

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

**Recuperação da Informação Jurídica em Sistema
Baseado em Casos**

**Dissertação submetida à Universidade Federal de Santa
Catarina para obtenção do grau de Mestre em Engenharia**

Tânia Cristina D'Agostini Bueno

**Orientador: Prof. Ricardo Miranda Barcia, Ph.D
Co-orientador: Christiane Gresse Von Wangenheim, Msc.**

Florianópolis, fevereiro de 1999

Recuperação da Informação Jurídica em Sistema Baseado em Casos

Tânia Cristina D'Agostini Bueno

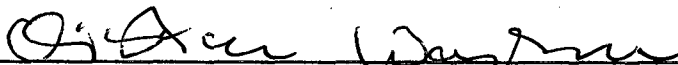
Esta dissertação foi julgada adequada para obtenção do título de "Mestre em Engenharia", especialidade Engenharia de Produção e aprovada em sua forma final pelo Programa de Pós-graduação.

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

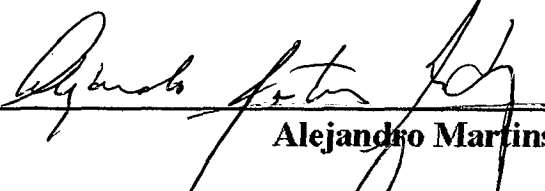
Banca Examinadora:



Ricardo M. Barcia, Ph.D, orientador



Christiane Gresse von Wangenheim, Msc, co-orientadora



Alejandro Martins, Dr.



Roberto Pacheco, Dr.

Agradecimentos:

Ao Professor Ricardo M. Barcia, pela sua experiente orientação e percepção nesta escalada espiritual do conhecimento.

Ao meu companheiro, Hugo C. Hoeschl, pelo seu carinho e pelas eletrizantes discussões jurídicas, que foram decisivas e constantes no desenvolvimento deste trabalho.

À Christiane G. von Wangenheim, pela excelente contribuição e orientação no desenvolvimento final desta abordagem.

Ao colega Eduardo da S. Mattos, pela sua dedicação e interesse na aproximação entre o Direito e a inteligência artificial.

Àqueles que direta ou indiretamente vêm auxiliando neste projeto, Professores Alejandro Martins, Roberto Pacheco e Rogério Cid Bastos.

Ao Grupo de Inteligência Artificial e Direito – GIAD, Professores Luis Adolfo Olsen da Veiga, Aires J. Rover e Marco Antonio Ferreira de Melo, pelo auxílio e interesse demonstrados no desenvolvimento desta área de pesquisa.

Ao incentivo financeiro através de Bolsa de Estudo concedido pela CAPES.

Àqueles que fizeram parte do desenvolvimento desta pesquisa, Cassiano Reis, Marcelo Eduardo de Oliveira, Marcio C. da Costa, Ilson Wilmar Rodrigues Filho, Antonio Carlos Marini Garcia e Rosina Weber, pelo esforço e dedicação empregados.

E em especial, aos meus pais, Peri e Dulce, pelo notável incentivo acadêmico e por me ensinarem que a imaginação é o princípio de tudo.

À minha família, principalmente àquelas pessoas que estiveram ao meu lado, auxiliando no cuidado de minhas filhas, pessoas sem as quais não seria possível concluir este trabalho.

Às minhas adoradas filhas, Oriana e Milena, as quais dedico este trabalho, pelos momentos de alegria e esperança que preenchem a minha vida.

Resumo

(Este trabalho descreve um modelo de recuperação de informação para amplas bases de textos jurídicos, com a finalidade de fornecer ao profissional do direito as jurisprudências mais semelhantes para a solução do seu problema.) No domínio do Direito, a jurisprudência tem um importante papel como fonte do direito; pois suas conclusões sustentam a aplicação da lei num caso concreto. Entretanto, os Tribunais brasileiros produzem uma quantia enorme de decisões todos os anos, tornando esta fonte do direito cada vez maior, o que impele os profissionais jurídicos a despender mais tempo na busca por uma decisão pertinente. Sofisticadas técnicas de IA são necessárias para minimizar o tempo de busca e melhorar a qualidade da informação recuperada. A meta principal de nossa abordagem baseada em casos é a recuperação de textos de jurisprudências com características semelhantes e adequados à solução de um problema jurídico, além da inclusão de novos documentos automaticamente na base de conhecimento. Isso é realizado com base em informações inicialmente fornecida pelo usuário em linguagem natural. Para possibilitar este processo de conhecimento textual informal, um vocabulário controlado e um *thesaurus* jurídico - baseado na terminologia jurídica usual - são integrados na recuperação e no processo de extração do conhecimento. Esta abordagem foi desenvolvida para a área do direito criminal e sua implementação foi realizada em casos de *Apelações Criminais*, do Tribunal de Justiça de Santa Catarina.

Abstract

This paper describes a model of retrieval of text documents in large juridical text bases. In the legal domain, jurisprudence has an important role as a juridical source; its decisions support the application of the Law to a concrete case. The problem is that Brazilian Courts produce an enormous amount of decisions every year, turning these text sources larger every time and forcing juridical professionals to spent more time in the search for a relevant decision. Sophisticated AI techniques are needed to minimize searches time and improves the quality and appropriateness of the retrieved information. The main goal of our case-based approach is the retrieval of adequate jurisprudencial texts with similar characteristics, based on initial information supplied by the user in natural language, besides including new documents automatically into the knowledge base. A major problem and issue of discussions in the juridical domain is the interpretation of the juridical language. Our approach is based on the theory of the juridical argument with parameter to define the important and discriminate information for the retrieval of relevant juridical texts. In order to enable the processing of informal textual knowledge in natural language, a controlled vocabulary and a juridical thesaurus based on the theoretical framework is integrated into the retrieval and extraction process. The approach is applied and validated in the area of the Brazilian criminal Law. Currently we are implementing a prototype including 3400 cases of Criminal Appeals originating from of the Court of Law of Santa Catarina.

Conteúdo

1 Introdução	10
1.1 Motivação.....	10
1.2 Objetivos.....	11
1.3 Importância.....	12
1.5 Estrutura.....	13
2 Domínio da aplicação: Pesquisa Jurisprudencial	14
2.1 A jurisprudência no sistema jurídico brasileiro.....	14
2.2 A importância pesquisa jurisprudencial.....	15
2.3 Sistemas e base de dados jurídicos.....	19
2.4 Conclusões.....	21
3 Fundamentos teóricos	23
3.1 Inteligência Artificial e Direito.....	23
3.2 Teoria da Argumentação jurídica.....	25
3.2.1 Análise dos casos.....	27
3.2.2 Requisitos formais da sentença penal.....	27
3.2.3 Requisitos retóricos da sentença penal.....	29
3.2.4 Uso funcional da linguagem jurídica.....	32
3.3 Raciocínio Baseado em Casos.....	35
3.3.1 Representação e indexação dos Casos.....	37
3.3.2 Recuperação (métrica da similaridade e seleção do melhor casos).....	39
3.3.3 Revisão e Adaptação.....	40
3.3.4 Reutilização, avaliação e aprendizagem.....	40
3.4 Recuperação da Informação.....	41
3.4.1 Soluções Integradas.....	42
4 Suporte Inteligente para Pesquisa Jurisprudencial	43
4.1 Recuperação de Acórdãos Criminais usando a teoria jurídica.....	43
4.1.1 Representação dos casos jurídicos.....	44
4.1.2 Índices para a recuperação.....	45
4.1.3 Conhecimento do domínio Jurídico.....	47
4.1.3.1 Vocabulário controlado.....	49
4.1.3.2 Dicionário de termos normativos.....	51
4.2 Extração Automática da Informação Textual.....	53
4.2.1 Extração dos valores de índices pela Localização.....	54
4.2.2 Extração dos valores de índices pela Inferência.....	56
4.2.3 Extração através de Vocabulário Controlado.....	57
5 Recuperação da informação Jurídica	58
5.1 Introdução.....	58
5.2 Ajuste da Situação.....	59
5.3 Cálculo da Similaridade.....	61
5.3.1 Medida da Similaridade Global.....	62
5.3.2 Medida da Similaridade Local.....	63
5.4 Seleção do Caso mais similar.....	65

6 Trabalhos Relacionados	66
6.1 Introdução.....	66
6.2 Prudentia.....	67
6.3 FAILQ.....	67
7 Conclusões Prudentia	70
7.1 Desenvolvimento Futuros.....	71
8 Referências Bibliográficas	72
ANEXO 1 - Artigo aprovado ICAIL99	76
JurisConsulta: Retrieval in Jurisprudential Text Bases using Juridical Terminology.....	77
ANEXO 2 - Artigo aprovado ENIA99	86
Uso da Teoria Jurídica para Recuperação em Amplas Bases de Textos Jurídicos.....	87

Lista de Tabelas

Tabela 1. Análise dos sistemas de busca de Jurisprudência.....	20
Tabela 2. Requisitos formais de uma Jurisprudência Criminal.....	29
Tabela 3. Descrição dos índices e a sua importância para a recuperação.....	47
Tabela 4. Estrutura de um Acórdão Criminal.....	53
Tabela 5. Exemplo de um Acórdão Criminal (Jurisprudência) representado em forma de caso.....	57

Lista de Figuras

Figura 1. Processo analógico.....	16
Figura 2. Formas de pesquisa jurisprudencial	19
Figura 3.Ciclo CBR.....	36
Figura 4. Representação do documento jurídico.....	45
Figura 5. Definição da Expressões Indicativas.....	49
Figura 6. Exemplo de um Acórdão Criminal estruturado e a localização das informação relevantes.....	55
Figura 7. Extrato da representação da legislação criminal usado para inferenciar valores do índice tipificação.....	56
Figura 8. Arquitetura do Sistema.....	59
Figura 9. Cálculo da Similaridade Global.....	62
Figura 10. Organograma dos Tribunais Brasileiros.....	71

Capítulo 1- Introdução

1.1 Motivação

No sistema legal brasileiro, a jurisprudência é uma fonte formal do Direito Positivo¹, e a sua importância reside no fato de fornecer uma adequação da lei ao caso concreto. Quando um profissional do direito realiza uma pesquisa jurisprudencial ele está buscando informações para reforçar o seu ponto de vista sobre a interpretação de uma norma jurídica, ele define argumentos persuasivos para fazer a analogia entre o seu problema atual e o anterior, já solucionado.

Porém, o número de jurisprudências cresceu em larga escala e complicou o acesso eficiente para este tipo de informação. Apesar da disponibilidade desses textos jurídicos em bancos de dados (como, por exemplo, o sistema Informa²), ou na Internet (por exemplo, Tribunais Federais³ e Tribunais Estaduais⁴, ou Órgãos públicos⁵), a dificuldade permanece. As buscas por informações jurídicas nesses sistemas requerem conhecimento jurídico e estão limitadas devido a problemas como a ambigüidade sintática e semântica, e também a incerteza existentes nos textos dos documentos⁶. Na prática, estes banco de dados e tecnologias baseadas na rede (Internet), recuperam freqüentemente um grande número de informações irrelevantes, além de exigirem uma repetida reformulação da busca para alcançar um resultado satisfatório. Então, o desenvolvimento de sistemas eficientes para a recuperação de informação jurídica usando técnicas da área de Inteligência Artificial (AI) é importante para todos os profissionais do direito, pois fornece benefícios significantes para a solução de problemas jurídicos.

¹ Reale, Miguel. Fontes e Modelos de Direito, p.203.

² Informa Jurídico: jurisprudência e legislação. Pro-link software. CD-1. 1996

³ Senado Federal. [Http://www.senado.gov.br](http://www.senado.gov.br). [20 de outubro de 1998].

⁴ Supremo Tribunal Federal. [Http://www.stf.gov.br](http://www.stf.gov.br). [10 de novembro de 1998].

⁵ Tribunal de justiça de santa catarina. [Http://www.tj.sc.gov.br](http://www.tj.sc.gov.br). [15 de novembro de 1998].

1.2 Objetivos

Para auxiliar o profissional do direito na busca pela decisão jurisprudencial pertinente ao seu problema jurídico, nós desenvolvemos um sistema inteligente para a recuperação destes textos legais. O protótipo do sistema mantém a representação e o armazenamento do conhecimento dos textos jurisprudenciais e recupera documentos com características semelhantes, baseado na informação inicial fornecida pelo usuário, em linguagem natural. Novos documentos são incluídos na base de conhecimento através da extração automática de informação relevante para o processo de recuperação. Para possibilitar o processamento do conhecimento informal em linguagem natural, um vocabulário controlado e um *thesaurus* jurídico - baseado na terminologia jurídica usual e na legislação são integrados na recuperação e no processo de extração. A teoria de argumentação jurídica⁷ ofereceu apoio teórico para interpretar e analisar o conteúdo dos Acórdãos Criminais. Baseado nesta estrutura teórica, nós definimos índices e expressões para a recuperação destes textos de jurisprudências. Nossa abordagem foi desenvolvida e validada na área do direito criminal brasileiro, nos casos de Apelações Criminais, oriundos do Tribunal de Justiça de Santa Catarina⁸.

Para a operacionalização de um sistema de busca jurisprudencial inteligente, nós aplicamos o Raciocínio Baseado em Casos (RBC). O RBC é uma poderosa tecnologia da Inteligência Artificial que usa experiências anteriores e semelhantes para a solução de um problema novo. O Raciocínio Baseado em Casos é baseado no princípio de analogia, assumindo que problemas semelhantes tem soluções semelhantes. Além disso, o RBC é uma técnica da IA muito adequada para a modelagem do conhecimento jurídico. Muitos teóricos da área (e.g., Rissland, Ashley, Branting, Skalak) defendem o uso dessa abordagem para o domínio jurídico.

⁶ Lenz, M. et al. Textual CBR, p.119.

⁷ Warat, Luiz A. Introdução Geral ao direito - I Interpretação da Lei , temas para uma reformulação, p. 39.

⁸ TJSC. Op. Cit. [15 de novembro de 1998].

1.3 Importância

As limitações da memória humana junto com crescente e contínuo armazenamento de jurisprudência, sugerem o desenvolvimento de um sistema de computational inteligente para apoiar os profissionais do direito na busca destas jurisprudências.

A eficiência de representar o raciocínio analógico faz do RBC a tecnologia indicada no desenvolvimento de um sistema inteligente por pesquisa jurisprudencial, diminuindo significativamente o tempo despendido na pesquisa.

A rapidez e a qualidade da pesquisa jurisprudencial realizada através das técnicas apresentadas nesta abordagem, garantem um desempenho mais eficiente do que a pesquisa executada pelos profissionais do direito fazem através de livros ou sistemas de banco de dados de texto.

A recuperação de jurisprudências, feita com qualidade, pode fornecer meios eficazes para que o julgador possa conciliar retoricamente os ditames de sua consciência com a expectativa das partes e dos órgãos judiciais superiores. Os advogados e promotores, por sua vez, poderão contar com mais um instrumento de orientação no sentido de selecionar os argumentos que colocarão nas mãos dos juízes com vista a suas pretensões.

A necessidade de atenção à produção de ferramentas, enfatizando que tal atividade, que gerará novos métodos e técnicas de armazenamento e manipulação de informações, embora não seja diretamente ligada à ciência jurídica, vai provocar fortes reflexos sobre o direito e o seu funcionamento e, acreditamos, sobre o verdadeiro sentido de Justiça.

1.4 Estrutura do trabalho

Este artigo descreve uma abordagem baseada em casos para a recuperação de textos de jurisprudenciais. No Capítulo 2, uma breve introdução no domínio de aplicação (pesquisa jurisprudencial) é apresentada. Os fundamentos teóricos da nossa abordagem são apresentados no Capítulo 3 e descreve o ramo da IA & Direito, a teoria da Argumentação Jurídica e outras concepções da teoria jurídica que embasam este trabalho, descreve ainda a técnica de IA utilizada para o desenvolvimento deste trabalho: o Raciocínio Baseado em Casos. No Capítulo quatro apresentamos a abordagem desenvolvida neste trabalho e descrevemos a representação de casos, a sua aquisição de documentos textuais, e o processo de recuperação. Uma comparação da nossa abordagem com outras pesquisas relacionadas é realizada no Capítulo cinco e as Conclusões são apresentadas no Capítulo 6.

Capítulo 2 – Domínio da aplicação: Pesquisa Jurisprudencial

2.1 A Jurisprudência no Sistema Jurídico Brasileiro

A palavra jurisprudência, segundo Enrique R. Aftalion⁹, pode ser empregada em duas concepções: a primeira é equivalente a *ciência ou conhecimento do direito (iuris prudentia)*; a segunda, que podemos considerar hoje a prevalente - e é a que nos interessa nesta abordagem - se refere, em termo gerais, *ao sentido uniforme das decisões dos órgãos judiciais do Estado*.

Podemos chamar, no âmbito do direito, o conjunto de precedentes (ou casos) de “Jurisprudência”. Na Itália e França, entre outros, a expressão “Jurisprudência” materializa as prescrições de caráter doutrinário, ou seja, as discussões acadêmicas, científicas e técnicas sobre o direito e a ciência jurídica. Aqui, pelas bandas nacionais, a mesma expressão significa o “conjunto reiterado das decisões judiciais”, e está consagrada como fonte subsidiária de normas jurídicas.

Os teorizadores do direito anglo-saxão (common law) advertem, de forma unânime e precisa, que toda sentença judicial cria direito. No direito latino americano e europeu, este assunto não é visto com igual unanimidade e nitidez. É muito controvertido o ponto relativo a questão das decisões judiciais constituírem uma fonte formal do direito (jurisprudência). Sem dúvida nenhuma, em torno desta questão gira um prejuízo ideológico generalizado nos direitos continental e latino americano: o temor de entregar demasiados poderes aos juizes. É indubitável que a índole criadora que tem a função judicial - enquanto individualização de normas gerais - não lhe é conferida por nenhuma teoria, pois é algo inerente à tarefa de interpretação (por ser um ato de compreensão). E com referência a influência que podem exercer os precedentes (jurisprudência) como fonte geral, ela pode ocorrer ou não, sendo evidente que no primeiro caso, contra esse fato a teoria nada pode fazer.

Portanto, a jurisprudência não é, no nosso direito, fonte obrigatória, já que o juiz pode deixá-la de lado com o devido fundamento. Não obstante, os juizes não ignoram que um *precedente* jurisprudencial considerado em conformidade com o pensamento jurídico predominante, possa ser deixado de lado, sem que isso signifique um atitude contra a comunidade.¹⁰

2.2 A Importância da Pesquisa Jurisprudencial

A pesquisa jurisprudencial é o maneira como os profissionais do direito realizam a busca por decisões jurisprudências, para aplicá-las na solução de um problema jurídico semelhante. Quando estes profissionais fazem uma pesquisa jurisprudencial, eles empregam o raciocínio analógico, comparando uma determinada situação presente com decisões passadas e observam se as conclusões e argumentos usados na decisões anteriores podem ser aplicadas na situação atual.

Na pesquisa jurisprudencial, a expressão pesquisa significa o ato de buscar jurisprudências. Ela não representa uma investigação científica que tem como objeto a jurisprudência. Expressão pesquisa jurisprudencial pode ser traduzida para para pesquisa jurídica como está definida em Cohen & Olson (1992)¹¹

Pesquisa de jurisprudência é a procura para situações semelhantes em uma fonte de passado decisões legais. Quando os humanos executam pesquisa de jurisprudência, eles empregam raciocínio analógico comparando uma determinada situação atual com decisões passadas. Eles examinam tantas decisões quanto eles podem e notam as afinidades entre eles e o assunto que motivaram a pesquisa. A Pesquisa de jurisprudencial resulta num jogo de situações jurídicas semelhantes que podem substanciar atividades judiciais. Na figura 1, é possível observar o processo do raciocínio analógico.

⁹ Aftalion Enrique R. Introduccion al Derecho, p. 361.

¹⁰ Aftalion Enrique R. Ob. Cit., p. 373.

¹¹ Weber, Rosina. Intelligent Jurisprudence Research, p. 16.

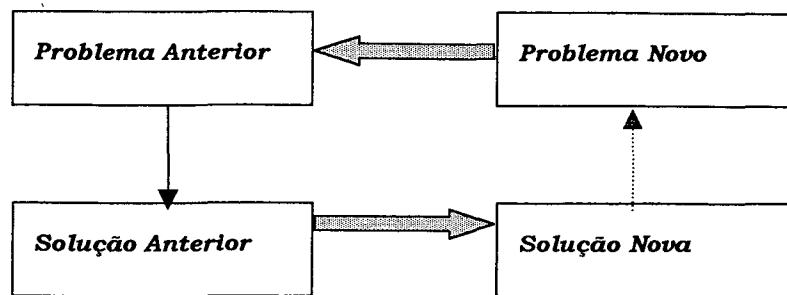


Figura 1. Processo analógico

[A eficiência em representar o raciocínio analógico faz do RBC a tecnologia indicada no desenvolvimento de um sistema inteligente para pesquisa de jurisprudência.]

No domínio jurídico existem diversas formas de aproveitar as experiências em casos anteriores para a tomada de posição diante de um estudo atual.

Porém, a análise de figuras de apoio ao raciocínio jurídico não pode ser realizada, aqui, sem a presença da analogia, um dos mais eficazes e pertinentes instrumentos de integração dos comando do direito¹².

Segundo Bobbio¹³:

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

.....

*“A **analogia** é certamente o mais típico e o mais importante dos procedimentos interpretativos de um determinado sistema, normativo: é o procedimento mediante o qual se explica a assim chamada tendência de cada ordenamento jurídico a expandir-se*

¹² Hoeschl, Hugo C. Tratamento Normativo da Telemática no Brasil, pág. 24.

além dos casos expressamente regulamentados." (Destacado do original).

A noção é de *utilidade indubitável*, e a *delimitação da análise da semelhança*, ponto de contato entre os casos, é necessária¹⁴:

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

Não pode ser confundida com a interpretação extensiva, nem ambas com o silogismo. Esse possui um mecanismo vertical de obtenção de conclusões, enquanto aquelas se valem de um recurso horizontal, mas, distintamente¹⁵:

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

¹³ Bobbio, Norberto. Teoria do ordenamento jurídico, p. 151.

¹⁴ Bobbio, Norberto. Idem, p. 151.

¹⁵ Bobbio, Norberto. Idem, p. 151.

A Jurisprudência estabelece uma correlação entre a Lei e o fato e determina o modo mais apropriado da aplicação da norma jurídica¹⁶. Por esta razão, o acesso a este tipo de informação é muito importante para o profissional do direito. A pesquisa Jurisprudencial fornece condições para a interpretação e aplicação das leis na resolução das questões judiciais. Por exemplo, o cliente de um advogado foi preso em flagrante por furtar uma residência à noite, entretanto, no julgamento, houve um erro na determinação da pena do crime. Neste caso, o advogado quer pedir a redução da pena, e conclui o *Habeas Corpus* é uma petição adequada para a solução do caso. Porém, o advogado desconhece a aplicação do *Habeas Corpus* para requerer a anulação do processo e a conseqüente diminuição da pena. O *Habeas Corpus* é um recurso aplicável para casos definidos na lei, entretanto estas aplicações são descritas de uma forma genérica. Neste contexto, pesquisa de jurisprudencial torna-se essencial para a solução do caso. A existência de uma decisão passada semelhante a este caso pode reforçar o argumento do advogado e embasar a sentença de um juiz.

Porém, a procura para decisões passadas apropriadas pode ser muito complicada. Por exemplo, procurando em livros de jurisprudencial, o usuário pode achar só alguns casos que aplicam ao problema presente, ou usando sistemas de banco de dados, ele poderá recuperar um número grande de casos irrelevantes. Em ambos os casos é necessário repetir a busca e focar em outros aspectos do problema. Outro fator relevante é a incessante e volumosa produção de decisões por parte dos Órgãos Jurídicos, tornando estas fontes do direito cada vez maiores. Este último fator força os profissionais do direito a despender mais tempo procurando uma decisão pertinente. Conseqüentemente, são requeridas técnicas sofisticadas e inteligentes para minimizar tempo de procura e melhorar a qualidade e conveniência da informação recuperada.

No próximo item, serão analisados a eficiência de alguns desses sistemas de Pesquisa Jurisprudencial.

¹⁶ Lênio L. Streak, *Súmulas no Direito Brasileiro: eficácia, poder e função*, pág. 11.

2.3 Sistemas e base de dados jurídicos para recuperação da informação.

Atualmente, os profissionais do direito podem ter acesso a vários bancos de dados jurídicos de jurisprudências, pois a maioria dos Tribunais brasileiros disponibilizaram a sua jurisprudência para o público em geral, com acesso via Internet ou através de sistemas comercializáveis. Entretanto, a recuperação da informação em muitos sistemas ainda requerem lógica Booleana ou a elaboração de uma sintaxe (questão) para a consulta. Isso dificulta o acesso ao usuário final, pois muitos deles tem uma experiência limitada na formulação de uma questão inicial para realizar uma busca ou refazer esta questão, depois que não atingiu o resultado esperado. A maioria das buscas nestas bases de dados é feita pelo usuário final através de profissionais intermediários (e.g., bibliotecários) que conhecem o funcionamento da base de dados, assim como a linguagem controlada e a indexação praticada nele. Quando uma consulta em linguagem natural esta sendo avaliada. Para o usuário final, quando a busca é realizada em base de dados textuais, o uso da linguagem natural é inevitável, porque nem recursos são avaliados por intermediários ou indexadores estão disponíveis; parece natural para o usuário buscar esta informação de uma forma direta.

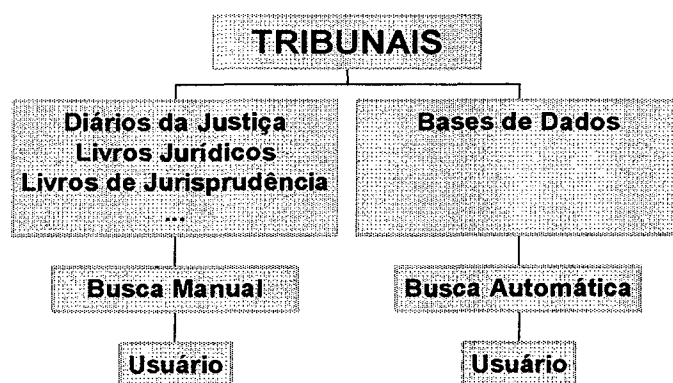


Figura 2. Formas de pesquisa de Jurisprudência

Foram analisados alguns sistemas de busca de jurisprudência (SBJ) enfocando a usabilidade (facilidade de uso) destes sistemas do ponto de vista do advogado. Os sistemas analisados são os seguintes SBJ: Tribunal Regional do Trabalho da 12ª Região –TRT12, Tribunal Regional do Trabalho da 3ª Região – TRT3, Tribunal de Justiça de Santa Catarina -TJSC, Revista dos Tribunais - RT, Informa Online, Supremo Tribunal Federal -STF. Na característica “Usabilidade da Tela de Pesquisa”, se destacaram os sistemas do STF, Revista dos tribunais e JurisConsulta. O STF reduz a tela de pesquisa a um único campo, bem dimensionado e documentado. A Revista dos Tribunais criou uma pesquisa básica e outra avançada, e através desta separação tornou simples sua utilização, o usuário realiza a pesquisa de acordo com sua necessidade. Mesmo utilizando vários campos a para a entrada de dados o JurisConsulta possui uma interface clara, seus campos são organizados de uma forma lógica, cada campo possui um rótulo explicativo e os botões possuem dicas dinâmicas para auxiliar o usuário.

	TRT 12 ^A	JurisConsulta	TJ-SC	Revista dos tribunais	Informa (Prolink)	STF
Usabilidade da Tela de Pesquisa	Regular	Ótima	Regular	Ótima	Ótima	Ótima
Forma de Pesquisa	Pares de Atributos	Linguagem Natural / Pares de Atributo-Valor	Palavras-Chave	Palavras-Chave	Palavras-Chave	Palavras-Chave
Recuperação	Busca Idêntica	Por similaridade	Busca Idêntica	Busca Idêntica	Busca Idêntica	Busca Idêntica
Pesquisa Boleana	Sim	Não	Sim	Sim	Sim	Sim
Pesos nas Palavras	Não	Sim	Não	Não	Não	Não
Resultado Apresentado	Ótimo	Ótimo	Fraco	Ótimo	Ótimo	Regular
Disponibilidade	Internet	Protótipo	Internet	Internet	CD	Internet
Domínio	Direito Trabalhista	Direito Penal	Todas áreas do direito	Todas áreas do direito	Todas áreas do direito	Todas áreas do direito

Tabela 1. Análise dos Sistemas de Busca de Jurisprudência

No item “Resultado Apresentado”, os melhores SBJ foram aqueles que forneceram como resultado da pesquisa: o número do Acórdão, a data de publicação, o local que foi julgado e a ementa do documento, permitindo ao usuário uma rápida triagem dos documentos pesquisados, e fornecendo rápido acesso ao documento completo. Destacaram-se os SJB, TRT 12^a, Revista dos Tribunais e Informa Online.

O divisor de águas entre os SBJ, são as características: “Pesos as Palavras” e “Forma de Recuperação”. Neste item somente o JurisConsulta se destaca, porque utilizam o mecanismo de recuperação por similaridade, enquanto os outros utilizam uma busca por palavras idênticas. Nestes mecanismos os casos recuperados são apresentados ao usuário com a mesma importância, não existe diferenciação entre os documentos apresentados. Isto ocorre porque na busca idêntica o documento deve possuir todas as palavras da pesquisa, ou possuir as palavras necessárias para satisfazer a expressão de busca. Assim, mesmo que o documento contenha a solução que o usuário procura ou seja similar ao problema que ele quer resolver, se ele não tiver usado a mesma palavra que esta no texto, este documento não será recuperado.

2.4 Conclusões

Além do crescimento incessante e em larga escala de documentos por parte dos órgãos jurídicos, existem outros problemas relacionados à pesquisa jurisprudencial.

Mesmo com a automatização da pesquisa jurisprudencial, que facilitou o acesso, alguns problemas persistem, pois os sistemas existentes para a recuperação deste tipo de informação utilizam normalmente a lógica booleana para a elaboração de uma sintaxe para a consulta e é muito comum o profissional do direito não ter este tipo específico de conhecimento. E quanto menor o conhecimento do domínio, mais dificuldade de encontrar a solução esperada. Isso faz com que frequentemente um

grande número de informações irrelevantes sejam recuperadas, além de exigirem um repetida reformulação da busca para alcançar um resultado satisfatório.

No domínio do direito brasileiro não há aplicações com técnicas de IA, o que compromete a utilidade da informação recuperada.

A ignorância sobre aplicação de métodos de IA (Inteligência Artificial) em base de dados muito amplas e heterogêneas, particularmente de documentos textuais faz com que a qualidade das informações recuperadas fique comprometida .

É por estas razões que o desenvolvimento de sistemas eficientes para a recuperação da informação jurídica, com técnicas de inteligência artificial, é muito necessário para os profissionais do direito.

Capítulo 3 - Fundamentos Teóricos

Neste capítulo nós apresentamos um breve resumo sobre o ramo da IA & Direito e da técnica de IA empregada para o desenvolvimento da presente abordagem – o RBC, além disso nós apresentamos aqui as teorias jurídicas utilizada para embasar a modelagem do conhecimento jurídico. O maior desafio dos teóricos da IA & Direito é a da linguagem empregada nos textos jurídicos: Como redefini-la? Como torná-la compreensível para a implementação computacional? Estas questões são o centro da nossa abordagem. Nós nos embasamos no uso funcional da linguagem jurídica e na teoria da Argumentação jurídica para a redefinição informações juridicamente relevantes presentes nos acórdãos judiciais. No Raciocínio Baseado em Casos – RBC, nós encontramos uma ferramenta adequada para a modelagem do raciocínio jurídico em sistemas para a recuperação de documentos jurisprudenciais. O RBC foi a técnica utilizada nesta abordagem por ser um processo baseado no princípio da analogia - similar ao realizado pelo profissional do direito na busca por decisões jurídicas passadas para resolver um problema jurídico atual.

3.1 Inteligência Artificial e Direito

A IA & Direito é um domínio interdisciplinar que origina-se, basicamente, da Inteligência Artificial (IA) e do Direito, mas também da filosofia¹⁷. O que motiva esta interdisciplinariedade é o fato que a construção de sistemas baseados no conhecimento jurídico envolve mais que técnicas de Inteligência Artificial.

A Inteligência Artificial (IA), numa tentativa de compreender a natureza da vida mental humana, embora tratando de questões tradicionais, introduziu elementos e ferramentas novas na abordagem de questões epistemológicas, possibilitando, através do computador, a elaboração de modelos explicativos para aspectos específicos da inteligência humana. A inteligência artificial procurou elaborar programas computacionais cuja aplicação prática consistia na construção de sistemas que pudessem executar um conjunto crescente de atividades para as quais se requer inteligência. O estudo e a modelagem de atividades cognitivas humanas e a explicação dos processos mentais responsáveis pelo comportamento inteligente passou a ser o centro

dos interesses dos pesquisadores da inteligência artificial, empenhados em explicitar alguns aspectos da nossa vida mental através da construção de modelos computacionais¹⁸.

Na resolução de um problema, a Inteligência artificial procura reduzir o tamanho do espaço de busca, isto é o tempo despendido na utilização de um procedimento que garanta a solução, caso ela exista. Essencialmente, os métodos heurísticos visam a explorar o espaço de busca com economia e eficiência. De alguma forma os métodos heurísticos e os resultados que proporcionam são um dos atrativos da IA, pois, não se conhecendo exatamente como o cérebro humano trabalha, e também, dispondo-se somente de máquinas que essencialmente foram projetadas para lidar com números, é necessário, antes de tudo, pensar como o conhecimento é estruturado e como podemos guardá-lo e manipulá-lo no computador¹⁹.

Com a mesma finalidade da IA, a *IA & Direito* é aplicada à ciência jurídica, fornecendo o conhecimento sobre as teorias e as opiniões sobre o que é a lei, quais são os problemas relacionados a ela, como eles são resolvidos e também, definindo a utilidade destes sistemas baseados no conhecimento jurídico. Muitos destes aspectos referem-se aos fundamentos teóricos do Direito, especialmente a disciplina "teoria jurídica", como sendo um campo do direito muito importante para a IA & Direito.

Para André Valente²⁰, a Teoria Jurídica é um meio e um fim para o aperfeiçoamento da qualidade dos sistemas baseados no conhecimento jurídico. Por esta razão, nós procuramos desenvolver a nossa abordagem dentro da ótica da teoria jurídica. Nós escolhemos a teoria da argumentação jurídica proposta por Warat²¹ e a sua visão sobre o uso funcional da linguagem jurídica, para definir quais os requisitos necessário para representar e recuperar informações relevantes de documentos jurídicos, mais especificamente de acórdãos criminais.

Na pesquisa jurisprudencial os problemas mais comuns são a quantidade enorme de documentos jurídicos e a necessidade de uma resposta adequada ao problema a ser resolvido. O estudo destas questões é um passo fundamental para

¹⁷ Valente, A. Legal Knowledge Engineering. A modeling Approach, p 2.

¹⁸ Science does it with feeling. The economist. 20 de Julho de 1996, p.71.

¹⁹ Rabuske, Renato Antônio. Inteligência Artificial. p. 07.

²⁰ Valente, A. op. cit., p.2.

²¹ Warat, Luiz Alberto. O Direito e sua Linguagem. p.35.

desenvolver um modelo para as tarefas jurídicas que possam ser efetivamente implementadas em sistemas baseados no conhecimento jurídico. A IA & Direito pode e deve resolver estas questões, fornecendo um caminho mais rápido para a solução do problemas do Judiciário, mesmo não sendo uma tarefa fácil, pois apenas muito recentemente os teóricos tem conseguido elaborar algo capaz de auxiliar nas tarefas mais complexas como a elaboração de sentenças, mas sem falsas esperanças ou exageros quanto aos limites da inteligência artificial.

No item seguinte nós discorreremos sobre a teoria jurídica utilizada para embasar nosso trabalho.

3.2 Teoria da Argumentação Jurídica

O desenvolvimento de um sistema baseado no conhecimento jurídico, como vimos no item anterior, requer o conhecimento da teoria jurídica. Diferentes teorias indicam modos alternativos de definir os critérios de relevância da lei. Para o desenvolvimento desta abordagem nós partimos de uma análise crítica feita por Luís Alberto Warrat sobre as práticas argumentativas do judiciário, da dogmática jurídica e das escolas de direito. Esta análise consiste basicamente, numa leitura ideológica da argumentação jurídica, apoiada numa abordagem semiológica aberta, antiformalizante, a qual procurará manter-se insensível às propostas neo-positivistas²².

Segundo Warrat, as diferentes teorias dogmáticas são formas irrenunciáveis de redefinição indireta e liberam o profissional do direito para eleger e aplicar as que lhe pareçam ideologicamente mais adequadas. Ele salienta a inutilidade da busca da natureza jurídica ou da essência dos termos definidos persuasivamente (e.g. abuso de poder, honestidade, motivo fútil, etc). São, por natureza, significados dependentes de um tentativa argumental. Nenhum argumento torna-se psicologicamente persuasivo enquanto não se mostra adequado à intuição dominante. A teoria da argumentação é em si mesma ideológica na medida em que se vale da noção de "opinião

geralmente aceita". Nenhum discurso jurídico será considerado significativamente legítimo se contradizer as formas axiológicas predominantes. Os argumentos são persuasivos no direito, quando não contradizem à ideologia de seus destinatários. Em última instância, a ideologia é o *Tribunal* através do qual se mede a eficácia ou ineficácia de um argumento jurídico.

Portanto, a moderna teoria da argumentação deve fundar-se sobre a tentativa de leitura dos argumentos e dos discursos teóricos como ideologia. Afinal, do ponto de vista intuitivo, sem qualquer reflexão crítica sobre o seu próprio comportamento, os juristas já vêm recorrendo a esta prática.

Toda vez que os critérios de relevância de um termo são alterados no seu uso contextual, ocorre uma redefinição. Para redefinir os juristas interferem diretamente sobre as significações dos termos, propondo novas regras designativas, ou realizando um forma indireta de interferência. Chamaremos modos diretos de redefinição às atividades definitórias dos juízes que explicitamente propõem nas sentenças mudanças dos critérios de relevância de certas palavras chaves contidas nas normas gerais. Tal mecanismo, salienta Warat²³, facilita a adequação da jurisprudência, às exigências, reais ou supostas, de um dado momento histórico. Na segunda hipótese, a redefinição indireta, pode este modo redefinitório aparecer como um conjunto de argumentos retóricos, como recurso às teorias, como adjetivação desqualificadora, apreciação axiológica dos fatos, alterações sintáticas das normas ou emprego de variáveis axiológicas.

Além disso, é interessante observar que, quando se diz que duas coisas ou situações têm algumas características comuns, quando se decide a inclusão de um caso a uma classe, incontestavelmente está sendo feita uma operação analógica. Definir é sempre fazer analogia, claro que de uma forma bastante encoberta. E isto é tão encoberto no direito penal, que a proibição da interpretação analógica é um dos pilares do

²² Warat, Luis Alberto. Op. Cit., p. 94.

²³ Warat, Luis Alberto. Op. Cit., p. 38.

princípio da reserva legal, esquema axiológico, eixo do processo penal democrático, apesar de que os juizes penais, quando interpretam, definem²⁴.

A seguir nós veremos uma forma de identificar no discurso jurídico estas formas persuasivas de reforçar a ideologia dominante através da redefinição dos termos normativos.

3.2.1 *Análise Retórica dos casos*

Quando um juiz entende que um conjunto de fatos configura uma situação normativa típica, alega sua comprovação empírica valorando os mesmos e argumentando sobre a possibilidade de valorá-los de outra forma. Retoricamente, pois, para solucionar um problema ideológico, apelando à coisificação de seu juízo de valor, apresenta o juiz suas valorações como dados susceptíveis de uma apreciação empírica²⁵.

As palavras que emprega a lei para descrever seus gêneros -e as que emprega a jurisprudência para descrever as espécies incluídas ou excluídas na lei- constituem *significações genéricas* referentes a verdadeiros *tipos empíricos*, claros e precisos no seu centro e muito confusos nos seus limites. Por outro lado, o caso mesmo com suas circunstâncias relevantes, pertence de fato a um tipo empírico ou gênero de conduta, e sempre implica, enquanto conduta, uma valoração. Além destes gêneros dados, na lei e a jurisprudência estão os *gêneros significativos* e as *palavras* do legislador ou do juiz. A própria significação, expressadas em vocábulos, é a fonte de sentido genéricos próprias da lei; porém na jurisprudência, pelo contrário, *a série de significações concordantes está referida*

Para entendermos o que é e o que determina uma *jurisprudência criminal*, inicialmente, cabe-nos analisar os requisitos formais de uma decisão criminal, partindo daí, numa análise retórica, definir outros aspectos desta mesma jurisprudência. No próximo item analisaremos os requisitos formais de um sentença penal.

²⁴ Warat, Luis Alberto. Op. Cit., p. 49.

²⁵ Warat, Luis Alberto. Op. Cit., p. 47.

3.2.2 *Requisitos formais da sentença penal*

Segundo Magalhães Noronha²⁶, “Sentença é a declaração judicial do direito ao caso concreto”. Segundo Frederico Marques²⁷, “Sentença é o ato de composição do litígio ou causa penal, em que o preceito normativo, imposto pela ordem jurídica, se transforma em preceito concreto e específico”. Ambas são concepções arraigadas no direito processual, onde a sentença é concebida estruturalmente como um silogismo em que a norma jurídica constitui a premissa maior, o fato *sub judice*, a premissa menor e a aplicação da norma ao fato, a conclusão. Tal concepção, conforme veremos no próximo item, não corresponde à realidade prática e encerra postura ideologicamente comprometida. Por enquanto, é bastante concuir que a falta ou defeito dos requisitos formais (veja tabela 1) pode acarretar nulidade da sentença (art. 800, do Código Penal). Isso significa que a forma da decisão de um juiz pode torná-la inválida.

O artigo 381, do Cód. de Processo Penal, estatui os requisitos formais da sentença penal:

"Art.381. A sentença conterá:

I - os nomes das partes ou, quando não for possível, as indicações necessárias para identificá-las;

II - a exposição sucinta da acusação e da defesa;

III - a indicação dos motivos de fato e de direito em que se fundar a decisão;

IV - a indicação dos artigos de lei aplicados;

V - o dispositivo;

VI - a data e assinatura do juiz".

Na tabela abaixo, esta a descrição sucinta do conteúdo deste artigo, para tornar uma sentença válida.

²⁶ Noronha, Magalhães. Curso de Direito Processual, Ed. Saraiva, 1978, p. 210.

²⁷ Marques, Frederico. Elementos de Direito Processual, III, p. 19.

INCISOS	IMPORTÂNCIA PARA A VALIDADE DE UMA SENTENÇA
I	Determina o âmbito pessoal de validade da sentença.
II	A finalidade é a demonstração por parte do juiz de que conhece o processo ou por ter participado de todas as suas fases, ou por ter examinado todas as suas peças.
III	Trata da parte mais difícil e importante da sentença. É aqui que o juiz reconstrói discursivamente o fato sub judice, procede à sua consunção ao direito que entende aplicável, acolhendo ou rejeitando os argumentos das partes e explicando, acolhendo ou rejeitando os argumentos das partes e explicando as razões pela qual decide
IV	O quarto é uma complementação do terceiro, está implícito.
V	É aqui que o juiz impõe qualitativamente a sanção penal, estabelece os limites objetivos da função sancionadora do Estado em relação ao réu, ou em outras palavras, determina o âmbito material de validade da sentença. O quinto requisito surge da necessidade de sintetizar a decisão poderíamos dizer que, se a fundamentação exigida pelo terceiro requisito constitui o núcleo retórico da sentença, o dispositivo encerra seu núcleo decisório
VI	O sexto quesito trata da autenticação da sentença.

Tabela 2. Requisitos formais de uma jurisprudência criminal.

A falta ou defeitos de um dos requisitos formais pode acarretar nulidade da sentença

Entretanto, os requisitos formais não são suficientes para recuperar todas as informações relevantes numa jurisprudência. Para Nilo B. de Brum²⁸ existem ainda, requisitos chamados retóricos que compõem estas jurisprudências. Embora não exista critério seguro que determine a reforma de uma sentença pela insuficiência de argumentação, este requisito é o que determina os motivos pelos quais uma decisão é tomada, uma sentença e reformada ou mantida, ou, até mesmo uma orientação dominante na jurisprudência dos tribunais é mantida. No próximo item explicaremos mais detalhadamente este requisito.

3.2.3 Requisitos Retóricos da Sentença Penal

²⁸ Brum, Nilo B.de. Requisitos Retóricos da Sentença Penal, p. 6.

Nilo de Bairros Brum²⁹ afirma que a sentença judicial exige mais que a mera fidelidade à lei e aos requisitos formais explícitos para impor-se à comunidade jurídica como norma individual vinculante, ou, em outras palavras, para que uma sentença receba adesão das partes ou dos tribunais, não basta que seu prolator observe religiosamente a lei material e os ritos processuais, nem mesmo é suficiente que cumpra a prescrições legislativas e as orientações doutrinárias e jurisprudenciais. A nova retórica advoga a razão argumentativa para o universo da linguagem não formalizável segundo a lógica cartesiana. Este universo significativo coloca problemas que somente podem ser resolvidos em termos de persuasão, eis que neles está pressuposta, de alguma forma, a liberdade de concordar ou discordar com as soluções propostas. Enquanto a razão de tipo cartesiano se desenvolve em nome de uma verdade absoluta, e portanto intemporal, a nova retórica é exatamente o campo de uma nova razão histórica e concreta.

Assim ele define requisitos retóricos da sentença como conjunto de argumentos essenciais da sentença, entendida esta como discurso persuasivo.

Então, como determinar quais são os argumentos essenciais numa jurisprudência?

Um juiz precisa justificar convincentemente os motivos pelos quais considerou um conjunto de provas a outro ou, mais genericamente, porque deu ao fato a configuração que deu, em detrimento de outras configurações possíveis e, em tese, igualmente razoáveis. Isto é atividade indiscutivelmente retórica que exige hábil manipulação de um vasto instrumental fornecido pela legislação processual, pelas teorias da prova, pelas presunções legais e extralegais, pelos precedentes judiciais e por um considerável elenco de falácias acreditadas na comunidade jurídica. Se a argumentação for persuasiva, isto é, se o fato foi reconstruído segundo os cânones aceitos pela comunidade jurídica, poderá o julgador obter o efeito de verdade indispensável para que sua decisão seja acatada e legitimada definitivamente; o mesmo vale para os profissionais do direito quando buscam extrair destas sentenças argumentos persuasivos para solucionar um problema jurídico.

²⁹ Brum, op. cit., pág. 84.

São argumentos essenciais numa jurisprudência, aqueles que podem convencer e persuadir a comunidade jurídica quanto a sua justa objetividade e legitimidade. Estes argumentos são fornecidos pela lei, pela chamada ciência do direito, mormente pelas diversas teorias da interpretação da lei e da avaliação das provas, e, ainda, pelas falácias acreditadas

Surge então os requisitos retóricos³⁰:

- **verossimilhança legitimada**: Se a argumentação for persuasiva, isto é, se o fato foi reconstruído segundo os cânones aceitos pela comunidade jurídica, poderá o julgador obter o efeito de verdade indispensável para que sua decisão seja acatada e legitimada definitivamente. Isto não quer dizer absolutamente que a versão eleita pelo juiz seja falsa, mas também não significa que seja verdadeira. O que se obtém através da reconstrução processual é a verossimilhança, pois, é o primeiro argumento contra a reforma da sentença judicial. Assim, para evitar-se a perpetuação dos litígios, a coisa julgada transforma a verossimilhança em uma presunção absoluta de verdade, embora, em alguns casos, admita-se o reexame dos fatos por meio da revisão do processo.
- **efeito de legalidade** : Obtida a verossimilhança fática, é, ainda, necessário dar-se uma solução jurídica ao caso. Como dizem os formalistas, é preciso aplicar o direito ao fato concreto. Na atual conjuntura das práticas judiciais, geraria grande insegurança na comunidade jurídica se os juizes passassem a decidir frontal e ostensivamente contra a lei sob a alegação de que o ordenamento legal é contraditório e lacunoso e que a norma legal é vaga, ambígua ou iníqua. Assim, qualquer que seja a decisão do juiz, deve ela parecer estar contida de alguma forma na previsão do legislador, pois não é boa estratégia retórica colocar-se contra a legalidade.
- **adequação axiológica**: Para atender a tal requisito, conta o órgão judicial com um cabedal de recursos fornecidos pelas teorias jurídicas, pelos múltiplos métodos de interpretação da lei e pela cobertura oferecida pelo próprio legislador que transfere a sua função ao órgão judicial através de expressões semanticamente vazias ou anêmicas, cujo significado somente pode ser obtido contextualmente. São exemplos

³⁰ Brum, Nilo. Op. Cit., p. 86.

dessas variáveis axiológicas "repouso noturno", "situação análoga", "mulher honesta", etc.

- **neutralidade judicial:** Neste campo de batalha ideológica, o juiz deve eleger parecendo neutro, racionalizar o irracional e dar cientificidade aos sentimentos. Assim, como último requisito retórico da sentença, podemos considerar o efeito de neutralidade judicial.

Estes argumentos devem atender às diversas linhas ideológicas que compõem o pensamento da comunidade jurídica em um espaço (aqui) e momento histórico(agora) determinados. Com isso, nós concluímos que é possível determinar quais são os argumentos essenciais existentes num acórdão, estabelecendo quais os argumentos usados com mais frequência num determinado Tribunal para usá-los como valores dos índices estabelecidos com base nos requisitos formais.

Segundo o autor, uma das consequências dessa desingenuização é a consciência da existência

Os tribunais possuem seus índices próprios para facilitar o armazenamento e a busca pelos acórdãos. Estes índices são normalmente muito úteis e eficazes, mesmo numa busca manual, por exemplo, se o profissional jurídico necessitar alguma jurisprudência sobre "*assalto a mão armada*", ele não encontrará num código a expressão referida onde a equivalente é "*roubo com emprego de grave ameaça*", mas é uma expressão muito usual no meio jurídico, e provavelmente será encontrada em qualquer índice jurisprudencial.

Para que se possa, claramente, expor explicar as formas retóricas da comunicação jurídica, será de grande utilidade iniciar com o estudo de seu relacionamento com os usos ou funções da linguagem natural.

3.2.4 Uso funcional da linguagem jurídica

Toda a evidência sugere que na busca feita pelo usuário final, a linguagem de indexação deve ser a linguagem natural, em vez de uma linguagem

controlada. Para uma busca interativa, esta linguagem deve ser diretamente acessível para o usuário formular a sua requisição.³¹

A linguagem não só permite o intercâmbio de informações e de conhecimentos humanos, como também funciona como meio de controle de tais conhecimentos. Ao fazermos tal afirmativa, estamos inserido-nos em uma corrente lingüístico-epistemológica geralmente conhecida como Positivismo Lógico ou Empirismo Contemporâneo. O positivismo Lógico realça o rigor discursivo como o paradigma da ciência, ou seja, a produção de um discurso científico requer uma análise preliminar em termos de linguagem³².

O Realismo Jurídico Norte- Americano, ao tentar rejeitar as teses do formalismo jurídico (que pretendia caracteriza a linguagem jurídica como geométrica, regida unicamente por critérios sintáticos), aplica a condição semântica de sentido, isto é, reivindicam as sentenças em substituição das normas jurídicas, pois apenas aquela possuem correspondência fática.

Do ponto de vista do Positivismo Lógico, a aplicação da condição semântica de sentido nos moldes do Realismo Norte-Americano é inconsistente porque, como veremos adiante, as condições semânticas só operam em linguagem de estrutura especificada, jamais em linguagem natural, baseada na condição pragmática de sentido. Nesta perspetiva, Warat reforça a idéia que as *sentença são normas e não fatos*.³³

A pragmática, conforme a definição proporcionada por Carnap, *é a parte da semiótica que estuda a relação dos signos com os usuários*. Sua problemática central gira em torno da análise dos modos de significar, usos ou funções da linguagem. Por esta razão, o emprego de um expressão ou termo num contexto pode provocar uma alteração na estrutura conceitual deste termo ou expressão.

Em primeiro, indagar sobre um uso lingüístico ou um modo de significar, é realizar uma análise em torno do significado emprestado às palavras e expressões pelos indivíduos que as estão usando. Parece evidente que não é possível

³¹ Lewis, David et al. Natural language Processing for Information Retrieval, p. 95.

³² Warat, Luis Alberto. O direito e a sua linguagem, p. 37.

³³ Warat, L. A. Idem, p.45.

interpretar o significado de um termo (ou expressão) sem atender ao uso em que ele parece, isto é, sem estabelecer seu significado contextual³⁴.

Sabemos bem, pela nossa experiência cotidiana, que a mesma frase dita por pessoas diversas, inserida ou não em um determinado contexto, pode assumir significados diferentes e muitas vezes opostos. De fato, é justo estudar a língua com métodos rigorosos que respondem as exigências da formalização e da exatidão, mas ao mesmo tempo não há necessidade de descuidar-se os múltiplos aspectos da comunicação verbal e o contexto social na qual ela se manifesta. A língua é acima de tudo troca comunicativa entre duas ou mais pessoas.

No estudo de uma língua determinada verificaremos que não há sentido autônomo - embora muitos linguístas o queiram - mas sentido dependente do contexto. Este é o erro da linguagem jurídica, o discurso jurídico não é formal é natural - todas as grandes conquistas do direito foram dentro da obscuridade da lei.

Se o emissor explicitar devidamente o que entende por um determinado termo ou frase, saberemos o que realmente significa aquele termo no contexto em que foi empregado, usando a intuição para entender a finalidade da linguagem, provavelmente verificaremos uma menor possibilidade de cairmos nas falácias da comunicação verbal.

Os estereótipos tipo *Ordem Pública e estado de necessidade* dão estabilidade ao direito, mas não asseguram a sua finalidade. No direito é imprescindível usarmos a significação contextual, não como fator de instabilidade, mas como condição de justiça e talvez, em determinados contextos, imprescindível será desprezarmos a significação de base para, desta forma, entendermos melhor o significado contextual dos textos jurisprudenciais.

Em suma, a análise com base pragmática (que extrapola o Positivismo Lógico) permite articular certas características do funcionamento significativo (persuasão, legitimação, antecipação), explicitando em grade parte as funções dos discursos na sociedade e mostrando como a identificação empirista entre

³⁴ Warat, L. A. Introdução Geral ao direito - I Interpretação da Lei, temas para uma reformulação, p. 127.

descrição e realidade é uma forma de despolíticação-politizadora do discurso científico, que cria um efeito de inquestionabilidade e realidade sobre o saber.

3.3 Raciocínio Baseado em Casos

O raciocínio baseado em casos constitui uma poderosa ferramenta de utilização das mais variadas experiências humanas. É muito usual, na história da humanidade, buscar-se uma experiência anterior para auxiliar na resolução de um problema atual, fazendo do passado um imenso banco de informações que nos ajuda a pensar sobre o presente e o futuro. Uma introdução e algumas referências gerais podem ser encontradas em: Riesbeck & Schank (1989), Kolodner (1993), Aamodt & Plaza (1994), Leake (1996) and Watson (1997).

O processo característico do RBC consiste em : identificar a situação atual, buscar a experiência mais semelhante na memória e aplicar o conhecimento desta experiência passada na situação atual. O RBC é uma técnica da IA muito adequada para a modelagem do conhecimento jurídico. Muitos teóricos da área (e.g., Rissland, Ashley, Branting, Skalak) defendem o uso dessa abordagem para o domínio jurídico.

Os advogados utilizam a técnica do Raciocínio Baseado em Casos quando eles usam casos para justificar argumentos. Os Tribunais usam este tipo raciocínio quando eles interpretam um novo caso na luz do caso anterior. Porém, discute-se acerca da forma de utilização das experiências passadas.

A jurisprudência, composta de casos e precedentes, constitui-se num bom campo de aplicação para os sistemas de RBC.

Para que possa ser utilizada, devemos estudar os referenciais já estabelecidos dessa área da IA, verificando no que e em que condições os institutos podem se aproximar e produzir um resultado concreto, com o desenvolvimento das bases iniciais para um sistema.

O processo básico do Raciocínio Baseado em Casos consiste na identificação de um problema a ser resolvido (caso de entrada); definição das principais características que identificam este problema; busca e recuperação na memória de casos com características similares; seleção de um ou mais dentre os casos recuperados; revisão destes casos para determinar a necessidade de adaptação; reutilização do caso adaptado para resolve o problema de entrada; avaliação da solução do problema de entrada; inclusão do caso adaptado na memória de casos (aprendizagem).

Podemos ainda, simplificar estes procedimentos em quatro etapas distintas, dependendo do enfoque de alguns autores : 1. representação e indexação; 2. recuperação; 3. revisão e adaptação; 4. reutilização, avaliação e aprendizagem. Veremos a seguir uma resumo de cada uma destas etapas.

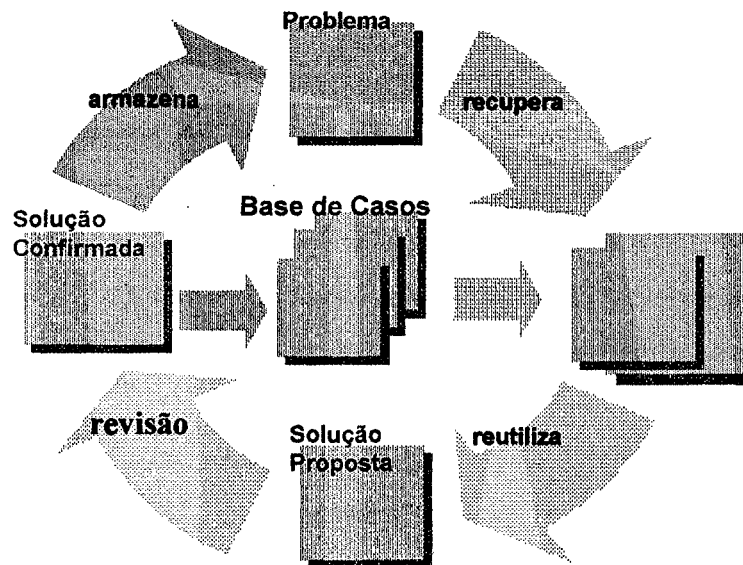


Figura 3. Ciclo CBR.

3.3.1 Representação e Indexação dos casos

A representação dos casos num sistema de RBC é essencialmente a representação do conhecimento. Há outros momentos em que algum conhecimento especialista é representado no sistema; a base de conhecimento está nos casos e os casos estão na memória de casos ou base de casos.

A representação da base de conhecimento num sistema de RBC consiste, então, em modelar os casos e definir o estilo de memória que organize estes casos. A execução destas etapas estão fortemente ligadas entre si e estas, por sua vez, dependem fortemente das outras etapas do sistema. As outras etapas do sistema, recuperação adaptação e aprendizagem devem ser projetadas simultaneamente à representação. Sempre orientando-se pelo objetivo da aplicação, avalia-se os benefícios da representação de conhecimento sugerida para todas as etapas complementares do sistema.

Como conhecimento nos sistema baseados em casos trata da descrição e organização dos casos na memória, a indexação é uma questão muito importante, pois a memória deverá ser indexada para proporcionar uma recuperação e reutilização eficientes.

A questão principal na escolha do vocabulário de indexação , está no fato que o referido vocabulário terá que representar uma interpretação da situação.

A indexação de um caso designa em qual circunstância o caso poderá ser recuperado.

O problema da seleção dos índices, pode ser dividido em **vocabulário de indexação e avaliação do índice**. É o vocabulário de indexação que nos diz quais os tipos de vocabulários podem ser usados para uma determinada classe de casos, é um *subconjunto dos vocabulários usados para a completa representação simbólica dos casos*, já a avaliação do índice é o processo de escolha das características de identificação para um caso particular.

Neste momento trataremos da primeira questão. Em primeiro lugar, na indexação das características foram observados os seguintes passos ³⁵: a) antecipação do vocabulário que o usuário poderá usar, a indexação tem que necessariamente antecipar a linguagem utilizada para formular a questão problema; b) utilização de conceitos pertencentes a área domínio da aplicação. A indexação tem que ser feita através de conceitos que são normalmente usados para descrever os itens que serão indexados, se eles são características superficiais ou alguma coisa mais abstrata; c) antecipação das circunstâncias nas quais o usuário quer recuperar alguma coisa (isto é, o contexto no qual será recuperado o caso) e se o vocabulário do usuário é apropriado para ser avaliado para descrever o item que ele/ela quer recuperar.

A comunidade do RBC propôs algumas orientações para a escolha de índices:

1. Os índices precisam representar o contexto, a soluções e os problemas do domínio;
2. Esta representação precisa ser proveitosa, isto é, ela precisa dirigir a proposta para a qual o caso será usado;
3. Os índices precisam ser abstratos o suficiente para tornar um caso proveitoso numa variedade de situações futuras.
4. Os índices precisam ser concretos o suficiente para tornar fácil o reconhecimento em situações futuras.

Nós sabemos que, índices precisam ser representativos, úteis, aplicáveis genericamente e reconhecíveis. O índice do caso deve descrever cada tipo de situação na qual ele tem o potencial de ser aplicado. Mas quais dos caminhos possíveis são obrigatórios para descrever a situação de um caso, a solução e o resultado .

Existem dois conjuntos de teorias que podem ser examinadas para determinar o conteúdo dos índices ³⁶:

- a funcional ou pragmática - ponto de vista que examina os tipos de casos que são avaliados e a tarefa que ele podem suportar para identificar as dimensões e símbolos que, se usados como índices, permitirão que os casos avaliados cubram a série de tarefas designadas.

³⁵ Kolodner, Janet L. Case-based reasoning, p. 194.

- a de lembrança - evidenciam conceitos particulares que ligam casos de um domínio ou conjunto de domínios, procurando ser o mais genericamente possível, mas tendo em mente que cada domínio provavelmente tem seu próprio vocabulário de domínio específico.

Os resultados de cada uma destas análises dizem-nos qual dimensão é importante para focalizar sobre a série de valores que cada um recebeu e o nível de detalhe que é vantajoso na representação.

Existem três coisas para se ter em mente enquanto símbolos e dimensões são escolhidos: 1. o vocabulário de indexação precisa ser geral o suficiente para cobrir a série de tarefas que o RBC é responsável e especificar o suficiente para fazer a diferenciação que é necessária para aquela tarefa. De qualquer maneira, não é necessário fazer o vocabulário mais geral ou mais específico que o requisitado para a tarefa designada; 2. o vocabulário de indexação precisa cobrir a série de casos que o RBC irá usar; 3. o vocabulário de indexação precisa antecipar o grau e direções nas quais o sistema irá expandir no futuro.

Os índices para um caso são um subconjunto de descrição ou representação de casos, por esta razão, o vocabulário que é avaliado para indexação de casos, pode ser tão bom quanto o representacional ou descritivo que está sendo avaliado.

As teorias do vocabulário de indexação são realmente representacionais - elas descrevem as dimensões representacionais que são necessárias para descrever os casos, ainda que os próprios casos possam ter mais detalhes neles que nos índices.

3.3.2 Recuperação dos casos (métrica da similaridade e seleção do melhor caso)

³⁶ Kolodner, Janet L. Ob. cit., p. 208.

Podemos recuperar casos buscando diretamente os índices das características, ou fazendo uma busca numa estrutura de índices, ou ainda fazendo uma busca em um modelo de conhecimento mais amplo. Quando a recuperação é do tipo que busca a similaridade diretamente comparando os índices, uma métrica de similaridade é usada para este cálculo. Em métodos de recuperação baseado em restrições ou classificação, uma métrica talvez seja necessária apenas para estabelecer o grau de similaridade dos casos e ordená-los. A seleção do *melhor caso* é a etapa final da recuperação. Segundo algum método de seleção, que pode ser perguntas ao usuário, heurística, regras, etc., realiza-se a escolha do caso cuja solução será utilizada como saída do sistema.

3.3.3 revisão e adaptação

Na resolução de um problema no sistema baseado em casos, velhas soluções são usadas como inspiração para a resolução de um novo problema. Isto porque as novas situações raramente encontram uma antiga que seja exatamente igual, deste modo, velhas soluções precisam ser fixadas para amoldar uma nova situação. Muito importante na *adaptação* são as metodologias para avaliar quais partes de uma solução antiga precisa adaptação para amoldar-se a uma nova.

3.3.4 reutilização, avaliação e aprendizagem

Assim que a adaptação é feita, a diretriz do caso escolhido pode ser então utilizada para resolver o problema de entrada. A avaliação tem por objetivo avaliar a qualidade da solução adaptada ao problema de entrada para definir se esta tem condições de ser adicionada à memória do sistema e pode ser automática ou com a participação do

usuário. A aprendizagem é esta inclusão do caso adaptado, reutilizado e avaliado à base de casos.

No presente trabalho, as duas últimas etapas foram excluídas, por duas razões: a primeira é referente ao escopo deste trabalho que é a recuperação de Acórdãos na íntegra, isto é, casos jurídicos que serão utilizados para fundamentar outras decisões judiciais; a segunda diz respeito a natureza jurídica destes Acórdãos judiciais, que são casos que não podem ser modificados e adaptados para serem incluídos no banco de casos do protótipo, pois não poderíamos denominar estes novos casos de Jurisprudência, cuja elaboração é de competência exclusiva dos Tribunais; seria necessário fazer uma outra base de dados específicos onde estariam os casos de entrada (que não são jurisprudências) revisados, adaptados e avaliados, para serem recuperados posteriormente para a solução de casos semelhantes.

3.5 Recuperação da Informação

Com o crescente aumento nos últimos anos no número de bancos de dados de texto disponível on-line, e a necessidade para técnicas mais aperfeiçoadas para ter acesso esta informação, houve um forte interesse na pesquisa feita na área de recuperação de informação (IR). Por muitos anos, a pesquisa de IR foi desenvolvida por uma comunidade pequena que obteve pouco impacto na indústria. A maioria das aplicações de recuperação de texto enfocava em bancos de dados bibliográficos. Além disso, os grandes serviços de informações eram baseado na lógica Booleana, com pouca atenção para os resultados de pesquisa em tópicos como modelos de recuperação, processamento do problema, pesagem dos termos, etc.

Dado a velocidade com que indústria adotou os resultados de pesquisa de IR dos anos setenta e oitenta, a comunidade de IR foi desafiada a identificar novas direções. Um problema é que as prioridades que os investigadores de IR colocam em assuntos de pesquisa necessariamente não é igual a prioridade de companhias e agências de governo que usam e vendem sistemas de IR. Entender essas prioridades e a

experiência operacional atrás delas é parte do processo de decidir quais assuntos são de importância fundamental e quais são passageiros.

3.5.1 Soluções Integradas

O problema mais importante do ponto de vista de companhias que usam e vendem sistemas baseados em textos é a integração com outros sistemas. Um sistema de recuperação de texto é uma ferramenta que pode ser usada para resolver parte dos problemas de administração de informação de uma organização. Porém, não é a solução mais completa. Tipicamente, uma solução completa requer outras ferramentas baseadas em texto, como as de extração, ferramentas para manipulação multimídia e ferramentas como OCR, um sistema de administração de banco de dados para dados estruturados, etc.

Um dos aspectos mais importantes de plataformas comuns em desenvolvimento é a integração de banco de dados administração e sistemas de IR. Uma integração efetiva destes sistemas, junto com multimídia, forneceria um sistema de informação que poderia ser usado para administrar muitas das aplicações que atualmente existem. Embora soluções parciais para esta integração existam, elas não resolvem um dos assuntos fundamentais que são aqueles sistemas de banco de dados que recuperam tudo usando lógica Booleana, considerando que textos requerem técnicas que envolvam tratamento de incerteza para uma recuperação efetiva.

Alguns destes desafios serão resolvidos a curto prazo, mas outros serão a base de projetos de pesquisa mais longos. As oportunidades e experiência contanto pela explosão de sistemas baseados em textos será inestimável para pesquisa de IR.

Capítulo 4 - Suporte Inteligente para a Pesquisa Jurisprudencial

Neste capítulo nós apresentamos uma abordagem baseada em casos para recuperação de textos jurisprudenciais com características semelhantes e adequados a solução de um problema jurídico, além da inclusão automática de novos documentos na base de conhecimento. Isso é realizado com apoio em informações inicialmente fornecidas pelo usuário em linguagem natural. Para possibilitar este processo de conhecimento textual informal, um vocabulário controlado e um dicionário jurídico - baseado na terminologia jurídica usual e na teoria jurídica- são integrados na recuperação e no processo de extração do conhecimento. Esta abordagem foi desenvolvida para a área do direito criminal.

4.1 Recuperação de Acórdãos Criminais usando a teoria jurídica

Para entendermos como um documento jurídico é estruturado e quais são as informações relevantes a serem recuperadas numa pesquisa jurisprudencial, é necessário prestar atenção nos problemas que usualmente acontecem no domínio jurídico.

Na pesquisa jurisprudencial os problemas mais comuns são a quantidade enorme de documentos jurídicos e a necessidade de uma resposta adequada ao problema a ser resolvido. O estudo destas questões é um passo fundamental para desenvolver um modelo para as tarefas jurídicas que possam ser efetivamente implementadas em sistemas jurídicos baseados no conhecimento.

No domínio jurídico, o objetivo da recuperação em textos jurídicos é fornecer informações que sejam bons argumentos para justificar a aplicação da lei em um caso concreto. O caso recuperado tem que auxiliar o profissional jurídico na

solução de um caso atual. Desta maneira, os Acórdãos Criminais têm que ser representados e armazenados de um modo acessível, os casos relevantes têm que ser recuperados e novos documentos jurídicos incluídos automaticamente na base de conhecimento. Isso é possível pelo emprego da teoria jurídica, que define os critérios de relevância das informações nos textos jurídicos.

Nas seções seguintes são descritas as partes principais de nossa abordagem baseada em casos:

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

4.1.1 Representação dos casos jurídicos

O conhecimento jurídico inserido numa jurisprudência é representado em forma de um caso que consiste no texto do documento original e um conjunto de índices em forma de pares atributo-valor (veja Figura 4). O uso de índices durante o processo de recuperação permite um acesso eficiente ao texto documental pela referência do conteúdo deles. Nós denominamos esta representação de formal, pois estes índices foram determinados em conformidade com os requisitos formais da sentença penal - exigência normativa do artigo 381 do Código de Processo Penal (veja Tabela 2, pág. 29) e utilizados conforme a sua relevância para a recuperação. Além desta representação dos casos jurídicos, o conhecimento do domínio é incluído em forma de um *vocabulário jurídico controlado* e um *dicionário de termos normativos*, usados para a valoração de alguns índices. É este conhecimento de domínio que permite a recuperação

dos documentos e o processo de extração automático, através da identificação de expressões indicativas e relevantes dos textos jurídicos em linguagem natural, juntamente com a modelagem explícita da semelhança destes termos jurídicos. Esta forma de representação nós denominamos de retórica, porque depende de uma análise retórica do conteúdo do Acórdão Criminal (veja item 3.2.3 *Requisitos Retóricos da Sentença Penal*).

Requisitos da Sentença Penal:

"art.381. a sentença conterà:
I - os nomes das partes ...
II- a exposição sucinta da acusação e da defesa;
III- a indicação dos motivos de fato e de direito em que se fundar a decisão;
IV - a indicação dos artigos de lei aplicados;
V - o dispositivo;
VI- a data e assinatura do juiz".

Estrutura de Representação dos Casos:

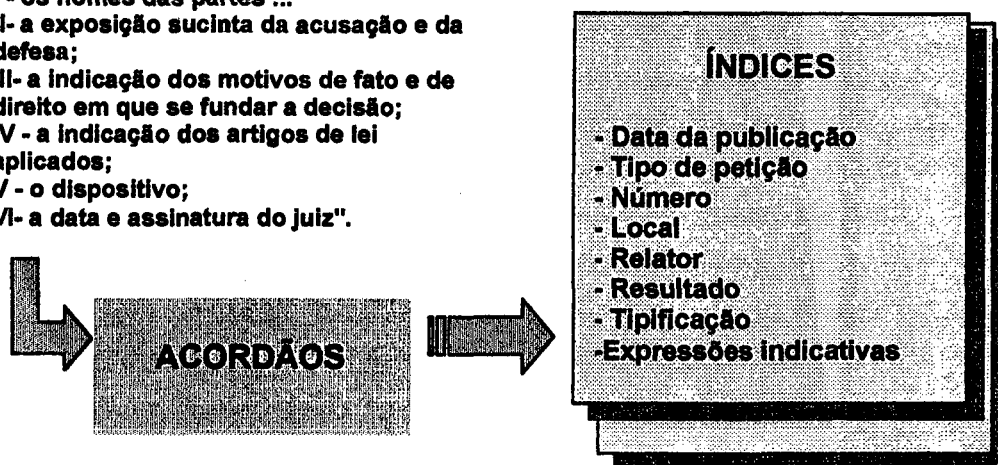


Figura 4. Representação de um documento jurídico em forma de caso.

4.1.2. Índices para a Recuperação

Para representar os textos jurídicos de uma maneira acessível, o Acórdão Criminal é dividido em uma representação estruturada de acordo com o conteúdo formal da sentença penal e o contexto das informações inseridas nele. Assim, um conjunto de índices é determinado para facilitar o processo de recuperação (veja a tabela 3).

No contexto jurisprudencial, o principal problema do desenvolvimento de sistemas baseados no conhecimentos jurídico é a determinação da informação juridicamente relevante a ser recuperada. O princípio da analogia sustenta a

maneira de reutilização dos casos jurídicos; por esta razão, os atributos dos documentos textuais, usados como índices para a recuperação, têm que indicar a utilidade das informações do caso na situação presente. E assim, na nossa abordagem, os índices são determinados pelos requisitos formais da sentença penal em conformidade com a importância deles para o contexto de recuperação da informação jurídica. Nós usamos esta forma de divisão como base para definir o processo de extração do conhecimento através da localização do atributo (veja item 4.2.1 *Extração de Valores de Índices pela Localização*).

Entretanto, só esta forma de representação não é suficiente para recuperar a informação jurídica relevante dos Acórdãos Criminais. A interpretação do conteúdo de uma jurisprudência não está vinculado somente ao seu conteúdo normativo, mas também as ideologias. Pois, afirmar que a lei possui um sentido exclusivo e verdadeiro é uma visão formalista, e não corresponde à realidade prática e encerra postura ideologicamente comprometida. Uma sentença não é uma interpretação exclusiva e verdadeira da lei, mas também uma interpretação que redefine o direito.

Um juiz precisa justificar convincentemente os motivos pelos quais considerou um conjunto de provas a outro ou, mais genericamente, porque deu ao fato a configuração que deu, em detrimento de outras configurações possíveis e, em tese, igualmente razoáveis. Isto é atividade indiscutivelmente retórica que exige hábil manipulação de um vasto instrumental fornecido pela legislação processual, pelas teorias da prova, pelas presunções legais e extralegis, pelos precedentes judiciais e por um considerável elenco de falácias acreditadas na comunidade jurídica. Se a argumentação for persuasiva, isto é, se o fato foi reconstruído segundo os cânones aceitos pela comunidade jurídica, poderá o julgador obter o efeito de verdade indispensável para que sua decisão seja acatada e legitimada definitivamente; o mesmo vale para os profissionais do direito quando buscam extrair destas sentenças argumentos persuasivos para solucionar um problema jurídico.

O resultado desta forma de representação são índices que representam o conteúdo jurídico dos textos da Jurisprudência Brasileira, como as *expressões indicativas* e a *tipificação*. Além disso, são representadas as informações

Número, Localização e Relator - estas informações podem ser usadas para limitar a procura, por exemplo, num certo período de tempo.

Na tabela a seguir, estão a descrição dos índices representativos do conteúdo formal e retórico de uma sentença penal e a importância para o processo de recuperação:

Índices		Descrição	Importância para a recuperação
Número do Acórdão		Número fornecido pelo Tribunal .	É usado para identificar o recurso
Data da Publicação		Refere-se a data de publicação no Diário da Justiça.	Com essa informação, o usuário poderá determinar qual o período da pesquisa (e.g. 11/11/1995 até 01/01/1999)
Local		Informa a cidade origem do processo.	Pode ser usado para definir a pesquisa por localidade.
Relator		Determina qual foi o juiz responsável pela elaboração da decisão.	Pode ser usado para recuperar os acórdãos proferidos por um determinado relator.
Tipo de recurso		Determina o tipo de recurso usado para rever uma decisão judicial	Serve para delimitar a pesquisa quanto ao tipo de procedimento legal usado para resolver uma questão jurídica.
Expressões jurídicas		Determina as informações jurídicas relevantes e discriminante numa sentença.	Usado para recuperar jurisprudência com tópicos similares.
Resultado		Indica se os Juizes negaram ou aceitaram o pedido feito no recurso	Usado para determinar o resultado do pedido.
Tipificação	Tipo geral	Define o enquadramento legal genérico, que é o tipo de crime que deu origem ao processo judicial, eg., homicídio, furto, estelionato, etc.).	Usado para diferenciar os diferentes tipos de crimes.
	Modalidade do crime	Define a intenção criminal da pessoa (e.g., crime cometido por uma pessoa quando o mesmo tem a intenção clara e objetiva de cometê-lo, ou quando, acontece acidentalmente, por culpa da pessoa.	Usado para definir a modalidade do crime cometido (doloso ou culposo)
	Qualificação do crime	Determina se um crime foi praticado em condições especiais(ex.: furto praticado à noite), ou se o agente possui uma característica especial que aumente ou diminua a pena (ex.: reincidência).	Usado para definir a qualificação de um crime (simples, qualificado ou privilegiado).
	Tentativa	Determina quando uma pessoa atinge ou não o seu objetivo criminal, por razões alheias a sua vontade (e.g., outra pessoa segura a arma no momento do disparo). intenção de matar a vítima.	Usado para determinar se uma pessoa alcançou ou não o objetivo de sua intenção criminal.
	Co-autoria	O mesmo crime pode ser cometido por duas ou mais pessoas. Estas pessoas(agentes) serão punidos conforme a sua participação no crime.	Usado para determinar o número de pessoas envolvidas num crime.

Tabela 3. Descrição dos índices e a sua importância para a recuperação

4.1.3 Conhecimento do Domínio Jurídico

que indicam quais são os termos relevantes para serem empregados na construção de argumentos persuasivos na solução de problemas jurídicos. É por esta razão que a teoria de argumentação jurídica fornece uma base teórica na construção de uma lista de expressões (termos-chave normativos) e de um dicionário de sinônimos destes termos (com base na terminologia usual dos Tribunais) para representar eficiência o conhecimento jurídico presente nestas decisões. Assim, esta forma de representação terá maior relevância jurídica que a extração de conhecimento feita exclusivamente baseado na informação contida no próprio texto jurisprudencial.

Os argumentos são persuasivos em Lei, quando eles não contradizem a ideologia de seus destinatários. Em última instância, a ideologia é o tribunal pelo qual a efetividade de um argumento jurídico é medida³⁸. Deste modo, para apoiar o processo de recuperação e uma extração automática dos índices relativo ao conteúdo jurídico dos textos da jurisprudência criminal, um vocabulário controlado e um dicionário de termos normativos foram desenvolvidos.

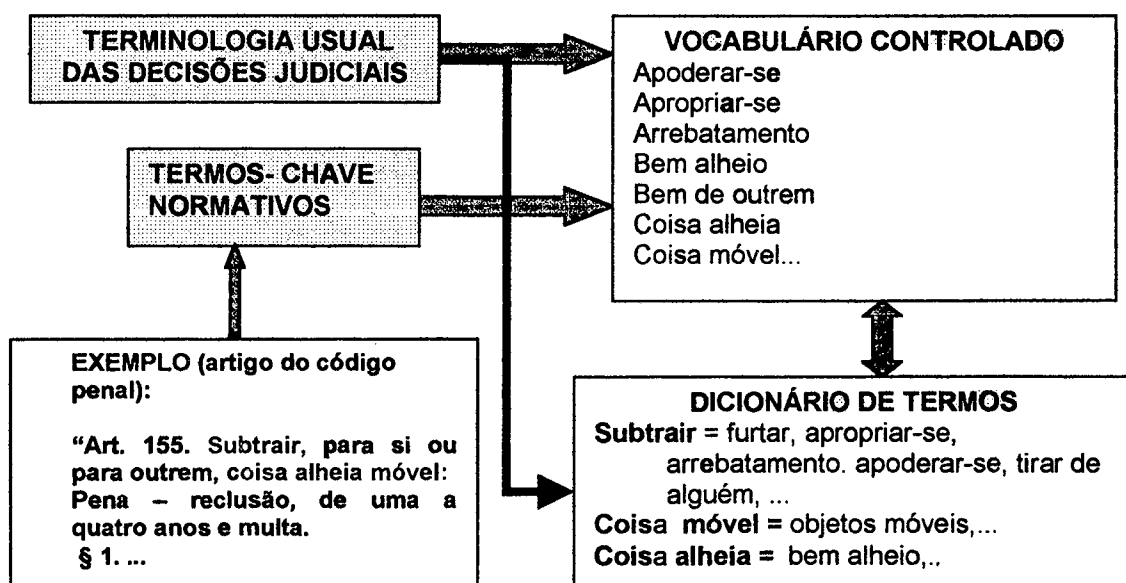


Figura 5. Definição da Expressões Indicativas.

³⁷ Warat, Luiz A. *Interpretação da Lei.: temas para uma reformulação*, pág. 90.

4.1.3.1 Vocabulário controlado

O vocabulário controlado define o valor do índice mais útil para a recuperação de um acórdão, enfocando no domínio de aplicação específico (Área Criminal). Este vocabulário controlado é constituído de *expressões indicativas* que, por sua vez, são definidas através dos termos-chave de uma norma. Por exemplo, o artigo 154 do Código Penal Brasileiro dispõe sobre o crime de violação de segredo profissional:

***“Art. 154. Revelar alguém, sem justa causa, segredo, de que tem ciência em razão de função, ministério, ofício ou profissão, e cuja revelação possa produzir dano a outrem:
Pena – detenção, de três meses a um ano, ou multa.
Parágrafo único. Somente se procede mediante representação.”***

Neste exemplo, podemos extrair os seguintes termos-chaves: - “*revelar segredo*”, “*produzir dano a outrem*”. São estes termos-chave que nós denominamos *expressões indicativas* e constituem os vocabulário controlado. Aqui, é possível observar que na extração dos termos-chave de uma norma, pode-se aumentar o número de termos, isso é necessário para que a *expressão indicativa* seja juridicamente relevante. Isto é, para a construção deste vocabulário controlado, foi necessário fundamentar a definição destes valores de índice em condições normativas que representam contextos e podem ser importante para a recuperação, como neste outro exemplo: “*agente menor de 21 anos*”; esta é uma expressão normativa que está inserida no art. 65, do Código Penal, como circunstância atenuante, e também no art. 115 do CP, para fins de redução do prazo prescricional, esta expressão é juridicamente relevante para a solução de questões que envolvam a menoridade do agente e portanto pode ser considerada uma *expressão indicativa*.

Do ponto de vista de uma estrita legalidade, o caso pode ser naturalmente incluído em uma classe normativa; no entanto, aos olhos do julgador, essa

inclusão pode parecer totalmente injusta. Normalmente, quando isso acontece, o juiz produz uma definição explicativa de caráter ideológico, apresentando como relevantes outras conotações para caracterização de algum termo-chave incluído na norma, e decide com isto uma mudança de linha de soluções³⁹. Muitas vezes, operar com os termos-chave das normas, não é, aos olhos do juiz, uma técnica de argumentação adequada. Então, ele deixa de lado as definições explicativas, e introduz termos de pura carga emotiva, produzindo uma definição persuasiva desses termos, podendo dirimir convincentemente neste caso, o conflito de interesses que tinha em mãos. É por esta razão que somente uma lista com os termos-chave normativos não é suficiente para representar as informações juridicamente relevantes de um caso. No próximo item, nós discorreremos sobre um dicionário de termos normativos, elaborado com a intenção de redefinir - no sentido de alterar as características de relevância de um termo- para possibilitar a aplicação de um termo normativo a situações antes não consideradas.

4.1.3.2 *Dicionário de Termos Normativos*

Toda vez que os critérios de relevância de um termo são alterados no seu uso contextual, ocorre uma redefinição. Para redefinir os juristas interferem diretamente sobre as significações dos termos, propondo novas regras designativas, ou realizando um forma indireta de interferência. Tal mecanismo, salienta Warat⁴⁰, facilita a adequação da jurisprudência, às exigência, reais ou supostas, de um dado momento histórico.

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

A linguagem natural, é a linguagem em que se inscrevem as normas; possui características estruturais e situacionais que dificultam a transmissão clara

³⁹ Warat, Luis Alberto. Op. cit. p.36.

⁴⁰ Warat, Luis Alberto. Op. cit. p. 38.

das mensagens. Estas características cumprem importante função retórica em relação às práticas tribunalícias. Assim, para considerarmos a similaridade entre qualquer expressão jurídica e os termos normativos relevantes (expressões indicativas) é necessário analisar que muitos termos são definidos persuasivamente, interferindo diretamente sobre as significações dos termos.

O dicionário de termos normativos define a semelhança lingüística, semântica e axiológica de condições que permitem a determinação da semelhança local entre valores de um índice. O dicionário foi desenvolvido por profissionais do direito com base na experiência deles/delas e fundamentado pelo conhecimento de domínio geral em direito criminal.

A adequação da terminologia usual à terminologia normativa foi realizada com base na terminologia usual do Tribunal de Justiça de Santa Catarina⁴¹, pois devemos levar em conta que cada Tribunal possui o seu próprio modo de redefinir a lei, direta ou indiretamente. Assim, para uma boa indexação é necessário usar os índices do próprio Tribunal do qual originaram-se as jurisprudências, por exemplo, suponhamos que no Tribunal **A**, a jurisprudência dominante defina que *propaganda difamatória* caracteriza o crime de *concorrência desleal*, já o Tribunal **B**, defende a idéia que a *propaganda difamatória* não caracteriza o crime de *concorrência desleal*, mas sim a *propaganda infundada*. No primeiro caso, haverá similaridade local entre a expressão “qualquer propaganda difamatória” e o termo normativo “concorrência desleal”, já no Tribunal B, haverá uma outra interpretação para o termo, assim como poderá haver Tribunais onde *propaganda infundada* seja sinônimo de *propaganda difamatória*.

Definida esta similaridade entre os termos, o dicionário permite recuperar casos relevantes que são semelhantes. Por exemplo, se a situação presente é caracterizada pela expressão coação moral, casos nas bases de conhecimento caracterizadas por expressões semelhantes, como por exemplo, *vis compulsiva* também são considerados como casos pertinentes, até mesmo se a expressão que coerção moral não é declarada. Isto evita a exclusão de um caso relevante, quando os valores de índice não são iguais, embora eles sejam bastante semelhantes. Outro exemplo, é o caso da

⁴¹ <http://www.tj.sc.gov.br>.

expressão “menoridade”, ela não é uma expressão normativa, no entanto na doutrina e jurisprudência é amplamente usada para definir os “menores de 21 anos”, os “menores de 18 anos”, os “menores de 14 anos”, etc. O termo “menoridade” é também um ótimo exemplo para ressaltar a importância do uso reiterado de uma expressão relevante. Assim, podemos definir que há similaridade local entre os termos normativos *menor de 21 anos*, *inimputáveis*, *menor de 18 anos*, *menor impúbere*, *menor púbere* e o termo usual *menoridade*.

4.2 Extração Automática da Informação textual

Pela análise de textos legais foi identificado um formato padrão. As jurisprudências são estruturadas de acordo com o tipo de informação fornecida,⁴² em quatro subestruturas básicas (veja Tabela 3). Algumas informações são facilmente localizadas no texto, embora a identificação de outras informações exigem apoio teórico e conhecimento do domínio.

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

Tabela 4. Estrutura de um Acórdão Criminal.

⁴² Weber, R. et al. Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97.

Então, nós desenvolvemos técnicas diferenciadas para a extração de conhecimento dos textos jurídicos (Acórdãos):

- Através da Localização.
- Através do uso de um Vocabulário Controlado e Dicionário de Termos Normativos.
- Através de Inferência.

4.2.1 extração de Valores de Índice através da Localização

Os documentos Jurisprudenciais são textos semi-estruturados. Eles seguem um formato definido pela lei . Os requisitos formais são definidos pelo art. 381, do Código de Processo Penal (veja, Tabela 2). Os índices para o processo de recuperação da informação jurídica foram definidos conforme a importância de cada um desses requisitos para o processo de recuperação. A dogmática processual também costuma dividir os requisitos do art. 381, do CPP, em três partes: relatório ou histórico(incisos I e II), motivação ou fundamentação (incisos III e IV) e conclusão ou dispositivo (incisos V e VI). Nós usamos esta forma de divisão como base para definir o processo de extração do conhecimento através da localização do atributo.

Parte das informações relevantes de um texto jurisprudencial são descritas explicitamente no texto numa posição predefinida. Por exemplo, a data da publicação, sempre é descrita a primeira linha do acórdão precedida da palavra-chave “DATA”. Todas as informações juridicamente relevantes representadas nas estruturas 1 e 3 (veja Tabela 4) podem ser extraídas por sua localização no texto. A relevância destes índices foi determinada com base no conhecimento especialista e nos requisitos formais da sentença penal. Isso inclui a identificação da subestrutura específica e das palavras-chave pertinentes. Para cada índice um algoritmo de extração específico foi desenvolvido com base neste conhecimento de domínio. A seguir, descrevemos estes algoritmos:

- *Data da Publicação*: é descrito na primeira linha do documento, precedido da palavra-chave “DATA”(veja Figura 6, estrutura 1).
- *Tipo de petição*: é descrito na segunda linha do documento, é a expressão anterior a

chave "DATA"(veja Figura 6, estrutura 1).

- Tipo de petição: é descrito na segunda linha do documento, é a expressão anterior a palavra-chave "n."(veja Figura 6, estrutura 1).
- Número: é descrito na segunda linha do documento, é a expressão *posterior* a palavra-chave "n."(veja Figura 6, estrutura 1).
- Local: é descrito na segunda linha do documento, é a expressão *posterior* a palavra-chave "de"(veja Figura 6, estrutura 1).
- Relator: é descrito na terceira linha do documento, é a expressão *posterior* a palavra-chave "Relator:"(veja Figura 6, estrutura 1).
- Resultado: é descrito na estrutura 3, são as expressões que indicam *conceder* ou *denegar* o pedido e suas variantes, posterior a palavra-chave "ACÓRDAO,"(veja Figura 6, estrutura 3).

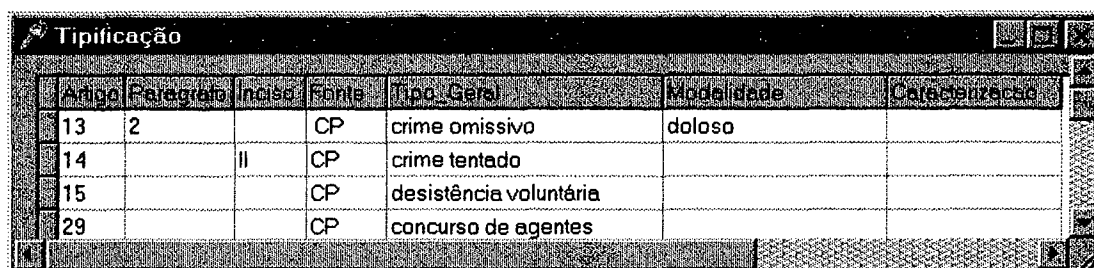
Estrutura 1.
DJ: 8.013 DATA: 17/05/90 PÁG: 08 Apelação criminal n. 20.824, de Urubici. Relator: Des. Emani Ribeiro.
Estrutura 2.
APELAÇÃO CRIMINAL. PEDIDO DE DESISTÊNCIA. HOMOLOGAÇÃO.
Estrutura 3.
Vistos, relatados e discutidos estes autos de apelação criminal n. 20.824, da comarca de Urubici, em que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

Figura 6. Exemplo de um Acórdão Criminal Estruturado e a Localização das informações relevantes.

4.2.2 Extração Através de Inferência

A determinação dos valores para o índice *Tipificação* requer a inferência de valores baseado em informações do texto jurisprudencial. O índice *Tipificação* é composto de um conjunto de índices (veja Tabela 3): *tipo geral*; *modalidade do crime*; *qualificação do crime*; *tentativa*; *co-autoria*. Para instanciar esses valores, as informações referentes a norma usada para a argumentação no documento jurídico são extraídas. A informação relevante é identificada no texto através de palavras-chave (por exemplo artigo, Art..) seguidos por um valor numérico. Para uma representação ser consistente e capaz de habilitar a inferência dos valores de índice relacionados às informações sobre a norma - que pode ser descrita diferentemente nos Acórdãos - foi delineado uma forma de representação de acordo com o Código Penal brasileiro. Com base nestas informações, os valores de índice são inferenciados. Por exemplo, se a informação extraída for “art. 121, § 2., IV c/c art. 14, ambos do Código Repressivo” (veja Figura 6, estrutura 4), os valores seguintes são inferenciados através do conhecimento de domínio representado na Tabela 3:

- *Tipo geral*: homicídio
- *Modalidade do crime*: doloso
- *Qualificação do crime*: qualificado
- *Tentativa*: sim
- *Co-autoria*: não



Artigo	Parágrafo	Inciso	Fonte	Tipo Geral	Modalidade	Caracterização
13	2		CP	crime omissivo	doloso	
14		II	CP	crime tentado		
15			CP	desistência voluntária		
29			CP	concurso de agentes		

Figura 7. Extrato do representação da legislação criminal usada para inferenciar valores do índice *Tipificação*.

4.2.3 Extração da Informação Através do Uso de Vocabulário Controlado

As informações relevantes num texto jurídico, que não podem ser identificadas unicamente pela sua localização no texto, são extraídas através do uso de um vocabulário controlado e um dicionário de termos normativos. Partes relevantes do Acórdão (veja Tabela 4, estrutura 2 e 4) são avaliadas para localizar expressões do vocabulário controlado. Sinônimos das *expressões indicativas* são identificados pela interação com o dicionário de termos jurídicos (veja item 4.1.3.2 *Dicionário de Termos Normativos*). A informação identificada e o índice *expressões indicativas* é instanciado com valores múltiplos. Por exemplo, no Acórdão exposto na Figura 5, três valores para este índice podem ser identificados: "*pedido de desistência do recurso*", "*homologação do pedido*", "*presentes os pressupostos*" (veja Tabela 3).

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

Tabela 5. Exemplo de um Acórdão Criminal (Jurisprudência) representado em forma de caso.

Capítulo 5 - Recuperação da informação Jurídica

5.1 Introdução

Para auxiliar o profissional do direito na busca pela decisão jurisprudencial pertinente ao seu problema jurídico, nós desenvolvemos um sistema inteligente para a recuperação destes textos legais. O protótipo do sistema mantém a representação e o armazenamento do conhecimento dos textos jurisprudenciais e recupera documentos com características semelhantes, baseado na informação inicial fornecida pelo usuário, em linguagem natural. Novos documentos são incluídos na base de conhecimento através da extração automática de informação relevante para o processo de recuperação. Para possibilitar o processamento do conhecimento informal em linguagem natural, um vocabulário controlado e um dicionário de termos normativos - baseado na terminologia jurídica usual são integrados na recuperação e no processo de extração. A teoria jurídica possibilitou uma análise mais científica do conteúdo dos Acórdãos Criminais. Baseado numa estrutura teórica, nós definimos índices e expressões para a recuperação destes textos de jurisprudências. Nossa abordagem foi desenvolvida e validada na área do direito criminal brasileiro, nos casos de Apelações Criminais, oriundos do Tribunal de Justiça de Santa Catarina ⁴³.

Após cada caso da base ser valorado, eles serão ordenados de forma decrescente de acordo com os valores obtidos com a métrica de similaridade, e os casos mais úteis, ou seja, com valores de similaridade mais altos, serão apresentados como o resultado da consulta. Os mais semelhantes são apresentados ao usuário que selecionar o mais apropriado na íntegra para melhor solução do seu problema jurídico. O processo de recuperação é dividido em:

⁴³ <http://www.tj.sc.gov.br>

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

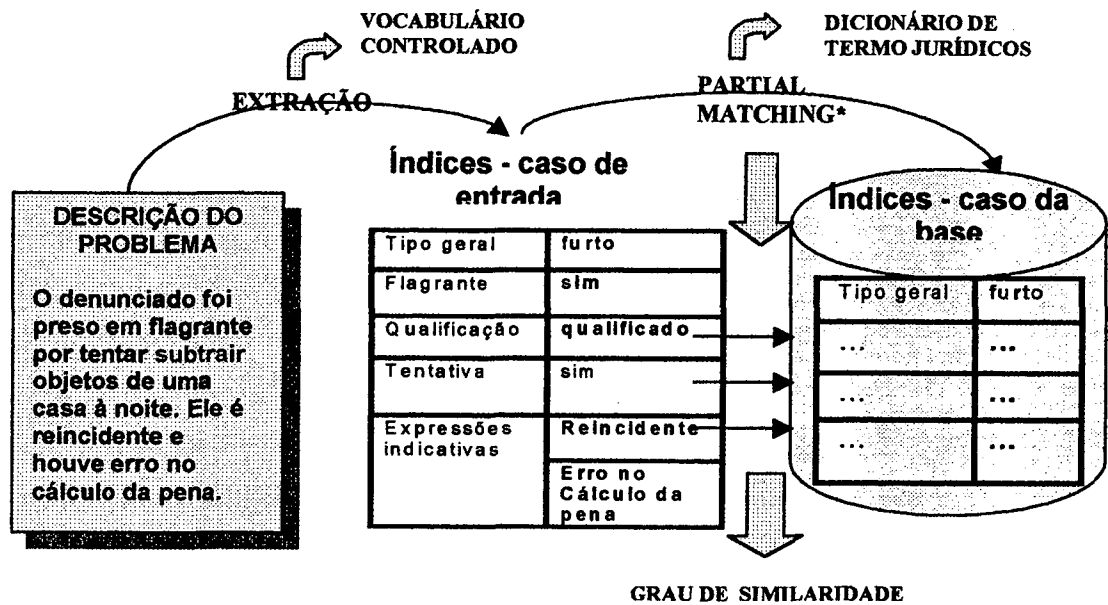


Figura 8. Arquitetura do Sistema

5.2 Ajuste da Situação

Um problema jurídico é descrito para iniciar o processo de recuperação. O ajuste da situação é realizado para identificar os valores atuais de cada índice na nova situação legal. Como o problema é descrito em linguagem natural, o sistema identifica as informações relevantes de uma forma automática através do uso do vocabulário controlado e do dicionário de termos normativos, também identifica as informações referentes a tipificação (tipo genérico, co-autoria, agravantes, etc.). Além disso, no ajuste da situação existem informações que o usuário pode escolher se fazem ou não parte da pesquisa, como o período da pesquisa, onde ele fornece a data inicial e a data final (e.g., de 10/10/1990 à 11/11/1997), o tipo de recurso (e.g., *Habeas Corpus*), o

relator (e.g., des. Moreira Alves) e o local (e.g., Ubirici), esses valores devem ser exatos, e são explícitos para o usuário.

Vejamos o seguinte exemplo:

"O acusado foi preso por cometer crime de homicídio. Depois de julgamento dele, o acusado foi submetido a exame de sanidade mental e foi verificado que ele estava louco no momento da ação criminal; o pedido é para anular o processo".

O advogado quer entrar com o recurso de *HABEAS CORPUS* para resolver este problema jurídico, logo ele informa na pesquisa a este tipo de recurso. Essa limitação não impede que o sistema recupere casos similares com tipos diferentes de recurso, mas com grau de similaridade inferior. Tudo vai depender do valor da similaridade definido no ajuste da situação.

Todas as evidencias sugerem que na consulta feita pelo usuário, a linguagem de indexação deve ser feita em linguagem natural,

Além disso, o ajuste da situação é apoiado, se necessário, pelo fornecimento de termos relacionados. Estes termos são inferenciados com base na descrição inicial do problema pelo usuário. Com base nessa descrição inicial, casos relevantes são identificados na base de casos e novos valores de índice são fornecidos como valores potenciais para o refinamento do ajusta da situação.

O objetivo desta abordagem é recuperar os casos mais úteis para a solução de um problema jurídico corrente. O problema é descrito pelo usuário e esta situação inicial é comparada com os casos na base de conhecimento. Como parte da comparação entre o caso da base de conhecimento e o problema fornecido o grau de similaridade é determinado pelo valor de similaridade e os mais similares são apresentados para o usuário, o qual pode explorá-los em detalhe e selecionar o mais apropriado para a solução do seu problema jurídico. Um outro exemplo, se o usuário fornecer o problema descrito a seguir:

“O acusado foi preso em flagrante. Ele estava tentando retirar algumas propriedades da casa da vítima à noite. Ele é um criminoso reincidente e o no seu julgamento ocorreu um erro no cálculo da pena.”

O sistema, então, extrairá as informações relevantes automaticamente da descrição em linguagem natural. Isto é realizado pela integração de um vocabulário jurídico controlado, constituído pela totalidade dos termos que integram o dicionário de termos normativos elaborado para um domínio jurídico particularizado. Com base nesta descrição, uma descrição formal é gerada pela definição dos índices para um caso específico. Esta descrição é parcialmente comparada com todos os casos na base de casos de maneira a identificar os mais similares. Para computar o valor de similaridade para cada caso, o dicionário de termos normativos é usado para indicar a similaridade entre termos individuais.

O próximo passo é definir quais são os casos mais úteis, dentre esses relevantes identificados no ajuste da situação. É o que veremos no próximo item.

5.3 Cálculo da Similaridade

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

Com base no ajuste da situação fornecido, os caso relevantes são recuperados de uma base de casos. Para identificar os casos mais úteis em relação ao problema inicial fornecido, o ajuste da situação é *comparado* parcialmente com cada caso na base de casos. A similaridade de cada índice do problema inicial com cada caso na base de casos é determinado pela *medida da similaridade local*. O grau total de

similaridade de um caso é calculado pela soma dos valores de similaridade local multiplicada por um fator de pesagem (veja, Figura 9).

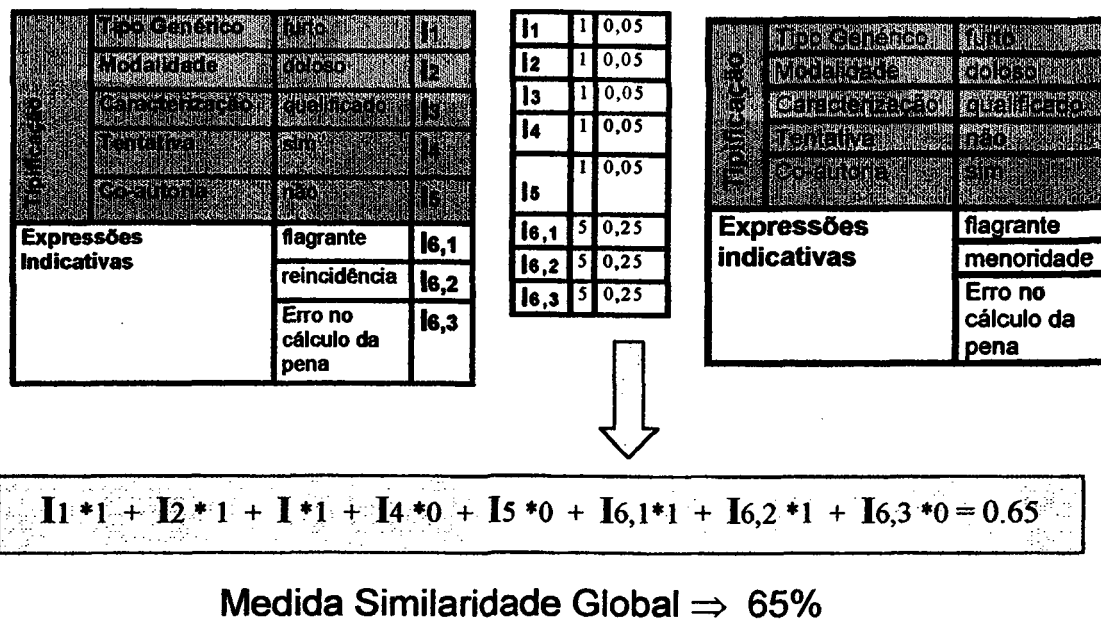


Figura 9. Cálculo da similaridade global.

5.3.1 Medida da Similaridade Global

A Medida de Similaridade Global é baseada no *vizinho mais próximo* :

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

Onde,

S = Ajuste da Situação

C_i = caso *i* da base de conhecimento

f_j = índice j

I_{sj} = índice j_{th} do ajuste da situação

I_{cij} = índice j_{th} do caso i

W_j = peso do índice j

O grau total de similaridade de um caso é calculado pela soma dos valores de similaridade local multiplicada por um fator de pesagem normalizado, por exemplo, os valores dos índices são normalizados com relação ao número total de valores deste índice. O índice *Expressões indicativas* é considerado multivalorado (mínimo 3 expressões, não existindo número limitado de valores), enquanto o índice *Tipificação* e seus sub-índices possuem um único valor. Assim, os valores dos sub-índices da *Tipificação* será sempre 1/5 do valor dos sub-índices da *Expressão indicativa*. No exemplo dado na Figura 9, a similaridade global terá o valor de 65% (sessenta e cinco por cento).

5.3.2 Medida da Similaridade Local

Com base no ajuste da situação fornecido, os caso relevantes são recuperados de uma base de casos. Para identificar os casos mais úteis em relação ao problema inicial fornecido, o ajuste da situação é *comparado* parcialmente com cada caso na base de casos. A similaridade de cada índice do problema inicial com cada caso na base de casos é determinado pela *medida da similaridade local*. O índice tipificação é composto por outros índices: *tipo geral* (f_1), *modalidade do crime* (f_2), *qualificação do crime* (f_3), *tentativa* (f_4), e *co-autoria*, (f_5). No exemplo da Figura 9, o *Tipo Geral* é o mesmo (furto), e por esta razão, o valor da similaridade local deste subíndice é 1, multiplicado por um fator de normalização. Ao contrário, no sub-índice *Tentativa* que possui valores diferenciados, o valor da similaridade local é zero, multiplicado por um fator de normalização.

Além disso, no sub-índice *Tipo geral* (f_1) a determinação do valor poderá ser auxiliada pelo dicionário de termos normativos. Por exemplo, se na descrição do ajuste da situação o *Tipo Geral* é descrito como " assassinato " e o *Tipo Geral* declarados no caso de é " homicídio " (um sinônimo para assassinato), então pelo uso do dicionário de termos normativos estas duas condições são consideradas como iguais. Sem o dicionário, a similaridade entre os dois valores poderia não ser considerada. Se isso ocorrer a similaridade entre os casos será considerada menor. Então, o valor um (1) é definido para o valor de índice do caso que é igual ao fornecido no ajuste da situação e também para o valor de índice que for sinônimo no dicionário de termos jurídicos, se não for igual o valor da similaridade é zero (0). A fórmula fica assim determinada:

$$f_1(gt_s, gt_{ci}) = \begin{cases} 1 & (gt_s = gt_{ci}) \vee (gt_s n gt_{ci}) \\ 0 & else \end{cases}$$

onde,

gt_s = tipo genérico do ajuste da situação

gt_{ci} = tipo genérico do caso

n = sinônimo do dicionário

O índice *expressões indicativas* é representado por um conjunto de valores (todos os valores encontrados com base no vocabulário controlado serão representados). Aqui, a similaridade local é calculada pela comparação de cada valor fornecido caso da entrada com cada valor do caso da base de casos, deste modo, é determinando o valor de similaridade máximo. Então, a soma dos valores de similaridade máximos é calculada e *dividida* pelo número de valores fornecidos no ajuste da situação para normalizar o valor de similaridade local deste índice. Pela integração do dicionário de termos normativos, como acontece no índice f_1 , é possível considerar ainda, os sinônimos no cálculo da similaridade local.

Assim, no cálculo da similaridade local do índice *expressões indicativas* (f_6), onde há multiplicidade de valores, a fórmula fica assim definida:

$$f_6(\{ie_{s1}, \dots, ie_{sn}\}, \{ie_{c1}, \dots, ie_{cn}\})$$

$$= \frac{\sum_k \max\{n(ie_{sk}, ie_{cil}) \mid ie_{cil} \in \{ie_{c1}, \dots, ie_{cn}\}\}}{|\{ie_{s1}, \dots, ie_{sn}\}|}$$

5.4 Seleção do melhor caso

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

Capítulo 6 - Trabalhos Relacionados

6.1 Introdução

Modelar textos jurídicos de um modo compreensível foi o objetivo de muitos trabalhos de pesquisa. Atualmente, na prática, existem uma grande variedade de sistemas de administração de banco de dados (e.g., Informa Jurídico)⁴⁴, e sistemas de recuperação de informação na Internet (e.g., Senado Federal, Tribunais Federais, e Tribunais Estaduais)^{45 46 47}. Porém, esses sistemas recuperam frequentemente um número grande de casos irrelevantes, considerando que casos importantes e similares à situação de interesse podem não ser achados. A recuperação de casos é limitada a um jogo de palavras-chave comuns e combinações destas palavras, sem considerar semelhança de semântica entre os termos e diferenças de importância dos índices para a determinação de casos relevantes.

Neste contexto, o RBC oferece um potencial significativo para a recuperação inteligente de documentos jurisprudenciais. Seus principais benefícios são o enfoque no conhecimento em forma de episódios individuais, em lugar de conhecimento de domínio genérico e, a recuperação baseada na similaridade. No entanto, na área jurídica, poucos são os sistemas baseados em casos visando a recuperação da informação. O Projeto Prudentia⁴⁸, que é um sistema para a recuperação da informação aplicado ao domínio jurídico.

⁴⁴ Informa Jurídico: jurisprudência e legislação. Pro-link software. CD-1. 1996

⁴⁵ Senado Federal. Avaliável Online. [Http://www.senado.gov.br](http://www.senado.gov.br). [20 de outubro de 1998].

⁴⁶ Supremo Tribunal Federal. Avaliável Online. [Http://www.stf.gov.br](http://www.stf.gov.br). [10 de novembro de 1998].

⁴⁷ Tribunal De Justiça de Santa Catarina. Avaliável Online. [Http://www.tj.sc.gov.br](http://www.tj.sc.gov.br). [15 de novembro de 1998].

⁴⁸ Weber, Rosina. Intelligent jurisprudence research, p. 104.

Os sistemas de RBC também são aplicados para a recuperação da informação em diferentes áreas, como o Projeto FallQ⁴⁹.

6.2 Projeto Prudentia

A nossa abordagem é orientada, especificamente, para a recuperação de documentos jurídicos em amplas bases textuais. Para tornar isso possível, nós desenvolvemos a metodologia de extração automática da informação, tomando como referência o trabalho desenvolvido no Sistema Prudentia. No Prudentia a extração automática da informação jurídica fornece uma lista de importante expressões que podem efetivamente representar os casos. O Prudentia é um sistema baseado em casos que auxilia os profissionais do direito a utilizar as fontes de pesquisa jurídica (Jurisprudência) de uma forma mais eficiente, através da recuperação de informações numa ampla e distribuída base de dados. No Prudentia o usuário descreve uma questão jurídica em linguagem natural como base da pesquisa. O sistema, então, traduz automaticamente a descrição literal num " caso problema" e pesquisa, na base de casos, por acórdãos com características similares. Estes acórdãos poderão ser vistos na íntegra e, posteriormente, utilizados como referência na solução de um conflito jurídico. Esta técnica permite uma recuperação de documentos textos com características semelhantes, baseando-se nas informações fornecidas pelo usuário em linguagem natural. O conhecimento de domínio, empregado na construção do sistema, permite a extração automática de índices dos documentos, tornando a busca mais rápida e a interface amigável.

A nossa abordagem está baseada no refinamento da representação de casos apresentada no Prudentia melhorada por uma maior integração do conhecimento de domínio(teoria jurídica), bem como o uso de um vocabulário controlado e um dicionário de termos normativos, definidos com base numa estrutura jurídico-teórica. Além disso, a nossa abordagem inclui também - além da medida de

⁴⁹ Lenz, M.; Burkhard, H.-D. CBR for Document Retrieval: The FallQ Project, p. 84.

semelhança global - o cálculo da similaridade local estimada para cada índice. Isto melhorou consideravelmente a recuperação, de uma forma efetiva e eficiente.

6.3 Projeto FallQ

O projeto FallQ utiliza a tecnologia de RBC para a recuperação de documentos. O objetivo do FallQ é fornecer suporte para o *Hotline staff*, respondendo em linguagem natural, as perguntas mais freqüentes. A estrutura dos casos no FallQ, foi definida de acordo com o documentos fornecidos. Os documentos são divididos em três estruturas. Um lista de palavras-chave podem ser extraídas automaticamente das *Questões mais freqüentes* (consideradas documentos). Cada caso é representado por um tipo de palavra-chave que classifica a informação .

O processo de recuperação e a estrutura de representação utilizados no FallQ são muito semelhantes aos aplicados neste trabalho. No entanto, além da área de aplicação ser muito diversa, a nossa lista de expressões (vocabulário controlado) e o thesaurus não foram construídos com base nos documentos jurisprudenciais, eles foram construídos através das listas de índices dos próprios Tribunais, das listas de indexação das leis e do conhecimento especialista.

O FallQ pode ser utilizado com base para trabalhos futuros, com relação ao tratamento gramatical para definir a similaridade local.

Capítulo 7 – Conclusões

7.1 Sistemas Baseados no Conhecimento Jurídico

A enorme quantidade de decisões produzidas pelos Tribunais, torna o domínio jurídico um campo fértil para o desenvolvimento de sistemas computacionais para a recuperação de dados. Para tanto, é necessário o emprego de técnicas de inteligência artificial para fornecer suporte adequado a pesquisa de acórdãos com características que respondam às necessidades do usuário, isto é, o sistema deve fornecer os acórdãos com características mais semelhantes a questão que originou a pesquisa jurídica.

Neste papel nós apresentamos uma aproximação baseada em casos para a recuperação de documentos de texto de bases de texto de jurisprudências criminais. A principal discussão no domínio jurídico é a interpretação da linguagem jurídica. Inúmeras teorias são formuladas para definir qual a verdadeira intenção do legislador. Nossa abordagem encontrou na teoria da Argumentação Jurídica parâmetros para definir qual a informação inserida numa jurisprudência é verdadeiramente relevante para a recuperação destes textos jurídicos pertinentes ao problema fornecido inicialmente pelo usuário. Com base numa estrutura teórica, os textos foram divididos em razão do seu conteúdo formal para melhor localização das informações e um vocabulário controlado e um dicionário de termos jurídicos foram construídos para apoiar a representação textos jurídicos e facilitar a interpretação de seu conteúdo. Nossa aproximação foi aplicada e validada a área da lei criminal brasileira. Atualmente nós estamos implementando um protótipo que inclui 3400 casos de Apelações Criminais que originam de do Tribunal de Lei de Santa Catarina.

A recuperação de jurisprudências, feita com qualidade, pode fornecer meios eficazes para que o julgador possa conciliar retoricamente os ditames de sua consciência com a expectativa das partes e dos órgãos judiciais superiores. Os

advogados e promotores, por sua vez, poderão contar com mais um instrumento de orientação no sentido de selecionar os argumentos que colocarão nas mãos dos juízes com vista a suas pretensões.

A necessidade de atenção à produção de ferramentas, enfatizando que tal atividade, que gerará novos métodos e técnicas de armazenamento e manipulação de informações, embora não seja diretamente ligada à ciência jurídica, vai provocar fortes reflexos sobre o direito e o seu funcionamento e, acreditamos, sobre o verdadeiro sentido de Justiça.

7.1 Desenvolvimentos futuros

Este trabalho está sendo desenvolvido na área do direito criminal brasileiro. A intenção é abranger também a área cível. Entretanto, o direito é um domínio muito amplo e mesmo na área Criminal é necessário aprimorar alguns aspectos da presente abordagem, tais como:

- A similaridade local pode ser aprimorada, através do tratamento gramatical dos valores dos índices, como o usado no projeto FallQ;
- Inclusão de outros tipos genéricos na lista do índice Tipificação. O presente trabalho inclui somente os tipos genéricos do Código Penal Brasileiro, existem ainda outras legislações que, embora versem sobre assuntos diferenciados da questão criminal, tratam de crimes especiais. São algumas delas: *Estatuto da Criança e do Adolescente, Código de Trânsito, Código Brasileiro de Telecomunicações, Código do Consumidor, Legislações ambientais, etc.*
- Atualização da lista de *expressões indicativas* do vocabulário controlado, para outros Tribunais, principalmente os Federais.

Estas modificações, embora exijam trabalho e dispêndio de tempo, tornarão o sistema apto para a recuperação de jurisprudência em qualquer instância da justiça brasileira (veja Figura 10).

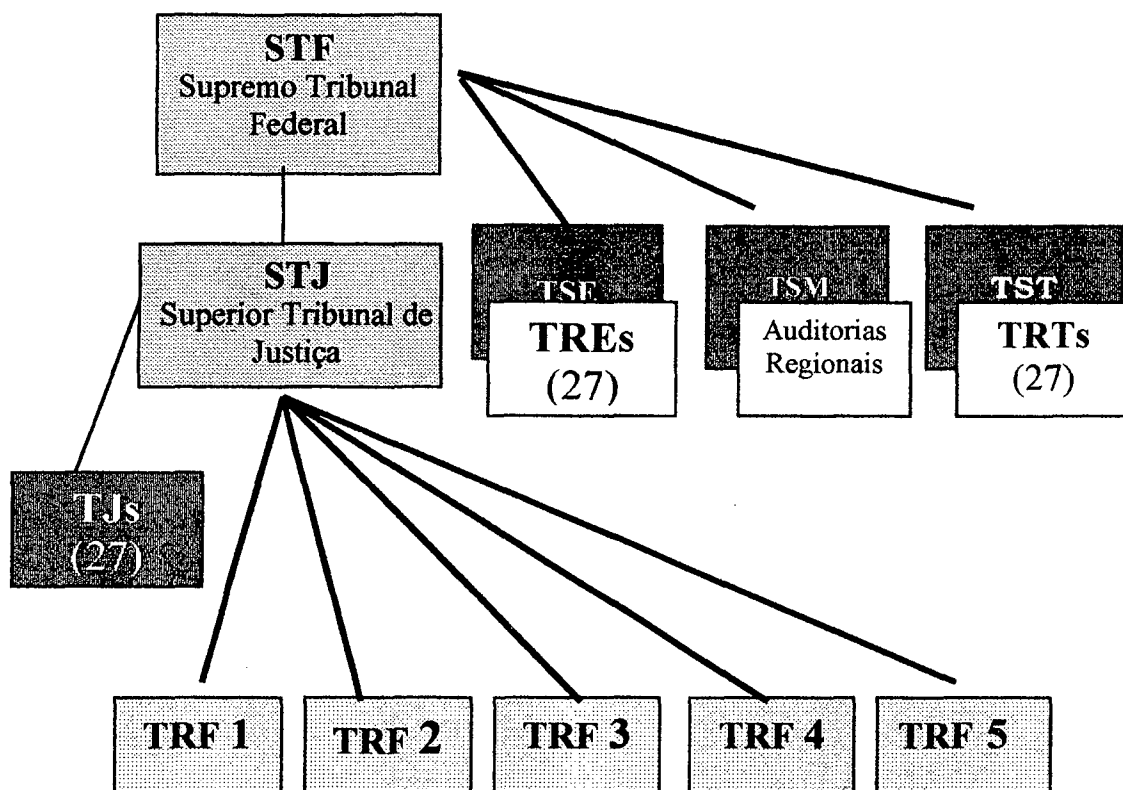


Figura 10. Organograma dos Tribunais Brasileiros.

8 Referências

- AMONDT, A., Plaza, E. (1994). Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches. *AI Communications*, 17(1), 1994.
- ASHLEY, Kevin D. (1988). Arguing by analogy in law: A case-based model. *Analogical reasoning: Perspectives of artificial intelligence, cognitive science, and philosophy*, ed. D. Hehnan. Boston: Kluwer.
- BARR, Avron & FEIGENBAUM, Edward. *The handbook of Artificial Intelligence*. Heuristech Press. 1981.
- BENCH-CAPON, Trevor. *Argument in Artificial Intelligence and Law*. *Jurix*. 1995.
- BOBBIO, Norberto. *Teoria do Ordenamento Jurídico*. 4^a ed., São Paulo, 1994.
- BRUM, Nilo B.de. *Requisitos Retóricos da Sentença Penal*. São Paulo: Ed. Revista dos Tribunais, 1980.
- BUENO, Tânia C. D'Agostini. *Recuperação Inteligente de Textos Legais*. Uma experiência na área da Inteligência Artificial aplicada ao Direito através da técnica de Raciocínio Baseado em Casos. Monografia. Universidade do Vale do Itajaí, Campus IV, Biguaçu. 1997.
- BUENO, Tânia C. D. *Retrieval in Jurisprudential Text Bases using Juridical Terminology*. Artigo aprovado na VII Conferência Internacional de Inteligência Artificial e Direito – ICAIL99. 14-17 de Junho de 1999, Oslo, Noruega.
- BUENO, Tânia C. D. *Uso da Teoria Jurídica para Recuperação em Amplas Bases de*

Textos jurídicos. Artigo aprovado no II Encontro Nacional de Inteligência Artificial – ENIA99, a ser realizado no Rio de Janeiro, de 19 a 21 de Julho de 1999.

DELMANTO, C. *Código Penal Comentado*. 3.ed. ampl. E atualizada.- Rio de Janeiro: Renovar. 1991.

HOESCHL, Hugo. O Ciberespaço e o Direito. *Revista Trimestral de Jurisprudência dos Estados*. Vol. 167. São Paulo:Ed. EJV, 1998.

INFORMA JURÍDICO: jurisprudência e legislação. Pro-link software. CD-1. 1996

JESUS, Damásio E. *Código de Processo Penal anotado*. 8. ed. ,atual. e aum. – São Paulo : Saraiva, 1989.

KOLODNER, Janet L. *Case-based reasoning*. Morgan Kaufmann Publishers, Inc. 1993. San Marco, CA94403.

LENZ M.; Hübner A.; Kunze M. Textual CBR. M. Lens, B. Boutsh-Sporl, H.-D. Burkhard, S. Wess (eds.). *Case- Based Reasoning Technology*. Springer Verlag. 1998.

LENZ, M.; BURKHARD, H.-D.. CBR for Document Retrieval: The FAILQ Project. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*. David Leake, Enric Plaza (ed.)-Berlin: Springer Verlag, 1997.

LEWIS, David, JONES, Karen S. Natural language Processing for Information Retrieval. *Communications of the ACM*. Vol. 39. N.I. January 1996.

RABUSKE, Renato Antônio. *Inteligência Artificial*. Florianópolis, Ed. da UFSC, 1995.

REALE, Miguel. *Fontes e Modelos de Direito*. São Paulo: Saraiva, 1994.

RISLAND, Edwina L.; DANIELS, Jody. Using CBR to drive IR. *AAAI*. Published in IJCAI-95. pg. 400-407, Montreal, Canada. 1996.

- SENADO FEDERAL .Available Online. [Http://www.senado.gov.br](http://www.senado.gov.br). [October, 20,1998].
- STREAK, L. L. *Súmulas no Direito Brasileiro: eficácia, poder e função*. Porto Alegre: Livraria do Advogado,1995.
- SUPREMO TRIBUNAL FEDERAL. Available Online. <http://www.stf.gov.br>. [November, 10,1998].
- TRIBUNAL DE JUSTIÇA DE SANTA CATARINA. Available Online. <Http://www.tj.sc.gov.br>. [November, 15,1998].
- UYTTENDAELE, Caroline, MOENS, Marie-Francine and 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) and Omsa (Tokyo).
- WARAT, L.A. *Interpretação da Lei.: temas para uma reformulação*. Porto Alegre: Sergio Antonio Fabris Editor, 1994.
- WARAT, Luis Alberto. *O Direito e sua Linguagem*. 2ª versão. 2ª ed. Sergio Antonio Fabris Editor. Porto Alegre. 1995.
- WEBER, R.; 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, R.; Martins, A., Barcia, R. . On Texts and Legal Cases. Accepted for presentation at the *AAAI-98 Workshop on Textual Case-Based Reasoning*, 1998.
- WEBER, Rosina. Intelligent jurisprudence research. Ph.D. Thesis, Graduate Program of Production Engineering at the Federal University of Santa Catarina, 1998.

WEBER, Rosina; Martins, Alejandro; Eduardo Mattos; Tania Bueno; Hugo Hoeschl; Roberto Pacheco;; Ricardo Barcia. (1998). Reusing Cases to the Automatic Index Assignment from Textual Documents. *6th German Workshop On Case-Based Reasoning - Foundations, Systems, and Applications -Berlin, March 6-8, 1998.*

WEBER-LEE, R.; Barcia, R.; Costa, M.; Rodrigues Filho, I.; Hoeschl, H. C.; Bueno, T.; Martins, A, and Pacheco, R. (1997a). A Large Case-Based Reasoner for Legal Cases. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97.* David Leake, Enric Plaza (ed.)-Berlin: Springer.

Anexo 1

Este trabalho foi submetido e aprovado para apresentação na VII Conferência Internacional de Inteligência Artificial e Direito – ICAIL99. 14-17 de Junho de 1999, Oslo, Noruega.

JurisConsulto: Retrieval in Jurisprudencial Text Bases using Juridical Terminology

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

Abstract

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

1. Introduction

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

However, the number of jurisprudencial cases has grown largely, complicating the efficient access to this type of information. Despite the availability of those juridical texts in databases (such as, e.g., Informa Jurídico [6]), or in the Internet (e.g., Federal Tribunals [13] and State Tribunals [14], or public organs [11]), the difficulty remains. The search in those systems requires juridical knowledge and is limited due to problems with syntactic and semantic ambiguity and existing uncertainties in the textual document [8]. In practice, these database and web-based technologies often retrieve a great number of irrelevant information and require a repeated reformulation of the search to achieve a satisfying result. Therefore, the development of efficient and effective systems for the retrieval of juridical information using techniques in the area of *Artificial Intelligence* (AI) is important for all juridical professionals and yields significant benefits for the solution of legal problems.

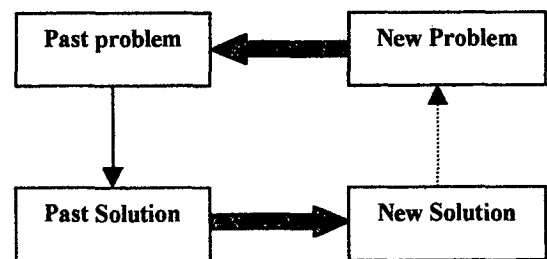


Figure 1. Process of analogy

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

interpret the law. For example, judges use this type of reasoning when interpreting a new case through its comparison with previous cases as a basis for decision.

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

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

2. Application Domain: Jurisprudential Research

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

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

According to Bench-Capon [2], jurisprudence is an excellent form to study argument. *Juridical argumentation theory* [16] provides the possibility for the analysis of the meaning of juridical norms and their individual meaning of justice. The

theory can help a judge to select arguments based on the Law, but considering different aspects (moral, political, social, personal, etc.). As a consequence, the Law is redefined in conformance with the Judge's interpretation, when the Judge decides a case and determines the application of a specific Law for the particular case. Those different possible interpretations of the Law are the main reason for the existence of divergent decisions. Juridical professionals can use the redefinition of the normative terms¹ as described in the jurisprudence for the construction and reinforcement of arguments, which facilitates the application of the Law to situations not considered before. Therefore, jurisprudence also plays an important role in the Brazilian Right.

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

3. Intelligent Support for Jurisprudential Search

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

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

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

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

¹ Normative terms represent juridical facts [16].

3.1. Representation of Legal Cases

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

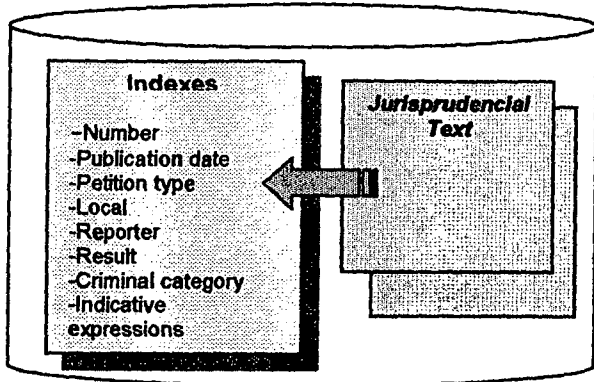


Figure 2. Representation of legal document.

The usage of indexes during the retrieval process allows an efficient access to the text documents by referring to their

content. Besides, the jurisprudential text cases, general domain knowledge is included in form of a controlled juridical vocabulary and a juridical thesaurus. Definitions of the content of norms are represented in form of rules. This domain knowledge supports the retrieval and extraction processes by identifying relevant terms in juridical documents given in natural language and explicitly modeling the similarity of juridical terms.

3.1.1. Retrieval Indexes

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

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

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

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

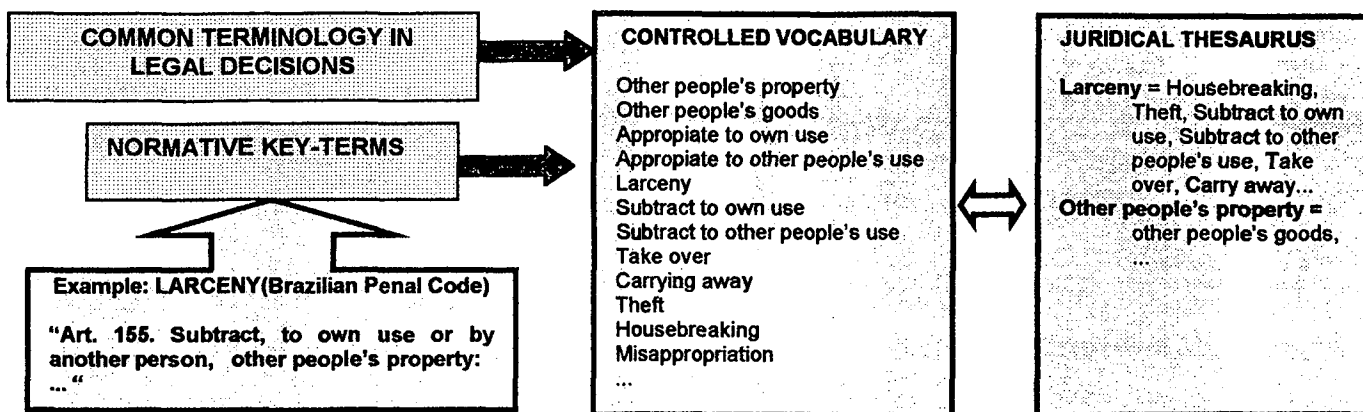


Figure 3. Controlled Vocabulary and Juridical Thesaurus.

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

3.1.2. Legal Domain Knowledge

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

3.1.2.1 Controlled Vocabulary

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

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

3.1.2.2. Juridical Thesaurus

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

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

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

3.2. Automatic Extraction of Textual Information

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

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

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

Table 2. Structure of a jurisprudencial document.

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

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

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

3.2.1. Extraction by Localization

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

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

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

3.2.2. Extraction by Inference

Information concerning some indexes, e.g., *Criminal Category*, is not explicitly stated in the legal documents, and can, therefore, not directly extracted from the texts. Those values need to be inferred based on information extracted from the legal text.

The index *Criminal Category* is composed of a set of indexes (see Table 1): *general term; crime modality; crime*

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

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

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

3.2.3. Extraction by Controlled Vocabulary

Information on the juridical argumentation in the petition is extracted by the usage of a controlled vocabulary. Relevant parts of the petition (see table 2, structure 2 and 4) are searched for expressions of the controlled vocabulary. Synonyms of relevant expressions are identified through the interaction of the juridical thesaurus. The information identified is extracted and the index *indicative expressions* is instantiated with the identified values. For example, from the petition shown in Figure 4, three indicative expressions can be extracted: "*Desistence request*" "*Ratification of request*" and "*Lawsuit conditions*" (see Table 3). As a consequence, the index *indicative expressions* can be instantiated with multiple values.

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

Figure 4. Localization of index values in the document.

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

Table 3. Example of a case representation

3.3. Retrieval of juridical cases

The objective of the system is to retrieve the most useful cases for the solution of a current juridical problem. The present problem is described by the user and the given situation assessment is compared with the cases in the knowledge base. As part of the comparison, the degree of similarity between each

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

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

3.3.1. Situation Assessment

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

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

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

3.3.2. Matching

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

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

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

Where,

S = situation assessment

C_i = case i of knowledge base

I_{sj} = j th index of situation assessment

I_{cij} = j th index of case i

W_j = weight of index j

f_j = local similarity measure for index j

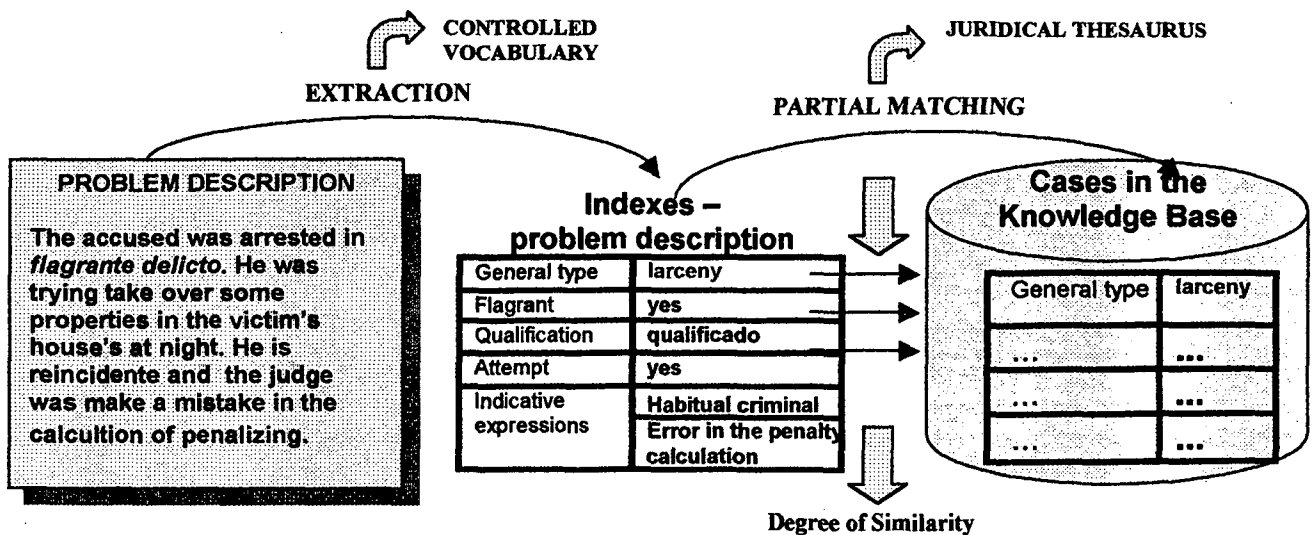


Figure 5. Retrieval Process.

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

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

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

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

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

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

3.3.3. Selection of the Most Useful Case

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

4. Discussion and Related Work

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

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

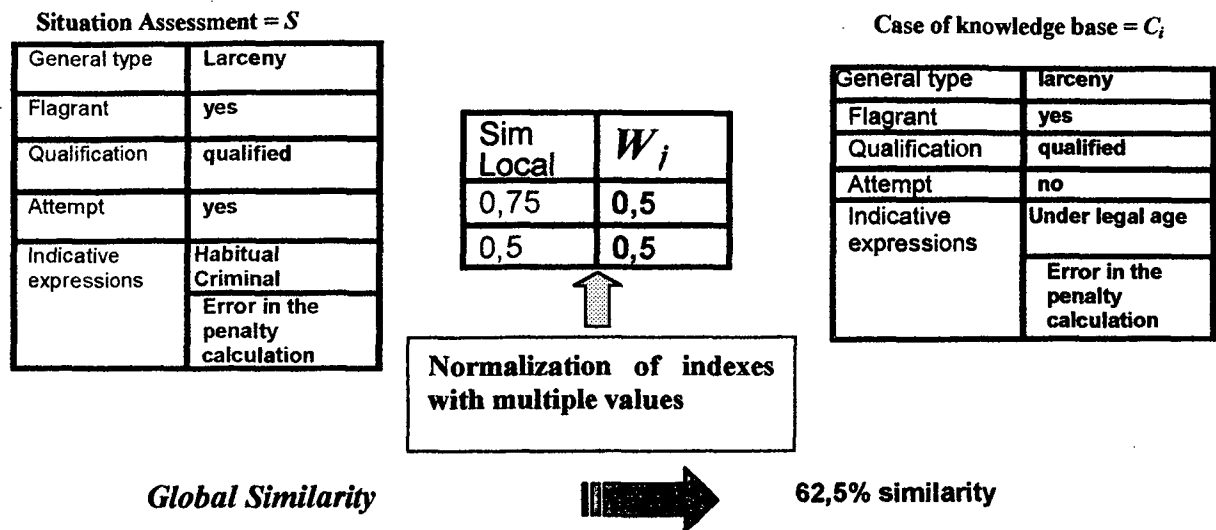


Figure 6. Calculation of similarity.

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

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

5. Conclusion

In this paper, we present a case-based approach for the retrieval of text documents from jurisprudential text bases. A major problem and issue of discussions in the juridical domain is the interpretation of the juridical language. Our approach is based on the theory of the juridical argument with defines important and discriminate information for the retrieval of relevant juridical documents. Based on the theoretical framework, a general legal domain knowledge has been modeled and

represented, such as, e.g., a controlled vocabulary and a juridical thesaurus, to support the representation of juridical documents and to facilitate the interpretation of its content.

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

Acknowledgements

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

References

- [1] Amond, A., Plaza, E. Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches. *AI Communications*, 17(1), 1994.
- [2] Bench-Capon, Trevor. Argument in Artificial Intelligence and Law. *Jurix*. 1995.
- [3] Delmanto, C. *Código Penal comentado*. 3.ed. ampl. E atualizada. - Rio de Janeiro: Renovar. 1991.
- [4] Jesus, Damásio E. de. *Código de Processo Penal anotado*. 8. ed. ,atual. e aum. - São Paulo : Saraiva, 1989.
- [5] Kolodner, J. *Case-Based Reasoning*. Morgan Kaufmann, Los Altos, CA. 1993.

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

Anexo 2

Este trabalho foi submetido e aprovado para ser apresentado no II Encontro Nacional de Inteligência Artificial – ENIA99, a ser realizado no Rio de Janeiro, de 19 a 21 de Julho de 1999.

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

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

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

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

1 Introdução

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

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

Nós utilizamos o Raciocínio Baseado em Casos (RBC) [Amondts-Plaza (1994)][Kolodner, (1993)], uma poderosa tecnologia da Inteligência Artificial para o desenvolvimento da presente abordagem [Rissland—Daniels, (1995)]. O RBC usa experiências anteriores e semelhantes para a solução de um problema novo. O RBC é baseado no princípio de analogia, assumindo que problemas semelhantes tem soluções semelhantes. Por esta razão, o RBC é uma técnica muito adequada ao domínio jurídico, pois utiliza o mesmo tipo de raciocínio

utilizado pelo juristas na solução de um problema [Bueno (1999)]. Para aplicar a técnica de RBC na recuperação de textos de jurisprudência, o domínio de aplicação (v.g., Acórdãos Criminais) precisa ser modelado de forma que características relevantes dos documentos sejam identificadas, assim como, a terminologia usada no domínio definida. Aqui, a teoria de argumentação jurídica [Warat, (1994)] ofereceu apoio teórico pela análise do significado do caso legal. Isso possibilita o processamento do conhecimento informal nas pesquisas por casos apropriados na base de casos em linguagem natural. A construção de um vocabulário controlado e de um dicionário jurídico permite o processamento dessa informação.

Com base nesta estrutura teórica, nós definimos a representação, o processo de recuperação e a extração automática dos atributos dos novos documentos a serem incluídos na base de casos.

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

2 Apoio Inteligente Para Recuperação em Amplas Bases de Textos Jurídicos

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

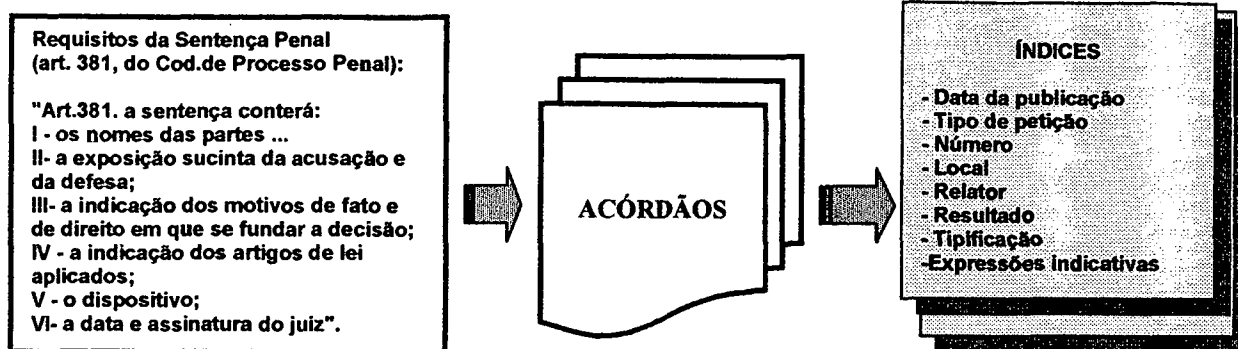


Figura 2. Representação de um Documento Jurisprudencial em Caso

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

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

2.1 Representação dos casos jurídicos

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

Pela análise dos textos jurídicos foi identificado um formato padrão. A jurisprudência é estruturada de acordo com o tipo de informação fornecida [Weber et al. (1997)], em 4 subestruturas básicas (veja tabela 1). Para reforçar esta forma de representação, o conhecimento do domínio é incluído em forma de um *vocabulário jurídico controlado* e um *dicionário de termos*. É este conhecimento de domínio que permite a recuperação dos documentos e o processo de extração automático, através da identificação de expressões indicativas e relevantes dos

textos jurídicos em linguagem natural, juntamente com a modelagem explícita da semelhança destes termos jurídicos.

2.1.1 Índices para a recuperação

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

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

No contexto da pesquisa jurisprudencial, o problema principal é a determinação dos índices e os valores deles usados para a recuperação. Como o uso de casos jurídicos está baseado no princípio de analogia, os atributos dos documentos de texto usados como índices para a recuperação têm que indicar a utilidade da solução do caso na situação presente. Na nossa abordagem, os índices são determinados pelos requisitos formais [Jesus, (1989)] e retóricos da sentença penal [Warat, (1995)] em conformidade com a importância deles para o contexto de recuperação da informação jurídica. Nós usamos esta forma de divisão como base para definir o processo de extração do conhecimento através da localização do atributo (veja item 2.2).

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

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

2.1.2 Conhecimento do Domínio Jurídico

No mundo jurídico, o profissional do direito, deve buscar argumentos persuasivos na solução de um caso concreto. Por esta razão, a jurisprudência é um lugar excelente para estudar o argumento jurídico [Bench-Capon (1995)]. Assim, para uma efetiva representação e recuperação destes casos jurídicos, é necessário considerar que um termo normativo pode ter várias interpretações nas decisões judiciais e todas indicam argumentos relevantes para a solução de problemas jurídicos. Do ponto de vista de uma estrita legalidade, um caso jurídico pode ser naturalmente incluído em uma classe normativa; no entanto, aos olhos do julgador, essa inclusão pode parecer totalmente injusta. Normalmente, quando isso acontece, o juiz produz uma definição explicativa de caráter ideológico, apresentando como relevantes outras conotações para caracterização de algum termo-chave incluído na norma, e decide com isto uma mudança de linha de soluções. É por esta razão que somente uma lista com os termos-chave normativos não é suficiente para

representar as informações juridicamente relevantes de um caso.

A Teoria da Argumentação Jurídica [Warat, (1994)] fornece a possibilidade de uma análise do significado da norma jurídica e seu significado individual de justiça. Segundo essa teoria, nenhum discurso jurídico será considerado significativamente legítimo se contradizer as formas axiológicas predominantes. Os argumentos são persuasivos no direito, quando não contradizem à ideologia de seus destinatários, isto é, a ideologia dos Tribunais.

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

2.1.2.1 Vocabulário controlado

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

Tabela 2. Estrutura de um caso jurídico

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

2.1.2.2 Dicionário de Termos Normativos

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

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

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

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

2.2. Extração automática da informação textual

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

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

Através de Inferência: A determinação dos valores para o índice *Tipificação* requer a inferência de valores com base no Código Penal [Delmanto, (1991)], baseado nas informações do texto jurisprudencial. O índice *Tipificação* é composto de um subconjunto de índices: *tipo geral; modalidade do crime; qualificação do crime; tentativa; co-autoria*. Por exemplo, se a informação extraída do texto for "art. 121, § 2., IV c/c art. 14, ambos do Código Repressivo", os valores são automaticamente inferenciados pelo o sistema – pelo uso de conhecimento de domínio em forma de regras.

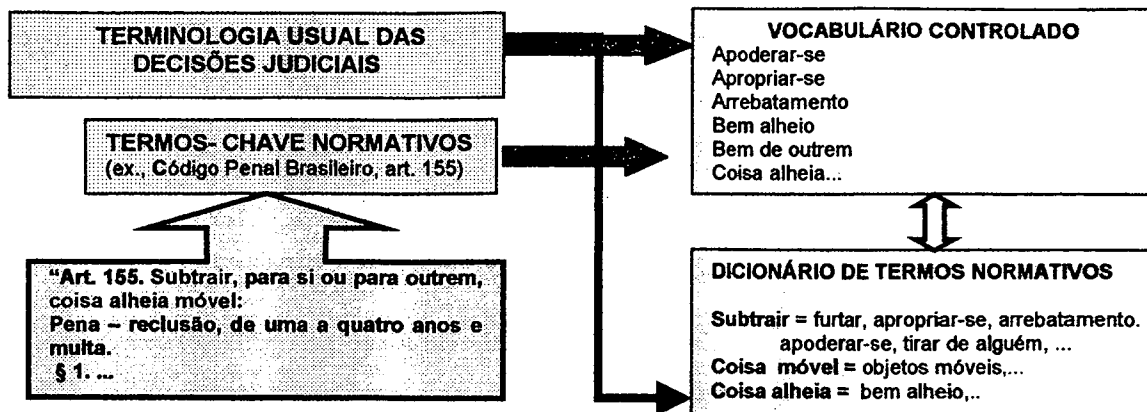


Figura 3. Construção do Vocabulário Controlado e do Dicionário de Termos normativos.

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

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

Tabela 3. Exemplo de um caso representado.

3 Recuperação de Casos Jurídicos

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

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

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

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

3.1 Ajuste da Situação

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

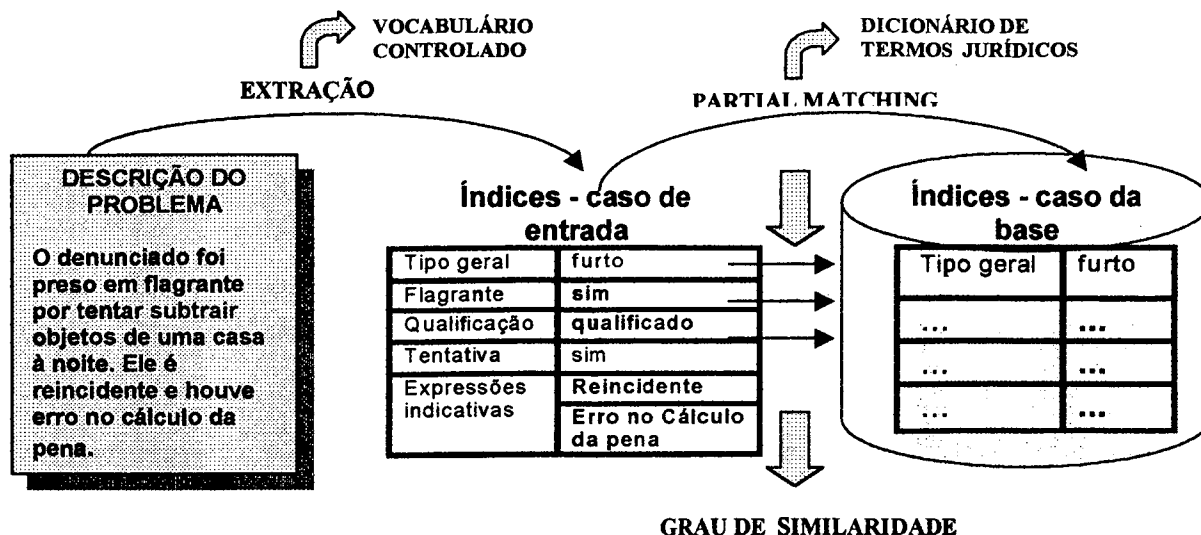


Figura 4. Processo de Recuperação.

3.2 Cálculo da Similaridade

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

3.2.1 Medida da Similaridade Global

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

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

Onde,

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

3.2.2 Medida da Similaridade Local

A similaridade entre cada índice da situação atual com cada caso na base de casos é determinado pelas medidas locais de similaridade. O valor da similaridade local para os índices: local, tipo de recurso, relator, resultado e tipificação é calculado através da seguinte fórmula:

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

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

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

Referências bibliográficas

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