uma etapa de adaptação é bem mais necessária para os sistemas de solução de problema do que para sistemas interpretativos. O desenvolvimento de um sistema de RBC em qualquer domínio é uma tarefa iterativa, compreendendo as tarefas do ciclo do RBC, como proposto por Aamodt e Plaza (1994), vide Figura 1.

São quatro as etapas principais no desenvolvimento de um sistema de RBC: **recuperar**, **reutilizar**, **revisar** e **reter**. Estas etapas conduzem e orientam o raciocínio empregado por um sistema RBC que é desempenhado através dos casos. Autores como Watson (1997) e Aamodt & Plaza (1994) referem-se a estas etapas como o ciclo do RBC. Outros autores como Kolodner & Leake (1996) referem-se ao ciclo de RBC através das tarefas de **recuperar**, **propor**, **criticar**, **justificar**, **avaliar**, **adaptar** e **armazenar**.

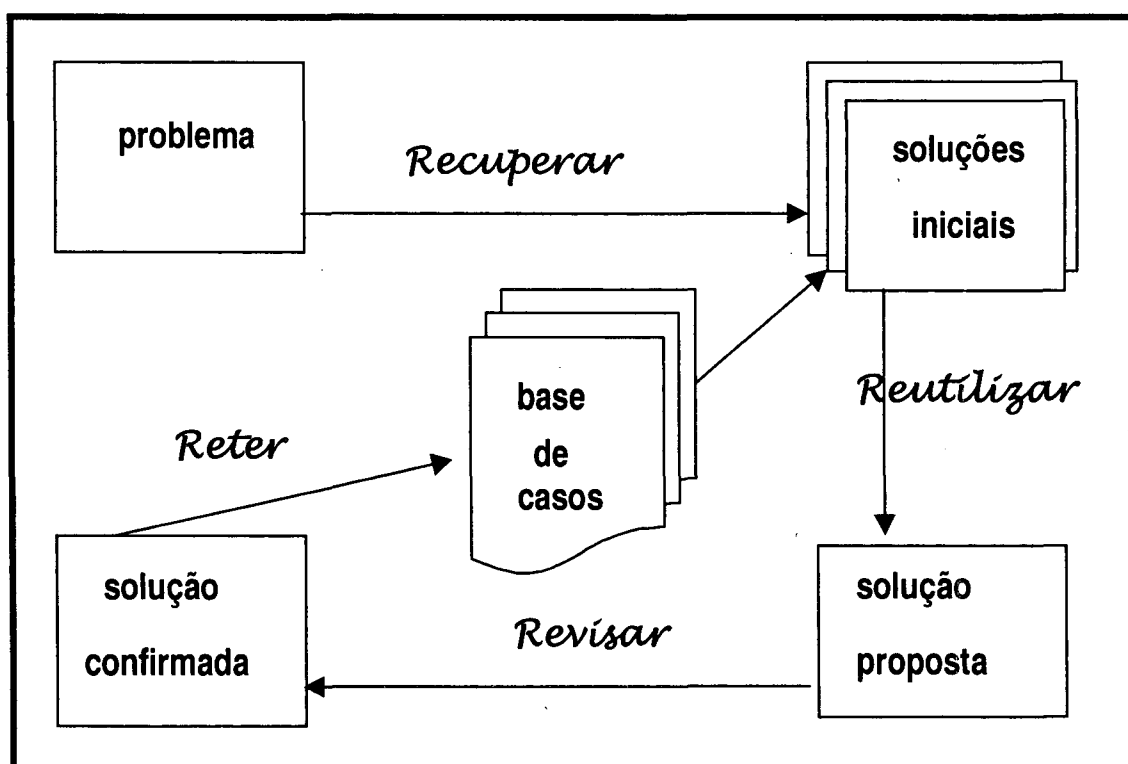


Figura 1 O ciclo do RBC.

Recuperar é o processo de retornar um ou mais casos da base de casos em resultado à comparação de um novo caso (caso alvo) com cada um dos casos da base (casos candidatos). Esta comparação é feita através de uma avaliação de similaridade. O resultado desta comparação é a seleção de um caso (ou uma combinação de casos) que sugere uma solução ao caso alvo. **Reutilizar** é a etapa pertinente ao aproveitamento do conteúdo presente no caso recuperado (adaptado ou não) no sentido de resolver o caso alvo. Na etapa **revisar**, a solução proposta é avaliada. **Reter** refere-se à adição desta nova experiência ou das experiências que inicialmente compõem a memória de conhecimento, podendo a adição de novos casos representar um mecanismo de aprendizagem.

A capacidade de lembrar um episódio prévio habilita sistemas de RBC a resolver diferentes tipos de problemas especialistas. Watson (1997) enumera as seguintes tarefas de aplicação: diagnóstico, avaliação, controle de processos, projeção, planejamento, projeto e configuração.

Independente da meta, a tarefa que os sistemas de RBC sempre executam é a comparação entre um novo caso alvo e os casos da memória. Esta é a essência da interpretação: identificar o que é relevante ao avaliar a similaridade e ordenar os resultados. Este é um dos pontos fortes da tecnologia de RBC e esta potencialidade é o que torna este tipo de sistema apropriado para o domínio do Direito.

Alega-se que os sistemas de RBC apresentam vantagens sobre os sistemas especialistas baseados em regras em virtude de seus reduzidos requisitos de aquisição de conhecimento (Watson, 1997). No domínio do Direito, isto não é necessariamente verdade desde que mesmo com a existência de uma base de dados, seu mapeamento para casos pode representar um gargalo para a engenharia do sistema. Contudo, com relação à representação do conhecimento, mesmo apresentando algumas dificuldades, é plenamente viável representar fatos e procedimentos através de uma base de casos; enquanto a representação baseada em regras é limitada em uma

determinada quantidade de regras que dificilmente conseguem representar todas as inter-relações de um domínio.

A próxima seção discute alguns tópicos no desenvolvimento de um sistema de RBC que influenciaram a presente pesquisa.

3.2 Conhecimento Especialista em RBC

Raciocínio baseado em casos, enquanto técnica de IA, incorpora conhecimento especialista para resolver tarefas. A distinção entre RBC e outras técnicas de IA reside no fato de que o conhecimento especialista está presente nos sistemas de RBC não somente no que chamamos de sua base de conhecimento (memória), como também em outras etapas no seu desenvolvimento. A escolha com relação a o que é o caso é uma questão que exige conhecimento especialista sobre o domínio para ser resolvida. A forma com que os casos são organizados na base de casos também utiliza conhecimento sobre o domínio. A definição de medidas de similaridade oriunda-se da identificação sobre o que faz um caso ser similar a outro, o que é um mérito totalmente dependente do domínio. O ajuste da situação também deve ser desenvolvido com conhecimento especialista. Conseqüentemente, utilizamos, encontramos, precisamos e representamos conhecimento especialista em diversas etapas de um sistema de RBC:

- na representação dos casos: somente um especialista sabe o que é relevante e o que é adequado sugerir como solução para um caso similar;
- ao obter a informação correta e capturar a necessidade do usuário;
- ao organizar os casos na memória;
- ao determinar como procurar casos similares na memória;
- ao identificar o que faz um caso ser similar a outro;
- ao escolher qual métrica de similaridade utilizar;
- ao apresentar os resultados ao usuário.

Entretanto, as necessidades de conhecimento especialista variam em cada aplicação e isto não quer dizer que os benefícios associados à aquisição de conhecimento não se efetivem. Isto simplesmente indica que diferentes sistemas de RBC comportam-se de forma distinta com relação ao conhecimento.

3.3 Casos e Base de Casos

As principais entidades envolvidas no processo de raciocínio de um sistema de RBC são os casos e a base de casos. O caso representa uma experiência ou uma interpretação de uma experiência. A base de casos consiste no conjunto de casos (base de casos) e os procedimentos de acesso a estes casos (Kolodner, 1993).

A representação dos casos compreende dois aspectos:

- o conhecimento que deve ser representado;
- qual formalismo adotar para representar os casos no sistema.

O caso deve representar conhecimento sobre o conteúdo e o contexto da experiência. A relevância de alguns aspectos varia em conformidade com o tipo de tarefa pretendida. Em sistemas de solução de problema, os casos devem englobar problemas, soluções e resultados da utilização das soluções (Kolodner, 1993). Sistemas interpretativos devem representar as experiências na forma que facilitar a interpretação das mesmas.

A representação dos casos também refere-se ao formalismo a ser adotado no programa. Muitos formalismos da IA podem ser empregados, tais como *frames*, representação tipo formulário e redes semânticas¹⁷. A escolha adequada depende da consideração de tópicos pertinentes à aquisição e às demais etapas do processo de desenvolvimento, tais como recuperação, adaptação e aprendizagem.

Na utilização da representação tipo formulário, os casos são representados por um conjunto de campos chamados *descritores*. Descritores são pares atributo-valor que caracterizam a informação contida em um caso. A parte

¹⁷ Para uma revisão, veja seção 2.2.2.

atributo de um descritor é também chamada dimensão (Kolodner, 1993). Seja a base de caso L com m casos, $x = \{x_1, x_2, \dots, x_m\}$ cada um definido por uma representação formulário com um conjunto de n descritores d , pode-se inferir que os casos x_i e x_l são similares ao nível da dimensão d_i .

Os descritores exercem a função de representar casos e alguns também são usados para guiar a recuperação, estes últimos são chamados índices.

3.4 Indexação

A indexação é a essência do raciocínio baseado em casos porque orienta a avaliação da similaridade. O conjunto de descritores que são usados como índices modelam a resposta para a pergunta, “*o que faz um caso ser similar a outro?*” - representando a relevância dos casos. A indexação determina o que deve ser comparado entre os casos para avaliar sua similaridade no intuito de recuperar casos que conduzam à tarefa principal - permitir a recuperação dos casos mais úteis para resolver ou interpretar o novo caso.

Kolodner (1993) apresenta o problema de indexação através de duas subtarefas: a definição do *vocabulário de indexação* e a *atribuição dos índices*. A identificação do vocabulário de indexação consiste em selecionar dimensões que, quando atribuídas, preenchem as funções dos índices. A última etapa refere-se à atribuição de valores para estas dimensões.

Ainda segundo Kolodner, os índices devem ser vaticinantes, suficientemente abstratos para serem genéricos e suficientemente concretos para serem reconhecidos. Um índice útil conduz o propósito dos casos que pode ser tanto associado a sua solução, ao aviso de um possível fracasso ou ao seu resultado. Para a seleção do vocabulário de indexação são propostas duas abordagens, *de lembrança* e *funcional*.

A abordagem de lembrança é intuitiva porque é a abordagem naturalmente empregada por especialistas humanos e engenheiros de conhecimento quando solicitados a selecionarem um índice: é a busca pelos aspectos que são levantados pelos especialistas do domínio quando resolvem problemas.

A abordagem funcional prevê três méritos com relação aos casos: (1) quais dimensões podem incorporar as tarefas pretendidas pelo sistema; (2) quais casos informam os valores para as dimensões; e (3) qual o nível de abstração que deve ser considerado (quando do desenvolvimento e vislumbrado possíveis expansões). A abordagem funcional é conduzida através da realização de cinco tarefas:

- selecione os casos suficientes para contemplar o universo de domínio;
- identifique as lições ensinadas pelas experiências, tais como condições e explicações que corroborem com soluções e suas conseqüências;
- identifique o contexto no qual as lições são ensinadas, sob quais circunstâncias as lições são válidas;
- descreve os índices que permitirão a recuperação quando cada contexto ocorrer procurando por situações genéricas que caracterizam seu conteúdo;
- incrementalmente, selecione as dimensões que podem representar os índices.

Ambas abordagens para a seleção do vocabulário de indexação visam decifrar as seguintes questões:

- quais dimensões são suficientes para descrever índices que permitam uma recuperação eficiente para cada tarefa que o sistema se propõe a desempenhar e que ainda incorpore a especificidade suficiente para distinguir tais tarefas;
- quais são os valores a serem atribuídos às dimensões para garantir que cada caso possa ser recuperado para a realização das tarefas propostas;
- qual o nível de detalhe destas dimensões que não comprometa expansões futuras.

O conhecimento especialista é *sine qua non* para um vocabulário de indexação vocabulário satisfatório na medida que seu desenvolvimento é

verdadeiramente uma tarefa de representação de conhecimento assim como a atribuição. Um dos tópicos levantado por Kolodner indica como pode ser útil a antecipação do vocabulário que seja familiar aos usuários; e, ainda mais importante, a antecipação das circunstâncias na qual um caso deve ser recuperado. Isto significa que a engenharia de conhecimento na etapa de indexação trata-se de vislumbrar como o usuário encara um problema e como e por que este usuário necessitaria lembrar algo.

O processo de indexação é uma oportunidade de superar a deficiência de experiências mal descritas e torná-las úteis e valiosas na realização da tarefa do sistema. Esta meta é conduzida pela correta interpretação da experiência a partir da perspectiva do especialista, permitindo a identificação do significado intrínseco e da correlação entre os entidades ativas participantes na experiência. Uma forma de buscar tais relações é tentar representar as correspondências entre as causas e conseqüências, razões e soluções.

De fato, a indexação pode representar um gargalo no desenvolvimento de sistemas de RBC como é o exemplo da necessidade de indexação automática para viabilizar o sistema. A indexação automática pode ser necessária em sistemas que comportam mecanismos de aprendizagem automática. Além disso, há domínios em cujo conhecimento está disponível somente em formato textual que exige grandes bases de casos. Os domínios do Direito, Economia e Medicina são exemplos nos quais um mecanismo de indexação automática é plenamente justificado.

Quando o conhecimento especialista está disponível apenas em formato de texto, temos o RBCT Raciocínio Baseado em Casos Textual, (Weber et al., 1998). Conseqüentemente, como a indexação é um problema no RBCT, a avaliação da similaridade também pode oferecer dificuldades no tratamento de novos casos alvo. Esta área de desenvolvimento é discutida a seguir.

3.5 Avaliação de Similaridade

O processo de avaliação da similaridade em sistemas de RBC refere-se à comparação de dois casos para avaliar como um está relacionado ao outro no que concerne a como estes casos podem compartilhar soluções e conseqüências. Esta avaliação é executada no nível dos atributos, associando-se valores cuja natureza determina a função de combinação a ser empregada. Funções de combinação comparam valores numéricos e alfanuméricos exatamente ou dentro de intervalos. A comparação entre textos pode ser feita por similaridade sintática ou semântica. Uma forma simples e direta de modelar similaridade é atribuindo-se um valor de uma (1) unidade para a função resultante similar e o valor zero (0) para o resultado não similar; e ainda atribuir valores intermediários entre estes. Por exemplo, em um sistema que selecione candidatos a um cargo, um atributo para experiência profissional valorado com experiência em bancos pode resultar em um valor intermediário para a função se a comparação for feita com um caso valorado com experiência em instituição financeira, o que pode ser implementado por meio de regras.

3.5.1 Métrica de Similaridade

A métrica de similaridade é uma função que mede numericamente os graus de similaridade entre dois casos. Uma métrica é normalmente necessária em sistemas nos quais os casos são comparados um a um e a medida de sua similaridade é o meio de distinguir entre os casos candidatos similares e não similares. Em sistemas que representam casos com redes, árvores ou grafos, uma métrica de similaridade é usada

somente para ordenar um conjunto de casos que tenham sido recuperados através de índices que orientam para um determinado local da rede.

A métrica de similaridade sintetiza a similaridade ao nível de cada atributo através de uma medida da importância que é usada para modelar a relevância de cada atributo na avaliação sintética da similaridade. Conseqüentemente, qualquer cálculo matemático orientado para avaliações sintética é válido, tais como a média ponderada e as integrais difusas. Contudo, ainda é possível encarar a avaliação de similaridade como um problema de reconhecimento de padrões dentro do espaço dos casos. Desta forma, algoritmos que realizam estes tipos de avaliação também podem ser usados, como os algoritmos de vizinho mais próximo (Cover & Hart, 1967; Okamoto & Satoh, 1995).

3.6 Ajuste da Situação

A etapa de ajuste da situação é necessária se o novo caso alvo não estiver representado da mesma forma que os casos da base; a meta é capturar a nova situação trazida pelo usuário e modelá-la na forma dos casos da base. Este processo pode ser bastante similar à engenharia do caso requerida na modelagem original dos casos da base, dependendo do formato original das experiências. Se a representação original dos casos equivaler-se ao formato com que o usuário informa o caso alvo, basta reaplicar a engenharia de caso. Por outro lado, se o usuário informar a nova situação em formato diferente, requerem-se novos métodos de engenharia de conhecimento. Apesar da meta ser a mesma, modelar a descrição nos termos da indexação empregada aos casos da base visando que o novo caso alvo: (1) possa ser recuperado quando adicionado a base de casos, e (2) possa ser comparado aos casos candidatos ao nível das mesmas dimensões. O processo de ajuste da situação é incremental e pode ser desenvolvido antes ou durante a recuperação em ciclos que refinam a indexação até a recuperação de um caso similar seja eficientemente concluída. Para verificar tal eficiência, Kolodner (1993) sugere a comparação dos resultados da recuperação realizada com este novo caso em comparação com os resultados das recuperações geradas por casos já adequadamente indexados. Se houver muitos casos similares ou nenhum caso similar¹⁸, isto é uma indicação de que a indexação talvez não esteja concluída.

As diferenças entre a representação do novo caso alvo e os casos da base não referem-se apenas à modelagem, como também à possibilidade do caso estar incompleto, impreciso ou simplesmente por não incorporar o mesmo

¹⁸ The fuzzy expression too many stes for amounts that avoid the identification of a distinctive set of useful cases.

conjunto de características que possam ter sido inferidas durante a representação do caso.

A extração de características superficiais a partir de uma nova descrição é de implementação simples, entretanto para inferir índices complexos é necessário o emprego de metodologias com alguma sofisticação. Neste caso, o conhecimento sobre o domínio pode ser representado através de regras viabilizando a inferência de características não superficiais.

Outra abordagem para inferir características é a *elaboração*, o processo onde uma recuperação inicial é realizada suportada exclusivamente por índices valorados por características superficiais. Assim, a partir da representação dos casos similares resultantes desta primeira iteração, é possível inferir que os mesmos tipos de índices possam ocorrer, facilitando a busca destes valores na nova situação.

O número de características superficiais em relação às características a serem inferidas pode variar determinando diferentes combinações que podem requerer diversas inferências até a valoração completa dos índices. Este problema é apresentado por Kolodner (1993) como o problema de controle.

3.7 Aprendizagem em RBC

A aprendizagem em sistemas de RBC pode ser empregada ao nível dos casos e da base de casos. As bases de casos podem ser estendidas por processos incrementais de aprendizagem se a tarefa e o projeto do sistema permitirem. A partir de um pequeno conjunto de casos semente (Leake, 1996), a base de casos pode crescer com novos casos. A geração destes novos casos oriunda-se de novos casos informados pelos usuários ou a partir de uma fonte externa. Os novos casos alvo são ou interpretados ou resolvidos pela tarefa do sistema. Nos sistemas de solução de problema, os novos casos podem submeter-se a uma fase de adaptação e somente após são adicionados à memória. Em sistemas interpretativos, os novos casos são adicionados mediante sua indexação. Quando somente são agregados casos reais à base, a indexação deve ser implementada automaticamente ou mediante a concordância de engenheiros de conhecimento.

A aprendizagem no nível dos casos acontece como expressão da aprendizagem com a experiência. A parte do caso destinada ao resultado do emprego de determinada solução ou interpretação serve a este propósito. Mantém-se no caso o registro de seu desempenho ao ser utilizado. Assim, tanto sucessos como fracassos são informados incrementando o conhecimento e as lições embutidas no caso. O registro do resultado de reutilização pode prevenir o usuário com relação às possíveis conseqüências de seu uso. Este procedimento é valioso porque para compensar a inclusão de informações no caso, o sistema evita a reutilização de sugestões menos favoráveis resultando no incremento da qualidade da recuperação.

3.8 Conclusão

Como uma ferramenta da IA, sistemas de RBC reproduzem o raciocínio humano para resolver problemas especialistas. As arquiteturas de implementação de sistemas de RBC variam em conformidade com as tarefas e metas do sistema. Estes sistemas diferem de outras técnicas de IA em virtude do conhecimento especialista estar representado em diversos módulos do sistema além de uma única base de conhecimento.

Há sistemas de RBC interpretativos e de solução de problemas, ambos representando a ação de relembrar uma experiência passada; os sistemas interpretativos lembram e comparam, enquanto os sistemas de solução lembram e adaptam. Conseqüentemente, sistemas de solução de problema constróem novos casos após adaptá-los permitindo um tipo de aprendizagem possível nos sistemas de RBC. Outras formas de aprendizagem referem-se à aprendizagem dentro dos casos e à aprendizagem de novos casos reais que podem ser implementados em sistemas interpretativos.

O ciclo do RBC compreende quatro processos relacionados a diferentes áreas de engenharia. Descrevemos com mais detalhe as áreas de engenharia nos sistemas de RBC que revisamos com a presente pesquisa.

O mais importante problema na engenharia de RBC tratado no presente trabalho refere-se à representação dos casos e, mais especificamente, à indexação. Estas áreas demandaram esforços de engenharia de conhecimento para sua manipulação dentro do sistema. Um gargalo surge quando o conhecimento disponível apresenta-se em forma textual. Os métodos de engenharia de caso precisam tratar com a extração de informação a partir dos textos. O próximo capítulo descreve alguns tópicos

Capítulo Quatro

Salton (1975), "recall is the proportion of relevant material actually retrieved; precision is the proportion of retrieved material actually relevant. Ideally, everything relevant is retrieved while at the same time everything extraneous is rejected, producing recall e precision values equal to 1."¹⁹

4. Aquisição de Conhecimento de Textos

Quando o conhecimento a ser utilizado em um sistema inteligente está disponível em forma de texto, o problema reside em extrair e mapear este conhecimento em algum tipo de formalismo de representação neste sistema, tal como casos. A extração de informação de textos é um tópico tratado pelo campo de Compreensão de Linguagem Natural descrito neste capítulo junto a outras importantes pesquisas relacionadas com o acesso ao conhecimento contido em textos.

¹⁹ Recall (índice de recuperação) é a proporção de material relevante realmente recuperado; o índice de precisão é a proporção de material recuperado que é realmente relevante. O ideal seria que tudo que é relevante fosse recuperado enquanto que tudo que não é relevante fosse rejeitado, produzindo valores para os índices de recuperação e precisão iguais a 1.

4.1 Compreensão de Linguagem Natural

As implementações de Compreensão de Linguagem Natural (LN) podem ser subdivididas em dois grupos (Allen, 1994): implementações baseadas em texto e baseadas em diálogo. Entre as implementações baseadas em texto, temos: busca de documentos num banco de dados de textos, extração de informação de textos, tradução de documentos, resumos de textos e compreensão de textos. As técnicas desenvolvidas para operar tais aplicações são conhecidas como técnicas de Processamento de Linguagem Natural e o outro campo de estudo que também desenvolve técnicas associadas à compreensão da linguagem é a Lingüística Computacional.

4.1.1 Lingüística Computacional

A Inteligência Artificial (IA) ocupa-se com modelagem de aspectos da cognição humana. A linguagem é o meio de comunicação humana e a Lingüística Computacional (LC) é o ramo das Ciências da Computação e da IA que trata com os aspectos computacionais da linguagem. A LC usa algoritmos e estruturas de dados das Ciências da Computação para examinar os seguintes tópicos: como identificar a estrutura das frases, como modelar o conhecimento e o raciocínio, e como usar a linguagem para realizar determinadas tarefas (Allen, 1994).

Apesar da teoria de LC representar um importante papel na compreensão de textos, o foco de sua pesquisa reside em seu aspecto prático. A modelagem de faculdades humanas em sistemas computacionais é o objetivo da LC que faz interseção com a metas da IA na viabilização da interação homem-máquina na resolução de problemas. Tradutores,

corretores ortográficos, e interfaces de linguagem natural são algumas metas de pesquisa da LC que vêm gerando técnicas para o tratamento da linguagem, principalmente destinados à compreensão da linguagem natural.

4.2 Extração de Informação

Algumas técnicas de processamento de LN originaram-se das pesquisas de compreensão de LN em aplicações baseadas em texto e em diálogos. O escopo da presente pesquisa limita-se às técnicas associadas ao recente ramo da compreensão da LN chamado Extração de Informação. Extração de Informação (EI) é a área associada com a extração de tipos específicos de informação de grandes volumes de textos irrestritos que contenham informação em determinado domínio (Lehnert, 1993, 1996). O sistema de EI deve ser alimentado com algumas diretrizes sobre o domínio que orientam sobre o que encontrar nos textos e o que extrair. No desempenho destas tarefas, os sistemas de EI não precisam compreender completamente o texto fonte; ao invés, os sistemas analisam o texto na busca de porções que possam conter a informação procurada.

Diferente de outras técnicas que também procuram por informações relevantes em textos, a saída de um sistema de EI é um banco de dados. Uma tarefa de EI prevê que um número de campos nos registros do banco de dados sejam preenchidos por um número limitado de possibilidades previamente determinadas ou por algum valor em aberto. Os valores dos campos são expressões do texto fonte. O banco de dados funciona como um modelo (vide seção 4.2.5).

A complexidade no desenvolvimento de um sistema de EI origina-se dos requerimentos de processamento de linguagem envolvidos no reconhecimento de palavras e análise de frases, bem como da compreensão ao nível de frase, e da análise de discurso até o nível de todo o documento. A pesquisa em EI foi alavancada pelos congressos *Message Understanding Conferences* (MUC) (Sundheim, 1991, 1992, Lehnert, & Sundheim, 1991).

O método empregado pelo laboratório de processamento de linguagem natural da Universidade de Massachusetts (Lehnert, 1996) ao participar dos MUCs (MUC-3, 1991, MUC-4, 1992, MUC-5, 1993) é bastante adequado para aplicações onde a precisão das classificações é mais importante do que o reconhecimento de cada documento. Os principais tópicos que foram objeto de pesquisa por este grupo foram análise de frase (Cardie e Lehnert, 1991), rotulagem semântica (*semantic tagging*) (Cardie, 1993), construção de dicionários (Riloff & Lehnert, 1993, Riloff, 1993), e classificação de texto (Riloff & Lehnert, 1992 e 1994).

4.2.1.1 Extração de Informação baseada em conhecimento

O papel de padrões lingüísticos é sustentar a interpretação de textos na Extração de Informação baseada em conhecimento. Em função da construção de padrões lingüísticos ser um gargalo mesmo em domínios limitados, propôs-se o uso de um mecanismo de aprendizagem indutivo para construir automaticamente uma base de conhecimento de padrões. O sistema automático é construído sempre que se identifica um padrão lingüístico desconhecido. Um pressuposto importante embasando esta pesquisa é o reduzido número de expressões normalmente utilizado para descrever uma informação dentro de um domínio limitado (Kim & Moldovan, 1995).

4.2.2 Template Mining

Template Mining ou mineração por modelos é uma técnica de processamento de LN que extrai dados de textos que possuem padrões que permitam o reconhecimento do que se deseja extrair ou de seus arredores. Um modelo contém informação sobre o que procurar no texto e é disparado a extrair determinadas partes devidamente indicadas. Lawson et al., (1996) descreve aplicações de *template mining* em domínios restritos

alegando que esta técnica é própria para áreas cujos textos são claros com frases objetivas e de natureza declarativa.

4.2.3 Text windowing

A técnica *text windowing* é do tipo orientada para *corpus* de textos que avalia palavras na busca de blocos de palavras que estejam relacionadas por sintática ou propriedades léxicas. Jacquemin (1996), descreve uma aplicação de *text windowing* em um método para selecionar trechos de textos motivados por propriedades léxicas, combinando informação conceitual em listas de termos com metaregras em filtros semânticos locais.

4.2.4 Documentos Auto-Explicativos

Consideramos adequado associar a técnica de *template mining* com a metodologia proposta por Branting & Lester (1996) para documentos auto-explicativos. Nesta metodologia, os textos são analisados e classificados por sua estrutura retórica. A ligação entre as técnicas se dá pelo aproveitamento das estruturas retóricas como fonte para a definição dos parâmetros dos modelos usados pela técnica de *template mining* para extração de dados.

4.2.5 Aquisição de Conhecimento de Textos

O trabalho publicado pelo Grupo de Engenharia de Conhecimentos de Textos da Universidade de Freiburg através de diversos artigos²⁰ descreve os esforços para analisar textos que apresentam novas formas de

²⁰ Some of these publications are: Schnattinger & Hahn, 1997; Hahn & Schnattinger 1997a, 1997b; Romacker, Schnattinger, Hahn, Schulz, & Klar, 1997; Hahn, Schnattinger, Romacker, 1996; Hahn, Klenner, Schnattinger, 1996a, 1996b, 1996c.

conhecimento. O grupo se utiliza de um *parser* de LN e almeja a expansão desta base de conhecimento. Da mesma forma que os grupos que tomaram parte dos MUCs²¹, eles também usam técnicas com modelos; entretanto, eles permitem que novos modelos sejam adicionados como resultado da aprendizagem de conceitos (Hahn, & Schnattinger, 1997).

O ponto central da pesquisa do grupo trata-se da aquisição de conhecimento de textos que ocorre com a aprendizagem de conceitos que alimenta um sistema de compreensão de linguagem natural. A aprendizagem de conceitos em uma plataforma de compreensão de linguagem natural é orientada para os recursos através do uso de um *Machine Readable Dictionary*²² MRD e é orientada por contexto. Os autores alegam que inferir o significado das palavras baseando-se em informações sobre o contexto é mais confiável do que procurar por seu significado em um MRD. A aprendizagem de conceitos é concebida com o desenvolvimento de uma abordagem de aprendizagem de raízes simbólicas. Um exemplo de aquisição de conceito é descrito em Hahn et al. (1996). O projeto do grupo visa duas aplicações práticas de aquisição de conhecimento de textos da língua alemã: artigos sobre testes de produtos de tecnologia de informação (100 documentos com 10^5 palavras) e artigos sobre descobertas médicas (120,000 documentos com 10^7 palavras).

O trabalho descrito por Mauldin (1991) usa a compreensão parcial de textos obtida através de um *parser* que realiza *text skimming*²³ para recuperação de informação conceitual, utilizando um banco de dados de *scripts* que, por sua vez, é alimentado por um método de aprendizagem e um MRD que aprimora o conhecimento léxico. A recuperação de informação executada pelo sistema FERRET é referida como recuperação de informação conceitual porque ao invés de realizar a busca através do uso

²¹ Message Understanding Conferences (Sundheim, 1991, 1992).

²² Um dicionário cujos conceitos são acessados dentro do sistema.

²³ Literalmente, filtragem de texto.

de palavras-chave (baseada em palavras) é usado conhecimento sobre os conceitos.

Um ponto de vista interessante sobre o problema de aquisição de conhecimento de textos é descrito em Futrelle & Zhang (1994) que apresenta técnicas de *bootstrap* que podem descobrir a estrutura de ordem da linguagem natural e definir classes de palavras presentes em *corpus* de textos. A definição de classes de palavras é baseada no princípio da substituição onde o significado de uma palavra é encontrado pela comparação dos contextos onde as palavras aparecem e onde elas podem ser substituídas por outra palavra da mesma classe.

4.2.6 Recuperação de Informação

Recuperação de Informação (RI) é "*a tarefa de encontrar documentos relevantes a partir de um corpus ou conjunto de textos em resposta a uma necessidade de informação de um usuário*" (Smeaton, 1997). Os limites da RI são claros; qualquer sistema que faça algo mais do que prover documentos ao usuário não é um sistema de recuperação de informação. Entre as últimas inovações da tecnologia de RI, Smeaton (1997) indica que as aplicações de RI começam a englobar técnicas tais como agentes com componentes de busca, além de outras tarefas associadas à RI como categorização, filtragem, roteamento e *clusterização*²⁴. Entretanto, a RI depende muito da linguagem natural na medida que tanto o documento a ser recuperado como a consulta feita pelo usuário são expressas em linguagem natural, dificultando a tarefa de RI.

O problema da aquisição de conhecimento de textos vem sendo questionado pela comunidade de RI em função da rápida pulverização de informação impulsionada pela Internet. A pesquisa sobre atividades

²⁴ Técnica que agrupa dados em *clusters*, grupos semelhantes entre si.

baseadas em *corpus* de textos tem sido encorajada, facilitando o desenvolvimento de soluções.

As soluções proporcionadas pela pesquisa em *corpus* de textos trazidas pela comunidade de RI são um aspecto da afinidade deste campo com RBC. O ponto fundamental desta afinidade trata-se do interesse da comunidade de RBC pelo fato da tarefa de RI ser essencialmente a recuperação de informação a partir de uma consulta. Na RI, a tarefa resume-se em combinar a consulta do usuário com um conjunto de documentos. De fato, o que se pretende é combinar representações de uma consulta com representações de documentos. Basicamente, os textos são representados em sistemas de RI por conjuntos de índices, e a comparação é baseada por uma equivalência no nível de palavra ou por uma equivalência conceitual.

A operação de recuperação em sistemas de RI objetiva computar graus de coincidência entre a consulta do usuário e os documentos para ordenar cada documento. Smeaton (1997) sugere algumas métricas heurísticas para tal ordenamento (coeficiente de Dice e de Cosine) e enumera áreas onde a pesquisa de recuperação de informação são bem ativas tais como a recuperação baseada em *clusters*²⁵, a recuperação pela combinação de diversas estratégias, indexação semântica latente, recuperação de passagem, e documentos de comprimento heterogêneo.

Motivado pela forte interdependência entre a RI e o processamento de linguagem natural, Smeaton (1995, 1995a, 1995b) questiona a real utilidade do processamento de linguagem e dos demais recursos lingüísticos para a RI. O autor alega que o processamento de linguagem natural oferece um auxílio modesto para a eficiência da RI pelo fato das técnicas de processamento de LN terem sido desenvolvidas visando aplicações de tradução automática e interfaces de linguagem natural. Uma

²⁵ Forma de recuperação que emprega a técnica de *clusterização*: agrupamento em grupos semelhantes entre si.

característica que limita a ajuda oferecida pelas técnicas de processamento de LN oriunda-se na complexidade destas técnicas tornando-as eficientes apenas quando aplicadas sobre uma pequena quantidade de textos. Uma solução para este problema é proposta em Zhai (1997) que apresenta um método para indexação de documentos testado num conjunto de documentos de 250 mega bytes. Neste método, o autor propõe um modelo probabilístico para realizar um *parsing* de expressões ao invés da indexação por palavras, demonstrando uma significativa melhora na recuperação.

Outras técnicas além das baseadas em linguagem natural também podem beneficiar a tecnologia de RI. O método *relevance feedback* (Haines e Croft, 1993) aprimora a qualidade da recuperação de informação inteligente ao modificar a consulta baseando-se na retroalimentação do usuário. A consulta é modificada através de uma modificação nos pesos que caracterizam seus termos motivada pela informação dada pelo usuário. O método *relevance feedback* é proposto em Haines e Croft (1993) como um aprimoramento do modelo de recuperação que utiliza Redes de Inferência (Turtle, 1991). Redes de inferência são um modelo de recuperação de informação baseado em probabilidade para raciocínio com incerteza. A rede de inferência é um grafo com quatro tipos diferentes de nós: para documentos, para a representação conceitual do conteúdo dos documentos, para as consultas, e o último para a informação desconhecida. A cada nova consulta, os nós são instanciados para cada documento do conjunto e as probabilidades são propagadas para inferir uma probabilidade associada à informação desconhecida; gerando, assim, um ordenamento dos documentos.

Uma aplicação do método de redes de inferência na recuperação de informação é o sistema de recuperação (Callan, Croft, e Harding, 1992) que vem sendo implementado com sucesso sobre uma base de 1 giga byte.

O sistema INQUERY contém um subsistema de *parsing* que comporta uma indexação sofisticada e uma complexa formulação nas consultas.

O método de *relevance feedback* do sistema INQUERY foi utilizado por Daniels & Rissland (1995) em uma pesquisa que também ressalta a importância do uso do paradigma de RBC na recuperação da informação. As autoras propõem um sistema híbrido de RBC-RI que realiza a busca por documentos similares em uma pequena base de conhecimento de casos em função das fortes necessidades de representação de conhecimento que ofereceria uma base maior. O resultado gerado pelo sistema baseado em casos é um conjunto de textos que sugerem uma lista de termos que é usada para definir uma consulta que conduz a recuperação de documentos baseada em textos.

4.2.7 Recuperação de Textos

A pesquisa em recuperação de informação de textos tem como marco histórico a publicação do artigo de Blair e Maron intitulado *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System* em 1985, que descreve um experimento para avaliar a eficiência da recuperação de um sistema de recuperação de textos. Neste artigo, os autores avaliaram o sistema STAIRS da IBM e propuseram um método para calcular o índice de recuperação (*recall*). As conclusões relatadas demonstraram a dificuldade em obter uma alto índice de recuperação usando um simples sistema de recuperação de textos com um grande banco de dados. Este trabalho foi atualizado por Blair (1996) quando ele discute o método antigo e como reduzir os cinco tipos principais de incerteza nas estimativas do índice de recuperação. A lição mais importante resultante deste trabalho refere-se à importância da informação textual. Alguns trechos do artigo reiteram: "Os documentos são onde os dados tornam-se conhecimento, são a inteligência de uma organização – a memória organizacional está mais presente em seus

documentos do que em seus bancos de dados. Se as organizações não detiverem um acesso razoável a sua informação textual, estas correm o risco de administrar por amnésia – de não estarem aptas a lembrar de seus triunfos passados (e construir a partir dos mesmos) nem de seus fracassos (e evitá-los no futuro).

4.2.7.1 Índices de Recuperação e Precisão

A avaliação da eficiência dos bancos de dados de texto pode ser efetuada através de dois parâmetros – índice de recuperação (*recall*) e o índice de precisão (*precision*) (Salton, 1975; Blair & Maron, 1985). O índice de recuperação é a proporção de documentos úteis do total de documentos na base. O índice de precisão indica a taxa de documentos úteis entre o total de documentos recuperados. Blair e Maron (1985) indicam que os sistemas de bancos de dados de texto limitam-se a um índice de recuperação que oferece apenas 25% de documentos relevantes, o que significa que o usuário deve ler todos os textos recuperados para concluir que somente 25% lhe serão úteis. Além disso, a faixa do índice de precisão varia causando uma baixa eficiência, o que pode ser extremamente perigoso em domínios como o Direito, por exemplo, onde questões de extrema relevância estão em jogo.

A baixa eficiência dos bancos de dados de texto originam-se principalmente dos métodos de indexação estatísticos empregados. Os métodos estatísticos não levam em conta o conhecimento, ou seja, eles selecionam termos dependendo apenas de sua freqüência de ocorrência. Em contrapartida, a recuperação baseada em similaridade empregada nos sistemas de RBC pode ser exclusivamente baseada em conhecimento. Um processo de indexação baseado em conhecimento garante mais eficiência na medida que o índice de recuperação aumenta pelo fato dos índices que orientam a similaridade serem escolhidos em função de conhecimento especialista, aumentando as chances de recuperação de documentos relevantes e úteis. Além disso, a indexação baseada em conhecimento

evita baixos índices de precisão na medida que decrescem as chances de recuperação de documentos sem relevância.

Uma comparação dos três tipos básicos de modelos para gerenciamento de textos é apresentada por Gelbart e Smith (1993) onde os autores introduzem o Modelo de Representação de Conhecimento FLEXICON (seção 2.3.1). Os modelos são: *exact match* (combinação exata), *best match* (melhor combinação) e modelos probabilísticos.

4.2.7.2 Modelo Booleano ou *Exact Match* (Combinação Exata)

A expressão combinação exata origina-se da busca pela perfeita combinação entre a consulta e os documentos do banco de dados que o modelo prevê. Este modelo é simplificado na medida que aceita grupos de palavras e não requer nenhuma estruturação determinada para os textos. Entretanto, as consultas têm uma estrutura distinta com palavras-chave e operadores Booleanos do tipo E, OU, e NÃO; gerando uma rigidez que não permite distinção no nível de relevância entre os documentos recuperados.

O maior problema de implementar o modelo Booleano no domínio do Direito reside na necessidade dos usuários utilizarem interfaces para construção de consultas com lógica Booleana, normalmente uma habilidade não dominada por profissionais da área.

Salton (1984) definiu um modelo que chamou de Booleano Estendido, onde usa operadores no intuito de sobrepujar a rigidez dos operadores tradicionais. Entretanto, esta proposta de solução torna a construção da consulta uma tarefa ainda mais difícil. Na busca de soluções para a complexidade da construção das consultas, Das-Gupta (1987) propôs informação semântica e sintática para localizar operadores E e OR embutidos em textos de linguagem natural. Ainda em 1990, JC Smith abordou o problema traduzindo consultas em linguagem natural para consultas Booleanas.

4.2.7.3 O Modelo de Vetor Espacial ou *Best Match*- Não Booleana

Ao contrário do modelo Booleano, o vetor espacial ou *best match* (melhor combinação) não aceita grupos de palavras. Ambos termos das consultas e os documentos são associados a pesos baseados na distribuição estatística (frequencial) dos termos no banco de dados no sentido de determinar sua importância. Desta forma, os documentos podem ser apresentados em ordem de relevância com relação à necessidade de informação representada pela consulta. O modelo representa tanto os documentos como as consultas através de vetores multidimensionais, cujas dimensões são palavras que representam textos. O grau de relevância dos documentos recuperados com relação à demanda do usuário é obtida pela comparação dos vetores da consulta com os dos documentos através de uma medida de correlação de similaridade. Gelbart & Smith (1993) alegam que o modelo obteve bons resultados em comparação a modelos de processamento de linguagem natural propostas na literatura: Dillon & Gray (1982), Fagan (1987), Smeaton (1991), e ainda um baseado em conhecimento de Gey & Chan (1989).

4.2.7.4 Modelos Probabilísticos

Os modelos probabilísticos são baseados no princípio do ordenamento de probabilidade (Robertson, 1977). Estes métodos seguem a abordagem de ordenar os documentos baseados na probabilidade da relevância com relação à consulta baseados na distribuição estatística dos termos nos textos.

4.3 Conclusão

A Extração da Informação é um novo campo da compreensão da linguagem natural que oferece resultados promissores no tratamento da aquisição automática de conhecimento de textos.

Da Lingüística Computacional, um campo que compartilha metas com a compreensão da linguagem natural, a análise de textos para definição de sua estrutura retórica pode proporcionar a informação necessária para a construção de um método de mineração por modelos.

A pesquisa pertinente à aquisição de conhecimento de textos tem sido enriquecida pelos estudos orientados à Recuperação de Informação de bases textuais. Em resposta ao modesto auxílio que as técnicas de processamento de linguagem natural podem proporcionar à RI, as pesquisas passam a propor combinações de técnicas que tendem a evitar estes métodos tradicionais.

Especificamente relacionado à recuperação de banco de dados de textos, três modelos de representação são apresentados oferecendo alternativas para o aprimoramento da recuperação. Com relação ao parâmetros de avaliação, os índices de recuperação e de precisão mantêm-se válidos.

Capítulo Cinco

5. Convertendo Textos em Casos: viabilização de uma aplicação prática

O objetivo deste capítulo é a descrição da metodologia que fornece os meios e o conhecimento, tornando viável a pesquisa jurisprudencial através de um sistema de raciocínio baseado em casos. A metodologia trata-se da conversão de textos em casos e baseia-se em conhecimento extraído de especialistas do domínio do Direito. A metodologia é descrita em dois níveis, na seção Desenvolvimento da Metodologia apresentamos os fundamentos e as tarefas da metodologia, onde fazemos uso do conhecimento extraído na etapa de aquisição de conhecimento para trabalhar os textos no sentido de definir métodos para implementar a metodologia. Na seção 5.2 Implementação da Metodologia, explicamos como implementar a metodologia, o que pode ser aproveitado e o que deverá ser repetido a cada nova implementação.

5.1 Desenvolvimento da Metodologia

O primeiro passo no desenvolvimento da metodologia é obter dos especialistas uma avaliação preliminar das experiências do domínio associando-as às tarefas do sistema. O passo seguinte é a definição dos atributos que compreenderão a representação dos casos de tipo formulário. Em seguida, a análise dos textos modela os textos em subestruturas funcionais. Depois, os atributos são associados às subestruturas nas quais os valores para estes atributos devam ser extraídos. Com estas etapas concluídas, podemos criar métodos para extrair estes valores das subestruturas indicadas utilizando-se conhecimento específico para esta extração. Neste ponto, a metodologia está pronta para implementação.

5.1.1 Avaliação e Representação dos Casos

Representação dos Casos como uma Tarefa Baseada em Conhecimento

Os descritores (vide 3.3) são atributos valorados que podem ser usados para representar casos, desta forma, a seleção dos descritores caracteriza a representação dos casos. A representação dos casos é uma tarefa essencialmente baseada em conhecimento e por isso, demanda uma etapa de aquisição de conhecimento. Os descritores podem descrever o caso, sua solução, bem como outras importantes informações sobre o conteúdo e contexto. Assim, os descritores podem assumir diferentes funções, tais como descrever o caso, indexar o caso para orientar a recuperação, descrever soluções, e armazenar informações sobre o resultado do uso do caso, (Kolodner, 1993). Examinemos a influência do conhecimento especialista na modelagem de descritores com as três seguintes funções: descrever o caso, indexar o caso e sugerir soluções ao caso. Os descritores

que descrevem um caso são usados também como índices e são essenciais para identificar o caso para o usuário como também para orientar a adaptação. Todas as variações nos descritores dependem do domínio da aplicação e da tarefa que o sistema se propõe a resolver e somente um especialista no assunto pode afirmar o que e como descrever os casos.

A indexação de um caso assemelha-se com a indexação de livros. Kolodner²⁶ ressalta que os índices devem: (1) usar o mesmo vocabulário que o usuário; (2) antecipar as circunstâncias da busca; (3) usar os conceitos que são normalmente empregados no domínio; e (4) representar uma interpretação de uma situação. O vocabulário usado no domínio, as circunstâncias de busca, os conceitos e a interpretação das situações representam todos aspectos do domínio do conhecimento; conseqüentemente, somente um especialista com profundo conhecimento do domínio pode realizar a tarefa de indexação.

A mais importante característica da recuperação nos sistemas de RBC é que esta deve ser guiada pela utilidade dos casos. A determinação de quais os descritores que serão úteis na solução de um problema similar não é uma tarefa trivial. Na extração dos valores para os atributos, fazemos uso da abordagem funcional e da abordagem de lembrança (veja seção 3.4), apesar dos descritores não serem necessariamente usados como índices. Alguns deles serão valorados com expressões que não podemos comparar automaticamente com expressões de outros textos. Estamos tentando otimizar a utilização dos textos recuperados, no sentido de oferecer lições ao usuário, contudo não podemos utilizar todas as lições para orientar a recuperação. Acima de tudo, as lições servem como soluções, apesar de também utilizarmos lições como índices.

Na busca por descritores, Kolodner ensina sobre a importância de identificarmos as lições que os casos ensinam e o contexto em que estas ocorrem. Esta lição representou uma grande contribuição no

desenvolvimento de nossa abordagem. Dentro de um domínio específico de conhecimento, os especialistas podem indicar quais as situações que são consideradas lições. Os descritores que descrevem as soluções também devem ser projetados por especialistas do domínio, na medida em que a identificação do que é uma solução para um determinado problema, necessariamente demanda conhecimento sobre este problema.

5.1.2 Atributos e Índices

A solução que viabiliza o desenvolvimento de um sistema de RBC para recuperar textos jurídicos sustenta-se no fato da representação dos casos ser uma tarefa baseada em conhecimento, especialmente no domínio do Direito. A etapa de representação do caso consiste, primariamente, na indexação dos casos. Desta forma, as abordagens existentes para realizar a indexação representam uma fundamentação complementar na solução. Por ser a definição do vocabulário de índices uma tarefa que exige conhecimento especialista sobre o domínio da implementação, podemos afirmar que esta tarefa deve ser realizada com especialistas humanos através de um processo de aquisição de conhecimento.

O processo de indexação engloba a definição do vocabulário de índices e a atribuição de índices. A aquisição de conhecimento necessária para implementar a metodologia de converter textos em casos pode ser dividida em dois estágios. O primeiro refere-se à aquisição de conhecimento no intuito de definir o conteúdo e o contexto que deve compreender os casos para cumprir a tarefa do sistema. O segundo estágio visa a definição dos atributos que englobam todos os tópicos que melhor representem as experiências descritas nas decisões judiciais sempre vislumbrando o cumprimento da tarefa proposta.

²⁶ A teoria apresentada sobre representação dos casos e indexação é retirada da obra de Janet Kolodner, *Case-Based Reasoning*, 1993.

Os atributos definidos como resultado deste processo são:

tipo. O número total de tipos de recursos no escopo do Tribunal de Justiça é por volta de 200. Alguns exemplos são recursos de *habeas corpus*, *apelações criminais*, etc.

número. O número de ordem da decisão.

relator. O nome do relator que redige a decisão.

local. A comarca onde foi iniciado o processo em resposta ao delito que o gerou.

página. Localiza a decisão no arquivo original.

data. Data da decisão.

fundamento (1,n). Fundamento é uma expressão que indica alguma sustentação tanto para o recurso como para a decisão.

tema (1,n). vide seção 5.1.2.1.

artigos secundários. Artigos ou leis que são considerados e/ou citados no recurso ou decisão além da tipificação principal.

tipificação. A categorização do delito que gerou o processo.

resultado. A decisão, que pode ser positiva, negativa ou neutra.

unanimidade. A informação se a decisão foi unânime ou por maioria de votos.

Mesmo com a definição dos atributos acima, o problema de representação dos casos ainda requeria uma solução para possibilitar a contemplação de todo o universo de experiências possíveis de ocorrerem no domínio. Esta solução foi a criação de índices de propósito múltiplo, como descrito a seguir.

5.1.2.1 Índices Jurídicos de Propósito Múltiplo

A interpretação dos especialistas com relação às decisões judiciais determinou o conjunto de atributos que foram definidos como vocabulário de índices para garantir uma recuperação eficiente. A definição do vocabulário de índices é uma das etapas do processo de indexação. A teoria de RBC sugere que no processo de engenharia de conhecimento seja vislumbrado todo o universo de casos que possam ocorrer na tentativa de buscar por todos os valores possíveis para os atributos definidos no sentido de preencher os requisitos de um bom índice.

Para escolher um índice, é preciso garantir que este índice receba um valor em cada caso. Na medida em que a tarefa do sistema em questão requer uma base bastante extensa, compreendendo diversos tipos de casos, fica praticamente inviável cumprir com esta determinação. Por outro lado, há diversos valores que estão presentes nas experiências, que precisam ser representadas como índices, para garantir uma avaliação de similaridade eficiente que resulte na recuperação das decisões mais úteis. Este dilema é uma restrição no processo de indexação que surge com a impossibilidade de valorar um mesmo atributo em todos os casos já que os casos representam experiências que podem variar em suas naturezas.

Nosso objetivo é superar esta restrição para representar adequadamente todos os casos a partir da interpretação dos especialistas. Ao avaliar os diferentes valores, que ocorrem nas descrições das experiências, observamos que as variações nas naturezas dos valores podem ser referenciais para classificá-las em grupos. Alguns objetivos específicos são motivações para alguns tipos de recursos. Por exemplo, são quatro as motivações para um recurso de *habeas corpus*: anulação, relaxamento, trancamento e revogação. Por outro lado, em apelações criminais, há outros tipos de motivações tais como as relacionadas à devida categorização de delitos em conformidade com seu grau de severidade, o que pode ser de extrema relevância na representação deste tipo de

experiência. Mesmo assim, ainda é possível encontrar casos de apelações criminais onde uma motivação como estas não seja aspecto muito relevante. Uma das alternativas para solucionar esta diversidade seria o uso de regras que atribuiriam diferentes pesos para diferentes motivações; contudo, há muitos outros índices cuja relevância na determinação da utilidade de uma decisão exige uma solução mais abrangente.

A solução encontrada foi a definição de índices de propósito múltiplo. Estes índices contemplam valores com múltiplos propósitos que podem ser classificados em diferentes classes. Seguindo a idéia do exemplo anterior, para utilizar um índice para a motivação do *habeas corpus*, seu valor poderia ser, por exemplo, *revogação* e na representação deste caso haveria um outro atributo, este destinado para a classe, cujo valor seria *motivação*, que não é um índice. Desta forma, nomeando este índice com o rótulo genérico *tema*, conseguimos valorar um número suficiente de índices, independente de sua natureza, superando a restrição na indexação de identificar todos os valores para todos os índices em todos os casos.

O índice *tema* destina-se a abarcar aspectos secundários ou circunstâncias presentes na interpretação dos casos. A complexidade destes casos origina-se do fato de que sua idealização surgiu em resposta à complementação da tarefa de definir o universo de atributos capazes de descrever o conteúdo e o contexto das experiências de decisões judiciais. Os valores para o atributo *tema* que aparecem nas decisões podem ser classificados em classes de alguma natureza, tal como: classe de testes solicitados (teste de sanidade mental, teste de dependência química), classe de motivação (solicitando revogação, anulação, relaxamento ou trancamento), classe do contexto externo (acidente de trânsito, redução de pena, paralisação dos serventuários, etc.), classe de qualificadora (arrombamento, requintes de crueldade), entre outros.

A definição de um conjunto de atributos para representar e indexar os casos compreende a representação do tipo formulário. O próximo passo é analisar a estrutura dos textos jurídicos.

5.1.3 Análise de Textos

O processo que denominamos análise de textos é decisivo na viabilização da metodologia. No domínio em questão, os textos que desejamos mapear em formato de casos são decisões judiciais proferidas por um tribunal de justiça: acórdãos. Os relatores destes tribunais são desembargadores que compartilham de uma formação bastante similar, interagem freqüentemente e ainda seguem algumas regras na redação dos acórdãos. A consequência disto são textos altamente estereotipados, o que permite o tratamento computacional, ou pelos menos, mais adaptáveis a métodos computacionais.

Branting e Lester (1996) propuseram uma metodologia para documentos auto-explicativos alegando que o conhecimento sobre as expressões ilocucionárias e sobre as estruturas retóricas de documentos complexos pode ser usado para indexação. A noção por trás das estruturas e expressões ilocucionárias proporcionaram a contribuição necessária para a definição de subestruturas funcionais nas decisões judiciais.

A etapa de aquisição de conhecimento subsidiou a tarefa de definição da estrutura retórica dos acórdãos bem como a identificação das partes onde as expressões ilocucionárias orientam a identificação de dimensões relevantes. Esta forma de examinar os textos foi muito útil e determinante no desenvolvimento da metodologia proposta.

Os acórdãos são analisados pelos especialistas do domínio para definir a estrutura retórica e identificar as partes que contêm as expressões ilocucionárias. Esta é mais uma tarefa baseada em conhecimento. Os relatores destes acórdãos seguem algumas diretrizes em sua confecção.

Por exemplo, a forma dos acórdãos é sempre a mesma, um cabeçalho seguido de uma ementa cujos dois últimos parágrafos sempre iniciam com as mesmas palavras: "Vistos" no primeiro e "ACORDAM", no segundo. Isto não apenas facilita a tarefa como possibilita o tratamento computacional. O formato destes acórdãos nos motivou a delinear subestruturas tais como *cabeçalho e ementa*.

Em uma outra etapa, os especialistas analisam os textos no intuito de assegurar quais as reais características de cada subestrutura, sempre na busca de exceções e diferentes formas de descrever uma mesma idéia. A estrutura retórica resultante dos acórdãos está definida através das subestruturas contidas na Figura 2.

- | |
|--|
| <p>1. Cabeçalho: características superficiais tais como data, local, relator e tipo.</p> <p>2. Ementa: varia em tamanho, inicia imediatamente após o cabeçalho e termina com dois parágrafos de fácil identificação. O primeiro relata quem está impetrando o recurso e o segundo apresenta o resultado. A ementa consiste de três partes: ementa:principal, ementa:impetrante, e ementa: resultado.</p> <p>:</p> <p>:</p> <p>3. Corpo: em sua conclusão freqüentemente encontra-se a decisões e seus fundamentos. É onde se dá a busca por expressões ilocucionárias. Os parágrafos introdutórios descrevem detalhes da situação, local, e indicam capitulação e tipificação, além de apontar para valores de fundamentos.</p> <p>:</p> <p>:</p> <p>:</p> <p>:</p> <p>4. Fechamento: Inicia com um parágrafo sobre os votos seguido de data, local da decisão e nomes dos participantes da decisão.</p> |
|--|

Figura 2 Estrutura retórica dos acórdãos.

Foram realizados testes amostrais para assegurar que cada subestrutura esteja realmente presente em cada acórdão. O desenvolvimento da análise de textos é uma etapa necessária somente uma vez, logo que o domínio é escolhido. A estrutura retórica dos textos somente requer revisão se a metodologia for aplicada para a modelagem de experiências de outro tipo

de tribunal, desde os tribunais de justiça tendem a descrever suas decisões da mesma forma.

5.1.3.1 Associando Atributos às Subestruturas

Logo em que as subestruturas e o conhecimento necessário para modelar a estrutura retórica em um programa computacional estão identificados, os especialistas já podem indicar em quais subestruturas extrair o valor para cada atributo. Os resultados estão na Tabela 4.

| atributo | subestrutura |
|---------------------|--------------------------------------|
| tipo | cabeçalho |
| relator | cabeçalho |
| data | cabeçalho |
| local | cabeçalho |
| número | cabeçalho |
| página | cabeçalho |
| tipificação | ementa:principal e corpo:tipificação |
| resultado | ementa:resultado |
| unanimidade | ementa:resultado |
| tema | ementa:principal e corpo |
| fundamento | ementa e corpo:conclusão |
| artigos secundários | todo o texto |

Tabela 4 Posição dos valores para os atributos na estrutura dos acórdãos.

A subestrutura *ementa* consiste de três partes: *ementa:principal*, *ementa:resultado* e *ementa:impetrante* que indicam diferentes porções de texto que são consistentemente e claramente distinguíveis dos demais. A Figura 1 ilustra com exemplos de subestruturas.

A distinção nítida e clara das subestruturas é um ponto de partida para empregar a técnica de mineração por modelos para extrair os valores dos atributos. Na próxima seção, a descrição dos demais requerimentos de conhecimento especialista.

5.1.4 Conhecimento para a Extração de Informação

A extração da informação é implementada através de métodos de mineração por modelos que procuram por expressões previamente conhecidas em posições determinadas no texto. Na medida que os textos

estão divididos em subestruturas e cada atributo está associado a uma dada subestrutura, cada aspecto pertinente a busca fica contemplado, exceto sobre o que deve ser procurado, o que é também obtido com conhecimento especialista. O resultado desta etapa de aquisição de conhecimento é uma lista de valores possíveis contidos nas decisões que representam o que será extraído pelo método de mineração por modelos (vide 4.2.2).

A aquisição de conhecimento para a definição do que ser extraído visa a enumeração de todos os valores possíveis que possam surgir na atribuição de cada caso. A aquisição é realizada para cada atributo através de um procedimento amostral. Por exemplo, a aquisição dos valores possíveis para o atributo *resultado* em recursos de *habeas corpus* foi feita a partir de uma amostra de 17 acórdãos. A garantia com relação à suficiência desta amostra ocorre apenas ao empregarmos o método ou, de forma mais definitiva, ao testarmos os resultados dos métodos.

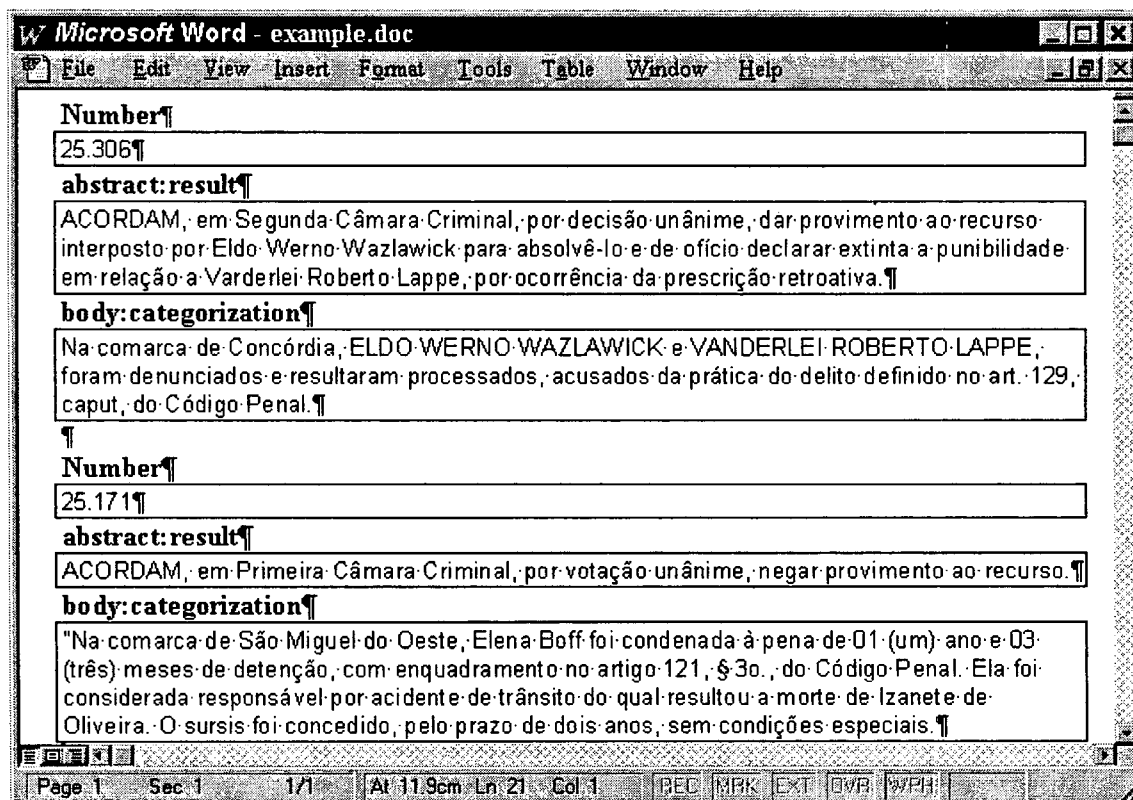


Figura 3 Exemplo de subestruturas.

O processo de desenvolvimento de métodos de mineração por modelos é incremental dentro de cada subdomínio. Primeiramente, implementamos o método baseando-nos exclusivamente nas orientações dos especialistas e executamos o método para avaliar sua eficiência. O resultado é avaliado através da proporção de decisões nas quais os valores para os índices foram extraídos com sucesso. Os textos nos quais não foram encontrados valores para atribuição são revisados no intuito de extrairmos novo conhecimento para ser representado no método.

O processo de extrair conhecimento para os métodos de mineração por modelos para o protótipo atual inicia-se com a reutilização do conhecimento adquirido no protótipo anterior. Por exemplo, a implementação do método para o índice *fundamento* inicia com a lista de valores atribuídos para o índice no protótipo anterior. Os especialistas revisaram a lista de valores para *fundamento* e sugeriram novos e também similares que poderiam aparecer em determinadas decisões. O método foi programado e executado resultando em 150 textos sem valor algum. Uma amostra de 10% destes 150 foi analisada pelos especialistas na busca de novos valores que foram agregados ao método. Este procedimento seguiu-se até que os especialistas pudessem garantir que os textos deixados sem valores realmente não traziam nenhum valor que justificasse o recurso.

Este processo incremental mostrou-se eficaz, apesar de oferecer uma desvantagem. O grande problema é a necessidade de ser confirmado por um especialista quando o texto não retorna nenhum valor para o índice. Uma possibilidade seria a de estabelecermos um limiar a partir do qual assumiríamos que os textos que não retornaram valores realmente não trazem esta informação. A eficiência do método repousa no fato de atribuir valores para 3.500 casos quando somente aproximadamente 100 (cem) textos foram examinados pelos especialistas. Isto representa menos de 5% do total de casos. Contudo, para um novo subdomínio este percentual certamente alcance todos os 5%. O máximo requerido em domínios bastante complexos sugere que o percentual máximo fica em 5%, apesar do

tamanho médio da amostra ser inferior. É interessante observar que o processo de busca pelos valores para os atributos é, de fato, uma forma de extrair amostras para inferir para toda a população. Apesar da metodologia ser estritamente baseada em conhecimento, é confortante observar que o resultado dos experimentos ficam em conformidade com a teoria estatística, com tendências a ser ainda melhor.

O importante é ressaltar que a definição dos atributos não demanda que os especialistas examinem um volume muito grande de textos. Sua capacidade de definir os atributos oriunda-se no seu conhecimento especialista sobre o domínio.

5.1.5 Outros Métodos: Elaboração

Uma das vantagens dos sistemas de RBC sobre outras tecnologias de IA refere-se à representação de conhecimento. Afirma-se que não é necessário que sejam representadas todas as inter-relações entre os elementos de um caso para que o sistema funcione. Testando e trabalhando com os protótipos do PRUDENTIA, constatou-se a veracidade desta alegação. Por exemplo, alguns casos fundamentados com insanidade, resultam na recuperação de casos associados a problemas de saúde mental tal como casos indexados com solicitação de sanidade mental como valor para o índice *tema*. Isto não acontece pelo fato de termos representado esta associação explicitamente, já que nosso sistema não avalia a similaridade semântica de cada valor dos índices. Isto acontece devido à existência de conceitos intrínsecos inerentes às experiências tal como em um caso onde o réu fora enviado a uma instituição psiquiátrica e foi alegada insanidade mental. Desta forma, comprovamos que um sistema de RBC realmente não exige representações explícitas de todas as correspondências entre conceitos para servir seu propósito.

Nesta altura, há um encorajamento para fazer uso das vantagens de representação de conhecimento de outra forma. Na medida em que as

decisões judiciais descrevem experiências que contemplam circunstâncias inerentemente correlacionadas, entendemos que estes casos são, por natureza, consistentes. Desta forma, é válido tentar procurar por valores para aspectos e circunstâncias secundárias em casos que sejam de alguma maneira similares a um determinado caso que desejamos atribuir valores. Esta é a sustentação para a reutilização de casos semente que possam indicar que tipos de valores procurar em um novo caso.

A reutilização de casos é motivada pelo método denominado *elaboração*²⁷. A idéia de reutilizar casos para sugerir possíveis valores pode ser estendida para aos novos casos que ainda não foram indexados. Este é um método alternativo para buscar e atribuir valores especialmente destinados a atributos que não são de muito fácil tratabilidade pelo método de mineração por modelos ou mesmo em subdomínios onde aumentam-se os requisitos de aquisição de conhecimento.

O uso do método de elaboração foi considerado para atribuir valores ao índice *tema* (Weber et al., 1998) em função da complexa variabilidade. Contudo, a aquisição de conhecimento aproveitada de protótipos anteriores resultaram em conhecimento suficiente para extrair valores através do método de mineração de forma satisfatória. A pesquisa sobre a reutilização de casos permanece com uma alternativa a ser empregada na atribuição de valores para o índice *tema* em outros subdomínios.

27 Em K. Branting, (1991), elaboração de casos consiste em inferir fatos que não estejam explicitamente declarados nos casos no intuito de aprimorar a combinação entre casos. A elaboração de casos foi chamada combinação de padrões baseada em conhecimento em (Porter et al., 1990) No contexto do raciocínio analógico, o processo de determinação de propriedades implícitas compartilhadas nos casos a partir de representações que diferem explicitamente vem sendo denominado de reformulação (Russel, 1986). Uma outra noção afim oriunda do campo de aprendizagem de máquina (machine learning) é a indução construtiva que foi definida como "qualquer forma de indução que gere novos descritores que não estavam presentes na informação original" (Dietterich e Michalski, 83). Kolodner (1993) descreve o processo de elaboração como ferramenta de suporte ao ajuste da situação, quando o valor de um índice deve ser inferido, é possível executar uma recuperação parcial buscando-se nos casos similares por valores possíveis para os atributos sem valor.

5.2 Implementação da Metodologia

Uma vez que os métodos de mineração por modelos foram projetados embutindo todo o conhecimento necessário, a execução destes métodos resulta na valorização baseada em conhecimento dos atributos na representação dos casos tipo formulário. Por outro lado, alguns atributos são valorados diretamente em função de sua ocorrência explícita em uma determinada posição fixa no texto original e no formato que deverá assumir no caso.

Apresentamos alguns pontos específicos da prática da conversão de textos em casos examinando cada método de valoração. O conhecimento necessário para identificar a subestrutura adequada e para extrair o valor devido está embutido em diversos métodos. Descrevemos a seguir cada método resultante do desenvolvimento da metodologia para extrair valores para cada índice, culminando na modelagem dos casos.

5.2.1 Extração de valores para a atribuição de índices

Os casos no PRUDENTIA são representados através de representações tipo formulário que compreendem os atributos conforme já descrito. A maioria dos atributos é simples, ou seja, recebe apenas um único valor; enquanto os índices *tema* e *fundamento* podem ter valores múltiplos (1,*n*). Os métodos que viabilizam a atribuição de índices de cada atributos são descritos a seguir.

5.2.1.1 Atributos Explícitos

Os valores para o *tipo* são extraídos diretamente sem a utilização de conhecimento especialista porque a segunda linha no cabeçalho dos

acórdãos inicia com a informação do *tipo*. É somente necessário que se confirme que a expressão completa foi extraída, o que é executado através da comparação com a lista que contempla os 204 tipos de recursos praticados no tribunal em questão. Este mesmo procedimento é executado para os outros atributos presentes na subestrutura cabeçalho: *relator*, *data*, *local*, *número* e *página*.

5.2.1.2 Tipificação

A *tipificação* refere-se à lei que originou o processo associado à decisão, *e.g.*, um crime ou uma contravenção. O formato deste atributo é um de uma lista de palavras que representam o título de um artigo ou de uma lei. A possibilidade de valorar este atributo com o número do artigo é uma alternativa futura. Por enquanto, utilizamos o número do artigo ou lei como um meio de extrair o valor textual para o atributo.

O método para o índice *tipificação* é inicialmente executado sobre a subestrutura *corpo:tipificação*. Esta subestrutura é um parágrafo que normalmente informa o artigo, a lei e a fonte (*e.g.*, Código Penal). O parágrafo normalmente apresenta uma seqüência informando a comarca onde o delito fora cometido seguido de uma expressão equivalente a "*por infringir os artigos 26 e 27, do Código Penal*".

O método extrai o artigo e sua fonte e traduz a informação ao título da *tipificação*. No sistema em questão, este processo atribuiu valores a cerca de 2.600 casos. Os demais textos não informam o artigo explicitamente então devemos passar para um segundo estágio.

O segundo estágio deste método processa-se através de uma busca na subestrutura *ementa:principal*. A busca é pelo título da *tipificação* cuja lista fonte é proporcionada pelos resultados do primeiro estágio e ainda complementando listas de versões anteriores. Este estágio resultou em mais 800 casos com valores para o índice *tipificação*. Os 100 textos restantes foram devolvidos para exame pelos especialistas. Novamente,

nos deparamos com o difícil problema de reconhecer quando um dado valor não é extraído por não estar presente no texto original. Isto pode ocorrer em recursos que são denegados sem maiores explicações. Este processo repete-se incrementalmente até que cada caso seja propriamente valorado.

5.2.1.3 Fundamento

O índice *fundamento* representa aspectos jurídicos ou materiais que substanciam um recurso ou sua decisão. Emprega-se uma estratégia diferente para extrair valores para os fundamentos, a busca direta nas devidas subestruturas por expressões de uma lista. Os valores mais importantes para os fundamentos são normalmente informados na subestrutura *ementa:principal* e os demais ocorrem na subestrutura *corpo:conclusão*. Assim, executamos o método inicialmente na subestrutura *ementa:principal* e em seguida na *corpo:conclusão*.

Uma importante lição extraída na aquisição de conhecimento refere-se ao fato de algumas expressões simplesmente não poderem significar qualquer coisa que não um fundamento, principalmente dependendo da porção do texto onde se encontram. Com este conhecimento, os especialistas puderam definir heurísticas onde podemos assumir que uma expressão tal como *primário* ou *negligência* necessariamente indica um fundamento se esta expressão aparecer nas subestruturas indicadas.

Há expressões com seu significado semântico garantido se extraídas da ementa, o que não podemos afirmar quando estas mesmas expressões aparecerem no na subestrutura *corpo*. Assim, este método contém um estágio intermediário para o tratamento de algumas palavras dependendo da subestrutura de onde foram extraídas. Exemplos seriam as expressões compostas com a palavra *culpa* e conjugações verbais do verbo *confessar* que revelam sua relevância dentro do contexto da decisão principalmente se mencionados na ementa. Estas mesmas expressões no *corpo* podem simplesmente ser parte de uma explicação, não necessariamente indicando um tópico importante associado ao conteúdo da decisão.

O atributo *fundamento* é múltiplo por ser destinado a contemplar diferentes aspectos da sustentação do recurso ou da decisão. Além disso, a busca em diferentes subestruturas resulta em diversos valores. Entretanto, é possível que, em alguns casos, haja apenas um ou dois valores para *fundamento*.

Após a extração de alguns valores de um conjunto de casos, os especialistas observaram a semelhança entre algumas palavras. A lista de expressões teve que ser revisada por especialistas que indicaram expressões que teriam sentidos sinônimos dentro do contexto jurídico. Por exemplo, as palavras *prisão*, *custódia*, *penitenciária*, e até variações do verbo *prender*. A lista de sinônimos aumenta a eficiência, aprimorando a recuperação de casos úteis de uma forma natural.

Na versão atual, o processo iniciou-se com a reutilização da lista de expressões de versões anteriores. O processo incremental foi repetido até que cada caso contemplasse os valores para o índice.

5.2.1.4 Artigos Secundários

O índice *artigos secundários* refere-se a artigos de diferentes códigos de leis que são mencionados durante a descrição das decisões. Isto pode ocorrer quando há um recurso de alteração na tipificação do delito em questão ou, por exemplo, quando nenhum argumento material é considerado em função de uma anulação promulgada devido a um vício formal. Os vícios formais são normalmente previstos em lei.

Os artigos das leis também podem indicar argumentos usados por alguma das partes na validação de alguma afirmação. Os diferentes códigos tornam necessário o uso de atributos bidimensionais, valorando o número do artigo ou lei numa dimensão, e o código ou Constituição fonte na outra, como, por exemplo, o artigo 12, da Constituição Federal.

Os valores para os *artigos secundários* não são procurados em alguma subestrutura específica, mas em todo o texto. Este método é implementado

após a extração da porção que informa a tipificação principal. Depois, seleciona-se ocorrências de números. A mineração é empregada na busca de fontes de leis que ocorram após de expressões como *artigo* e suas variações com o emprego de caracteres curinga.

5.2.1.5 Unanimidade

Na subestrutura *ementa*, o último parágrafo inicia com uma frase onde o valor para a *unanimidade* pode ser extraído, denominamos este parágrafo de *ementa:resultado*. Na ocorrência de opiniões divergentes, a frase conterá uma expressão similar a "*por maioria de votos*". Há em torno de dez variações para expressar tal idéia, facilitando o processo. O valor para *unanimidade* é Booleano porque ou é unânime ou não.

5.2.1.6 Resultado

Há diferentes formas de expressar se um recurso foi concedido ou denegado, e estas formas variam de acordo com o tipo de recurso. A subestrutura que informa o resultado é fortemente estereotípica. A aquisição de conhecimento foi feita para o segundo protótipo que continha recursos de habeas corpus e apelações criminais. Como as formas variam segundo o tipo de recurso, implementamos somente um método. O método seleciona a subestrutura *ementa:resultado* e verifica o tipo de recurso para escolher a base de conhecimento específica para o tipo em questão.

Este é o único método que se manteve implementado em Amzi! Prolog²⁸, já que o uso da programação lógica neste método resultou mais eficaz que o uso de caracteres curinga na identificação de expressões. Além de regras e da definição de diferentes bases de conhecimento, a extração desta informação não seria tratável por um método de mineração por modelos. As expressões resultantes da aquisição de conhecimento realmente demandam um processamento de linguagem natural.

²⁸ Amzi!Prolog 3.3Mar96 Copyright ©94-95 Amzi!inc.

O primeiro requisito para as regras relacionadas ao resultado é o tipo de recurso já que depende do tipo a forma que o resultado é expresso. Por exemplo, em recursos de *habeas corpus*, o verbo empregado para expressão a aceitação do recurso é *conceder* enquanto o verbo *denegar* é utilizado para expressar a rejeição. Em outros recursos, são empregados verbos tais como *prover* e *conceder*. Este conhecimento foi extraído na etapa de aquisição de conhecimento para a elaboração deste método. Este conhecimento reduz o problema de uma maneira que podemos inferir regras como a que segue: "SE o tipo é *habeas corpus* ENTÃO procure na subestrutura *ementa:resultado* pelos verbos *conceder* e *denegar*. Na Figura 4, a interface²⁹ mostra dois exemplos da subestrutura *ementa:resultado* onde o resultado é informado. O comando resultado dispara as regras que retornam o valor do resultado rejeitado (*denegado*) no acórdão 10881 e *concedido* no 10886.

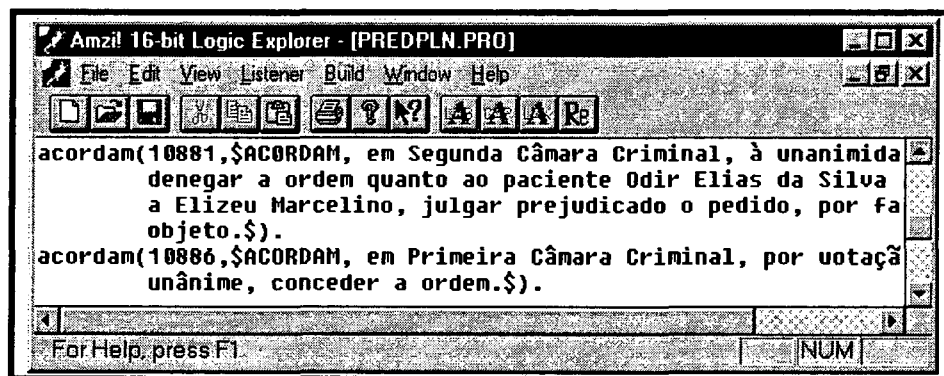


Figure 1 Figura 4 Partes de dois acórdãos onde o resultado é extraído.

Este exemplo demonstra o uso do conhecimento especialista na orientação da busca dos valores adequados no texto. O sistema é projetado para retornar um aviso se nenhum valor for encontrado. Sempre que uma nova expressão for usada por um relator impedindo que o sistema dispare uma regra, o mecanismo de aviso informa sobre o fracasso e uma nova regra pode ser criada. Este mecanismo garante a eficiência e facilita a manutenção do sistema.

²⁹ Amzi!Prolog 3.3Mar96 Copyright ©94-95 Amzi!inc.

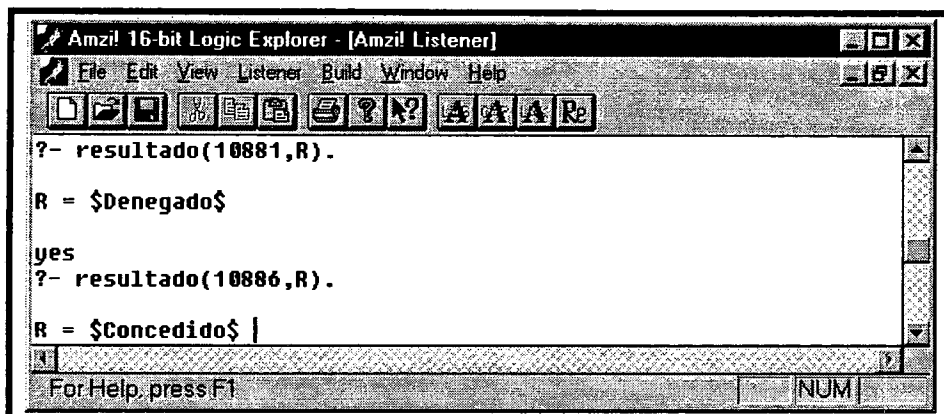


Figure 2 Figura 5 Leitura do valor para resultado em dois acórdãos no Amzi! Prolog.

5.2.1.7 Tema

O índice *tema* destina-se a representar aspectos secundários ou circunstâncias que complementam a modelagem dos casos. A complexidade deste índice é oriunda do fato deste índice ter sido definido para complementar o universo de atributos para descrever o conteúdo e contexto das experiências de decisões judiciais (vide seção 5.1.2.1 Índices Jurídicos de Propósito Múltiplo). O método que atribui valores para este índice completa a tarefa de atribuição automática de índices da mesma forma que a definição deste atributo complementou a tarefa de representação dos casos.

| nome do atributo | método de extração |
|---------------------|---|
| tipo | direto |
| relator | direto |
| data | direto |
| local | direto |
| número | direto |
| página | direto |
| tipificação | artigo da lei confirmado pela lista de tipificações |
| resultado | regras demo selecionam a base de regras |
| unanimidade | pequeno conjunto de expressões |
| tema | lista de expressões e elaboração |
| fundamento | lista de expressões |
| artigos secundários | busca por artigos e fontes |

Tabela 5 Resumo dos métodos empregados na mineração de valores para os atributos.

Na versão atual, o índice *tema* foi valorado através da técnica de mineração por modelos empregada sobre as subestruturas *ementa.principal* e *corpo*. O processo incremental de aquisição de

conhecimento foi realizado em 5% dos textos. Este é o único índice para o qual foi designado um método alternativo, a reutilização de casos por elaboração 5.1.5. A seguir, descrevemos sucintamente como implementar o método de elaboração para a atribuição de valores para *tema*.

Reutilizando Casos

Para o atributo *tema*, introduzimos um meio de realizar a atribuição automática de índices através da implementação de um outro sistema de RBC que usa a própria experiência para aprender a atribuir valores para os índices. Este sistema usa aprendizagem incremental na reutilização de casos sementes que foram modelados de forma semi-automática. Estes casos semente produzem o conhecimento necessário para encontrarmos os valores adequados para os índices nos novos casos.

A motivação para o desenvolvimento deste método oriunda-se no processo de elaboração. A partir de uma base de casos com 138 casos semente, podemos gerar automaticamente uma base com 3.500 casos dentro de um mesmo subdomínio. Estes 138 casos semente foram valorados automaticamente com exceção do índice *tema*.

O problema é encontrar valores possíveis para o atributo *tema* em 3.500 casos que estão representados através dos demais atributos. Uma vez que avaliamos a similaridade de cada novo caso em comparação com os 138 da base de casos já pronta, estamos reutilizando conhecimento embutido nesta base de casos para inferir sugestões com relação aos tipos de valores que devemos procurar nos novos casos. Desta forma, temos que implementar um sistema para desempenhar a aprendizagem incremental visando o crescimento automático da base de casos original.

A base de 138 casos engloba um total de cerca de trezentos (300) valores para o atributo *tema* que estão representados em termos de conceitos, objetos, contextualização, e em termos de suas relações com o caso do qual o índice pertence.

Os casos de entrada são compostos de novas decisões que foram automaticamente modeladas e indexadas exceto pelo valor do índice *tema*. Quando um novo caso é informado, o sistema busca na base por decisões similares, contando apenas com os demais índices para a avaliação de similaridade que orienta esta busca. A saída da recuperação é um conjunto de casos similares que são apresentados ao usuário, indicando os valores atribuídos para o índice *tema*. O usuário escolhe os casos e os valores que o sistema deve tentar encontrar no novo caso. Este é um processo iterativo que termina com um conjunto de sugestões. O usuário é convidado a revisar estes novos valores encontrados no novo caso antes de serem atribuídos e os casos agregados à base. Como resultado, podemos reutilizar o conhecimento embutido nos casos e além disso, usar o conhecimento presente nos novos casos. Isto é possível em função do conteúdo das decisões judiciais serem de alguma forma interrelacionadas. Assim, ao invés de tentar representar simbolicamente estas inter-relações, inferimos sobre eles fazendo uso do contexto de novos casos.

5.3 Avaliação e Conclusão

Como o presente capítulo destina-se a descrever os métodos empregados na conversão de textos em casos, esta metodologia fica sujeita a uma avaliação a partir da qual podemos tirar algumas importantes conclusões. Assim, dedicamos esta seção para resumir a metodologia e discutir os resultados obtidos com estes métodos.

A conversão de textos em casos é a solução para o problema de implementar um sistema inteligente para a pesquisa jurisprudencial. Esta solução é essencialmente baseada no uso de conhecimento especialista e técnicas de mineração por modelos para a extração de informação.

No desenvolvimento da metodologia, empregamos as abordagens observacional e retrospectiva ³⁰ para aquisição de conhecimento. Para dar início à representação dos casos, especialistas usam a abordagem de lembrança³¹ para determinar os atributos que descrevem os casos. Depois, os especialistas examinam os textos no intuito de atribuir valores para cada atributo contemplado nos casos. Por exemplo, o índice resultado de um recurso foi sugerido como uma característica relevante e com o emprego da abordagem retrospectiva de aquisição de conhecimento, os especialistas puderam indicar que o valor para o índice resultado é informado sempre no mesmo parágrafo. Os especialistas mostraram que dependendo do tipo do acórdão (*habeas corpus*, apelação criminal, mandado de segurança) o resultado, bem como outros elementos são informados de forma diferente. Uma nova etapa de aquisição de

³⁰ Para métodos de aquisição de conhecimento veja *Expert Systems: design e development*, capítulo 17, de John Durkin.

³¹ Métodos de Indexação são descritos na seção Indexação, 3.4.

conhecimento é realizada para buscar sinônimos para enriquecer o método.

Como resultado destes experimentos, identificamos importantes etapas para orientar este processo:

- determine as características básicas usando a abordagem de lembrança;
- leia textos para identificar sua estrutura, *i.e.*, qual a função de cada parágrafo (subestrutura);
- associe os valores dos atributos a cada subestrutura;
- determine valores para os atributos buscando-as na subestrutura determinada do texto;
- projete métodos para extrair os valores de cada subestrutura;
- identifique sinônimos;
- extraia os valores mapeando textos em casos.

Estes passos compreendem etapas de representação de caso, análise de texto e métodos de extração de informação para cada atributo. Os processos pertinentes a estes passos são: representação dos casos, análise de textos, associação atributo-subestrutura. Os métodos são: identificador de subestrutura, extrator de tipificação, extrator de fundamento, extrator de artigos secundários, extrator de unanimidade, extrator de resultado e extrator de tema.

Os processos são avaliados pelo funcionamento global do sistema de RBC resultante, já que a engenharia destes processos é decisiva para o sucesso do sistema. Vejamos a avaliação dos métodos, que demonstraram eficiência e espaço para aprimoramentos.

O **identificador de subestruturas** objetiva a identificação automática das subestruturas previamente definidas. Algumas subestruturas como a *ementa:principal* e cabeçalho apresentam limites bem explícitos. Opostamente, a subestrutura *corpo:conclusão* tem limites difusos que são

de difícil identificação até mesmo pelos especialistas. A imprecisão presente nos limites extremos de algumas subestruturas caracterizam problemas em aberto que, quando resolvidos, podem contribuir para melhorar o resultado final do sistema.

O método **extrator de tipificação** demanda revisão. A primeira etapa do método extrai artigos de leis e seus códigos fonte e é absolutamente eficaz. Entretanto, a segunda etapa ainda pode resultar em atribuições incorretas.

O **extrator de fundamentos** é bastante eficaz. A única pendência com relação a este método trata-se da definição de um limiar para evitar etapas desnecessárias de aquisição de conhecimento em textos que efetivamente não informem valores para o índice. Este método usa um dicionário que traduz os termos sinônimos. Esta lista de sinônimos é revisada por especialistas no processo incremental de desenvolvimento.

O **extrator de artigos secundários** depende de um resultado eficiente do extrator de tipificação. Este método não requer um identificador de subestrutura.

O **extrator de unanimidade** é o mais simples. Este método se aproxima da valoração de atributos explícitos que não oferecem muitas variações na forma de ocorrência.

O **extrator de resultado** é muito eficiente e fortemente dependente da aquisição de conhecimento que deve ser repetida em cada novo subdomínio.

O **extrator de tema** demonstrou-se eficiente na forma em que foi implementado no protótipo atual. Este é o único método com uma solução alternativa. Assim como o extrator de fundamentos, este método requer um tradutor de sinônimos, cuja lista deve ser revisada em cada iteração no processo de implementação do mesmo.

Todos os métodos, com exceção do extrator de artigos secundários, têm sua eficiência associada à precisa identificação das subestruturas. Por outro lado, se o método resultar ineficiente, o processo incremental garante seu sucesso. Atribuições incorretas acontecem muito raramente permanecendo este um foco de atenção.

A implementação da metodologia será estendida aos demais domínios dentro do PRUDENTIA. A aquisição de conhecimento para os métodos será repetida, embora sempre aproveitando-se a experiência prévia. A extração conclui o processo de conversão.

A implementação da metodologia para um novo domínio exige apenas a repetição do desenvolvimento da metodologia. Desde que os textos sejam estereotipados, mantendo uma estrutura retórica regular, o desenvolvimento e a implementação da metodologia é viável em qualquer domínio.

Capítulo Seis

6. O Sistema

Este capítulo destina-se a apresentar o sistema PRUDENTIA. Iniciamos descrevendo a pesquisa jurisprudencial e o domínio do Direito. Depois, apresentamos a arquitetura do sistema e a implementação de etapas tais como a representação dos casos, indexação e ajuste da situação. Na seção 6.4, PRUDENTIA: descrevemos as versões do sistema para definirmos os próximos passos. Na seção 6.5, exemplos ilustram o funcionamento do sistema e são comparados com consultas em sistemas de banco de dados. Em 6.6, discutimos questões de validação e concluímos em 6.7.

6.1 PRUDENTIA

O sistema PRUDENTIA busca por decisões judiciais que possam ser úteis ensinando lições a uma nova situação informada pelo usuário. O sistema retorna situações similares que são encontradas através do emprego de raciocínio analógico simulado pela inferência do RBC. O sistema PRUDENTIA ilustra a viabilidade de um sistema de RBC ser aplicado na Pesquisa Jurisprudencial Inteligente decorrente da metodologia descrita no capítulo anterior. A versão atual do sistema engloba 3.500 decisões judiciais que são a fonte de conhecimento da pesquisa jurisprudencial. O sistema realiza o raciocínio analógico ao comparar a nova situação jurídica com as decisões da base de casos, retornando um conjunto de situações similares.

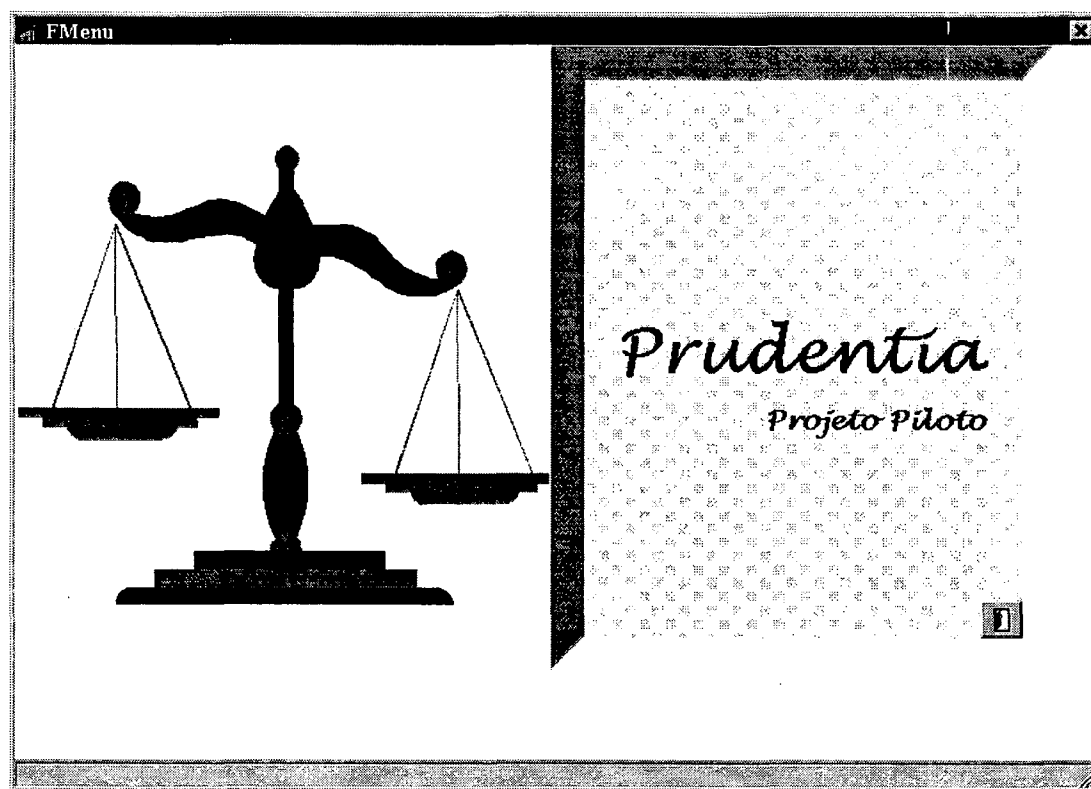


Figura 6 Interface introdutória do PRUDENTIA.

6.2 Pesquisa Jurisprudencial e o Domínio do Direito

O sistema realiza a mesma tarefa que operadores do Direito ao procurar por casos jurídicos na jurisprudência. Quando esta tarefa é desempenhada por especialistas humanos, eles conduzem a busca comparando a interpretação de uma dada situação jurídica com interpretações das decisões encontradas na pesquisa. Os especialistas procuram por situações similares que possam oferecer caminhos para a nova situação. Os especialistas humanos empregam o raciocínio analógico (Durkin, 1994) nesta tarefa. Desta forma, RBC é a tecnologia mais indicada para realizar a tarefa de pesquisa por ser o único paradigma inteligente que simula o raciocínio analógico. O resultado do desenvolvimento deste sistema de RBC equivale a equipar um especialista humano com a capacidade e velocidade de memória de um computador.

A pesquisa jurisprudencial é necessária em várias atividades dentro do domínio do Direito. O Direito engloba diversas atividades desde julgamento e consultoria à administração das atividades jurídicas e ensino. As principais atividades dos operadores do Direito podem ser classificadas em campos da atividade jurídica: planejamento jurídico, argumentação, julgamento e arbitragem, administração, consultoria, ensino, suporte à elaboração das leis, entre outras.

Na atividade de julgar, os juízes e desembargadores fazem uso da pesquisa jurisprudencial como parte do processo de definir e fundamentar suas decisões. Por outro lado, advogados de defesa e promotores tentam provar suas teorias de argumentação sustentando-se na jurisprudência. Entre a maioria destas atividades jurídicas, a pesquisa jurisprudencial representa

uma importante ferramenta que contribui com a correção de cada tarefa. Uma pesquisa jurisprudencial eficaz resulta no fortalecimento de uma sociedade justa.

6.2.1 O Escopo

O alvo da presente pesquisa é a jurisprudência brasileira. O sistema legal brasileiro oriunda-se do Direito Romano, que também é praticado na maioria dos países europeus e caracteriza-se pelo emprego de códigos como a principal fonte de Direito. Este sistema difere do sistema americano e britânico, o *Common Law*, cuja fonte principal de Direito são as decisões judiciais. No sistema brasileiro, a jurisprudência é uma das fontes de Direito, mas não a principal.

Como fonte das decisões judiciais, o escopo do presente focaliza os acórdãos produzidos pelo Tribunal de Justiça de Santa Catarina (TJSC) no período de 1990 a 1996. Os registros do TJSC deste período consistem em 17.2 Mb de 3,447 acórdãos digitalizadas. Estes registros possuem cerca de 2.5×10^6 palavras, com 10^7 caracteres. Estas decisões são a entidade fundamental de nossa aplicação, elas descrevem as experiências que são modeladas como casos no sistema de RBC.

6.3 Arquitetura do Sistema

A arquitetura básica sobre a qual construímos o sistema PRUDENTIA está esquematizada na Figura 7.

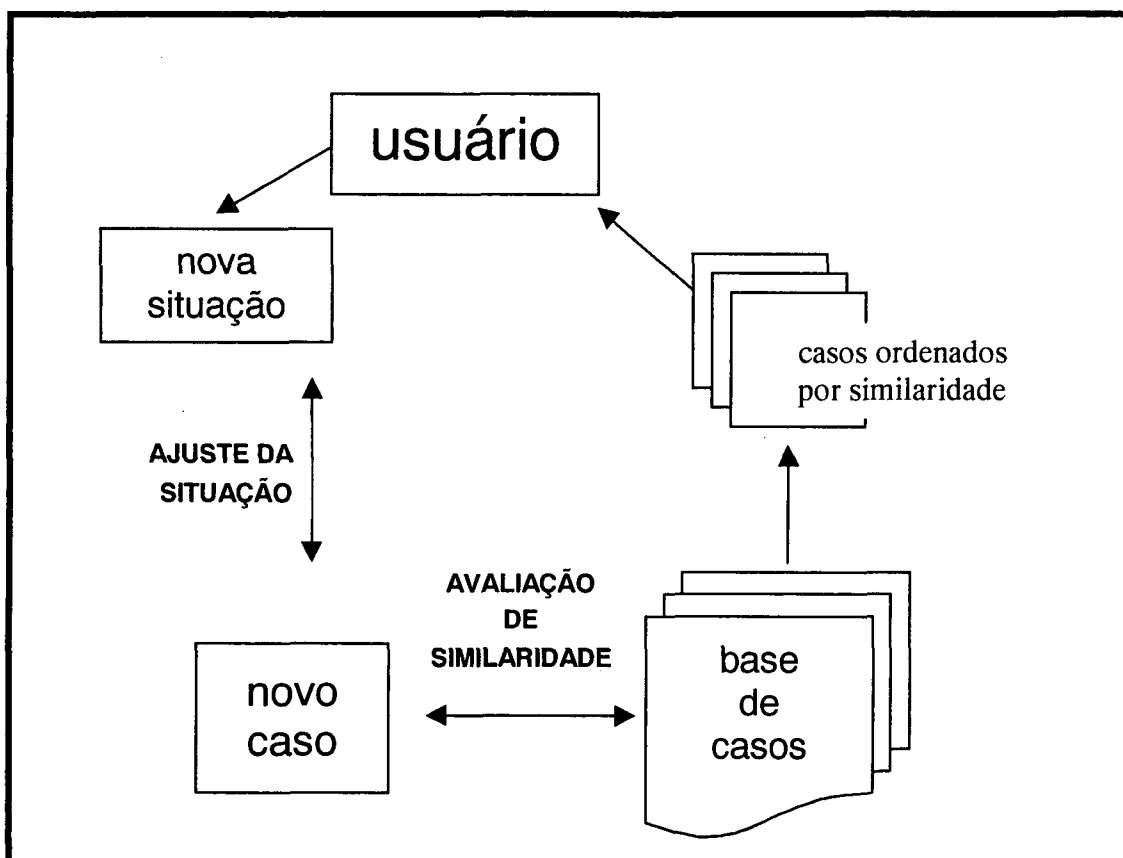


Figura 7 Arquitetura do sistema.

O raciocínio no sistema PRUDENTIA inicia com a identificação de uma nova situação jurídica por um operador do Direito qualquer na prática de alguma atividade jurídica depara-se com uma nova situação que exige uma pesquisa jurisprudencial. Este profissional acessa o sistema PRUDENTIA com sua interpretação com relação a esta nova situação em mente. O sistema inicia o processo tentando apreender a nova situação da mente do usuário através da etapa de *ajuste da situação*. Os métodos de

ajuste da situação procuram inferir valores para atribuir aos atributos em uma representação tipo formulário, modelando a nova situação da mesma forma que os casos na base de casos. Então, o sistema pode comparar a nova situação jurídica - doravante referenciada caso alvo - com cada caso candidato da base de casos. Uma métrica de similaridade mede o valor de cada similaridade para ordenar os casos candidatos para serem oferecidos como resultado da iteração. As próximas seções descrevem as etapas do desenvolvimento do sistema PRUDENTIA.

6.3.1 Casos e a Base de Casos

A base de casos compreende o conjunto de casos e os mecanismos usados para ligá-los à arquitetura do sistema. Na teoria de RBC, a estrutura organizacional refere-se à forma com que os casos são organizados na base. No sistema PRUDENTIA utilizamos casos organizados em uma estrutura plana. A estrutura plana de organização dos casos é implementada em um banco de dados relacional, o que permite um grande volume de casos na base. Os casos são modelados por uma representação tipo formulário, *i.e.*, um conjunto de pares atributo-valor.

6.3.2 Atributos e Índices

O problema da engenharia de conhecimento parte da descrição textual das experiências jurídicas que requerem mapeamento para uma representação estruturada que deve seguir as diretrizes apontadas pelo conhecimento especialista. Estas diretrizes impõem metas e restrições no intuito de manter a representação estruturada em conformidade com uma interpretação especialista da experiência jurídica. Na prática, a correta representação deve resultar em uma forma que quaisquer especialistas percebam o mesmo conteúdo ao lerem a descrição textual do que ao lerem a representação estruturada.

A representação estruturada tipo formulário consiste em um conjunto de atributos que englobam o conteúdo e o contexto da experiência cujo conhecimento deseja-se representar. Cada atributo valorado representa um descritor que sustenta a representação do caso. Os atributos no sistema PRUDENTIA são: *tipo*, *relator*, *data*, *local*, *número*, *página*, *tipificação*, *resultado*, *unanimidade*, *tema(1,n)*, *fundamento(1,n)*, *artigos secundários*. Veja seção 5.1.2 para maiores detalhes.

Dos atributos acima, alguns são tratados como índices *default*, embora outros atributos também pudessem ser utilizados como índices. O vocabulário de índices é a essência da recuperação. Como índices *default* temos: *fundamento(1,n)*, *tema (1,n)*, *tipo*, e *tipificação*.

6.3.2.1 Índices Essenciais

Os índices essenciais são indispensáveis para garantir o sucesso da avaliação de similaridade e para assegurar uma recuperação eficiente. Os índices *fundamento* e *tema* são os essenciais devido a sua natureza e relevância na recuperação. Os índices como *tipo*, *local* ou *tipificação* podem ser usados como índices primários como em uma operação de banco de dados objetivando particionar a base.

6.3.2.2 Índices Básicos

O subconjunto de índices básicos consiste de valores para os atributos *tipificação* e *tipo*, e pelo menos mais dois valores dos índices essenciais. A definição deste subconjunto define o mínimo necessário para iniciar o processo de ajuste da situação. Assim, os índices básicos são os valores mínimos para representar uma experiência no presente contexto.

6.3.3 Avaliação da Similaridade

O que faz um caso ser similar a outro e como avaliamos esta similaridade em um sistema computacional? A resposta para esta pergunta também é respondida pelos especialistas. Eles indicam níveis de importância de um índice com a relação aos outros. Os especialistas comparam os casos e determinam níveis de similaridade entre cada dois casos para referenciar a calibragem dos pesos. Todos estes são procedimentos no processo de extrair conhecimento consistente sobre a similaridade para definir pesos capazes de modelar a importância dos descritores. Este é um processo incremental que varia com diferentes combinações de índices.

A avaliação de similaridade é modelada com conhecimento especialista com relação a dois diferentes aspectos. Primeiro, o valor de similaridade é atribuído quando dois valores são comparados através de uma função de combinação. Como abordagem básica, uma função de combinação retorna valores um (1) e zero (0) para similar e não similar, respectivamente. Em um segundo plano, dependendo da orientação do conhecimento, os valores resultantes podem variar no intervalo fechado $[0,1]$ para níveis graduais de similaridade.

O índice *tipo* recebe valores graduais porque os especialistas concordam que um caso representando uma experiência de um tipo diferente de recurso pode, de fato, ser útil a uma determinada situação jurídica. Supondo um caso alvo sobre um recurso de *habeas corpus*, o índice *tipo* recebe valor de similaridade 1,0 em cada caso candidato com o mesmo valor de *tipo* e recebe o valor 0,7 para candidatos com valor *apelação criminal*. A avaliação de similaridade para os demais índices está comentada na seção 6.3.3.1.

O segundo aspecto da similaridade refere-se ao peso determinado para os índices que expressam a importância dos mesmos no total da avaliação de

similaridade. Estes valores são formas de modelar a importância relativa dos índices. Esta modelagem é gerenciada pela métrica de similaridade que retorna uma medida. No sistema PRUDENTIA, esta métrica é obtida por uma avaliação sintética que reúne pesos e valores de similaridade resultantes das funções de combinação. Contudo, a avaliação de similaridade no PRUDENTIA faz uso de uma abordagem mais específica para o tratamento dos índices de propósito múltiplo.

6.3.3.1 Avaliação de Similaridade dos Índices de Propósito Múltiplo

A meta mais importante da avaliação da similaridade é a modelagem da capacidade do especialista. A meta do sistema é achar os casos mais úteis e ordená-los para o usuário. A referência para um caso útil é um caso similar. A interpretação das decisões judiciais pode, de fato, indicar alguns aspectos direta ou indiretamente pertinentes à utilidade, indo além da similaridade. Como já mencionado, com relação ao tipo de recurso, um acórdão de um *habeas corpus* pode ser tão útil quanto um de apelação criminal para um terceiro recurso. Desta forma, se a tipificação do crime for, por exemplo, homicídio, não faz muita diferença qual o tipo de recurso de onde pode se extrair uma boa lição.

A importância das circunstâncias, na avaliação de similaridade entre dois casos, pode aumentar com relação à tipificação, se as circunstâncias destes casos forem muito similares. O uso de importâncias variáveis exige um tratamento especial na medida que os métodos tradicionais não lidam com tais idiosincrasias. A definição de valores variáveis de importância é a primeira abordagem que surge, apesar de que uma implementação desta idéia requer bastante cuidado, na medida que estamos lidando com índices de propósito múltiplo que podem assumir diferentes valores.

É importante reiterar que cada índice de propósito múltiplo é apenas um índice; ou seja, seu valor contribui como apenas um índice. Ilustremos uma inferência usando o índice de propósito múltiplo *fundamento(1,n)*. A

cardinalidade indicada pelo intervalo $(1, n)$ indica que o número máximo de valores que o índice pode ter em um caso alvo será n . No protótipo atual, $n = 6$. Os casos candidatos da base têm até quatorze (14) valores para *fundamento* e sete (7) valores para *tema*. Desta forma, quando estabelecemos um peso de oito (8) unidades para este índice, isto não quer dizer que há seis possibilidades de combinação cada uma contribuindo com oito unidades para o valor total da similaridade. Opostamente, isto significa que há somente um único valor de similaridade (resultante das funções de combinação) que corresponderá ao peso de valor oito. Por outro lado, isto também não significa que um dado caso alvo deva comportar seis valores. Uma forma alternativa a esta abordagem seria variar o peso do índice ou fixar o número possível de valores a serem atribuídos.

A questão de determinar pesos em uma avaliação sintética ganha relevância na medida em que o valor resultante da similaridade contribui com a recuperação de um determinado caso quanto mais próximo este valor estiver da unidade. Opostamente, este valor diminui a medida final da similaridade na medida que tende a zero. Em resumo, nosso problema trata-se de manter livre o número de valores para o índice *fundamento* seguindo-se as duas diretrizes:

- (1) a importância relativa decresce na medida em que mais similaridades são identificadas entre dois casos, e
- (2) a variação no número de valores para um índice não exercer grande influência no valor total da similaridade.

A maneira com que lidamos com esta representação é esticar os limites do valor da similaridade para aumentar a importância relativa dada ao índice *fundamento*³². O incremento é confirmado pela visão dos especialistas com relação a decisões onde a importância relativa da

³² Apesar de ilustrarmos a explanação com o índice *fundamento*, a mesma abordagem é empregada para o índice *tema*.

tipificação decresce quanto mais valores similares são encontrados em dois casos sob avaliação.

O índice *fundamento*, da forma que é implementado no protótipo atual, pode assumir os valores apresentados na Tabela 6.

| fundamento (1,n) | similaridade | peso | contribuição para o valor total de similaridade |
|--------------------------|--------------|------|---|
| um valor similar | 1.0 | 8.0 | 8.0 |
| dois valores similares | 1.2 | 8.0 | 9.6 |
| três valores similares | 1.4 | 8.0 | 11.2 |
| quatro valores similares | 1.6 | 8.0 | 12.8 |
| cinco valores similares | 1.8 | 8.0 | 14.4 |
| seis valores similares | 2.0 | 8.0 | 16 |

Tabela 6 Intervalo de valores de similaridade para *fundamento*.

Em um primeiro momento, pode parecer como se houvésemos estipulado um peso de dezesseis (16) unidades ao índice *fundamento* quando os seis valores forem similares. Entretanto, o número que é adicionado à soma dos pesos permanece fixado em oito (8) unidades. O efeito final é que a importância relativa dos outros índices diminui, demonstrando que os objetivos são alcançados.

6.3.3.2 Integrais Difusas

A medida de similaridade sintetiza a importância relativa dos índices e o significado da similaridade em uma medida de similaridade. Os métodos matemáticos tradicionais para sintetizar este tipo de informação precisam obedecer axiomas de aditividade e continuidade. Entretanto, os especialistas sugerem que a importância combinada de dois descritores pode não ser necessariamente aditiva. Por exemplo, a importância de um *fundamento insanidade* pode ter valor 0,8 e a importância do *tema atestado de saúde mental* pode ser 0,6; entretanto, sua importância combinada não deve ser 1,4. A Teoria dos Conjuntos Difusos ³³ apresenta métodos que possibilitam a definição de funções de agregação que não somam linearmente os pesos das características, representando um

³³ Sobre Teoria dos Conjuntos Difusos veja Kaufmann, 1975, Zadeh, 1975, e Wang, 1992. Para uma aplicação em RBC, veja Baldwin, 1993.

método alternativo de modelar aspectos subjetivos das interpretações das decisões judiciais.

Propomos o uso da Teoria de Medidas Difusas (*Fuzzy Measure Theory*), (Sugeno, 1977) para realizar a avaliação sintética. Perante um problema de múltiplos critérios, como o da função de similaridade, as avaliações sintéticas tradicionais fracassam em virtude de sua natureza aditiva. Na vida real, os efeitos dos fatores são interativos. O uso da avaliação sintética, calculada através de integrais difusas considera todos os atributos e suas inter-relações. Uma abordagem para empregar a Integral de Sugeno foi proposta por Weber et al. (1995) e permanece uma atraente alternativa para a modelagem da avaliação da similaridade.

6.3.4 Ajuste da Situação

O objetivo do ajuste da situação é avaliar a descrição da nova situação e modelá-la na mesma forma que os casos da base. Cada nova sessão no sistema PRUDENTIA inicia com o processo de ajuste da situação. O sistema procura identificar a nova situação trazida pelo usuário objetivando a modelagem desta situação na mesma forma que os casos da base.

A necessidade de uma etapa de ajuste da situação origina-se na natureza de sistema interpretativo que é o PRUDENTIA. A dificuldade em projetar os métodos de ajuste da situação reside nos obstáculos oferecidos pela compreensão da linguagem natural, já que o usuário traz na mente uma interpretação do tópico que motivou a necessidade de pesquisa. Como os usuários do sistema são também operadores do Direito, necessita-se outra etapa de aquisição de conhecimento para projetar meios para superar as dificuldades impostas pela linguagem natural e alcançar a mente do usuário.

Os especialistas indicaram como os operadores do Direito tendem a descrever as situações jurídicas. Um problema é que, apesar dos

descritores serem extraídos de acórdãos, estes não seguem pressupostos jurídicos ou mesmo a teoria do domínio. Assim, é necessário extrair do usuário os valores para cada atributo que compreenda a representação do caso.

Fundamentalmente, objetivamos encontrar os valores para cada índice na nova situação jurídica. Os valores superficiais tais como o tipo de recurso são perguntados diretamente ao usuário. Descrevemos o método para atribuir valor ao atributo *tipificação* e propomos a combinação com um método de elaboração com listas de palavras para realizar a atribuição de valores para *fundamento e tema*.

Um dos métodos de ajuste da situação visa a atribuição de valores para o atributo *tipificação*. A atribuição deste atributo no caso alvo pode ser problemática. De fato, a *tipificação* está associada a um crime ou a uma contravenção (o enquadramento criminal quando na área criminal) ou ao assunto do processo. De acordo com Código Penal, que representamos por uma árvore de tipificações, o usuário final pode informar a *tipificação* com alguma expressão em qualquer nível da árvore. O nível mais alto é a área, *e.g.*, civil ou criminal. Supondo uma *tipificação roubo*, o nível mais alto é *criminal* e ainda há um nível intermediário cujo valor é *crimes contra a propriedade*. O módulo de ajuste da situação recebe a expressão informada pelo usuário e procede a busca por este valor na árvore de tipificações. Assim que a expressão é encontrada, a interface solicita confirmação do usuário com relação aos níveis mais altos da árvore (vide Figura 8). Em algumas situações o número de ramos abaixo de uma expressão é numeroso e torna-se necessária um refinamento da *tipificação* para um nível mais específico.

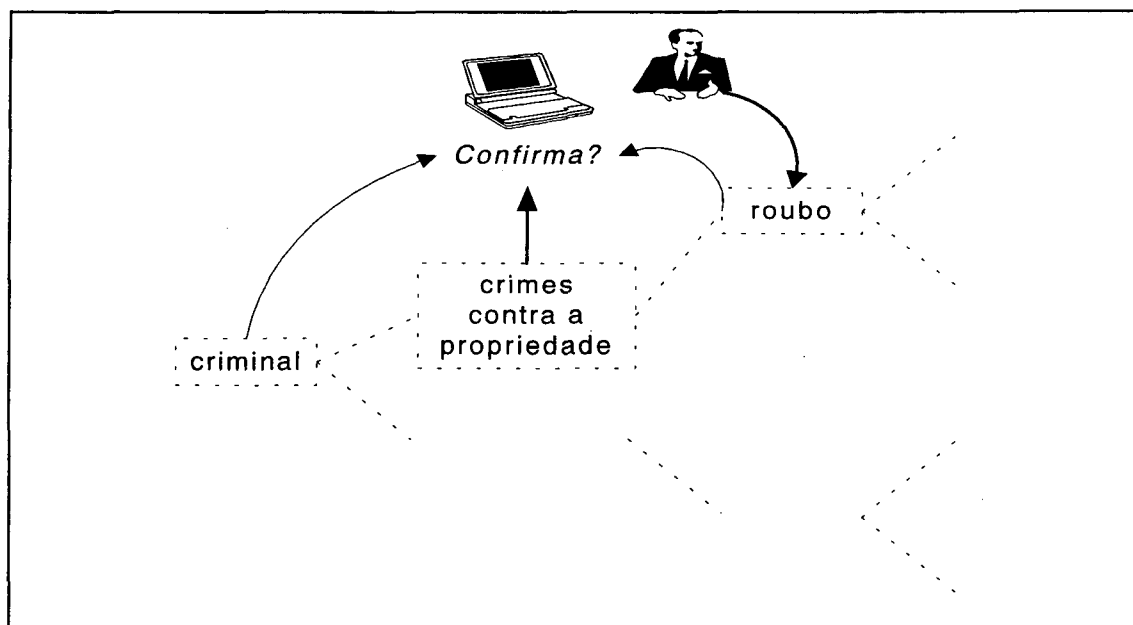


Figura 8 Atribuição do índice *tipificação* com auxílio da árvore de tipificações.

A partir da constatação de que os operadores do Direito não são familiarizados com linguagens de programação e nem mesmo com a prática de construção de consultas com operadores lógicos, é importante que seja projetado um eficiente método de ajuste da situação exigindo o mínimo do usuário jurista.

Escolhemos utilizar um método de elaboração para gerar uma lista de expressões para que o usuário escolha os valores que sejam pertinentes à nova situação jurídica que o sistema procura reconhecer. Este processo é conduzido por meio de uma lista de expressões que realmente valoram atributos nos casos da base. Entretanto, a lista pode ultrapassar 200 expressões, tornando problemático e demorado para o usuário encontrar as expressões devidas numa lista tão grande. Assim, utilizamos o método de elaboração para reduzir o tamanho da lista de expressões. O método de elaboração é executado inicialmente com uma iteração que recupera os casos mais similares a partir de um caso alvo incompleto construído com as expressões extraídas do resumo da situação informado pelo usuário.

Descrição do Caso

Informe o tipo de recurso :

Informe o enquadramento :

Descreva seu caso :

Glossário genérico de conceitos que normalmente são usados nos acordãos :

Ação privada
 Ação pública
 acervo probatório
 Acidente com morte
 Acidente de trânsito

Número de Casos : 138

Figura 9 Usuário redige um resumo da nova situação.

6.3.4.1 Metodologia

A interface inicial pergunta ao usuário jurista para redigir um breve resumo da situação que motivou a pesquisa e para preencher os valores respectivos à *tipificação* e *tipo de recurso* conforme ilustra a Figura 9. Um glossário genérico apresenta a lista de expressões originada com os valores atribuídos aos índices essenciais.

O sistema busca expressões que apareçam no resumo que também pertençam à lista de valores para *fundamentação* e *tema* na busca de compor os índice básicos para realizar uma primeira iteração. Os casos similares resultantes são selecionados para prover valores dos índices *tema* e *fundamento* para compor o novo glossário reduzido: o glossário específico. Esta nova lista de expressões é mais curta, facilitando a escolha de valores que estejam em conformidade com a nova situação. Este é um procedimento incremental que objetiva auxiliar o usuário com sugestões de expressões que podem ocorrer tanto na nova situação quanto nas experiências da base.

Pode acontecer que o usuário não informe os índices básicos no resumo inicial. Assim, ainda é possível fazer uso dos únicos atributos valorados (tipo e *tipificação*) e usá-los como índices primários como uma consulta de banco de dados que particiona a base de casos gerando a redução do glossário de palavras.

O método de ajuste da situação continua reduzindo a lista até que o sistema atinja os requisitos mínimos para a constituição de um caso alvo. O sistema pode, então, iniciar a recuperação sem considerar qualquer das iterações prévias pertinentes ao ajuste da situação.

6.4 PRUDENTIA: Estágios de Desenvolvimento

Inicialmente, desenvolvemos um protótipo usando apenas acórdãos em recursos de *habeas corpus* em processos de homicídio no intuito de demonstrar o potencial do uso de um sistema de raciocínio baseado em casos para recuperar casos jurídicos. Os índices foram escolhidos com o objetivo de captar os pontos fortes e fracos dos textos com relação à possível utilidade dos mesmos. O primeiro protótipo foi testado com 22 casos e foi desenvolvido com uma ferramenta de aplicativos.

A reação dos especialistas jurídicos ao primeiro protótipo nos motivou a desenvolver um sistema capaz de englobar todos os tipos de decisões judiciais. Os especialistas sugeriram descritores e algumas características para as interfaces. Eles também sugeriram um mecanismo para realizar novas recuperações baseadas em um conjunto menor de descritores que poderiam ser escolhidos pelos usuários. A necessidade do uso de conhecimento especialista sobre o domínio da aplicação no desenvolvimento de um sistema de RBC ficava cada vez mais evidente, principalmente para as etapas de avaliação da similaridade e ajuste da situação.

O segundo protótipo foi desenvolvido para a área criminal para dois tipos de recursos: *apelações criminais* e *habeas corpus*. A base de casos contempla 138 casos que foram modelados semi-automaticamente.

O terceiro protótipo contempla um conjunto de 3.500 experiências que foram automaticamente convertidas em casos. Estes casos representam a experiência de todas as apelações criminais que foram submetidas ao TJSC no período de 1990 a 1996.

Os próximos estágios referem-se à geração da base de casos para todos os recursos de *habeas corpus* para o mesmo período e, a seguir, para as demais recursos dentro da área criminal. Esta etapa aumentará o tamanho da base para aproximadamente 10.000 casos e o conhecimento já extraído será reutilizado. As etapas de aquisição de conhecimento, ainda necessárias, destinam-se a extrair circunstâncias associadas às novas tipificações dos demais subdomínios. Estes próximos estágios serão principalmente caracterizados pela atenção ao tempo de execução e precisão.

O próximo grande esforço de aquisição de conhecimento acontece no início da modelagem dos acórdãos da área civil. A inclusão dos acórdãos civis representa uma nova implementação da metodologia de conversão. Conforme explicado na seção 5.2, o processo incremental inicia-se com implementações anteriores, passando por novas etapas de aquisição de conhecimento enquanto necessário. Contudo, é possível antever que a terminologia jurídica relacionada ao índice *tema* será renovada. Na atribuição de valores para fundamentos, a maioria dos princípios formais são reutilizados. Com relação aos valores circunstanciais, novos tópicos serão considerados, desde que talvez não seja considerada a arma do crime, mas intenção de causar dano é a mesma intenção de cometer um crime.

6.5 Exemplos

Uma nova sessão no sistema PRUDENTIA inicia motivada por alguma situação surgida quando algum operador do Direito exerce uma atividade jurídica. Este jurista é o usuário do sistema de Pesquisa Jurisprudencial Inteligente que procura lições e soluções para um novo episódio jurídico que leva em mente.

A primeira meta do sistema é captar esta nova situação que está na mente do usuário. Conforme mostrado na Figura 9, seção 6.3.4.1, o usuário é primeiro solicitado a preencher valores para o *tipo* de recurso e a *tipificação*, e digitar um breve resumo desta situação que origina a sessão.

6.5.1 Exemplo "porte"

Neste exemplo, o resumo da nova situação informado relata:

"O réu apela de sua condenação por porte de drogas em função da pequena quantidade de cocaína apreendida o que indica ausência de fins comerciais."

A partir do resumo digitado pelo usuário, o método de ajuste da situação no sistema é capaz de atribuir os seguintes valores para os índices:

tipo = apelação criminal

tipificação = tráfico de entorpecentes

fundamento(1) = fins comerciais

fundamento (2) = apreensão

tema (1) = quantidade de droga

tema (2) = cocaína

Neste exemplo, o sistema considera suficientes os índices atribuídos e utiliza os mesmos para criar o caso alvo. A recuperação inicia-se e os resultados estão na Figura 10.

The screenshot shows a window titled "FCasosEsc" with a header "Casos Seleccionados". Below the header is a table with the following columns: Acórdão, Posicao, Valor Similar, DJ, Data, Pag, Localizacao, and Re. The table lists eight cases, with the first two having a similarity value of 100.00. Below the table, there is a text area containing the following text:

CRIME CONTRA A SAÚDE PÚBLICA. INCORRÊNCIA DE TRÁFICO DE ENTORPECENTE. DESCLASSIFICAÇÃO PARA PORTE PARA USO PRÓPRIO. INEXISTÊNCIA DE PROVA SEGURA ACERCA DA PRIMEIRA HIPÓTESE. Indemonstrado nos autos a exteriorização do tráfico, a capitulação correta da infração está no artigo 16 da Lei n. 6.368/76.

Recurso provido para esse fim.

Vistos, relatados e discutidos estes autos de apelação criminal n. 33.458, da comarca de Joinville (1a. Vara), em que é apelante Valmir José Dionísio, sendo apelada a Justiça, por seu Promotor:

ACORDAM, em primeira Câmara Criminal, por votação unânime, dar provimento ao recurso para, desclassificado o crime de tráfico de entorpecente para o de porte para uso próprio, condenar o apelante a seis (06) meses de detenção, como incurso nas sanções do artigo 16 da Lei n. 6.368/76, mantida a multa fixada na sentença.

At the bottom of the interface, there are navigation buttons: a set of four arrows (left, right, double left, double right), a button labeled "Caso Escolhido", and a button labeled "← Voltar".

| Acórdão | Posicao | Valor Similar | DJ | Data | Pag | Localizacao | Re |
|---------|---------|---------------|-------|----------|-----|-------------|-----|
| 33.458 | 1ª | 100.00 | 9.327 | 27/09/95 | 09 | Joinville | Nã |
| 33.698 | 2ª | 100.00 | 9.349 | 31/10/95 | 08 | Itajaí | Na |
| 27.327 | 3ª | 97.16 | 8.500 | 19/05/92 | 13 | Itajaí | Nã |
| 27.747 | 4ª | 97.16 | 8.604 | 16/10/92 | 06 | São José | Err |
| 25.439 | 5ª | 97.16 | 8.076 | 24/08/90 | 13 | Joinville | Err |
| 27.156 | 6ª | 97.16 | 8.267 | 07/06/91 | 11 | Lages | Na |
| 28.703 | 7ª | 95.74 | 8.595 | 05/10/92 | 06 | Chapecó | So |
| 26.790 | 8ª | 95.74 | 8.207 | 12/03/91 | 16 | Capital | Th |

Figura 10 Casos recuperados, exemplo "porte".

Há dois casos com medida de similaridade 100. A interface na Figura 10 mostra os oito casos mais similares. A flecha preta na parte superior esquerda indica o caso selecionado. Na segunda metade da interface, é mostrada parte do acórdão original. Ao deslizar através da lista de casos para a direita, visualiza-se os descritores dos casos. Este é um mecanismo muito importante porque uma das reclamações dos usuários dos sistemas disponíveis de bancos de dados de texto é sobre a necessidade de ler todo o acórdão a fim de determinar sua possível utilidade. A representação dos casos permite oferecer ao usuário o mesmo resultado que uma rápida leitura do texto.

Referindo-se o presente exemplo a um cliente que visa uma apelação, um dos aspectos a verificar é se ambos os casos com similaridade 100 têm um resultado positivo, ou seja, tiveram sua apelação concedida. Normalmente,

um recurso com resultado positivo é mais provável de ensinar a direção para o usuário que deseja ter seu recurso concedido; enquanto que resultados negativos podem prevenir com relação a um possível fracasso. A Figura 11 confirma que o primeiro caso na ordem é o único com resultado positivo, fazendo deste caso uma boa escolha na condução da pesquisa.

The screenshot shows a window titled "FCasosEsc" with the subtitle "Casos Seleccionados". It contains a table with three columns: "Resultado1", "Unanimidade1", and "Tema4". The first row is highlighted, showing a positive result, unanimous agreement, and the topic "perícia". Below the table, there is a detailed text block for the selected case, followed by navigation buttons: "Caso Escolhido" and "Voltar".

| Resultado1 | Unanimidade1 | Tema4 |
|------------|--------------|-----------------------|
| positivo | sim | perícia |
| sem | sim | erva |
| negativo | sim | continuidade delitiva |
| negativo | sim | narcotráfico |
| positivo | sim | entorpecente |
| negativo | sim | droga |
| negativo | sim | pena |
| negativo | sim | perícia |

CRIME CONTRA A SAÚDE PÚBLICA. INCORRÊNCIA DE TRÁFICO DE ENTORPECENTE. DESCLASSIFICAÇÃO PARA PORTE PARA USO PRÓPRIO. INEXISTÊNCIA DE PROVA SEGURA ACERCA DA PRIMEIRA HIPÓTESE.
 Indemonstrado nos autos a exteriorização do tráfico, a capitulação correta da infração está no artigo 16 da Lei n. 6.368/76.
 Recurso provido para esse fim.
 Vistos, relatados e discutidos estes autos de apelação criminal n. 33.458, da comarca de Joinville (1a. Vara), em que é apelante Valmir José Dionísio, sendo apelada a Justiça, por seu Promotor:
ACORDAM, em primeira Câmara Criminal, por votação unânime, dar provimento ao recurso para, desclassificado o crime de tráfico de entorpecente para o de porte para uso próprio, condenar o apelante a seis (06) meses de detenção, como incurso nas sanções do artigo 16 da Lei n. 6.368/76, mantida a multa fixada na sentença.

Navigation buttons: [Left Arrow] [Right Arrow] [Case Selected] [Return]

Figura 11 Resultados dos casos mais similares.

Este exemplo levanta a questão sobre o uso do atributo resultado como índice ou não. Uma recuperação englobando o índice resultado recupera com valores maiores os casos que apresentarem o mesmo resultado que o caso alvo. Estas são duas formas de combinar os índices que podem ser opções para os usuários escolherem.

O caso selecionado é definido por um pressionar do mouse sobre um caso. Isto permite a visualização da representação tipo formulário do caso selecionado como mostra a Figura 12. A visualização da representação tipo formulário permite ao usuário entender e interpretar o documento mesmo sem lê-lo. Com tal interpretação baseada em conhecimento da experiência,

o usuário pode decidir por reutilizar ou não a mesma. Neste exemplo, o usuário considera o contexto da interpretação merecedor de utilização.

Caso Recuperado

Acórdão : 33.458 Posição : 1ª Similaridade : 100.00 Tipo de Recurso : Apelação Criminal

Tipificação : TRÁFICO

Resultado : positivo Unanimidade : sim

Fundamentos :

| | |
|----------------|------------------|
| agravante | antecedentes |
| apreensão | atenuante |
| caracterização | comercialização |
| confissão | desclassificação |
| flagrante | intenção |
| materialidade | prova |
| | |

Temas :

| |
|--------------|
| cocaína |
| entorpecente |
| erva |

Acórdão Voltar

Figura 12 A representação tipo formulário do caso selecionado.

Do texto do acórdão, o usuário seleciona o trecho transcrito na Figura 13 porque este ensina uma importante lição. O texto sugere a necessidade de testemunhas e instrumentos (como uma balança de precisão) para autorizar a condenação, podendo ser aproveitado na situação atual.

... Além, da quantidade da droga ser ínfima, assevera que faltam outros elementos a autorizar um decreto condenatório com base no art. 12 da Lei n. 6.368/76, tais como a identificação de pessoas que tivessem adquirido ou pudessem adquirir a droga e instrumentos pertinentes ao tráfico.

Figura 13 Trecho extraído do texto do acórdão.

Contudo, a pesquisa ainda pode continuar de duas maneiras, procurando lições em outro caso por ser similar ou por seus atributos. Examinando os atributos que resumem o conteúdo dos outros casos, observa-se que o quinto caso da lista (Figura 14) também tem resultado positivo e o valor

para o *fundamento* (2) = anulação da decisão. Isto já sugere utilidade deste caso. É possível visualizar a ementa do quinto caso pela janela da interface que mostra parte do acórdão, Figura 14. A ementa sugere que o crime não ficou caracterizado pela falta de consciência da formação da sociedade para traficar. Ainda, ao examinarmos os demais atributos deste caso, encontramos o valor *tema* (6) = *violação de princípio*. Assim, do trecho da decisão, o usuário pode extrair outra lição que pode ser útil na sustentação de seu argumento, qual seja: "*a violação do princípio da correlação entre a denúncia e a sentença anula a decisão*".

Este exemplo demonstra como os operadores do Direito podem pesquisar por jurisprudência usando o sistema PRUDENTIA. O exemplo também salienta a utilidade de manter atributos cujos valores são lições e não somente índices para guiar a recuperação. Agora vamos comparar a mesma pesquisa com um sistema de banco de dados de texto.

The screenshot shows a window titled 'FCasosEsc' with a header 'Casos Seleccionados'. Below the header is a table with four columns: 'Fundamento1', 'Fundamento2', 'Fundamento3', and 'Tema1'. The table contains ten rows of case attributes. The fifth row is highlighted, showing 'agravante' for Fundamento1, 'anulação da decisão' for Fundamento2, 'apreensão' for Fundamento3, and 'cocaína' for Tema1. Below the table, the text of the decision for the selected case is displayed, starting with 'CRIME CONTRA A SAÚDE PÚBLICA. NARCOTRAFICÂNCIA. COCAÍNA. FORMA ASSOCIATIVA NÃO COMPROVADA. NÃO FICANDO CARACTERIZADO O CRIME DE ASSOCIAÇÃO PARA A PRÁTICA DA NARCOTRAFICÂNCIA, QUE SE CORPORIFICA QUANDO OS AGENTES TÊM CONSCIÊNCIA DA FORMAÇÃO DA SOCIEDADE PARA TRAFICAR. HÁ QUE SE ABSOLVER QUEM FOI DENUNCIADO, TÃO-SOMENTE, PELA VIOLAÇÃO DO ART. 14 DA LEI DE TÓXICOS. CONTUDO, QUEM FOI DENUNCIADO POR INFRAÇÃO AO ART. 12 E TEM EM SEU PODER QUASE UM QUILO DE COCAÍNA E RESTA PROVADO QUE SE DEDICA A SUA COMERCIALIZAÇÃO NÃO PODE ESCAPAR À REPRIMENDA POR IMPERATIVO LEGAL. Vistos, relatados e discutidos estes autos de apelação criminal n. 25.439, da comarca de Joinville (1a. Vara), em que são apelantes a Justiça, por seu Promotor, Lizomar Albano e Jorge Francisco de Mira Júnior, sendo apelados

At the bottom of the window, there are navigation buttons: a set of four arrows (left, right, double left, double right), a button labeled 'Caso Escolhido', and a button labeled '← Voltar'.

| Fundamento1 | Fundamento2 | Fundamento3 | Tema1 |
|---------------------|---------------------|-----------------|---------------------|
| agravante | antecedentes | apreensão | cocaína |
| apreensão | autoria | caracterização | balança de precisão |
| apreensão | autoria | caracterização | ameaça |
| apreensão | comercialização | denúncia | cocaína |
| agravante | anulação da decisão | apreensão | cocaína |
| anulação da decisão | apreensão | comercialização | apelar em liberdade |
| agravante | apreensão | autoria | cocaína |
| agravante | antecedentes | apreensão | cocaína |

Figura 14 Atributos dos casos recuperados.

6.5.1.1 Comparação

O primeiro aspecto a comparar refere-se à seleção de palavras-chave na construção da consulta. A filosofia baseada em conhecimento do PRUDENTIA gera um glossário que permite que palavras sejam identificadas automaticamente pelo sistema. Isto faz com que o sistema identifique palavras deste glossário ao usuário digitar um breve resumo da nova situação. O procedimento mais próximo em um sistema de banco de dados é a busca manual nas listas de palavras que ocorrem em cada campo dos registros.

No banco de dados, o usuário deve construir uma consulta. Para utilizar o sistema de banco de dados, necessita-se construir uma consulta. Utilizemos para a consulta os mesmos valores atribuídos aos índices no PRUDENTIA. A consulta fica como mostra a Figura 15.

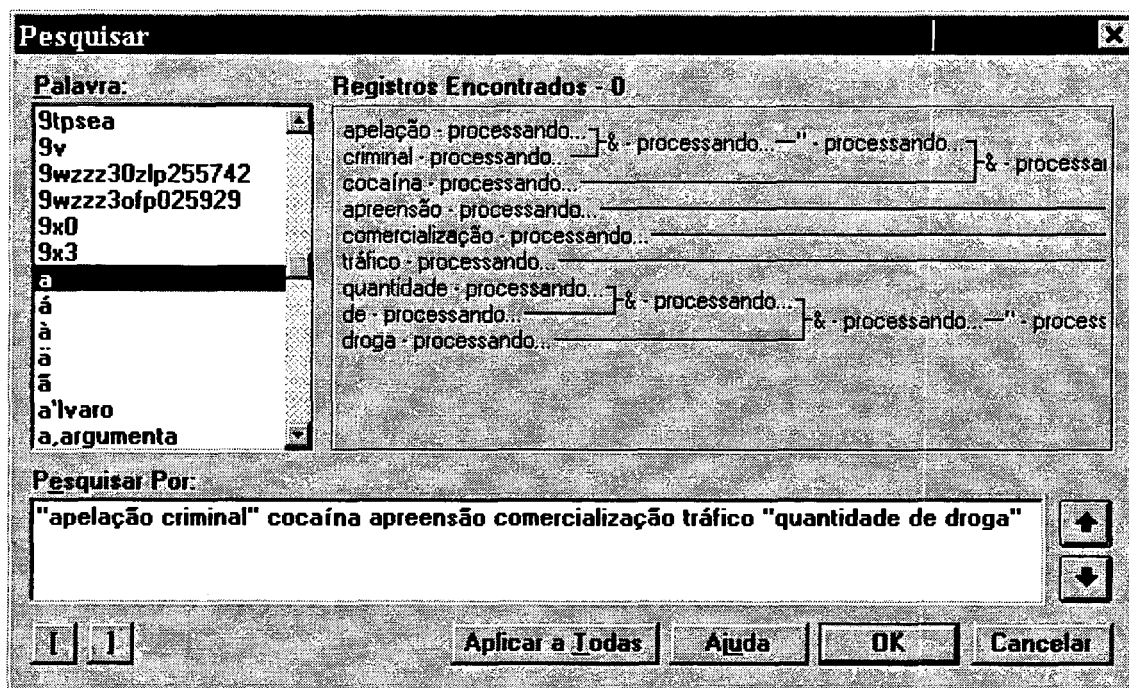


Figura 15 Consulta para o exemplo "porte"³⁴.

Executando esta consulta no sistema de banco de dados utilizado pelo TJSC (FOLIO INFOBASE®), resultam 23 documentos conforme indicado

³⁴ Interface de busca no FOLIO INFOBASE®.

na barra inferior da Figura 16. O usuário agora tem 23 acórdãos para ler no intuito de decidir sobre sua possível utilidade com relação à situação inicial. Em conformidade com a avaliação de Blair e Maron (1985) este resultado de 23 documentos indica que devem haver cerca de 100 decisões que podem ser úteis nesta base. Podemos concordar com esta estimativa em virtude de:

- possíveis erros de redação que não foram considerados;
- documentos de outros tipos de recursos e outras tipificações não são recuperados, já que não há combinação parcial;
- palavras-chave são usadas todas com o mesmo grau de importância;
- possíveis erros na construção da consulta.

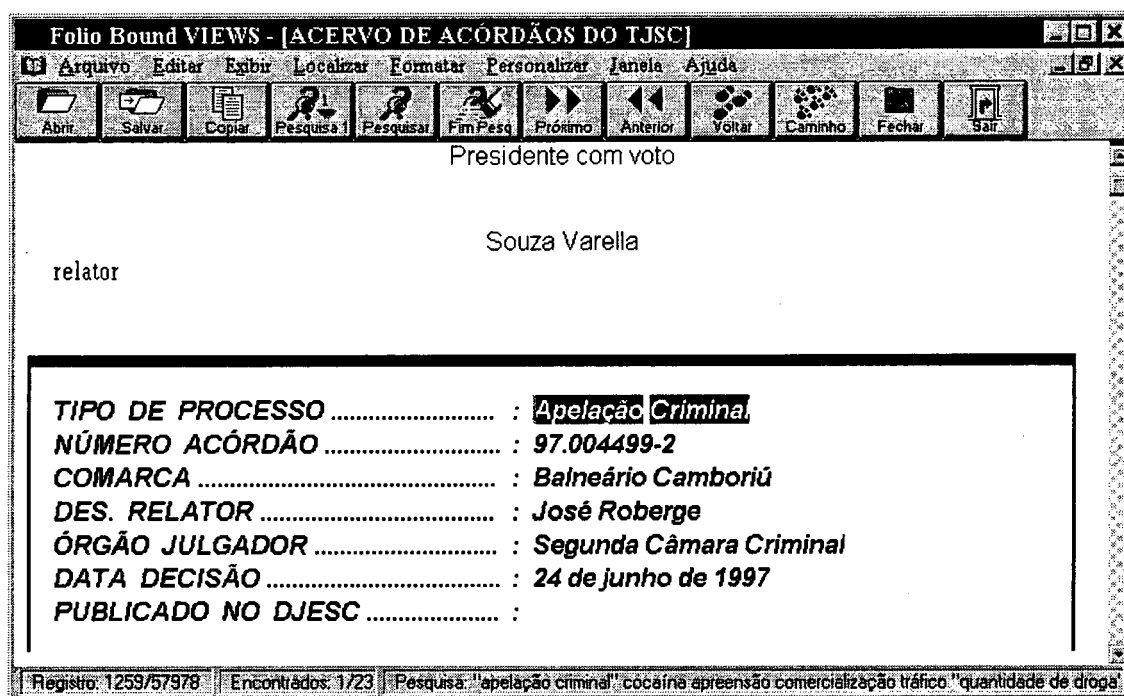


Figura 16 Resultados da consulta no banco de dados de texto.

Uma solução para os possíveis erros de digitação e redação é o uso de caracteres curinga que, quando disponíveis, retardam o tempo de busca significativamente. O sistema de banco de dados demonstrado, por exemplo, não permite o emprego de mais de um caracter curinga na mesma consulta. Além disso, mesmo que o sistema permitisse o uso indiscriminado de caracteres curinga, o usuário deveria examinar a lista

de ocorrências de cada campo manualmente, na busca das variações nas ocorrências de cada palavra; uma tarefa exaustiva que poucos usuários se engajariam.

A consulta básica dos bancos de dados implica no uso de um conector E que limita a recuperação somente de documentos que necessariamente contenham todas as palavras presentes na consulta. Este formato impede a recuperação de qualquer documento que possua qualquer valor alternativo para os índices. Conforme a visão dos especialistas, um documento que contenha o mesmo conjunto de expressões com até duas ou três alterações ainda pode ser útil. Uma solução para este problema seria o uso de conectores tais como OU e XOU normalmente disponíveis. O uso do XOU também é exclusivo e o uso do conector OU acarreta um dilema de custo-benefício, na medida em que mais documentos podem ser recuperados, baixando o índice de precisão dos resultados.

O grande problema se confirma como sendo a equivalência na participação dos índices na consulta. As circunstâncias não contribuem homogeneamente na construção de um conteúdo e por isso as consultas não podem gerar a recuperação dos documentos ordenados por sua relevância.

Por fim, ainda lembramos que mesmo um único erro de digitação é suficiente para reduzir o índice de recuperação de uma consulta. As chances de cometer erros aumenta quanto maiores forem as consultas. O sistema de banco de dados demonstrado, por exemplo, nem ao menos comporta espaço suficiente na tela para mostrar resultados intermediários de grandes consultas. Concluimos que este sistema não foi projetado considerando este propósito.

6.5.2 Exemplo "abandono material"

A situação jurídica que origina esta nova sessão tem os seguintes valores de índices:

tipo = *apelação criminal*

tipificação = *abandono material*

fundamento (1) = *justa causa*

tema (1) = *prisão civil*

A interface mostrando os casos recuperados está na Figura 17, onde a quinta coluna é dedicada à *tipificação* do delito. Neste exemplo, os primeiros sete casos recuperados são tipificados com abandono material com exceção do quinto. Isto enfatiza a questão sobre as razões que possam ter causado a recuperação deste caso. Mesmo antes de visualizar os valores dos atributos ou o acórdão, concluímos que este deva ser um caso em que o conteúdo seja tão similar que sua importância cresce em relação à importância do índice *tipificação*. Isto mesmo já seria uma indicação da utilidade deste caso para a nova situação que motivou a pesquisa.

O quinto caso está rotulado com o número 28.271 conforme a primeira coluna, e doravante o referenciamos como acórdão 28.271. A *tipificação* deste caso é estelionato. Ao ler o conteúdo do acórdão, verifica-se que a apelação pela nulidade é concedida em virtude do réu não ter sido intimado pelo fato de que o mesmo já se encontrava sob custódia por uma prisão civil. Esta lição já justifica a recuperação deste caso entre os mais úteis.

| FCasosEsc | | | | | | |
|--------------------|-----|----------|-------------------|-------------------|--------------------------|----------------|
| Casos Selecionados | | | | | | |
| Acordao | Pos | ValorSim | Recurso | Tipificacao1 | Fundamento1 | Fundamento2 |
| 28.412 | 1ª | 100.00 | Apelação Criminal | ABANDONO MATERIAL | extinção da punibilidade | falta de compr |
| 24.680 | 2ª | 84.00 | Apelação Criminal | ABANDONO MATERIAL | antecedentes | anulação da de |
| 28.457 | 3ª | 84.00 | Apelação Criminal | ABANDONO MATERIAL | caracterização | confissão |
| 27.502 | 4ª | 60.00 | Apelação Criminal | ABANDONO MATERIAL | confissão | conjunto proba |
| 28.271 | 5ª | 60.00 | Apelação Criminal | ESTELIONATO | anulação da decisão | defesa |
| 28.641 | 6ª | 60.00 | Apelação Criminal | ABANDONO MATERIAL | | |
| 32.387 | 7ª | 60.00 | Apelação Criminal | ABANDONO MATERIAL | culpabilidade | |
| 28.425 | 8ª | 48.80 | Apelação Criminal | ESTELIONATO | anulação da decisão | atenuante |

APELAÇÃO CRIMINAL. ARGÜIÇÃO DE NULIDADE PROCESSUAL. INEXISTÊNCIA. RETORNO DOS AUTOS À PROCURADORIA GERAL DE JUSTIÇA PARA ANÁLISE DO MÉRITO DO RECURSO.
 Vistos, relatados e discutidos estes autos de apelação criminal n. 28.271, da comarca de Criciúma (2a. Vara), em que é apelante Dailton José Luiz, sendo apelada a Justiça, por seu Promotor;
 ACORDAM, em Segunda Câmara Criminal, por decisão unânime, rejeitar a preliminar argüida e remeter os autos à douda Procuradorla Geral de Justiça, para que se manifeste sobre o mérito.
 Custas na forma da lei.
 Na comarca de Criciúma, DAILTON JOSÉ LUIZ foi denunciado e processado pela prática do crime do art. 171, § 2o., VI, do Código Penal, tendo em conta que no dia 29 de julho de 1987, emitiu sem a devida provisão de fundos, o cheque n. 678373-2, do BESC daquela cidade, no valor de Cr\$ 40.000,00, causando dano econômico à empresa

Figura 17 Casos recuperados do exemplo "abandono material".

6.5.2.1 Comparação

Supondo agora utilizar os elementos da situação do segundo exemplo para construir uma consulta no sistema de banco de dados de texto do TJSC. Novamente, fazemos uso dos valores dos índices usados para representar a nova situação no PRUDENTIA como palavras-chave da consulta, a saber: apelação criminal, abandono material, justa causa e prisão civil.

Há duas formas de usar palavras-chave em uma consulta de banco de dados, como índices primários que particionam a base e como simples palavras-chave, procuradas por suas ocorrências em toda a base.

Os valores indicando a *tipificação*, o *fundamento* e *tema* podem aparecer em qualquer documento independente do tipo de recurso. Assim, para aumentar a precisão e diminuir o número de documentos pouco relevantes recuperados, deve-se utilizar o *tipo* de recurso como índice primário; ou seja, fazendo a busca somente nos documentos cujo campo tipo tiver o valor especificado. Por outro lado, usando o tipo de recurso como palavra-

chave, causa a recuperação de todos os documentos que contenham a expressão deste tipo de recurso, mesmo os cujo campo tipo tenham outro valor.

A escolha de utilizar campos específicos como índices primários sempre implica em uma decisão sobre a relação custo-benefício que decresce o índice de precisão em favor de um melhor índice de recuperação. Assim, para obter um resultado próximo ao obtido com o PRUDENTIA, a busca não poderia excluir outros valores de *tipificação*, por exemplo; porque isto impediria a recuperação de um caso útil como o acórdão 28.271 descrito no exemplo acima.

6.6 Verificação

De acordo com Watson (1997), a **validação** de um dado sistema refere-se ao exame de sua eficiência com relação a tarefas propostas, enquanto **verificação** trata do grau de correção com o qual este sistema realiza as tarefas propostas. A validação do PRUDENTIA está sustentada na crescente oferta de sistemas para pesquisa jurisprudencial no mercado brasileiro. Apresentamos agora os tópicos pertinentes à verificação.

The screenshot shows a software window titled "FCasosEsc" with a title bar containing standard window controls. The main content area is titled "Casos Seleccionados" in a stylized font. Below the title is a table with the following columns: "Acordao", "Posicao", "ValorSimilar", "DJ", "Data", "Pag", "Localizacao", and "Re". The table contains eight rows of data. The first row is highlighted, indicating it is the selected case. Below the table, there is a text area containing the details of the selected case, including the type of appeal, the facts, the legal reasoning, and the final decision. At the bottom of the window, there are navigation buttons: a set of four small arrows (back, forward, search, etc.), a button labeled "Caso Escolhido", and a button labeled "Voltar".

| Acordao | Posicao | ValorSimilar | DJ | Data | Pag | Localizacao | Re |
|---------|---------|--------------|-------|----------|-----|----------------------|-----|
| 26.206 | 1ª | 140.00 | 8.156 | 21/12/98 | 03 | Concórdia | Th |
| 32.728 | 2ª | 74.40 | 9.323 | 21/09/95 | 21 | Maravilha | Nã |
| 28.082 | 3ª | 72.80 | 8.500 | 19/05/92 | 11 | Criciúma | WI |
| 27.083 | 4ª | 69.60 | 8.256 | 22/05/91 | 11 | Concórdia | Álv |
| 25.333 | 5ª | 69.60 | 8.156 | 21/12/90 | 03 | Palmitos | Alb |
| 26.754 | 6ª | 69.60 | 8.383 | 22/11/91 | 18 | Criciúma | Th |
| 26.478 | 7ª | 64.80 | 8.248 | 10/05/91 | 07 | Braço do Norte | Re |
| 26.552 | 8ª | 64.80 | 8.287 | 12/03/91 | 14 | São Francisco do Sul | Ro |

Apelação criminal.
Abaloamento de veículo.
Em marcha à ré o acidente é previsível e evitável.
Culpa caracterizada - Embora irrelevante, para caracterizar a responsabilidade do agente que a lesão resultante seja leve, grave ou gravíssima, na aplicação da pena o juiz deve levar em consideração a gravidade do mal causado.
Sentença mantida.
Vistos, relatados e discutidos estes autos de apelação criminal n. 26.206, da comarca de Concórdia (1a. Vara), em que é apelante Elói Fidelis, sendo apelada a Justiça, por seu Promotor:
ACORDAM, em Segunda Câmara Criminal, por votação unanime, negar provimento ao recurso.

Navigation buttons: [Back] [Forward] [Search] [Case Selected] [Return]

Figura 18 Caso alvo como parte da base de casos.

Precisão da recuperação é verificada com o resultado dado pelo sistema quando o caso alvo é parte da base de casos. No sistema PRUDENTIA, em função dos índices de propósito múltiplo, o valor máximo de similaridade não é 100, já que os valores de similaridade não são apresentados em termos de percentuais. Desta forma, ao alimentar o sistema com um caso

alvo que seja equivalente a um contido na base, o valor de similaridade chega a 140, como ilustrado na Figura 18.

A precisão da recuperação não é tão relevante em um sistema interpretativo como em um sistema de solução de problema. Os casos da base são interpretações de decisões judiciais e duas experiências jurídicas diferentes podem ter interpretações bastante similares. Entretanto, os métodos de conversão automática de textos em casos ainda não resultaram em um caso igual a outro.

Consistência da recuperação não é um tópico pertinente no PRUDENTIA porque o sistema não é utilizado para avaliar similaridades parciais na base de casos. Esta é uma consideração futura, para quando a base aumentar, mesmo que a quantidade limitada de índices compense.

Duplicação de Casos ocorre quando dois casos distintos recebem o mesmo valor de similaridade com relação ao caso alvo. A base do sistema não possui dois casos com os mesmos índices, apesar de ser até freqüente que dois casos distintos recebam o mesmo valor de similaridade (vide Figura 10 e Figura 17). Contudo, no contexto da pesquisa jurisprudencial, o mesmo valor de similaridade não tem um efeito que reduza a precisão ou eficiência da realização desta tarefa.

Cobertura dos casos é difícil manter em um sistema da natureza do PRUDENTIA já que todos os casos representam experiências reais e conseqüentemente não podem ser manipulados. Todos os esforços neste sentido foram feitos durante a seleção das características e do nível de abstração dos valores. Assim, qualquer esforço no sentido de equalizar a distribuição dos valores das características provocaria perda de eficiência no sistema com relação à utilidade dos casos.

Verificação de Eficiência em sistemas de RBC destinados à recuperação conceitual é um problema em aberto. A utilização plena dos parâmetros dos bancos de dados de textos exigiria o conhecimento prévio do número de casos verdadeiramente úteis contidos na base. O teste de

amostras com avaliações de especialistas parece o método mais aceitável. Contudo, não há limites para a interpretação de um operador do Direito com relação ao que pode ser uma informação útil.

Tempo de recuperação é um parâmetro cujos resultados causaram surpresa porque ao comparar o sistema PRUDENTIA com um sistema bancos de dados, o segundo foi mais lento. Um tempo médio de recuperação na versão atual são 45 segundos na opção que mostra os 100 casos mais similares. Durante a inferência, toda a coleção de casos é comparada com o caso alvo, mas o que atrasa a recuperação é a preparação dos casos para serem levados à tela. O tempo médio pode tender a até 2 ou 3 minutos para mostrar os 1.000 casos mais similares.

Ordenamento de recuperação é um aspecto que intensifica a qualidade da recuperação, o que depende da correção na modelagem da importância a partir do conhecimento especialista. Este é um parâmetro que pode ser testado com informação alimentada por especialistas.

Consistência da base de casos é indicada por iterações que resultam em casos recuperados com valores graduais de similaridade. Uma iteração em que não haja um caso recuperado com um alto valor de similaridade ou em que muitos casos tenham valores muito próximos sugerem inconsistência na base³⁵.

Índices de precisão e de recuperação são os parâmetros utilizados na Recuperação de Informação e são difíceis de serem estendidos à recuperação em RBC. O índice de recuperação depende do volume de casos úteis na base. O índice de precisão determina os casos úteis entre o conjunto de casos recuperados.

³⁵ Kolodner (1993) sugere que isto pode indicar que o caso alvo ainda não está construído em seu ajuste de situação.

A recuperação baseada em casos conta com a eficiência da recuperação baseada em similaridade. Se o sistema for construído corretamente, podemos esperar bons resultados.

6.7 Conclusão

Este capítulo apresentou o sistema PRUDENTIA, um sistema interpretativo de raciocínio baseado em casos para auxiliar operadores do Direito na tarefa de pesquisar jurisprudências, através da reprodução computacional do raciocínio analógico. A viabilidade deste sistema foi conquistada com a metodologia de conversão automática de documentos jurídicos em casos.

A representação dos casos no sistema PRUDENTIA diferencia-se dos demais no que concerne à definição dos índices de propósito múltiplo; índices que podem receber $1, n$ valores na avaliação da similaridade. A contribuição destes índices no valor total da similaridade permanece a mesma, apesar do valor de similaridade ter seu limite superior relaxado. Isto resulta na possibilidade de haver um valor de similaridade acima de uma unidade para ser ponderado. Por exemplo, o índice *fundamento* pode assumir uma maior importância do que o índice *tipificação* se houverem suficientes valores similares para o índice *fundamento*³⁶. Esta é a modelagem da interpretação de um especialista do domínio, com relação a decisões judiciais cujas descrições comportam circunstâncias ou base argumentativa que incrementem sua importância em detrimento da *tipificação*. Este procedimento pode levar a um valor máximo de 140 unidades para similaridade.

O ajuste da situação faz uso de glossários gerados a partir de expressões que realmente constam nos casos da base. A lista de expressões auxilia no processo de identificar as variadas formas de expressar idéias e na construção de um dicionário de sinônimos que unifica tais formas.

³⁶ Exemplo "abandono material" na seção 6.5.2 ilustra esta afirmação.

Exemplos de sessões com o sistema PRUDENTIA demonstram o poder da recuperação baseada em similaridade, em comparação com aplicações de bancos de dados de textos. No domínio do Direito, é difícil antecipar quando um acórdão poderá contribuir com alguma informação útil, mas a interpretação do conhecimento jurídico facilita esta tarefa.

Em um sistema de RBC interpretativo, não é necessário considerar o sucesso de uma dada experiência. Nas descrições das decisões judiciais, um atributo é o resultado, que indica se o referido recurso foi ou não concedido. Este atributo é importante de ser conhecido pelo usuário quando da avaliação da utilidade de um caso com o intuito de impedir a repetição de falhas ensinadas por casos fracassados. Isto explica o porquê do resultado ser um atributo e não um índice.

As comparações com os bancos de dados de textos, apresentadas após os exemplos, visam ilustrar a relevância da presente pesquisa como uma solução alternativa no suporte à pesquisa jurisprudencial. Resumimos os tópicos mais importantes a seguir.

- erros de redação ou digitação nos textos originais impedem a recuperação de todos os documentos de um conjunto;
- a falta de combinação parcial impede que um documento seja recuperado se um único valor diferir;
- o fato de todos os índices terem o mesmo grau de importância impede uma avaliação global de similaridade, resultando em um conjunto de documentos sem ordenamento de relevância;
- dificuldades na construção de consultas podem prejudicar buscas em virtude de falhas além de manter leigos afastados de métodos computacionais;

- buscas por campos que particionam a base impedem a combinação de um documento a partir de um valor diferente do que o valor no índice primário;
- caracteres curinga, quando disponíveis, retardam o tempo da busca e aumentam a complexidade da consulta;
- a busca com palavras-chave requer o uso de conectores que são complexos e lentos.

Estes aspectos contribuem para a baixa qualidade oferecida pelos bancos de dados de textos que podem ser substituídos com a introdução da recuperação baseada em casos da Pesquisa Jurisprudencial Inteligente.

Capítulo Sete

7. Conclusões

O macro objetivo de desenvolver um sistema para pesquisa jurisprudencial capaz de melhorar a qualidade dos resultados gerados pelos sistemas de banco de dados de texto foi alcançado e demonstrado pelos exemplos executados com a versão atual do PRUDENTIA.

O fato de não haverem sistemas comerciais baseados de IA destinados ao domínio do Direito sugere que este domínio não seja muito adequado para ser modelado dentro dos formalismos tradicionais utilizados nas pesquisas de IA. Uma importante conclusão do presente trabalho diz respeito à manutenção dos fatos da vida juntos com a teoria do domínio como uma forma de simplificar a representação do conhecimento jurídico. A abordagem de representação de conhecimento adotada na presente pesquisa é a modelagem da interpretação da maneira com que os especialistas jurídicos enxergam o mundo. Ao invés de modelar o Direito como um objeto real, optamos por encarar o Direito como um alvo abstrato que é perseguido pelos profissionais que nele atuam.

A teoria de RBC contribuiu com duas idéias fundamentais nesta pesquisa. Uma delas foi a abordagem proposta por Kolodner (1993) de procurar

pelas lições que as experiências passadas podem ensinar no contexto em que estas acontecem como caminhos de se identificar a utilidade de um caso passado. Estas diretrizes contribuíram com a escolha dos atributos que acabaram por intensificar a utilidade dos casos recuperados.

A outra direção importante proposta na literatura refere-se à representação de documentos proposta por Branting & Lester (1996) que nos chamou a atenção para o exame da estrutura dos textos no sentido de avaliar mais profundamente os objetivos e as expressões nos textos. Tais idéias consideradas em conjunto com o entendimento sobre a importância da utilização do conhecimento especialista foram decisivos no embasamento e direcionamento deste trabalho.

A diretriz que orienta que um caso deve representar uma experiência cresce em significado na medida que trabalhamos exaustivamente com um sistema de RBC. Quando as experiências estão devidamente representadas estas embutem significados subjetivos que garantem o sucesso do conceito de RBC.

Na pesquisa sobre as alternativas para o tratamento de textos nos sentimos desencorajados a utilizar técnicas tradicionais de processamento de LN (veja 4.2.6), tais como análise de frases e análise semântica pelo fato destas técnicas terem sido desenvolvidas para tarefas como tradução automática, tornando-as pouco úteis no suporte à recuperação de documentos. Influenciados por tal idéia, conduzimos a pesquisa no sentido de evitar estas técnicas o que nos gerou um resultado positivo. O uso da técnica de mineração por modelos (*template mining*) para extrair expressões para indexar e representar os casos foi suficiente para nossos propósitos. Até mesmo o método alternativo que propusemos para extrair os complexos valores para o índice *tema* com o uso da técnica de elaboração (5.1.5) foi desprezado na medida em que a técnica de *template mining* foi suficientemente satisfatória para nossos objetivos. Combinamos a

mineração por modelos com o uso de listas de palavras (léxicon), radicais de palavras e caracteres curinga.

O entendimento de que a representação dos casos é uma tarefa especializada foi fundamental para o desenvolvimento da abordagem baseada em conhecimento para construir casos a partir de textos jurídicos. A abordagem é demonstrada através do sistema de RBC implementado no complexo domínio do Direito. Como resultado desta abordagem, nós vencemos a dificuldade da modelagem de casos, o que vem impedindo o deslançar de sistemas de RBC em problemas práticos, inclusive facilitando o desenvolvimento de sistemas cujo conhecimento está disponível em grandes *corpus* de textos.

O esforço de engenharia de conhecimento empreendido no desenvolvimento da metodologia que converte textos em casos é parcialmente reutilizável. O aproveitamento do conhecimento específico adquirido não é possível em diferentes domínios; contudo, tanto a arquitetura quanto as diretrizes de aquisição são totalmente reutilizáveis. É possível beneficiar-se, mesmo em outros domínios, com o conhecimento adquirido sobre, por exemplo, como perceber a presença de uma lição no texto. O trabalho orientado por conhecimento não dispensa a etapa de extração; entretanto, na medida que algumas regras são extraídas, estas funcionam em todos os textos - fato que foi comprovado desde os primeiros experimentos.

Os índices de propósito múltiplo são responsáveis por resolverem um importante problema na modelagem dos casos. Eles permitem a modelagem de uma importância variável de um índice com relação ao conteúdo do caso quando da avaliação de similaridade. O relaxamento dos valores limites da similaridade permite o aumento da importância relativa de dado índice, quando necessário. Os índices de propósito múltiplo também representam uma alternativa quando não for possível reconhecer um valor para um índice em todos os casos da base. Robustos sistemas de

RBC podem fazer uso deste mecanismo para aumentar o alcance e a capacidade de representação de um pequeno conjunto de atributos.

Os acórdãos usados no desenvolvimento da metodologia são descrições de decisões proferidas pelo Tribunal de Justiça de Santa Catarina. As decisões são prolatadas pelos relatores que são desembargadores que possuem uma formação característica específica, produzindo textos altamente estereotipados. A estrutura consistente e a forma homogênea destes acórdãos foram elementos decisivos no sucesso da metodologia.

A forma estruturada dos textos sugere uma extensão simplificada para outros domínios. Basta que os textos alvo deste outro domínio possuam uma estrutura fixa e forma homogênea. Além disso, para empregar a metodologia em questão é necessário que se tenha conhecimento prévio sobre o domínio. Estas características permitem o uso de técnicas por modelos da Extração de Informação e a reutilização do conhecimento extraído nas etapas de aquisição. Esta conclusão choca-se diretamente com a afirmação colocada por Daniels & Rissland (1995) alegando que textos não são tratáveis por métodos baseados em conhecimento.

Com relação à implementação da tecnologia de RBC, uma questão trouxe dificuldade ao processo. Na definição do vocabulário de índices, não há dúvidas na identificação das dimensões mais indicadas, como, por exemplo, a importância de representarmos a *tipificação* de um crime. Entretanto, manter-se dentro das diretrizes que orientam para que os índices sejam suficientemente genéricos para facilitar a recuperação e, ao mesmo tempo, específicos para garantir a utilidade, pode ser uma tarefa bastante complexa. Tentamos sempre guiar nossas escolhas baseando-nos nas possibilidades oferecidas pelo conteúdo das experiências, mantendo-nos dentro das restrições de implementação. No exemplo específico da *tipificação*, entendemos que alcançaremos melhores resultados, garantindo recuperação e utilidade, se substituírmos este índice por outros dois que contemplem aspectos genéricos e específicos da mesma informação.

Um outro importante tópico na tecnologia de RBC refere-se à questão entre os conceitos útil e similar. Ao desenvolver um sistema de RBC, o objetivo é recuperar os casos mais úteis. Todavia, é difícil antecipar cada particularidade de uma experiência que possa vir a ser útil e em quais situações realmente estes serão úteis. A alternativa usada pelos engenheiros de RBC é focalizar a similaridade entre as experiências como a mais forte indicação de utilidade. No sistema PRUDENTIA, a representação dos casos orientada pelo conhecimento especialista empregada sobre experiências fortemente estereotipadas facilita a antecipação de utilidade. Assim, temos fortes razões para esperar uma recuperação que proporcione altos níveis de utilidade.

Bancos de dados de textos empregam métodos estatísticos de indexação na tarefa de recuperação de documentos. A cada iteração disparada por uma consulta, estes sistemas recuperam os textos integrais forçando os usuários a lerem todo o texto para verificarem sua utilidade. Isto é ainda mais problemático se considerarmos os grandes volumes de textos que podem ser recuperados a partir de uma consulta em função da baixa eficiência dos bancos de dados de textos. Como elemento da metodologia apresentada no capítulo 6, a representação tipo formulário das decisões tem efeito de um resumo da interpretação de um especialista com relação a estas experiências. Os atributos valorados compreendem uma interpretação - um resumo - que visa transmitir ao usuário o mesmo conteúdo que o texto original. Este resumo é mostrado ao usuário quando os casos são recuperados, servindo como resumo do caso. Isto proporciona uma significativa economia de tempo já que o usuário não necessita ler toda a decisão para avaliar sua utilidade na pesquisa.

Nos últimos anos, o mercado brasileiro tem oferecido um número crescente de sistemas de banco de dados dirigidos à pesquisa jurisprudencial, o que demonstra a familiaridade dos operadores do Direito com soluções computacionais. Contudo, tal familiaridade não muda a realidade com relação às dificuldades destes profissionais de operarem sistemas que

exijam, por exemplo, construção de consultas lógicas. Há um aspecto promissor com relação à postura dos operadores do Direito, que é sua preocupação com a produção de material jurídico de formas a facilitar sua reutilização futura. De fato, esta parece ser uma preocupação muito natural se lembrarmos a diretriz básica da escrita, que é a utilização futura do texto que é produzido. É óbvio que os relatores estejam dispostos a contribuir com um estilo de escrita na qual eles concordem que venha a facilitar o aproveitamento de sua produção jurisprudencial.

O desenvolvimento de um sistema inteligente é um processo permanente de tomada de decisão. Ao desenvolver o PRUDENTIA, sempre buscamos tomar decisões baseando-nos na teoria de engenharia de conhecimento; assim, apenas seguimos diretrizes do domínio na medida em que estas não violassem os limites da engenharia de conhecimento.

As aplicações de sistemas inteligentes no domínio do Direito oferecem melhores resultados quando baseadas na teoria de RBC. Entendemos que a presente pesquisa representa mais uma instância no fortalecimento da posição de sistemas de RBC como a ferramenta de IA com o melhor potencial nos sistemas jurídicos. Além disso, esperamos que a metodologia para conversão de textos em casos torne-se um marco na implementação de aplicações práticas para o domínio do Direito. Esta metodologia pode ser generalizada para ser implementada em diferentes domínios fazendo de RBC uma poderosa tecnologia no tratamento de problemas práticos com vastas fontes de conhecimentos.

7.1 Desenvolvimentos Futuros

O próximo marco na implementação do PRUDENTIA é alcançado quando sua base de casos contemplar todas as decisões judiciais do tribunal de justiça escolhido; ou seja, a implementação da metodologia na representação dos 204 tipos de recursos que o tribunal contempla. Neste ponto, somente pequenos detalhes restarão para a disponibilização do PRUDENTIA como ferramenta comercial. Neste sentido, os próximos passos são:

- ajustar o protótipo atual;
- converter os textos da área criminal;
- converter os textos da área civil.

Cada um destes passos envolve ações distintas. Os ajustes necessários compreendem atualizações nos dicionários de sinônimos e nas listas de tipificações.

Um método a ser revisado é o extrator de tipificações, pois o método requer uma calibragem e ainda o índice será alterado para duas dimensões. O identificador de subestruturas também necessita revisão, principalmente com relação às subestruturas que apresentem limites difusos.

Um aspecto ainda pendente na metodologia de converter textos em casos refere-se à aquisição de conhecimento dos valores para os índices *fundamento* e *tema*. Uma lista de expressões alimenta o método de mineração por modelos que é testado e o conjunto de textos que não produzem valores são revisados. Normalmente, selecionamos uma amostra destes textos para repetir a etapa de extração de conhecimento. Assim as novas palavras são adicionadas para novo teste. A dificuldade trata-se da

necessidade de revisar os textos para garantir a inexistência destes valores. Ao invés de repetir este processo de ler as amostras de textos, no intuito de garantir a inexistência dos valores, sugerimos o uso de um limiar. Este valor pode resultar de uma pequena porcentagem de textos que tendem a não informar valores combinados com uma média de valores realmente informados nos textos. Isto evitaria a tendência a definir valores não tão importantes pela motivação de encontrar valores de atribuição.

Um importante requisito para a precisão desta pesquisa refere-se à ferramenta de programação dos métodos. Foi demonstrada a viabilidade dos métodos implementando-os no *WordBasic* - uma linguagem Microsoft embutida no *Microsoft Word* ³⁷. Os métodos nos quais empregamos o *WordBasic* são os que extraem palavras com o uso de caracteres curinga. O mercado oferece algumas alternativas que devem ser confrontadas além da possibilidade de desenvolver um programa *ad hoc*.

Ao final da fase de ajustes, planejamos revisar todo o sistema através de testes que assegurem que não haja mais espaço para melhorias, além da extensão para outros domínios.

A extensão para outros subdomínios refere-se, inicialmente, à implementação da metodologia para as demais recursos dentro da área criminal, tais como embargos, exceção de suspeição criminal e recursos extraordinários. A conversão dos acórdãos da área civil ainda exige esforços de engenharia do conhecimento, já que até mesmo as questões materiais mudam em função da diferente natureza das decisões nesta área (seção 6.4).

³⁷ Copyright Microsoft Corp.

8. Referências

- (1992). American heritage dictionary of the English language, third edition copyright 1992 by Houghton Mifflin Company. Electronic version licensed from e portions copyright 1994 by InfoSoft International Inc. The dictionary of cultural literacy Edited by E.D. Hirsch, Jr., Joseph F. Kett, e James Trefil.
- Aamodt, A. e Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, e System Approaches. *Artificial Intelligence Communications*, 7 (1), 39-59.
- Allen, J. (1994). *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, Inc.
- Ashley, Kevin D. e Rissland, Edwina L. (1988a). Compare e Contrast, A Test of Expertise. *Proceedings of a Workshop on Case-Based Reasoning*, 31-36.
- Ashley, Kevin D. e Rissland, Edwina L. (1988b). Waiting on weighting: A symbolic least commitment approach. *Proceedings of AAAI-88*. Cambridge, MA: AAAI Press/MIT Press.
- Ashley, Kevin D. (1990). *Modeling Legal Argument: reasoning with cases e hypotheticals*. A Bradford book. The MIT Press, Cambridge, Massachusetts.
- Bain, W. (1984). *Toward a model of subjective interpretation*. Yale University, Department of Computer Science Technical Report no. 324.

- Bain, W. (1986). *Case-based reasoning: A computer model of subjective assessment*. Ph. D. diss., Department of Computer Science, Yale University.
- Bain, W. (1989). Judge. In *Inside case-based reasoning*, ed. C.K. Riesbeck e R.C. Schank. Northvale, NJ: Erlbaum.
- Bench-Capon, T.J.M. (1995) Argument in Artificial Intelligence e Law. *JURIX 1995*.
- Blair, D.C. e Maron, M.E. (1985). An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. *Communications of the ACM*, 28, 3, 280-299.
- Blair, David, C. (1996). STAIRS Redux: Thoughts on the STAIRS Evaluation, Ten Years After. *Journal of the american society for information science*, 47(1), 4-22.
- Brachman, R. (1979). On the epistemological status of semantic networks. In N. Findler (ed.) *Associative Networks, Representation e Use of Knowledge by Computer*, Academic Press.
- Branting, L. Karl e Lester, James C. (1996) Justification Structures for Document Reuse. *Advances in Case-Based Reasoning: third European Workshop, EWRBC-96*. Lausanne, Switzerle, November 14-16, Ian Smith; Boi Faltings (ed.), Berlin; Springer, 76-90.
- Branting, L. Karl (1988). The role of explanation in Reasoning from Legal Precedent. *Proceedings of a Workshop on Case-Based Reasoning (DARPA)*, Florida, May 10-13. Janet Kolodner(ed.) Morgan Kaufmann Publishers, 94-103.
- Branting, L.K. (1989). Integrating Generalizations with Exemplar-Based Reasoning. *Proceedings of a Workshop on Case-Based Reasoning*, 224-229.
- Branting, L. Karl (1991a). Exploiting the Complementarity of rules e precedents with reciprocity e fairness. *Proceedings of a Workshop on Case-Based Reasoning (DARPA)*, Washington, D.C., May 8-10, 39-50.
- Branting, L. Karl (1991b). Reasoning with Portions of Precedents. *Proceedings of the Third International Conference on Artificial Intelligence e Law*, Oxford, Engle, June 25-28, 1991.

- Branting, L. Karl (1991c). Building explanations from rules e structured cases. *International Journal of Man-Machine Studies*, 34, 797-837.
- Bing, Jon (1980) Legal Norms, discretionary rules e computer programs. In B. Niblett (ed.) *Computer Science e Law*, Cambridge, Engle: Cambridge University Press, 1980, 119-146.
- Burton, S.J. (1985). *An Introduction to Law e Legal Reasoning*. Little, Brown e Company.
- Callan J.P., Croft W.B. e Harding S.M., (1992). The INQUERY Retrieval System. *Proceedings of the 3rd International Conference on Database e Expert Systems Applications*, 78-83. In Center for Intelligent Information Retrieval at the University of Massachusetts, Amherst. Available online <http://ciir.cs.umass.edu/info/psfiles/irpubs/ir.html> [September 6, 1996].
- Cardie C. e Lehnert W. (1991) A Cognitively Plausible Approach to Understanding Complex Syntax. *Proceedings of the Ninth National Conference on Artificial Intelligence*, 117-124.
- Cardie, Claire (1993). A Case-Based Approach to Knowledge Acquisition for Domain-Specific Sentence Analysis. *Proceedings of the Eleventh National Conference on Artificial Intelligence*, AAAI Press/MIT Press, 798-803.
- Chen, Ming-Syan; Han, Jiawei e Yu, Phillip S. (1996). Data Mining: an Overview from a Database Perspective. *IEEE Transactions on Knowledge e Data Engineering*, 8, 6.
- Coval, S.C. e Smith, J.C. (1986). *Law e its presuppositions: Actions, agents e rules*. London: Routledge e Kegan Paul.
- Computational Linguistics FAQ. Available <http://www.cs.columbia.edu/~acl/nlpfaq.txt> [April 25,1997].
- Cover, T.M. e Hart, P.E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13, 1, 21-27.
- Cyre, Walling (1997). Capture, Integration, e Analysis of Digital System Requirements with Conceptual Graphs. *IEEE Transactions on Knowledge e Data Engineering*, 9, 1, Jan-Feb.
- Daniels, J. J. e Rissland, E. L. (1995). A Case-Based Approach to Intelligent Information Retrieval. *Proceedings of the SIGIR '95 Conference SIGIR '95 Seattle WA USA 1995 ACM*.

- Das-Gupta, P. (1987). Boolean Interpretation of Conjunctions for Document Retrieval. *Journal of the American Society for Information Science*, 38.
- Dean, Thomas, Allen, James e Aloimonos, Yiannis. (1995). *Artificial Intelligence Theory e Practice*. Addison-Wesley Publishing Company, Menlo-Park, CA.
- Dillon, M. e Gray, S. (1982). FASIT: A Fully Automatic Syntactically Based Indexing System. *Journal of the American Society of Information Science*, 34.
- Durkin, John (1994). *Expert Systems: design e development*. Prentice-Hall, Inc., New Jersey.
- Fagan, J. (1987). *Experiments in Automatic Phrase Indexing for Document Retrieval: A Comparison of Sintactic e Non-Syntactic Methods*. Ph.D. Thesis, Technical Report 87-868, Cornell University, Computer Science Department.
- Ferns, William; Dologite, Dorothy; Mockler, Robert e Pfeffer, Robert (1996). An Expert System for Juvenile Delinquent Disposition Advisement. *Applied Artificial Intelligence*, 10, 329-352.
- FLAIR Project. (1997). Available online <http://www.flair.law.ubc.ca/project/index.html>
- Futrelle, R.P. e Zhang, X. (1994). Large-Scale Persistent Object Systems for Corpus Linguistics e Information Retrieval. *Digital Libraries: Proceedings of the First Annual Conference on the Theory e Practice of Digital Libraries*, June 19-21, 1994 - College Station, Texas, USA. [Online] Available <http://www.csdl.tamu.edu/DL94/paper/futrelle.html> [April 16, 1997].
- Gardner, A. (1984) *An Artificial Intelligent Approach to Legal Reasoning*. Ph.D. thesis, Stanford University.
- Gelbart, D. e JC Smith (1991). Beyond Boolean Search: FLEXICON, a legal text-based intelligent system. *Third International Conference on Artificial Intelligence e Law*, ACM, Oxford, UK, 225-234.
- Gelbart, Daphne e Smith, J.C. (1993). FLEXICON: An Evaluation of a Statistical Ranking Model Adapted to Intelligent Legal Text Management. *The Fourth International Conference on Artificial Intelligence e Law, ICAIL-93*, Vrije Universiteit Amsterdam, The Netherles, ACM, 142-151.

- Gey, F. e Chan, W. (1989). Comparing Vector Space Retrieval with the RUBRIC Expert System. *SIGIR FORUM*, 23, 1.
- Giarratano, J. e Riley, G. (1993). *Expert Systems: Principles e Programming*, PWS Publishing Company, Boston, MA.
- Ginsberg, Matt (1993). *Essentials of Artificial Intelligence*. Morgan Kaufmann Publishers, Inc. San Francisco, CA.
- Greinke, A. (1994). *Legal Expert Systems: A Humanistic Critique of Mechanical Legal Inference*. [Online] Available gopher://infolib.murdoch.edu.au:70/00/.ftp/pub/subj/law/jnl/elaw/efereed/greinke.txt. [March 26,1997].
- Hafner, Carole. (1981). Representation of Knowledge in a legal information retrieval system. In R. Oddy, S. Robertson, C. Van Rijsbergen, e P. Williams (eds.) *Information Retrieval Research*, London, Butterworths e Co.
- Hage, J.C. (1995). Teleological Reasoning in Reason Based Logic. *Proceedings of the Fifth International Conference on IA e Law*, University of Maryle, ACM Press, 11-20.
- Hahn, U.; Klenner, M.; Schnattinger, K. (1996a) A quality-based terminological reasoning model for text knowledge acquisition. N. Shadbolt, K. O'Hara e G. Schreiber (eds.), *Advances in Knowledge Acquisition: Proceedings of the 9th European Knowledge Acquisition Workshop (EKAW '96)*. Nottingham, U.K., May 14-17, 1996. Berlin: Springer, 131-146 (Lecture Notes in Artificial Intelligence, 1076).
- Hahn, U.; Klenner, M.; Schnattinger, K. (1996b) Learning from texts: a terminological metareasoning perspective. S. Wermter, E. Riloff e G. Scheler (Eds.), *Connectionist, Statistical e Symbolic Approaches to Learning for Natural Language Processing*. Berlin: Springer, 453-468 (Lecture Notes in Artificial Intelligence, 1040).
- Hahn, U.; Klenner, M.; Schnattinger, K. (1996c). Automated knowledge acquisition meets metareasoning: incremental quality assessment of concept hypotheses during text understanding. *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*, (ed. by B.R. Gaines e M. Musen). Banff, Alberta, Canada, November 9-14, 2, 58-1 - 58-20.

- Hahn, U.; Schnattinger, K. e Romacker, M. (1996d) Automatic knowledge acquisition from medical texts. *Proceedings of the 1996 AMIA Annual Fall Symposium* (formerly SCAMC) --- Beyond the Superhighway: Exploiting the Internet with Medical Informatics (Ed. by J.J. Cimino). Washington, D.C., October 26-30, 1996. Philadelphia, PA: Hanley e Belfus, pp. 383-387 (Symposium Supplement of the Journal of the American Medical Informatics Association).
- Hahn, Udo e Schnattinger, Klemens (1997a). Deep Knowledge Discovery from Natural Language Texts. *Proceedings of the 3rd Conference on Knowledge Discovery e Data Mining, KDD'97*.
- Hahn, Udo e Schnattinger, Klemens (1997b). A Qualitative Growth Model for Real-World Text Knowledge Bases. *Proceedings of the 5th Conference on Computer-Assisted Information Searching on the Internet, RIAO'97*.
- Haines, D. e Croft, W.B., (1993). Relevance Feedback e Inference Networks. *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research e Development in Information Retrieval*, 2-11. In Center for Intelligent Information Retrieval at the University of Massachusetts, Amherst. Available online <http://ciir.cs.umass.edu/info/psfiles/irpubs/ir.html> [September 6, 1996].
- Hunter, Dan (1996). Commercializing neural networks. *The Journal of Information, Law e Technology (JILT)*. Available online <http://elj.warwick.ac.uk/elj/jilt/artifint/2hunter/> [April 1, 1997].
- Jacquemin, C. (1996). What is the tree that we see through the window: a linguistic approach to windowing e term variation. *Information Processing e Management*, 32, 4, 445-458.
- Kedar-Cabelli, Smadar. (1984). *Analogy with purpose in legal reasoning from precedents*. Report LRP-TR-17, Laboratory for Computer Science Research, Rutgers University, July 1984.
- Kim, Jun-Tae e Moldovan, Dan. I. (1995). Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transactions of knowledge e data engineering*, 7, 5.
- Klahr, Philip (1996). Global Case-Base Development e Deployment. *Advances in Case-Based Reasoning: third European Workshop; proceedings EWRBC-96*, Lausanne, Switserle, November 14-16, 1996. Ian Smith; Boi Faltings (ed.)-Berlin; Springer, 1996.

- Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann, Los Altos, CA.
- Kolodner, J. e Leake, D. (1996). A tutorial introduction to RBC. *Case-Based Reasoning: Experiences, Lessons, e Future Directions*. David Leake (ed.) AAAI Press/The MIT Press, Menlo Park, California.
- Kolodner, J. e Simpson, R. L. (1988). *The MEDIATOR: A case study of a case-based reasoner*. Georgia Institute of Technology, School of Information e Computer Science Techn. Report GIT-ICS-88/11.
- Kruppenbacher, T.A. (1984). *The application of artificial intelligence to contract management*, 1. Master's thesis, Department of Civil, Environmental e Architectural Engineering, University of Colorado.
- Lawson, M.; Kemp, N.; Lynch, M. e Chowdhury, G. (1996). Automatic Extraction of citations from the text of english-language patents –an example of template mining. *Journal of Information Science*, 22, 6, 423-436.
- Leake, David. (1996). *Case-Based Reasoning: Experiences, Lessons, e Future Directions*. AAAI Press/The MIT Press, Menlo Park, California, 1996.
- Lehnert, W. (1993) *Natural Language Processing Overview* (from the 1993 Research Brochure for the Department of Computer Science at the University of Massachusetts, Amherst). in Natural Language Processing Laboratory, University of Massachusetts. Available Online <http://www-nlp.cs.umass.edu/~nlpgroup/nlproj.html> [August 20,1996].
- Lehnert, W. (1996) *Information Extraction*. Short introduction to information extraction systems. In Natural Language Processing Laboratory, University of Massachusetts. Available Online [http://www-nlp.cs.umass.edu/~nlpgroup/nlpie.html#what is IE?](http://www-nlp.cs.umass.edu/~nlpgroup/nlpie.html#what%20is%20IE%3F), [August 20,1996].
- Lehnert, W. e Sundheim, B., (1991). A Performance Evaluation of Text-Analysis Technologies. *IA Magazine*, 81-94. In Center for Intelligent Information Retrieval Online, Information Extraction Bibliography. Available Online <http://ciir.cs.umass.edu/info/psfiles/tepubs/tepubs.html>, [August 20,1996].
- Liebowitz, J. (1988). It's clearly evident, isn't it? (Legal expert systems). *Expert Systems for Information Management*, 1, 2, 126-138, Summer.

- Luger, G. e Stubblefield, W. (1993). *Artificial Intelligence: Structures e strategies for complex problem solving*. The Benjamin.Cummings Publishing Company, Inc. Redwood City, California.
- Mase, B.; Linoff, G. e Waltz, David L. (1992). Classifying News Stories using Memory Based Reasoning. *Proceedings of the SIGIR*, Copenhagen, 59-65.
- Mauldin, Michael (1991). Retrieval performance in FERRET: A conceptual Information retrieval system. *The 14th International Conference on Research e Development in Information Retrieval*, Chicago, october, ACM SIGIR.
- McCarty, L.T. e Sridharan, N.S. (1982). *A computational theory of legal argument*. Report LRP-TR-13, January, Laboratory for Computer Science Research, New Brunswick, N.J..
- McCulloch, W. S. e Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-137.
- Meldman, J. A. (1975). *A preliminary study in computer-aided legal analysis*. PhD thesis, Departments of Electrical Engineering e Computer Science, MIT, August 1975.
- Michaelsen, R. (1982). *A knowledge-based system for individual income e transfer tax planning*, PhD thesis, University of Illinois, Accounting Dept., Champaign-Urbana, 1982.
- Minsky, Marvin. (1975). A framework for representing knowledge. In: Winston P. (ed.) *The Psychology of computer vision*. McGraw-Hill.
- Newell, A. e Simon, H.A. (1961). GPS, a program that simulates human thought. In Billing, H. (ed.) *Lernenden Automaten*, pages 109-124. R. Oldenbourg, Munich, Germany. Reprinted in (Feigenbaum e Feldman, 1963, pp. 279-293).
- Okamoto, S. e Satoh, K. (1995). An Average-Case Analysis of k-nearest Neighbor Classifier. *Case-Based Reasoning Research e Development*, Manuela Veloso e Agnar Aamodt (eds.), First International Conference, ICRBC-95, Sesimbra, Portugal, Springer, 253-264.
- Popp, W. G. e Schlink, B. Judith (1975). A computer program to advise lawyers in reasoning a case. *Jurimetrics Journal*, 5, 4, 303-314, Summer.

- Prakken, H. e Sartor, G. (1995). On the Relation Between Legal Language e Legal Argument. *Proceedings of the Fifth International Conference on IA e Law*, 1-10. Univ. of Maryle, ACM Press.
- Quillian, M.R. (1968). Semantic Memory. *Semantic Information Processing*, M. L. Minsky (ed.), MIT Press, Cambridge, Mass.
- Ram, Ashwin. (1991). Interest-Based Information Filtering E Extraction In Natural Language Understanding Systems. *Bellcore Workshop on High-Performance Information Filtering*, Morristown, NJ, November, 1991.
- Riesbeck, C.K. e Schank, R. C. (1989). *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, Publishers, New Jersey.
- Riloff, E. (1993). Automatically Constructing a Dictionary for Information Extraction Tasks. *Proceedings of the Eleventh National Conference on Artificial Intelligence*, AAAI Press/MIT Press, 811-816.
- Riloff, E. e Lehnert W. (1992). Classifying Texts using Relevancy Signatures. *Proceedings of the Tenth National Conference of Artificial Intelligence*, 329-334.
- Riloff, E. e Lehnert W. (1993) Automated Dictionary Construction for Information Extraction from Text. *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications*, 93-99.
- Riloff, E. e Lehnert W. (1994). Information Extraction as a Basis for High-Precision Text Classification. *ACM Transactions on Information Systems*, 12, 3, 296-333.
- Rissland, E. L. e Skalak, D. B. (1991). CABARET: Statutory Interpretation in a Hybrid Architecture. *International Journal of Man-Machine Studies*, June, 34, 839-887.
- Rissland, E. L.; Skalak, D. B., e Friedman, M. T. (1993) BankXX: A Program to Generate Argument through Case-Based Search. In *Proceedings of the Fourth International Conference on IA e Law ICAIL-93*, Amsterdam, The Netherles, June, 117-124.
- Rissland, E., Skalak, D. e Friedman, M. (1996) Using Heuristic Search to Retrieve Cases that Support Arguments. In *Case-Based Reasoning: Experiences, Lessons, e Future Directions*. David Leake, ed. AAAI Press/The MIT Press.
- Robertson, S.E. (1977). The Probability Ranking Principle in IR. *Journal of Documentation*, 33, 294-304.

- Russel, Stuart J. e Norvig, P. (1995). *Artificial Intelligence: a modern approach*. Prentice-Hall, Inc., Upper Saddle River, N.J.
- Salton, Gerard (1975). *Dynamic Information e Library Processing*. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
- Salton, Gerard (1984). The Use of Extended Boolean Logic in Information Retrieval. *SIGMOD Conference*, 277-285.
- Schank, R. (1982). *Dynamic Memory: A theory of learning in computers e people*. New York, Cambridge University Press.
- Schank, R., Kass, A. e Riesbeck, C. (1994). *Inside case-based explanation*. Lawrence Erlbaum Assoc., Hillsdale, N.J.
- Schneider, M., Keel, A. , Langholz, G. e Chew, G. (1996). *Fuzzy Expert System Tools*. John Wiley e Sons Ltd., Engle.
- Simpson, R. L. (1985). *A computer model of case-based reasoning in problem solving: An investigation in the domain of dispute mediation*. Georgia Institute of Technology, School of Information e Computer Science Technical Report no. GIT-ICS85/18. Atlanta.
- Smeaton, Alan (1991). Information Retrieval Research: How it Might Affect the Practicing Lawyer. In *Law, Decision-Making e Microcomputers*. S. Nagel ed.
- Smeaton, Alan (1995) Linguistic Approaches to Text Management: An Appraisal of Progress. *Journal of Document e Text Management*, 2, 2, 67-80, 1995. In Alan Smeaton's Online Publications. [Online] Available <http://lorca.compapp.dcu.ie/~asmeaton/publist.html> [September 12,1996].
- Smeaton, A.; Kelledy, F.; O'Donnell, R.; Quigley, I.; Richardson R. e Townsend, E. (1995a). Information Retrieval Applications: Using Linguistic Resources or Using Language Processing. *Proceedings of IA'95 Language Engineering Conference*, pp.65-74, June1995, Montpellier, France. In Alan Smeaton's Online Publications. [Online] Available <http://lorca.compapp.dcu.ie/~asmeaton/publist.html>, [September 12,1996].

- Smeaton, A.; Kellely, F.; O'Donnell, R.; Quigley, I.; Richardson, R. e Townsend, E. (1995b). Low Level Language Processing for Large Scale Information Retrieval: What Techniques Actually Work. Proceedings of the Workshop on Terminology, *Information Retrieval e Linguistics*, CNR, Rome, October 1995. In Alan Smeaton's Online Publications. [Online] Available <http://lorca.compapp.dcu.ie/~asmeaton/pubs-list.html>, [September 12,1996].
- Smeaton, A (1997). Information Retrieval: Still Butting Heads with Natural Language Processing?, *Information Extraction*. M.T Paziienza (Ed.), Springer-Verlag Lecture Notes in Computer Science (in press), 1997, prepared for a Summer School in Rome in 1997. In Alan Smeaton's Online Publications. [Online] Available <http://lorca.compapp.dcu.ie/~asmeaton/pubs-list.html>, [May 21,1997].
- Smith, J.C. e Deedman, Cal. (1987). The Application of Expert Systems Technology to Case-Based Law. *The First International Conference on Artificial Intelligence e Law, ICAIL-87*, Boston, MA, ACM, 84-93.
- Smith, J.C. e Gelbart, Daphne (1990). Towards a Comprehensive Legal Information Retrieval System. *Proceedings International Conference on Database e Expert Systems Applications*, Vienna, July , pp. 121-125.
- Smith, J.C. (1976). *Legal obligation*. London: Athlone Press, 1976.
- Smith, J.C. (1984). *Liability in negligence*. Toronto: Carswell, 1984.
- Smith, J.C. (1997). The Use of Lexicons in Information Retrieval in Legal Databases. *International Conference on Artificial Intelligence e Law, ICAIL-97*, Melbourne, Australia, ACM, 29-38.
- Smith, J.C. (1997b) (a work in progress) An Introduction to Artificial Intelligence e Law: or, Can Machines Be Made to Think Like Lawyers? In *Artificial Intelligence e Law, The Focus of My Research e Writing by JC Smith*. [Online] Available <http://www.flair.law.ubc.ca/jcsmith/logos/noos/machine.html>.
- Smith, J.C., Gelbart, Daphne e Graham, Deborah (1992). Building Expert Systems in Case-Based Law. *Expert Systems with Applications*, 4, 335-342.
- Sowa, J. F. (1984). Conceptual Structure: Information Processing. In *Mind e Machine*. Addison-Wesley, Reading, MA.

- Sugeno M. (1977) Fuzzy Measures e Fuzzy Integrals - A Survey. In *Fuzzy Automata e Decision Processes*, eds. Gupta Madan M., Saridis, George N. e Gaines, Brian R. , North-Holland, New York.
- Sundheim, B.M. (1992). (ed.) *Proceedings of the Fourth Message Understanding Conference* (Defense Advanced Research Projects Agency, Morgan Kaufmann, Los Altos, CA).
- Sundheim, B.M. (1991). (ed.) *Proceedings of the Third Message Understanding Conference* (Defense Advanced Research Projects Agency, Morgan Kaufmann, Los Altos, CA).
- Sycara, E.P. (1987). *Resolving Adversarial conflicts: An approach to integrating case-based e analytic methods*. Georgia Institute of Technology, School of Information e Computer Science Technical Report no. GIT-ICS-87-26. Atlanta.
- Turtle, H.R. e Croft, W.B. (1991). Evaluation of an Inference Network-Based Retrieval Model. *ACM Transactions Information Systems*, 3 , 187-222.
- Turtle H. R. (1991). *Inference Networks for Document Retrieval*. Ph.D. dissertation.. In Center for Intelligent Information Retrieval at the University of Massachusetts, Amherst. Available online <http://ciir.cs.umass.edu/info/psfiles/irpubs/ir.html>. [March 12, 1997].
- Uyttendaele, Caroline, Moens, Marie-Francine e Dumortier, Jos. (1996). SALOMON: Automatic Abstracting of Legal Cases for Effective Access to Court Decisions. *JURIX'96*.
- Valente, A. (1995). *Legal Knowledge Engineering: A Modelling Approach*. IOS Press, (Amsterdam) e Omsha (Tokyo).
- Waltz, David L. e Pollack, Jordan B. (1985). Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation. *Cognitive Science*, 9, 51-74.
- Waterman,D.A. (1986). *A guide to expert systems*. Addison-Wesley Publishing Company.
- Waterman, D. A. e Peterson, M. (1980). Baseado em regras models of legal expertise. *Proceedings of the First Annual National Conference on Artificial Intelligence*.
- Waterman, D. A. e Peterson, M. (1981). *Models of legal decision-making*. Rep. R-2717-ICJ, The Re Co.

- Watson, Ian (1997). *Applying Case-Based Reasoning: techniques for enterprise systems*. Morgan Kaufmann Publishers, Inc. San Francisco, California.
- Weber-Lee, R., Barcia, R., Khator, S. (1995) Case-Based Reasoning for Cash Flow Forecasting using Fuzzy Retrieval. *Case-Based Reasoning Research e Development*, Manuela Veloso e Agnar Aamodt (eds.), First International Conference, ICRBC-95, Sesimbra, Portugal, Springer, 510-519.
- Weber, Rosina (1997) *PRUDENTIA: Enabling a real world application of Case-Based Reasoning to jurisprudence research*. Qualifying Examination approved by the Graduate Program of Production Engineering at the Federal University of Santa Catarina. In Activities. Available online <http://www.eps.ufsc.br/~rosina/html/activities.html>, [October, 1997].
- Weber-Lee, Rosina; Barcia, R.; Costa, M.; Rodrigues Filho, I.; Hoeschl, H. C.; Bueno, T.; Martins, A, e Pacheco, R. (1997a). A Large Case-Based Reasoner for Legal Cases. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on RBC, ICRBC97*. David Leake, Enric Plaza (ed.)-Berlin: Springer.
- Weber, Rosina; Barcia, R.; Pacheco, Roberto; Martins, Alejandro; Hoeschl, Hugo; Bueno, Tania; Costa, Marcio; Rodrigues Filho, Ilson. (1997b). Representing Cases From Texts In Case-Based Reasoning. *III Congresso Internacional de Engenharia Industrial e XVII ENEGEP*, 6 a 9 de Outubro de 1997, Canela, RS, Brasil.
- Weber, Rosina; Martins, Alejandro; Eduardo Mattos; Tania Bueno; Hugo Hoeschl; Roberto Pacheco; Christiane G. von Wangenheim; Ricardo Barcia. (1998). Reusing Cases to the Automatic Index Assignment from Textual Documents. *6th German Workshop On Case-Based Reasoning - Foundations, Systems, e Applications - Berlin*, March 6-8, 1998.
- Weber, Rosina; Martins, Alejandro, Barcia, Ricardo (1998) On Texts e Legal Cases. Accepted for presentation at the *AAAI-98 Workshop on Textual Case-Based Reasoning*.
- Zadeh, L.A. (1975). The Concept of a Linguistic Variable e its Application to Approximate Reasoning-I. *Information Sciences*, 8, 199-249.

8.1 Bibliografia

- Blair, David e Maron, M. A. (1990). Full-text information retrieval: further analysis e clarification. *Information Processing e Management*, 26, 3, 437-447.
- Branting, L. Karl e Bruce W. Porter.(1991) Rules e Precedents as Complementary Warrants. *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, Anaheim, California, July 14--19, 1991.
- Cohen, Morris e Olson, Kent. (1992). *Legal Research in a Nutshell*. St. Paul, Minn. West Publishing Co.
- Dormont, Agnès e Gruselle, Jean Pierre (1993). A Constructivist Approach to Tense e Aspect: from Text to Polytyped Strings. *Int'l Conf On Tools with AI*, Boston, MA, 484-485.
- Dubois, D., Prade, H. e Yager, R.R. (editors) (1996). *Fuzzy Information Engineering: a guided tour of applications*. John Wiley e Sons, Inc., New York.
- Gardner, Claudia (1996). IBM Data Mining Technology. Build Date: 04/11/96 18:47:20 Build Version: 1.3.0. Book Path: D:\DEMO_BOOKS\DATAMINE.BOO IBM BookManager® BookServer Copyright 1989, IBM Corporation. All rights reserved. [Online] . <http://booksrv2.raleigh.ibm.com/cgi-bin/bookmgr/bookmgr.cmd/BOOKS/datamine/1.0> Available[April 30,1997].
- Grabisch, Michel (1997). Fuzzy Measures e Integrals: A Survey of Applications e Recent Issues. In *Fuzzy Information Engineering. A Guided Tour of Applications*. Didier Dubois, Henri Prade, Ronald R. Yager. John Wiley e Sons, Inc.
- Hastings, John D.; L. Karl Branting, e Jeffrey A. Lockwood. (1995) Case Adaptation Using an Incomplete Causal Model. *Case-Based Reasoning Research e Development*, Manuela Veloso e Agnar Aamodt (eds.), First International Conference, ICRBC-95, Sesimbra, Portugal, Springer, 181-192.

- Herman, Donna e Ceela, Gerald (1989). A very fast prototype retrieval system using statistical ranking. *SIGIR Forum*, 23, 4.
- Klir, George J. e Yuan, Bo (1996). *Fuzzy Sets e Fuzzy Logic Theory e Applications*. Prentice-Hall, Upper Saddle River, New Jersey,
- Kolodner, J. (1989). Judging Which is the Best Case for a Case-Based Reasoner. Proceedings of a Workshop on Case-Based Reasoning, 77-81.
- Kupiec, Julian; Pedersen, Jan e Chen, Francine (1995). A Trainable Document Summarizer. *SIGIR'95*, WA, ACM, 68-73.
- Owens, Cristopher (1993). Integrating Feature Extraction e Memory Search. *Machine Learning*, 10, 117-145.
- Piatetsky-Shapiro, Gregory. (Sep 1995). Knowledge Discovery e Data Mining FAQ. [Online] Available [HTTP://info.gte.com/~kdd/](http://info.gte.com/~kdd/). [May 1st, 1997].
- Rissland, E. e Ashley, K. (1987) A Case-Based System for Trade Secrets Law. In Proceedings: *First International Conference on Artificial Intelligence e Law*, Northwestern University, 1987.
- Romacker, Martin; Schnattinger, Klemens; Hahn, Udo; Schulz, Stefan e Klar, Rüdiger (1997). A Natural Language Understanding System for Knowledge-Based Analysis of Medical Texts. In R. Klar e O. Opitz (Eds.), *Proceedings of the 20th Annual Meeting of the GfKL*, Freiburg, Germany, March 6-8, 1996. Berlin etc.: Springer.
- Schnattinger, Klemens e Hahn, Udo (1997) Intelligent Text Analysis for Dynamically Maintaining e Updating Domain Knowledge Bases. In *IDA'97 - Proceedings of the 2nd International Symposium on Intelligent Data Analysis*.
- Swami, Arun (1995). Data Mining with Silicon Graphics Technology. In *Silicon Graphics Technology*. [Online] Available [HTTP: http://www.sgi.com/Technology/data-mining.html](http://www.sgi.com/Technology/data-mining.html) [April 23, 1997].
- Tata, C., Wilson, J.N. e Hutton, N. (1996). Representations of Knowledge e Discretionary Decision-Making by Decision-Support Systems: the Case of Judicial Sentencing. [Online] Available <http://elj.warwick.ac.uk/elj/jilt/artifint/2tata/>. [March 26, 1997].

- Thomas F. Gordon. (1987). Some Problems with Prolog as a Knowledge Representation Language for Legal Expert Systems. In C. Arnold, editor, Yearbook of Law, Computers e Technology, pages 52-67, Butterworths, London, 1987.
- Wang, Z. e Klir, G. (1992). Fuzzy Measure Theory. Plenum Press, New York.
- Weber-Lee, Rosina, Pacheco, R. , Martins, A., Barcia, R. (1996) Using Typicality Theory to Select the Best Match. EWRBC'96. *Advances in Case-Based Reasoning*. Ian Smith; Boi Faltings (eds.) third European Workshop; proceedings/ EWRBC-96, Lausanne, Switzerle, November 14-16, 1996. -Berlin; Springer, 1996.
- Whitaker, Leslie A., Stottler, Richard H., Henke, Erea, e King, James A. (1990). Case-Based Reasoning: Taming the similarity heuristic. Proceedings of the Human Factors Society, 312-315.