

**TÂNIA CRISTINA D'AGOSTINI BUENO**

**ENGENHARIA DA MENTE:  
UMA METODOLOGIA DE REPRESENTAÇÃO DO CONHECIMENTO  
PARA CONSTRUÇÃO DE ONTOLOGIAS EM SISTEMAS BASEADOS  
EM CONHECIMENTO**

**TESE DE DOUTORADO**

**FLORIANÓPOLIS**

**2005**

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

**TÂNIA CRISTINA D'AGOSTINI BUENO**

**ENGENHARIA DA MENTE:**  
**UMA METODOLOGIA DE REPRESENTAÇÃO DO CONHECIMENTO**  
**PARA CONSTRUÇÃO DE ONTOLOGIAS EM SISTEMAS BASEADOS**  
**EM CONHECIMENTO**

Tese submetida à Universidade Federal de Santa Catarina – UFSC para obtenção do grau de Doutor em Engenharia.

**Orientador: Professor Ricardo Miranda Barcia, Ph.D**

**FLORIANÓPOLIS**  
**2005**

**TÂNIA CRISTINA D'AGOSTINI BUENO**

**ENGENHARIA DA MENTE:**

**UMA METODOLOGIA DE REPRESENTAÇÃO DO CONHECIMENTO  
PARA CONSTRUÇÃO DE ONTOLOGIAS EM SISTEMAS BASEADOS  
EM CONHECIMENTO**

Esta Tese foi julgada adequada para obtenção do título de “Doutor em Engenharia”, especialidade Engenharia de Produção, e aprovada em sua forma final pelo Programa de Pós-Graduação.

Florianópolis, 12 de janeiro de 2005.

**Professor Edson Pacheco Paladini**

**Coordenador do Programa de Pós-Graduação em Engenharia de Produção**

**Banca Examinadora:**

---

**Ricardo M. Barcia, Ph.D, Orientador**

---

**Carlos Augusto M. Remor, Doutor**

---

**Jesus Cardeñosa Lera, Ph.D**

---

**Christiane Coelho de Souza Reinisch  
Coelho, Doutora**

---

**Walter Félix Cardoso Júnior, Doutor**

---

**Alejandro Rodrigues Martins, Doutor**

“Não posso provar a você que Deus existe, mas meu trabalho provou empiricamente que o ‘padrão Deus’ existe em cada homem, e que esse padrão (*pattern*) é a maior energia transformadora de que a vida é capaz de dispor ao indivíduo. Encontre esse padrão em você mesmo e a vida será transformada”.

Carl G. Jung

À

**Lara, Milena e Oriana**

## AGRADECIMENTOS

À equipe da WBSA – Sistemas Inteligentes S. A. e à equipe do Instituto de Governo Eletrônico e Inteligência Jurídica – IJURIS – pessoas de grande poder de criação e com um maravilhoso senso de trabalho em equipe. Agradeço pela participação na concepção, implementação e viabilização econômica deste projeto – à Suíte de Engenharia do Conhecimento.

À equipe da SAEI, do Gabinete de Segurança Institucional da Presidência da República, pelo equilíbrio emocional, coesão e pelo compartilhamento de muitos conhecimentos que proporcionaram uma intensa participação da equipe na concepção e representação do conhecimento do sistema KMAI®, especialmente ao Comandante Cunha Couto, ao Tenente Coronel Rufino Sales e ao Comandante Camilo.

Pelo incentivo financeiro da Bolsa de Estudo concedida pela CAPES.

Ao Professor Ricardo M. Barcia, por aceitar o desafio de orientar esta Tese.

Ao meu marido, Hugo C. Hoeschl, pelo seu apoio, por me ouvir e auxiliar na conclusão das muitas idéias, que permeiam este trabalho, com o seu conhecimento.

Aos amigos e colegas de trabalho, André Bortolon, Cristina S. Santos e Eduardo Mattos, pelo conhecimento compartilhado na construção da Suíte de Engenharia do Conhecimento, que influenciaram diretamente nas grandes inovações do sistema.

À colega e amiga Christianne C. S. R. Coelho, pelo incentivo e a excelente contribuição no desenvolvimento final desta abordagem.

Aos professores, Alejandro Martins, Carlos Augusto Mongliott Remor e Jesus Cardeñosa Lera, pela disponibilidade e interesse em ler e comentar este trabalho.

Às superpoderosas garotas da equipe de Engenharia do Conhecimento da WBSA, em especial Aline Junckes.

À minha mãe, Dulce Bueno, filósofa, pela sua orientação fundamental sobre os princípios da Filosofia, que me permitiram compreender alguns aspectos da conexão existente entre a Filosofia e a Inteligência Artificial na organização do conhecimento.

E, em especial, ao meu pai, Peri Backer Bueno, pelo carinho e atenção, indispensáveis na construção de um conhecimento.

Àqueles que direta ou indiretamente contribuíram na organização desta pesquisa, Aline Nicolini, Marcos Carlson e ao pessoal da Secretaria do PPGEP, especialmente Neiva e Rosimeri.

À minha família, principalmente àquelas pessoas que estiveram ao meu lado, minha avó, Alzira T. D'Agostini, a quem dedico este trabalho, e Sônia Webler, que ajudaram no cuidado de minhas filhas, tanto nas viagens de cunho acadêmico que fiz como nos trabalhos desenvolvidos diariamente.

Às minhas adoradas filhas, Oriana, Milena e a pequena Lara, que tantas vezes me indagaram sobre a relevância de se fazer uma Tese, agradeço seus questionamentos, pois foram essas dúvidas que priorizaram alguns aspectos desta pesquisa.

Aos grandes matemáticos, filósofos, pensadores, artistas, pesquisadores que souberam compartilhar seus conhecimentos e permitiram a evolução da sociedade humana.

Àquelas pessoas que de uma maneira atuante perpetuam o seu amor pela humanidade.

Enfim, às musas inspiradoras, para que este trabalho continue...

## RESUMO

Na era da informação digitalizada, pesquisadores de diferentes áreas do conhecimento enfrentam uma nova questão: a organização semântica dos dados. A matéria tornou-se importante pelo fato das informações digitalmente disponíveis estarem dispostas, em sua maioria, como dados não-estruturados. Todavia, sabe-se que a estruturação de dados é um problema complexo que poderá ser resolvido através da construção de modelos formais e de linguagens da Ciência da Computação, sobre as quais é preciso observar que essas áreas específicas possuem uma cultura própria e um modo singular de se comunicar. Observando os fatores envolvidos, nesta pesquisa, desenvolveu-se uma metodologia para a compreensão e representação computacional dessa maneira de comunicação, o que permitiu, a cada comunidade ou ambiente de trabalho, a viabilidade de expressar o seu domínio de conhecimento. Denominou-se esta metodologia de *Engenharia da Mente*, cujo conceito ‘Ontológico’ tornou-se uma importante ‘chave’ para a estruturação de dados e para a construção de Sistemas Baseados em Conhecimento.

A *Engenharia da Mente* é definida pela Engenharia do Conhecimento e Engenharia de Ontologias como um processo de sincronização do conhecimento desenvolvido com a finalidade de fixar modelos conceituais em relação ao objetivo e à aplicação do modelo de gestão baseado na Inteligência Artificial (IA), e identificar e sistematizar habilidades intelectuais da equipe de desenvolvimento do sistema, auxiliando-os na percepção do problema com qualidade e criatividade.

A *Engenharia da Mente* identifica a emoção como um aspecto da inteligência humana indispensável na etapa da representação do conhecimento na IA, mais especificamente no desenvolvimento de Sistemas Baseados em Conhecimento (SBC), em ambientes de trabalho cooperativos na Web. Assim, quanto maior a sincronicidade entre o conhecimento do especialista e o engenheiro do conhecimento, maior a efetividade do sistema.

Os fundamentos desta metodologia passam pela perspectiva da neurociência – um dos campos mais respeitados e concorridos da ciência na atualidade – compreendendo a estrutura da linguagem e do inconsciente humano, em conjunto com os mecanismos utilizados pela mente, para definir o conhecimento relevante. Este processo engloba o estudo de pessoas, processos e tecnologias, através do compartilhamento do conhecimento, da visualização e da definição de relevância. A sincronização desses fatores tem por objetivo permitir a compreensão do conhecimento, ou expertise, de um determinado domínio na sua totalidade, através de um sistema computacional – mais especificamente uma ferramenta de Engenharia de Ontologia – que atue como um mecanismo de aquisição de conhecimento.

Neste estudo, a metodologia da *Engenharia da Mente* foi aplicada para auxiliar dois processos distintos: a Engenharia do Conhecimento e a Engenharia de Ontologias. Esses processos foram aplicados numa série de sistemas que serão apresentados nos resultados deste trabalho. A Engenharia de Ontologias recebeu um enfoque especial através do desenvolvimento de uma ferramenta: a Suíte de Engenharia do Conhecimento.

**Palavras-chaves:** Engenharia do Conhecimento. Engenharia de Ontologias. Sistemas Baseados em Conhecimento. Gestão do Conhecimento.

## ABSTRACT

In the era of the digitalized information, researchers of different areas of the knowledge face a new subject: the semantic organization of the data. The matter became important for the fact of the information digitally available be disposed, in majority, as no-structured data. Though, it is known that the structuring of data is a complex problem that can be solved through the construction of formal models and of languages of the Computer Science, on which it is necessary to observe that those specific areas possess an own culture and a singular way of communicating. Observing the involved factors, in this research, it was developed a methodology for the understanding and computational representation of that sorts things out of communication, what allowed, to each community or work atmosphere, the viability of express their knowledge domain. Was called this methodology of *Mind Engineering*, whose concept 'Ontological' became an important 'key ' for the structuring of data and for the construction of Knowledge Based Systems.

The *Mind Engineering* is defined by the Knowledge Engineering and Ontological Engineering as a process of synchronization of the knowledge developed with the purpose of fastening conceptual models in relation to the objective and to the application of the administration model based on the Artificial Intelligence (AI), and to identify and to systematize intellectual abilities of the team of development of the system, aiding them in the perception of the problem with quality and creativity.

The *Mind Engineering* identifies the emotion as an aspect of the indispensable human intelligence in the stage of the representation of the knowledge in the Artificial Intelligence, more specifically in the development of Systems Based on Knowledge (SBC) in cooperative work atmospheres in the Web. As larger the sincronicidade between the Knowledge Specialists and the Knowledge Engineers, larger the effectiveness of the system.

The foundations of this methodology go by the perspective of the neuroscience - one of the fields more respected and competed of the science at the present time - understanding the structure of the language and of the human unconscious, together with the mechanisms used by the mind, to define the relevant knowledge. This process includes the people's study, processes and technologies, through the sharing of the knowledge, of the visualization and of the definition of relevance. The synchronization of those factors has for objective to allow the understanding of the knowledge, or expertise, of a certain domain in the totality, through a computational system - more specifically a tool of Ontological Engineering - that acts as a mechanism of knowledge acquisition.

In this study, the methodology of the *Mind Engineering* was applied to aid two different processes: the Knowledge Engineering and the Ontological Engineering. Those processes were applied in a series of systems that will be presented in the results of this work. The Ontological Engineering received a special focus through the development of a tool: the Suite of Engineering of the Knowledge.

**Word-keys: Knowledge Engineer. Ontological Engineering. Systems Based on Knowledge. Knowledge Management.**



## LISTA DE FIGURAS

Figura 1 – Ciclo do Raciocínio Baseado em Casos (RBC).....	31
Figura 2 – Análise da Estrutura Retórica .....	36
Figura 3 – Pesquisa Contextual Estruturada.....	37
Figura 4 – Engenharia da Mente aplicada à Construção de Ontologias em Sistemas Baseados em Conhecimento .....	70
Figura 5 – Interface do Sistema JurisConsulta .....	94
Figura 6 – Módulo de Indexação Automática do Sistema JurisConsulta .....	94
Figura 7 – Processo de Recuperação Baseado em Casos do Sistema JurisConsulta.....	95
Figura 8 – Definição das Expressões Indicativas no Sistema JurisConsulta baseada na Teoria da Argumentação Jurídica.....	96
Figura 9 – Medida da Similaridade Global do Sistema JurisConsulta .....	98
Figura 10 – Interface do Sistema Olimpo® .....	102
Figura 11 – Interface <i>AlphaThemis</i> ®, mostrando a saída dos dados e ressaltando o uso de pesos dinâmicos .....	104
Figura 12 – Primeira versão da Suíte de Engenharia do Conhecimento.....	107
Figura 13 – Processos do Sistema KMAI® .....	109
Figura 14 – Estrutura do Sistema KMAI .....	110
Figura 15 – Análise Textual do Sistema KMAI .....	111
Figura 16 – Gráfico dos Resultados.....	112
Figura 17 – Apresentação dos Filtros para Geração dos Gráficos .....	113
Figura 18 – Geração de Gráficos por Assunto .....	113
Figura 19 – Cadastro de Usuários.....	114
Figura 20 – Módulo SC – INFO .....	115
Figura 21 – Ontologias no Sistema KMAI.....	116
Figura 22 – Exemplo de uma Ontologia e suas relações na Suíte de Engenharia do Conhecimento.....	126
Figura 23 – Módulo de Cadastro de Novas Ontologias.....	127
Figura 24 – Escolha de Tema e Subtema para inserção.....	128
Figura 25 – Exclusão de termo do Dicionário .....	129
Figura 26 – Exclusão de termo do Tema .....	129
Figura 27 – Exclusão de termo do Subtema.....	130
Figura 28 – Visualização das relações e possibilidade de edição.....	131
Figura 29 – Interface de alteração de relação de sinônimos.....	132
Figura 30 – Relatório dos Subtemas .....	133

Figura 31 – Cadastro de Novo Tema .....	134
Figura 32 – Cadastro de novo subtema .....	135

### **LISTA DE TABELAS**

Tabela 1 – Exemplo de Resultado do Cálculo de Similaridade do Sistema Olimpo® .....	102
Tabela 2 – Valor dos Pesos das relações no Cálculo da Similaridade .....	120
Tabela 03 – Resumo dos Resultados da Avaliação de Ferramentas de Engenharia de Ontologias .....	139
Tabela 04 – Resultado da Avaliação da Suíte de Engenharia do Conhecimento .....	140

### **LISTA DE QUADROS**

Quadro 1 – UNL – <i>Universal Natural Language</i> .....	48
Quadro 2 – Tipos de inventários realizados dentro de uma instituição .....	78
Quadro 3 – Tipos de inventários de pessoas .....	79
Quadro 4 – Vocabulário Controlado no contexto da Segurança Pública .....	85
Quadro 5 – Leis da Percepção aplicada à construção de Ontologias Suíte de Engenharia do Conhecimento .....	88
Quadro 6 – Módulos da Suíte de Engenharia do Conhecimento .....	124
Quadro 7 – Objetivos da Aplicação da Engenharia da Mente .....	142

### **LISTA DE GRÁFICOS**

Gráfico 1 – Avaliação das Ferramentas de Engenharia de Ontologia .....	138
--	-----

## LISTA DE ABREVIATURAS

AC: Aquisição de Conhecimento  
COAF: Conselho de Controle de Atividades Financeiras  
EC: Engenharia do Conhecimento  
EO: Engenharia de Ontologias  
ES: Engenharia de *Software*  
FATF: *Financial Action Task Force on Money Laundering*  
FS: Fábrica de *Software*  
FTP: *File Transfer Protocol*  
GAFI/FATF: Grupo de Ação Financeira sobre Lavagem de Dinheiro  
IA: Inteligência Artificial  
IR: *Information Retrieval*  
KBS: *Knowledge-Based Systems*  
KMAI®: *Knowledge Management with Artificial Intelligence*  
ODE: *Open Dynamics Engine*  
ONU: Organização das Nações Unidas  
OWL: *Ontology Web Language*  
PCE®: Pesquisa Contextual Estruturada  
RBC: Raciocínio Baseado em Casos  
RC: Representação de Conhecimento  
RC2D®: Representação do Conhecimento Contextualizada Dinamicamente  
RDF: *Resource Description Framework*  
SBC: Sistemas Baseados no Conhecimento  
SE: Sistemas Especialistas  
UNL: *Universal Networking Language*  
URI: *Uniform Resource Identifier*  
UW: *Universal Words*  
XML: *Extensible Markup Language*

## SUMÁRIO

### CAPÍTULO 1

<b>INTRODUÇÃO</b> .....	<b>14</b>
1.1 MOTIVAÇÃO .....	18
1.2 OBJETIVOS .....	19
1.2.1 Objetivos Específicos .....	20
1.3 RESUMO DA METODOLOGIA.....	20
1.4 ESTRUTURA DO TRABALHO .....	24

### CAPÍTULO 2

<b>2 REVISÃO DA LITERATURA</b> .....	<b>26</b>
2.1 ENGENHARIA DO CONHECIMENTO .....	26
2.1.1 Aquisição e Representação do Conhecimento na Inteligência Artificial .....	27
2.1.1.1 A Representação do Conhecimento na Técnica Raciocínio Baseado em Casos (RBC) .....	30
2.1.1.2 RCSD® e PCE® .....	34
2.2 ENGENHARIA DE ONTOLOGIAS.....	39
2.2.1 Relações da Engenharia de Ontologias .....	41
2.2.1.1 Ontologias e Inteligência Artificial .....	42
2.2.1.2 O que são Ontologias?.....	43
2.2.1.3 Ontologias e Representação do Conhecimento.....	44
2.2.2 Modelos de Desenvolvimento de Ontologias .....	45
2.2.2.1 UNL – <i>Universal Networking Language</i> .....	46
2.2.2.2 <i>Semantic Web</i> .....	48
2.2.2.3 <i>WordNet</i> .....	50
2.3 ENGENHARIA DA MENTE: A SINCRONICIDADE DO CONHECIMENTO .....	51
2.3.1 Bases da Sincronicidade .....	55
2.3.1.1 Teorias da Linguagem .....	55
2.3.1.2 O Método Científico e o Imaginário .....	59
2.3.2 Conclusão .....	65

## **CAPÍTULO 3**

<b>3 AS METODOLOGIAS DA ENGENHARIA DA MENTE.....</b>	<b>68</b>
3.1 FASES DA ENGENHARIA DA MENTE.....	70
3.1.1 Fase 1: Compartilhamento do Conhecimento.....	70
3.1.2 Fase 2: Visualização.....	71
3.1.3 Fase 3: A Definição de Relevância.....	72
3.2 ENGENHARIA DA MENTE APLICADA À ENGENHARIA DO CONHECIMENTO.....	74
3.2.1 Fase 1: O Compartilhamento do Conhecimento na Engenharia do Conhecimento...	75
3.2.1.1 Identificando os aspectos subjetivos para formalização da equipe de EC.....	75
3.2.1.2 Uniformização do Vocabulário.....	76
3.2.1.3 Inventário do Conhecimento Institucional ou do Domínio de Aplicação.....	77
3.2.1.4 Inventário de Pessoas.....	78
3.2.2 Fase 2: Visualização na Engenharia do Conhecimento.....	80
3.2.3 Fase 3: Identificação de Relevância na Engenharia do Conhecimento.....	80
3.3 ENGENHARIA DA MENTE APLICADA À ENGENHARIA DE ONTOLOGIAS..	81
3.3.1 Fase 1: Compartilhamento do Conhecimento na Engenharia de Ontologias.....	81
3.3.1.1 Identificação do Vocabulário Controlado do Domínio.....	82
3.3.2 Fase 2: Visualização na Engenharia de Ontologias.....	85
3.3.2.1 Criação de Domínios e Subdomínios (Temas e Subtemas).....	86
3.3.2.2 O Desenvolvimento do Dicionário e de suas Relações.....	86
3.3.3 Fase 3: Identificação de Relevâncias na Engenharia de Ontologias.....	89
3.3.3.1 Validação dos Termos.....	91
3.4 CONCLUSÃO.....	92

## **CAPÍTULO 4**

<b>4 A METODOLOGIA APLICADA EM SISTEMAS.....</b>	<b>93</b>
4.1 SISTEMAS BASEADOS EM CONHECIMENTO DESENVOLVIDOS COM A TÉCNICA RC2D E PCE.....	93
4.1.1 JurisConsulta®.....	93
4.1.2 Olimpo®.....	100
4.1.3 <i>Alpha Themis</i> ®.....	103
4.1.4 Sistema KMAI®.....	108
4.1.4.1 Análise Textual.....	111
4.1.4.2 Análise Gráfica.....	112
4.1.4.3 Nota Informativa.....	114

4.1.4.4 Ambiente de Administração .....	114
4.1.4.5 Editor de Ontologias.....	115
4.1.4.6 SC – INFO .....	115
4.2 A CONSTRUÇÃO DE ONTOLOGIAS NO SISTEMA KMAI.....	116
4.2.1 Ontologias e Cálculo da Similaridade do Sistema KMAI.....	119
4.3 SUÍTE DE ENGENHARIA DO CONHECIMENTO .....	122
4.3.1 Módulos da Suíte de Engenharia do Conhecimento .....	124
4.3.1.1 Módulo de Criação de Novas Ontologias.....	125
4.3.1.2 Módulo de Exclusão de Ontologias.....	128
4.3.1.3 Módulo de Busca e Edição .....	130
4.3.1.4 Módulo Relatório .....	132
4.3.1.5 Módulo de Administração .....	133
4.3.2 Avaliação da Suíte de Engenharia do Conhecimento .....	136
4.3.2.1 Resultado da Avaliação da Suíte EC.....	140
4.3.3 CONCLUSÃO .....	141
<b>RESULTADOS .....</b>	<b>143</b>
<b>CONSIDERAÇÕES FINAIS E DESENVOLVIMENTOS FUTUROS.....</b>	<b>145</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>147</b>
<b>ANEXOS .....</b>	<b>152</b>

## CAPÍTULO 1

### INTRODUÇÃO

Mistificada no seu nascimento e posteriormente renegada ao campo da ficção científica, a Inteligência Artificial (IA) retomou o fôlego recentemente em aplicações para a recuperação de informações em larga escala, gestão da informação e resolução de problemas. Os algoritmos, as técnicas e as metodologias de Inteligência Artificial estão aprimorando os resultados dos sistemas computacionais convencionais. Entretanto, os sistemas inteligentes ainda não são capazes de, por exemplo, associar que: se João está em São Paulo seu pé também estará lá, a menos que se informe isso de maneira explícita. Ou seja, para o computador só existe o que é declarado numa linguagem que ele possa entender.

Garry Kasparov, campeão mundial de xadrez, sentiu o quão inteligente pode ser esta forma de organização do conhecimento. Mas quem foi inteligente? O *Deep Blue* ou os seus programadores? Que conhecimento venceu o maior enxadrista do mundo?

Ao começarem a pensar sobre o que é a inteligência, os teóricos da IA focaram suas atenções nos aspectos cognitivos, como memória e resolução de problemas. Porém já existiam pesquisadores que, cedo, reconheceram que os aspectos não cognitivos também são importantes para definir a inteligência.

A década de 1970 marcou a passagem da IA para a ‘vida adulta’. Com o aparecimento dos primeiros Sistemas Especialistas (SE), a tecnologia permitiu o desenvolvimento de sistemas com desempenho intelectual equivalente ao de um ser humano adulto, o que abriu perspectivas de aplicações comerciais e industriais.

A gradativa mudança de metas – desde o sonho de construir uma Inteligência Artificial de caráter geral comparável a do ser humano até os bem modestos objetivos atuais de tornar os computadores mais úteis, empregando ferramentas que auxiliem as atividades intelectuais de seres humanos – coloca a Inteligência Artificial na perspectiva de uma

atividade que praticamente caracteriza a espécie humana: a capacidade de utilizar representações externas, seja na forma de linguagem, seja através de outros meios (HILL, 1989)<sup>1</sup>. Para Bittencourt (1998)<sup>2</sup>, sob esse ponto de vista, a computação, em geral, e a IA, em particular, são o ponto culminante de um longo processo de criação de representações de conhecimento, que se iniciou com as primeiras pinturas rupestres. Essa nova perspectiva coloca os programas de Inteligência Artificial como produtos intelectuais no mesmo nível dos demais, ressaltando questões cuja importância é central para os interesses atuais da IA – por exemplo: expressar as características individuais e sociais da inteligência utilizando computadores de maneira a permitir uma maior produtividade.

Os sistemas computacionais baseados em técnicas da Inteligência Artificial possuem um diferencial com relação a outros sistemas computacionais: eles permitem uma constante evolução e adaptação do conhecimento. Sem essa característica, não se poderia considerar um determinado sistema, inteligente. No entanto, definir o conhecimento a ser colocado dentro de um sistema é o grande gargalo no desenvolvimento de sistemas inteligentes, principalmente porque o processo é basicamente experimental e depende muito da habilidade do engenheiro do conhecimento. Por essa razão, a construção de um sistema baseado em IA é diferenciado e chama-se Engenharia do Conhecimento (EC), uma metodologia em que a participação do especialista e do engenheiro do conhecimento é constante em todas as fases da construção de um *software*, da concepção aos testes finais.

Uma grande dificuldade, em particular, relaciona-se com a definição da terminologia que se emprega para nomear os conceitos e as relações (Rezende, 2003)<sup>3</sup>. Contudo, não é a única dificuldade. Se, no processo de aquisição do conhecimento, existe uma infinidade de métodos, nenhum se volta à compreensão e à estruturação do processo de conhecimento das pessoas envolvidas, sejam os especialistas, sejam os engenheiros do conhecimento. Um dos modelos que mais se aproxima a essa abordagem foi o elaborado pelo *CommonKADs* (Schreiber et al, 2002)<sup>4</sup>, no qual se destacam os modelos que fazem parte de todo o processo de Engenharia do Conhecimento envolvido na construção de Sistemas Inteligentes Baseados em Conhecimento.

---

<sup>1</sup> Hill, W.C. The Mind at AI: Horseless carriage to clock. *The AI Magazine*, pages 29-41, Summer 1989.

<sup>2</sup> Bittencourt, G. Inteligência artificial – Ferramentas e teorias. Op. Cit. (1998).

<sup>3</sup> Rezende, Solange O. (org.). *Sistemas Inteligentes: Fundamentos e Aplicações*. Barueri-SP: Manole, 2003.

<sup>4</sup> Schreiber, G.; Akkermans, H.; Anjewierden, A.; Hoog, R.; Shadbolt, N.; de Velde, W. V.; and Wielinga, B. *Knowledge Engineering and Management: the CommonKADS Methodology*. MIT Press. Cambridge. Massachussets. 2002.



Recentemente, alguns trabalhos, como o desenvolvido por Guarino e Welty (2000)<sup>5</sup>, enfocam as Ontologias do ponto de vista do desenvolvimento de Sistemas Baseados no Conhecimento (SBC) e discutem como se podem usar os princípios da Ontologia formal e da Engenharia Ontológica na prática da Engenharia do Conhecimento.

A noção de Ontologia se popularizou na área de integração inteligente da informação, recuperação de informação na Internet e gerência do conhecimento. A razão deve-se, em parte, ao que promete: uma compreensão compartilhada e comum de algum domínio que possa ser comunicada através das pessoas e dos computadores (Duineveld et al, 2000)<sup>6</sup>. Os diferentes desenvolvimentos de abrangência mundial, que utilizam Ontologias, têm referência nos trabalhos em redes colaborativas baseados na Web, como a *WordNet*<sup>7</sup>, *SemanticWeb*<sup>8</sup> e UNL - *Universal Networking Language*<sup>9</sup>.

Na teoria de Descartes, do século dezessete, o modelo de consciência separa a nossa mente de todo o mundo em torno de nós. Sabe-se que o modelo de Descartes não serve para a estruturação do modelo de Ontologias, nem para a organização das pessoas que irão trabalhar no desenvolvimento de Ontologias. Os líderes experientes sabem que emoções como a incerteza, o medo, a confusão e a perda de autoconfiança são sentimentos comuns que podem preceder o surgimento de uma novidade e fazem parte da maneira como uma decisão é tomada. Isso significa que as emoções fazem parte da estruturação do conhecimento.

Um grande passo nessa direção foram as mais recentes descobertas da neurociência, que renovaram as expectativas em torno da IA. O cérebro humano responde a uma série de questões de forma matemática e, muitas das coisas que se julgavam impossíveis a uma máquina realizar, têm-se conhecimento que, na atualidade, são possíveis de se reproduzir com aceitável precisão. ‘Descobriram-se’ leis que afetam a inteligência humana do mesmo modo que Mendeleev descobriu a Tabela Periódica dos Elementos. Ele não inventou os elementos químicos, tampouco seus pesos atômicos, mas demonstrou que eles existiam na Natureza e que possuíam ‘características’ exatas, que obedeciam a uma ordem matemática e objetiva.

---

<sup>5</sup> Guarino, Nicola and Chris Welty. 2000. A Formal Ontology of Properties. In, Dieng, R., and Corby, O., eds, Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management. Springer-Verlag LNCS Vol. 1937:97-112. October, 2000.

<sup>6</sup> Duineveld, A. J., Stoter, R., Weiden, M. R., Kenepa, B. and Benjamins, V. R.: Wondertools. A comparative study of ontological engineering tools. In the International Journal of Human-Computer Studies, July 2000.

<sup>7</sup> Wordnet. Disponível em: <http://www.cogsci.princeton.edu/~wn/>. Acesso em: 19 jan. 2004.

<sup>8</sup> Semantic Web. Disponível em: <http://www.w3.org/2001/sw/>. Acesso em: 19 jan. 2004.

<sup>9</sup> UNL – Universal Networking Language. Disponível em: <http://www.unl.ias.unu.edu/unlsys/index.html>. Acesso em: 19 jan. 2004.

No presente trabalho, procurou-se demonstrar que existem elementos na representação do conhecimento aplicáveis tanto para a Engenharia do Conhecimento quanto para a Engenharia de Ontologias. Esses elementos constituem o que se denomina *Engenharia da Mente*, uma metodologia capaz de identificar habilidades intelectuais da equipe de Engenharia do Conhecimento e Ontologias, a fim de utilizá-las no desenvolvimento de estratégias para a percepção do problema de forma criativa e permitir a qualidade na construção de Sistemas Baseados em Conhecimentos.

Segundo Damásio (2000)<sup>10</sup>, a metodologia da *Engenharia da Mente* permite avaliar, como parte do conhecimento a ser inserido no sistema, tanto o conhecimento tácito e explícito das instituições quanto sua combinação com as emoções – indispensáveis para aquilo que se denomina inteligência.

Nesta pesquisa, essa metodologia foi aplicada na Engenharia do Conhecimento com o objetivo de identificar o conhecimento relevante a ser representado na criação de Ontologias e na própria Engenharia de Ontologia; e, com isso, fornecer um vocabulário comum de uma área, definindo, com níveis diferentes de formalidade, o sentido dos termos e o relacionamento entre eles.

Levantam-se questões relativas à psicanálise, à lingüística e à neurociência que indicam o significado dos símbolos que, supostamente, existem dentro dos indivíduos. Foram aplicados e avaliados procedimentos para conhecer e reconhecer nas pessoas esses símbolos e compartilhar os seus próprios. Observou-se que eles são importantes para a representação do conhecimento na Inteligência Artificial e na construção de Ontologias, em que, ao incluir tais procedimentos na Engenharia do Conhecimento e na Engenharia de Ontologia, proporcionam uma maior sincronia das ações e, conseqüentemente, a construção de uma base de conhecimento com resultados mais efetivos, num menor tempo em relação àqueles nas quais foram utilizadas as metodologias mais conhecidas nesses processos.

Além disso, a metodologia possibilitou ao sistema baseado em Ontologias indexar automaticamente novos documentos à base de conhecimento, com maior efetividade que outras metodologias de representação e aquisição do conhecimento existente na IA. Com a aplicação da metodologia, verificou-se que o computador passou a compreender melhor a ‘intenção’ do usuário do sistema e a interpretar os dados textuais com maior exatidão.

---

<sup>10</sup> Damásio, António. O mistério da Consciência: do corpo e das emoções ao conhecimento de si. Tradução Laura Teixeira Motta. 5ª edição. São Paulo: Companhia das Letras, 2000.

## 1.1 MOTIVAÇÃO

A *Engenharia da Mente* é definida pela Engenharia do Conhecimento e na Engenharia de Ontologias como um processo de sincronização do conhecimento desenvolvido com o objetivo de fixar modelos conceituais em relação ao objetivo e à aplicação do modelo de gestão baseado na Inteligência Artificial, e identificar e sistematizar habilidades intelectuais da equipe de desenvolvimento do sistema, auxiliando-os na percepção do problema com qualidade e criatividade.

A metodologia da *Engenharia da Mente* é baseada em três premissas:

- Compartilhamento do conhecimento;
- Visualização;
- Identificação de relevâncias.

Para o desenvolvimento deste trabalho, esse processo foi aplicado nas seguintes atividades:

- Engenharia do Conhecimento;
- Engenharia de Ontologias.

A metodologia da *Engenharia da Mente* produz resultados diferenciados para cada área de aplicação. O desenvolvimento de cada uma das premissas e o seu impacto são o escopo principal desta pesquisa.

Na Engenharia do Conhecimento (EC), a aplicação da metodologia resulta na sincronização de conhecimentos entre o especialista, o engenheiro do conhecimento e o analista de sistema. Através dessa sincronização, obtêm-se as informações relevantes para a construção do Sistema Baseado em Conhecimento (SBC), visando compreender o objetivo do sistema e utilizar todo o conhecimento disponível digitalmente, a fim de construir a base de conhecimento orientada para este escopo. A aplicação da metodologia pode variar o grau de conhecimento sobre o domínio e sobre a aplicação de técnicas de Inteligência Artificial para a percepção e solução dos problemas com efetividade e qualidade. O objetivo da aplicação desta metodologia na EC é permitir que engenheiros do conhecimento e especialistas elaborem algoritmos e encontrem soluções tecnológicas para a sua área de atuação baseado

em conhecimentos de Inteligência Artificial. A metodologia também tem por objetivo uma atuação constante e síncrona com a equipe de programadores.

Na Engenharia de Ontologias, a aplicação da metodologia da *Engenharia da Mente* tem por objetivo que o especialista do domínio e o engenheiro do conhecimento compreendam o contexto de aplicação e definam as Ontologias dentro da modelagem do domínio, bem como o uso das Ontologias (integração a sistemas baseados no conhecimento) e o seu gerenciamento (evolução e integração). Isso permitirá que o especialista do domínio elabore Ontologias dentro do ambiente da Suíte de Engenharia do Conhecimento e defina o vocabulário controlado e suas relações.

Para ter validade, uma Ontologia precisa ser aceita e utilizada pelos usuários de um sistema. O principal obstáculo é que cada pessoa tem uma visão individualizada de uma Ontologia. As relações (conexões) entre as palavras (conceitos) diferem não só pelo contexto em que as empregam, mas também pela forma peculiar que cada indivíduo observa aquele contexto. Assim, a tão requisitada questão da coerência na IA, e também na construção de Ontologias, somente será viável quando todos os especialistas e engenheiros do conhecimento conceberem a solução de um problema de uma mesma forma, isto é, visualizarem o objeto e os objetivos do sistema de forma síncrona e efetiva.

Enfim, a metodologia da *Engenharia da Mente* proporciona um meio de aguçar as capacidades cognitivas dos Engenheiros do Conhecimento. Longe de ser um encarceramento formal da realidade, ela é uma afinação do instrumento cognoscitivo do sujeito para que comece a pesquisar a realidade.

## 1.2 OBJETIVOS

O principal objetivo da metodologia da *Engenharia da Mente* é permitir a qualidade na representação do conhecimento baseada em Ontologias e no desenvolvimento de Sistemas Inteligentes.

A metodologia trabalha as questões subjetivas que podem influenciar os processos de Engenharia do Conhecimento e Engenharia de Ontologias. A metodologia foi organizada em três fases principais: Compartilhamento do Conhecimento, Visualização e Definição de Relevância. Essas fases foram associadas aos processos de Engenharia do Conhecimento e

Engenharia de Ontologias, gerando uma visão compartilhada da base de conhecimento e permitindo um constante aprimoramento dos resultados na definição, no desenvolvimento e nos resultados dos Sistemas Inteligentes desenvolvidos com a técnica.

### 1.2.1 Objetivos Específicos

- Descrever os elementos principais da Inteligência Artificial;
- Descrever os domínios de aplicação da Engenharia do Conhecimento e da Engenharia de Ontologias, e apresentar as principais teorias e desenvolvimentos mundiais;
- Descrever a ferramenta de construção de Ontologias e os resultados de sua aplicação no Sistema KMAI®;
- Relacionar áreas da Inteligência Artificial, da Engenharia de Ontologia, da Teoria Psicanalítica e da própria Neurociência que identificam aspectos cognitivos essenciais para uma efetiva representação do conhecimento;
- Descrever a metodologia e o resultado da sua aplicação na Engenharia do Conhecimento e na Engenharia de Ontologias, nos sistemas JurisConsulta, *AlphaThemis*, Olimpo e KMAI.

### 1.3 RESUMO DA METODOLOGIA

A trajetória do desenvolvimento deste trabalho iniciou-se nos estudos da Inteligência Artificial, especificamente na aplicação da técnica denominada Raciocínio Baseado em Casos (RBC)<sup>11</sup> (ver item 2.1.1) para sistemas de buscas inteligentes na área jurídica<sup>12</sup>.

No desenvolvimento de sistemas inteligentes, existe um procedimento denominado Engenharia do Conhecimento (EC). Esse processo é o responsável pela construção do Sistema

---

<sup>11</sup> Ver detalhes em Kolodner, Janet L. Case-based reasoning. Morgan Kaufmann Publishers, Inc. 1993. San Marco, CA 94403.

<sup>12</sup> Ver detalhes em Bueno, Tânia Cristina D'Agostini. O Uso da Teoria Jurídica para Recuperação em Amplas Bases de Textos Jurídicos. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 1999.

Baseado em Conhecimento (SBC), pela forma de representação do conhecimento e também de aquisição.

Inicialmente, empregou-se a metodologia de representação e aquisição do conhecimento mais utilizada<sup>13</sup>. A técnica de IA denominada RBC foi associada à Teoria da Argumentação Jurídica<sup>14</sup> para auxiliar a forma de representação dos casos no sistema Jurisconsulto. Essa concepção permitiu definir qual seria o conhecimento relevante num texto jurídico e como representá-lo. A associação de técnicas possibilitou a realização de um processo automático de indexação de textos com base em um vocabulário controlado e num dicionário de termos normativos construído através da relevância dos termos definidos persuasivamente, denominados termos-chaves normativos<sup>15</sup>.

Na seqüência da evolução, diante da necessidade de tornar o processo de aquisição mais célere, incorporaram-se técnicas de Recuperação da Informação (IR - *Information Retrieval*)<sup>16</sup>, associando a relevância dos termos com a freqüência das palavras, somadas ao vocabulário controlado e ao dicionário de termos normativos, o que originou a metodologia de representação do conhecimento denominada Representação do Conhecimento Contextualizada Dinamicamente (RC2D)<sup>17</sup>.

A RC2D permitiu a representação automática de casos em Sistemas Baseados em Conhecimento, seja na área jurídica<sup>18 19</sup>, seja na área de gestão da informação<sup>20</sup>.

O desenvolvimento de sistemas para gestão do conhecimento tornou necessária a constante atualização de novas informações na base de conhecimento, isto é, um sistema que permitisse o aprendizado contínuo, a exemplo dos Sistemas Baseados em Caso. A solução foi

<sup>13</sup> Ver detalhes em Kolodner, Janet L. Op. Citada.

<sup>14</sup> Ver detalhes em Warat, Luis Alberto. O Direito e a sua Linguagem. 2ª Versão. Sergio Antonio Fabris Editor: Porto Alegre, 1995.

<sup>15</sup> Warat, Luis Alberto. Op. Citado.

<sup>16</sup> Baesa-Yates, R. and Ribeiro-Neto B., Modern Information Retrieval. Addison Wesley, Reading, Mass. 1999.

<sup>17</sup> Hoeschl, Hugo César. Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU. Tese de Doutorado. Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina. Florianópolis: 2001.

<sup>18</sup> Bueno, Tânia C. D., et al. Retrieval in Jurisprudencial Text Bases using Juridical Terminology. Proceedings of 7th International Conference in Intelligence Artificial And Law - ICAIL, 1999, Oslo. ACM, 1999.

<sup>19</sup> Hoeschl, Hugo C., et. al. Olimpo: Contextual Structured Search to improve the representation of UN Security Council with Information Extraction methods. *Proceedings da 8a International Conference on Artificial Intelligence and Law*, ICAIL - 2001, St. Louis. New York: ACM SIGART, 2001. p.217 – 218.

<sup>20</sup> Ribeiro, Marcelo S.; Mattos, Eduardo da S., Bueno, Tânia C. D.; Hoeschl, Hugo C. KMAI- Knowledge Management With Artificial Intelligence. The Symposium on Professional Practice in AI in the First IFIP International Conference On Artificial Intelligence Application And Innovations. Toulouse, 2004.

construir uma ferramenta de fácil utilização, para que o próprio usuário do sistema pudesse inserir novas informações à base. Criou-se inicialmente um módulo independente para definir os sinônimos das expressões representativas do domínio, no sistema Jurisconsulto (ver Capítulo 4). As relações se aprimoraram no sistema *AlphaThemis*, no qual havia termos secundários e conexos, com pesos. Era o início da Suíte de Engenharia do Conhecimento e da utilização de Ontologias para a representação do conhecimento.

No entanto, uma grande mudança ocorreu no desenvolvimento desses sistemas. Para melhorar sua qualidade e efetividade, adotou-se o processo de Engenharia de *Software* (ES)<sup>21</sup>. Para tanto, criou-se uma Fábrica de *Software* (FS) e substituiu-se o processo de Engenharia do Conhecimento (EC), que ficou como uma célula isolada da FS. A mudança não gerou os efeitos esperados. Os especialistas do domínio, que antes participavam ativamente da implementação, ficaram praticamente isolados do desenvolvimento – sua função era aprovar os resultados e fornecer informações na etapa de EC. Observou-se que, na Fábrica de *Software*, os programadores criavam um *software* diverso do projetado pela célula de Engenharia do Conhecimento.

Além disso, quando era realizado o processo de Engenharia do Conhecimento, o *design* do *software* era determinado no momento da definição dos atributos. Na Fábrica de *Software*, ele ficou para uma etapa de finalização do sistema, de responsabilidade somente da célula de identidade visual. Com isso, freqüentemente, a interface do sistema necessitava ser adequada ao ‘imaginado’ pelos engenheiros do conhecimento que, por sua vez, tinham que se adequar ao ‘imaginado’ pelo usuário final, e muitas vezes com alterações consideravelmente complexas. Na metodologia da Engenharia de *Software*, esse processo de ‘manutenção’ é comum e previsível. Porém, para a equipe envolvida neste trabalho, o processo consumiu um tempo três vezes maior, o que exigiu aumentar a equipe para realizar a mesma demanda, além do desgaste emocional.

Antes da implantação da Fábrica de *Software*, os sistemas JurisConsulto, *AlphaThemis* e Olimpo finalizavam o seu desenvolvimento com menos erros e mais inovação, e a equipe de

---

<sup>21</sup> A Engenharia de *Software* é um rebento da Engenharia de Sistemas e de Hardware. Ela abrange três elementos fundamentais – métodos, ferramentas e procedimentos – que possibilitam ao gerente o controle do processo de desenvolvimento de software e oferece ao profissional uma base para a construção de software de alta qualidade, de forma produtiva. Os métodos de engenharia de software proporcionam os detalhes de “como fazer” para construir um software. Os métodos envolvem um amplo conjunto de tarefas que incluem: planejamento e estimativas de projeto, análise de requisitos de software e sistema, projeto da estrutura de dados, arquitetura de programa e algoritmos de processamento, codificação, teste e manutenção. Os métodos da engenharia de software muitas vezes utilizam notações gráficas ou orientadas a linguagens especiais e introduzem um conjunto de critérios para definição de qualidade de software.

testes centrava-se mais na avaliação da base de conhecimento e na busca de novos mecanismos para melhorar a performance do sistema. Tinha-se a sensação de ter realizado o melhor em todas as etapas. Resolveu-se, então, unir os procedimentos à Engenharia do Conhecimento e à Engenharia de *Software*, observando quais eram os elementos essenciais para produzir os resultados esperados.

Enquanto isso, a representação do conhecimento nesses sistemas passou a utilizar um modelo baseado em Ontologias, muito mais complexo que o vocabulário controlado, aprimorando o modelo de recuperação da informação baseado no contexto de aplicação do sistema KMAI. Como existem hoje três grandes redes colaborativas de abrangência mundial (UNL, a *Semantic Web* e a *WordNet*) que utilizam a Internet para o desenvolvimento de Ontologias, as mesmas foram empregadas como referência. Observaram-se as dificuldades comuns a estes desenvolvimentos, principalmente a UNL, na qual a participação desta equipe de pesquisa foi mais concreta e assertiva. Para tal, construiu-se um ambiente cooperativo na Web que permitiu o compartilhamento síncrono de pessoas e conhecimentos para criação e desenvolvimento de Ontologias. O resultado do esforço foi a construção da ferramenta denominada Suíte de Engenharia do Conhecimento (ver item 4.3).

O uso de ferramentas para suportar o processo da Engenharia de Conhecimento possibilitou a organização de uma base estabelecida no relacionamento entre expressões relevantes de um contexto. Não obstante, a experiência mostrou que, freqüentemente, o gargalo da estrutura de compartilhamento de Ontologias se encontra mais no processo social do que na tecnologia (Benjamins, 1998)<sup>22</sup>. Assim, produzir um ambiente de construção de Ontologias colaborativo na Web não foi suficiente para reproduzir os resultados esperados. Por isso, criou-se um ambiente Web que permitiu o compartilhamento, a visualização e a organização do conhecimento de forma síncrona e não hierárquica.

Uma etapa tornou-se essencial para a sincronização da equipe: a da ‘Visualização’. Observou-se que os membros da equipe sentiam-se mais confortáveis quando o desenho da Interface era projetado antes de se começar a implementação; o mesmo aconteceu na fase de construção dos mapas conceituais do domínio a ser trabalhado. Os mapas tornaram-se a base do processo de Engenharia de Ontologias.

---

<sup>22</sup> Benjamins, V.R. The ontological engineering initiative (KA)<sup>2</sup>, *Formal Ontology in Information systems*. IOS Press, Amsterdam. 1998.



Na Engenharia do Conhecimento, a ‘Visualização’ diminuiu as surpresas com relação ao usuário. O usuário, ao observar o *design* da Interface, via a interpretação que a equipe de Engenharia do Conhecimento e de Desenvolvimento deu aos seus desejos; assim, ele podia sugerir com segurança as alterações necessárias. Esse aspecto ‘emocional’ é um dos resultados do processo de Visualização.

A seguir, observou-se a necessidade do programador na FS ter conhecimentos de Inteligência Artificial para conseguir interpretar corretamente e adequadamente a análise de requisitos e o diagrama de casos de uso<sup>23</sup>. Notou-se também que o conhecimento do domínio influenciava os resultados. Por isso, decidiu-se por uma etapa anterior à ‘Visualização’, que foi denominada ‘Compartilhamento do Conhecimento’. Dessa maneira, todos os integrantes da equipe do projeto obtinham um conhecimento mínimo sobre a IA, sobre as linguagens de programação, sobre o domínio de aplicação, etc. Foi nesta fase que se pôde observar que alguns procedimentos funcionavam como na Engenharia do Conhecimento.

Na seqüência, observou-se que esses procedimentos incentivavam a criatividade da equipe (as melhores inovações criadas pela equipe surgiram do intercâmbio de informações e exercício de visualização). Inovação exige investimento; por isso foi preciso estabelecer prioridades de desenvolvimento. A ‘Identificação de Relevâncias’ tornou-se o terceiro procedimento da metodologia.

Assim, ‘Compartilhamento do Conhecimento, Visualização e Identificação de Relevância’ tornaram-se a base da metodologia desenvolvida neste trabalho, para as áreas de Engenharia do Conhecimento e Engenharia de Ontologias.

A metodologia desenvolvida nesta pesquisa, denominada *Engenharia da Mente*, foi adequada ao processo da aquisição de conhecimento, visando permitir que o especialista do domínio e o engenheiro do conhecimento trabalhem em sincronia, seja em ambientes cooperativo de trabalho na Web, seja fora dele.

É importante ressaltar que a equipe de pesquisadores formavam, no início, um grupo de pesquisa junto ao Programa de Pós-graduação em Engenharia de Produção e Sistemas (PPGEP/ UFSC). No decorrer da pesquisa, constituíram um Instituto de Pesquisa Privado (IJURIS) e, recentemente, a empresa de Tecnologia WBSA, onde as pesquisas realizadas no Instituto foram aplicadas e adequadas ao mercado de tecnologia da informação.

---

<sup>23</sup> Diagrama de casos de uso descreve a seqüência das interações entre atores e o sistema necessário para realizar a entrega do serviço que satisfaça ao objetivo.

## 1.4 ESTRUTURA DO TRABALHO

A Estrutura do Trabalho está dividida em quatro Capítulos mais os Resultados e as Conclusões. No Capítulo 1, encontram-se a Introdução, os Objetivos, a Motivação e um resumo da Metodologia aplicada nesta pesquisa.

O Capítulo 2 apresenta a Revisão da Literatura, com os elementos teóricos que conceituam e definem a Engenharia do Conhecimento, a Engenharia de Ontologias e a Engenharia da Mente, compreendendo a sincronicidade entre as duas anteriores.

No Capítulo 3, descrevem-se detalhadamente as Metodologias da Engenharia da Mente. O capítulo 4 apresenta a Metodologia aplicada em Sistemas Baseados em Conhecimento e na construção de Ontologias. Ao final, descrevem-se os Resultados da aplicação da metodologia, as Conclusões e Desenvolvimentos Futuros,

Os Anexos I, II e III trazem três artigos referentes à tese, publicados em congressos internacionais.

## CAPÍTULO 2

### 2 REVISÃO DA LITERATURA

#### 2.1 ENGENHARIA DO CONHECIMENTO

Engenharia do Conhecimento (EC): termo usado para descrever o processo global de desenvolvimento de Sistemas Baseados no Conhecimento (SBC). Tipicamente envolve uma forma especial de interação entre os construtores do sistema, chamados engenheiros do conhecimento (equipe multidisciplinar, necessariamente composta por analista(s) de sistemas) e um ou mais especialistas em alguma área. Multidisciplinar por natureza, a Engenharia do Conhecimento compreende tipos de pesquisas difíceis de classificar numa abordagem bem delimitada (VALENTE, 1995)<sup>24</sup>.

O objetivo do processo de EC é capturar e incorporar o conhecimento fundamental de um especialista do domínio, bem como seus prognósticos e sistemas de controle. O processo envolve reunião de informações, familiarização do domínio, análise e esforço no projeto. Além disso, deve-se codificar, testar e refinar o conhecimento acumulado.

Embora se pesquise a EC há certo tempo, ainda não existem caminhos padronizados que levem a um desenvolvimento seguro e preciso, o que exige muita criatividade e envolve mais arte que ciência.

Basicamente, a EC abrange questões sobre processos de aquisição e representação do conhecimento. O engenheiro do conhecimento extrai dos especialistas procedimentos, estratégias e regras práticas para solução de problemas, e constrói o conhecimento obtido em um sistema inteligente. O resultado é um programa que soluciona problemas à maneira dos especialistas humanos.

---

<sup>24</sup> Valente, André. *Legal Knowledge Engineering: A modelling approach*. IOS Press:Amsterdam. 1995. Pág. 8.

Ao se identificar a Aquisição de Conhecimento (AC), processo pelo qual acontece a extração do conhecimento, como uma geradora de dificuldades no processo de construção de SBC, ela passa a se constituir numa importante área de pesquisa na Inteligência Artificial.

Os Sistemas Baseados no Conhecimento (SBC) possuem características específicas. Segundo Newell (1982)<sup>25</sup>, um SBC deve conter a descrição do sistema sob duas perspectivas distintas: a do conhecimento, processável pelo homem, e a simbólica, processável pelo computador. O autor distinguiu a importância de separar a análise e a modelagem do método de resolução do problema e a atividade de representar esse método em um formalismo que seja computacionalmente eficiente.

Assim, os SBCs possuem como principais características uma base de conhecimento e um mecanismo de raciocínio capaz de realizar inferências sobre essa base e obter conclusões a partir desse conhecimento. É na base de conhecimento que fica o conhecimento do especialista do domínio. Na base, o conhecimento também é organizado e disponibilizado de forma permanentemente acessível, a fim de ser facilmente recuperado e utilizado por todos os usuários do sistema. Rezende (2003)<sup>26</sup> ressalta ainda que os SBCs requerem conhecimento sobre a habilidade, a experiência e as heurísticas usadas pelo especialista, além de uma profunda interação com este profissional. Por essa razão, a utilização de SBCs é crescente, principalmente nas organizações com amplas bases de dados para o gerenciamento do conhecimento.

Na construção de um SBC, existem diferentes formas de se representar o conhecimento. A etapa inicial de construção de um SBC é denominada Engenharia do Conhecimento, na qual se definem as metodologias de aquisição e representação do conhecimento. A seguir, descrevem-se algumas das metodologias e suas aplicações.

### 2.1.1 Aquisição e Representação do Conhecimento na Inteligência Artificial

Primeiramente, deve-se esclarecer que Representação de Conhecimento (RC) são métodos utilizados para modelar os conhecimentos de especialistas em algum campo, de forma eficiente, e colocá-los pronto para serem acessados pelo usuário de um sistema

---

<sup>25</sup> Newell, A. The Knowledge Level. *Artificial Intelligence*. V. 18, 1982. Pág. 87-127.

<sup>26</sup> Rezende, Solange O., Pugliesi, Jaqueline B., Varejão, Flávio M. *Sistemas Inteligentes: fundamentos e aplicações*. Organização, Solange Oliveira Rezende – Barueri, SP: Manole, 2003. Pág. 16.

inteligente. Ou seja, RC é uma combinação de estruturas de dados e procedimentos interpretativos que, se usada corretamente em um programa, terá uma conduta inteligente.

O primeiro passo em direção à inteligência (artificial), desde o desenvolvimento da lógica, foi dado por Aristóteles (384-322 a.C.), quando o filósofo começou a explicar e a codificar certos estilos de raciocínio dedutivo, que chamou de silogismos.

Muitos foram os matemáticos e lógicos que contribuíram para a formação do que se denomina hoje Inteligência Artificial. Segundo Margaret A. Boden *apud* Arís et al (2003), foi em 1950, quando Turing escreveu sobre a possibilidade das máquinas adquirirem inteligência, que nasceu a IA. Outros autores dizem que a IA nasceu em 1943, quando MacCulloch e Pitts propuseram um modelo de neurônio do cérebro humano e animal. Essa idéia, junto à hipotética máquina de Turing, e as concepções da Teoria da Informação de Shannon, conduziram Wiener ao desenvolvimento da Cibernetica.

A revolução da IA começou em 1956, na cidade de Hannover, quando se reuniram dez cientistas representantes de diferentes áreas do conhecimento, como a matemática, a neurologia, a psicologia e engenharia eletrônica. O ponto em comum desse grupo multidisciplinar era o fato de que todos usavam o computador para simular diferentes aspectos da inteligência humana. O termo ‘Inteligência Artificial’ se deve a McCarthy, um dos integrantes do grupo<sup>27</sup>. Entretanto, a IA surgiu de idéias filosóficas, científicas e tecnológicas bem anteriores a 1956 – como a lógica, de 23 séculos atrás.

Nos primeiros anos da IA, os modelos conceituais de base de dados foram marcados, em sua maior parte, por um modelo improvisado e inconsistente, o que gerou muitos problemas práticos de integração de base de dados, que se enfrenta inclusive nos dias de hoje. Os primeiros investigadores concentravam-se excessivamente na resolução de problemas gerais, e os processos de raciocínio eram simulados mediante técnicas uniformes – como a busca heurística e a demonstração de teoremas – que resultaram ineficazes em soluções de problemas de dimensões consideráveis. Tudo isso levou a uma série de fracassos e decepções que não encontrou outra linha de escape até os anos 1970, quando começaram a se desenvolver os primeiros Sistemas Especialistas, posteriormente denominados Sistemas

---

<sup>27</sup> Os principais integrantes do grupo eram: John McCarthy; Marvin Minsky (fundador do laboratório de IA do MIT); Claude Shannon (laboratórios Bell); Nathaniel Rochester (IBM); Allen Newell (primeiro presidente da AAI); Herbert Simon (Premio Nobel da Carnegie-Mellon University).

Baseados no Conhecimento. No final dos anos 1980, ocorreu um retorno aos estudos das Redes Neurais (ARÍS et al, 2003)<sup>28</sup>.

Davis, Shorbe e Szolovits (1993)<sup>29</sup> definem Representação do Conhecimento (RC) como algo que substitui o objeto ou fenômeno real de modo a permitir a uma entidade determinar as conseqüências de um ato pelo pensamento em vez da realização deste ato.

Pode-se entender a RC como uma forma sistemática de estruturação e codificação do que se sabe sobre uma determinada aplicação. Contudo, segundo Rezende (1993)<sup>30</sup>, ao contrário de uma codificação qualquer, uma RC deve apresentar as seguintes características:

- Ser compreensível ao ser humano, pois, caso seja necessário avaliar o estado de conhecimento do sistema, a RC deve permitir a sua interpretação;
- Abstrair-se dos detalhes de como funciona internamente o processador de conhecimento que a interpretará;
- Ser robusta, isto é, permitir sua utilização mesmo que não aborde todas as situações possíveis;
- Ser generalizável, ao contrário do conhecimento em si que é individual. Uma representação necessita de vários pontos de vista do mesmo conhecimento, de modo que possa ser atribuída a diversas situações e interpretações.

A RC é um dos problemas cruciais de IA, pois não existe uma teoria geral de representação do conhecimento, embora pesquisadores de IA estudem muitas técnicas de representação do conhecimento.

De fato, o desempenho e o grau de ‘inteligência’ de um SBC estão diretamente ligados à qualidade e usabilidade do conhecimento contido no sistema. Segundo autores como Alty (1984), Fikes (1985), Woods (1986) e Wah (1989), a ênfase sobre o conhecimento nos SBCs fez com que a área de RC assumisse papel fundamental, levando ao desenvolvimento de vários formalismos de representação do conhecimento (REZENDE, 2003b)<sup>31</sup>.

---

<sup>28</sup> Arís, Enrique P., González, Juan L. S., Rubio, Fernando M. *Lógica Computacional*. Thomson Editores Spain: Madrid, 2003.

<sup>29</sup> Davis, R., H. Shorobe, & P. Szolovits. What is a Knowledge representation? *AI Magazine* 14 (1), 17-33. 1993.

<sup>30</sup> Rezende, Solange Oliveira (org.). *Sistemas Inteligentes, Fundamentos e aplicações*. São Paulo: Manole. 1993.

<sup>31</sup> Rezende, Solange Oliveira (org.). *Sistemas Inteligentes, Fundamentos e aplicações*. São Paulo: Manole. 2003b.

Trabalhar com RC em IA envolve o projeto de várias classes de estruturas de dados para armazenamento de informações, bem como o desenvolvimento de procedimentos para manipulação dessas estruturas.

A representação é um conjunto de convenções sintáticas e semânticas que possibilita descrever coisas. Os símbolos, que podem ser usados e as maneiras de arranjá-los, são a representação sintática específica, enquanto que a representação semântica específica, o significado incorporado nos símbolos, são representados pela sintaxe. Para saber qual o tipo de representação usar em um determinado problema, considera-se o uso final do conhecimento que pode ser a aquisição de mais conhecimento ou a recuperação de conhecimento ou o ‘raciocínio’ sobre esse conhecimento para se obter uma solução.

#### 2.1.1.1 A Representação do Conhecimento na técnica Raciocínio Baseado em Casos (RBC)

O Raciocínio Baseado em Casos constitui uma poderosa ferramenta de utilização das mais variadas experiências humanas. É um tipo de raciocínio muito usado, 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 ajuda as pessoas a pensar sobre o presente e o futuro. A técnica tem os seus fundamentos nas pesquisas da ciência cognitiva (Schank, 1982)<sup>32</sup>, e foi amplamente difundida na Inteligência Artificial (Kolodner, 1993 e Amond; Plaza, 1994)<sup>33 34</sup>.

O processo característico do RBC consiste em: identificar a situação atual, buscar a experiência mais semelhante na memória e aplicar o conhecimento dessa experiência passada na situação atual. O RBC é uma técnica da IA muito adequada para a modelagem do conhecimento. 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 desses casos para determinar a necessidade de adaptação; reutilização do caso adaptado para resolver o

<sup>32</sup> Schank, R. *Dynamic Memory: A theory of learning in computers and people*. New York, Cambridge University Press, 1982.

<sup>33</sup> Kolodner, Janet L. *Case-based reasoning*. Morgan Kaufmann Publishers, Inc. 1993. San Marco, CA 94403.

<sup>34</sup> Amond, A.; Plaza, E., “Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches”. *AI Communications*, 17(1), 1994.

problema de entrada; avaliação da solução do problema de entrada; inclusão do caso adaptado na memória de casos (aprendizagem). (Ver figura 1):

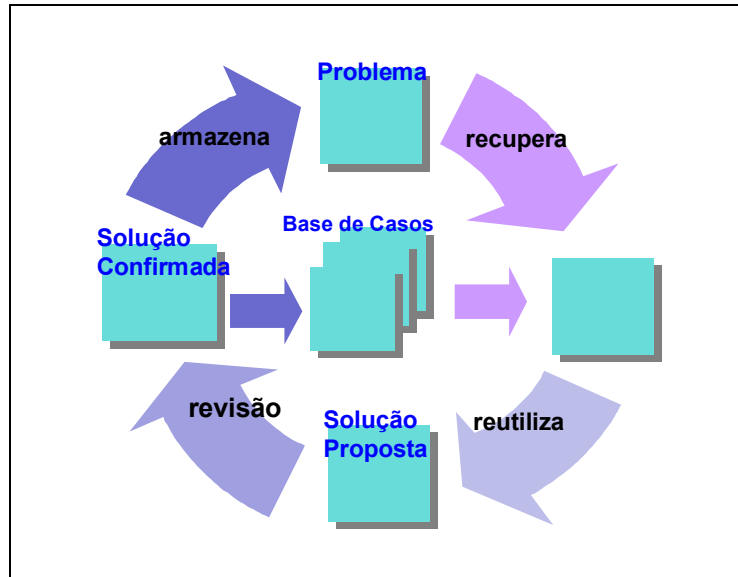


Figura 1 – Ciclo do Raciocínio Baseado em Casos (RBC).

Esses procedimentos podem ser simplificados 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.

A representação dos casos num sistema de RBC é essencialmente a representação do conhecimento. Existem outros momentos em que algum conhecimento especializado 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 esses casos. A execução das etapas está fortemente ligada entre si, e esta, por sua vez, depende intrinsecamente das outras etapas do sistema. As outras etapas do sistema, recuperação, adaptação e aprendizagem, devem ser projetadas simultaneamente à representação. Orientando-se sempre pelo objetivo



da aplicação, avaliam-se os benefícios da representação de conhecimento sugerida para todas as etapas complementares do sistema.

Como o conhecimento nos Sistemas 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 eficiente. 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**. O vocabulário de indexação diz qual o tipo de vocabulário pode ser usado 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.

Segundo Kolodner (1993)<sup>35</sup>, com relação ao vocabulário, na indexação das características, foram observados os seguintes passos:

- 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 à 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 quer recuperar.

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

- 1) Os índices precisam representar o contexto, as soluções e os problemas do domínio;
- 2) A representação precisa ser proveitosa, isto é, ela precisa dirigir a proposta para a qual o caso será usado;

---

<sup>35</sup> Ver detalhes em Kolodner, Janet L. *Case-based reasoning*. Morgan Kaufmann Publishers, Inc. San Marco, CA94403. p. 164. 1993.

- 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.

Em resumo, os índices precisam ser representativos, úteis, aplicáveis genericamente e reconhecíveis, possibilitando descrever cada tipo de situação no qual ele tem o potencial de ser aplicado.

Segundo Kolodner (1993)<sup>36</sup>, existem dois conjuntos de teorias que podem ser examinadas para determinar o conteúdo dos índices:

- 1) A teoria funcional ou pragmática: ponto de vista que examina os tipos de casos que são avaliados e a tarefa que ele pode suportar para identificar as dimensões e símbolos que, se usados como índices, permitirão que os casos avaliados cubram a série de tarefas designadas.
- 2) A teoria 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 dessas análises dizem 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 temas para se ter em mente enquanto os 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.

---

<sup>36</sup> Ver detalhes em Kolodner, Janet L. Ob. cit. Pág. 208.

Os índices para um caso são um subconjunto de descrição ou representação de casos. Por essa 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.

#### 2.1.1.2 RC2D® e PCE®

O RC2D®<sup>37</sup> é uma metodologia de representação de conhecimento cujo enfoque está centrado no modelo teórico elaborado na evolução acima descrita. É um processo dinâmico de aquisição do conhecimento de textos, definido através da elaboração de um vocabulário controlado e um dicionário de termos, associado a uma análise de frequência das palavras e expressões indicativas do contexto (Hoeschl, 2001)<sup>38</sup>. Outros desenvolvimentos semelhantes, como a UNL (*Universal Networking Language*), a *Semantic Web* e a *WordNet* trabalham com a marcação do texto integral, definindo a sua classificação sintática. No RC2D, isso não é necessário; apenas as partes realmente significativas são representadas. Essa técnica realiza comparações entre o contexto dos documentos, possibilitando a realização de uma busca mais precisa e com maior qualidade; foi desenvolvida especialmente para apoiar implantações que utilizem algoritmos com Inteligência Artificial (Hoeschl, 2001)<sup>39</sup>.

Segundo Hoeschl et al (2003), um sistema que utilize o PCE:

[...] permite, dentre outras atividades, que se interrogue o sistema sobre uma nova demanda [...] e obtenha-se uma listagem dos documentos referentes ordenados por similaridade de conceitos e contexto [...]. Esta tecnologia aprimora o clássico raciocínio baseado em casos, englobando aos seus algoritmos o banco de dados (informações estruturadas) e a mineração de textos (não-estruturadas)<sup>40</sup>.

---

<sup>37</sup> Todos os direitos reservados da marca e uso da aplicação estão reservados a WBSA – Sistemas Inteligentes S.A. Copyright 2001.

<sup>38</sup> Hoeschl, Hugo C., et. al. Olimpo: Contextual Structured Search to improve the representation of UN Security Council with Information Extraction methods. Op. Cit.

<sup>39</sup> Hoeschl, Hugo C., et. al. Op. cit.

<sup>40</sup> Hoeschl, Hugo. C. et al, 2003. Structured Contextual Search For The Un Security Council. *Proceedings of the fifth International Conference On Enterprise Information Systems*. Anger, France, v.2. p.100 – 107.

Todos os sistemas citados apresentam características em comum: um domínio de aplicação específico e documentos semi-estruturados (jurisprudência dos Tribunais Brasileiros e Resoluções do Conselho da ONU). Surgiu, então, um novo desafio: organizar bases de conhecimento para recuperação inteligente de fontes e assuntos diversos, originários da Internet. Estudos estratégicos em Gestão do Conhecimento, aliados a técnicas de Inteligência Artificial (RBC e Mineração de Dados), permitiram o desenvolvimento de um sistema de Gestão de Conhecimento (Ribeiro et al, 2004)<sup>41</sup>, o que possibilitou a aplicação da metodologia ora desenvolvida neste trabalho.

O RC2D® consiste num processo dinâmico de análise do contexto geral que envolve o problema enfocado. Ele realiza comparações entre o contexto dos documentos, possibilitando a realização de uma busca mais precisa e com mais qualidade. Além disso, os documentos são recuperados através de índices pré-determinados, que podem ser valorados pelo usuário quando da consulta.

A técnica foi desenvolvida para aplicação no Sistema Olimpo®. O aperfeiçoamento da interface de análise se dá através da utilização do RC2D®, que realiza o processo de representação por diversas vezes de acordo com a consulta do usuário. (Ver figura 2, pág. 36).

Segundo Hoeschl (2001), repetindo esse procedimento empírico e de forma cíclica, são geradas a pré-lista 1 e 2. A primeira é formada por expressões encontradas no teor das resoluções, depois de minuciosa leitura e análise dos documentos. A segunda é construída pela adaptação da consulta do especialista às expressões encontradas nos documentos da base do conhecimento. Depois, são testadas e analisadas estatisticamente, de forma a que se analise sua real função como elemento de referência para recuperação e indexação documental. Expressões que ocorrem com uma frequência muito alta, próxima a 100% das resoluções, têm pouca utilidade para diferenciá-las. Aquelas expressões que aparecem somente em um ou dois documentos, salvo se forem extremamente caracterizadoras, consideram-se também pouco úteis ao estabelecimento de contexto. Então, procura-se eliminar os extremos, selecionando o conjunto de expressões que possuam significância na comparação documental.

---

<sup>41</sup> Ribeiro, Marcelo Stopanovski; Mattos, Eduardo da Silva; Bueno, Tânia Cristina D'agostini; Hoeschl, Hugo Cesar. KMAI- Knowledge Management With Artificial Intelligence. The Symposium on Professional Practice in AI in the First IFIP International Conference On Artificial Intelligence Application And Innovations. Toulouse, 2004.

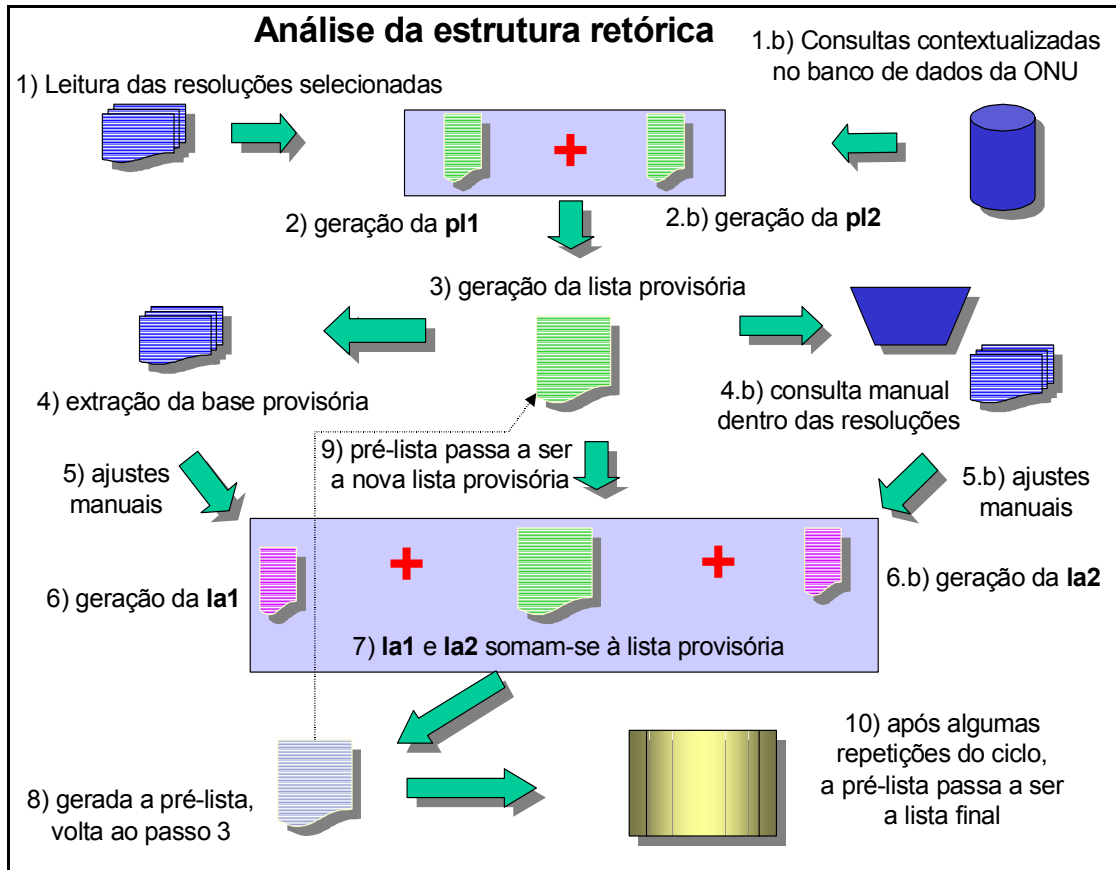


Figura 2 – Análise da Estrutura Retórica.

As duas listas são somadas e o procedimento efetuado, até que se obtenha uma lista final de expressões-chave que passará por um processo de validação, permitindo a realização da consulta de modo flexível e eficiente, apresentando níveis satisfatórios de similaridade contextual.

A PCE® é uma metodologia que permite uma busca em linguagem natural através do contexto das informações contidas na base do conhecimento, rompendo, assim, o paradigma de busca por meio de palavras-chave e conectores lógicos, possibilitando ao usuário descrever um número de caracteres elevados por cada consulta, permitindo, dessa maneira, uma concepção mais elaborada da busca.

Citando Hoeschl (2001)<sup>42</sup>:

<sup>42</sup> Hoeschl, Hugo César. Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU. Op. cit.

A pesquisa é considerada 'contextual' e 'estruturada' pelas seguintes razões: 1) É levado em consideração o contexto dos documentos armazenados quando da formação de estrutura retórica do sistema; 2) Este contexto norteia o processo de ajuste da entrada bem como da comparação e seleção dos documentos; 3) Quando da elaboração da consulta, a entrada não está limitada a um conjunto de palavras, ou à indicação de atributos, podendo assumir o formato de uma questão estruturada pelo conjunto de um longo texto somado à possibilidade de acionamento de pesos dinâmicos sobre atributos específicos, que funcionam como 'filtros' e fazem uma seleção preliminar dos documentos a serem analisados.

Para isso, segundo Hoeschl (2001), devem-se atribuir pesos a esses índices, para então serem ativados os 'filtros' que diminuem a área de varredura do sistema, permitindo uma recuperação mais especializada. (Ver figura 3).

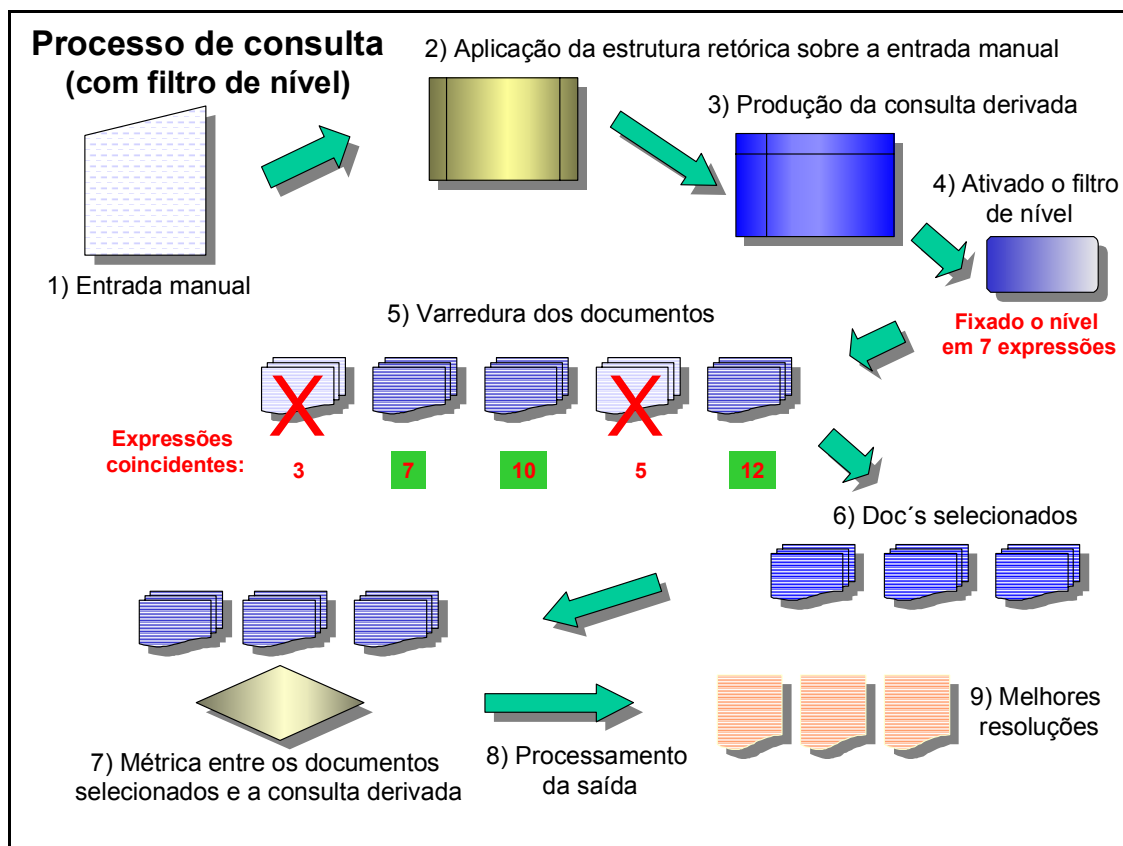


Figura 3 – Pesquisa Contextual Estruturada.

A PCE é 'pesquisa' no sentido de configurar um sistema que pesquise informações, ou as busque, ou as recupere. É 'contextual' porque o conhecimento, por ela utilizado, é representado contextualmente; assim, a pesquisa é realizada na base, levando em consideração o contexto apresentado quando da solicitação da pesquisa. É 'estruturada' em função de analisar tanto o contexto da solicitação quanto aquele dos documentos nos quais a pesquisa é

realizada. Para que a técnica funcione adequadamente, é fundamental que a representação do conhecimento seja contextualizada, e que ocorra dinamicamente.

Hoeschl (2001) classifica o processo RC2D em uma etapa teórica e outra empírica:

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

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

Para Hoeschl (2001), a PCE assume uma autonomia axiológica em relação ao RBC e à IR (*Information Retrieval*):

Não se pode afirmar que esta técnica consiste em mera recuperação de informação (IR), eis que está qualificada por procedimentos inteligentes. Da mesma forma, não se pode afirmar que o trabalho ora sob análise está restrito aos círculos do RBC, eis que vai além dele. Decorrência lógica é que não se pode enquadrá-lo como RBC textual. Seu perfil está mais próximo de um sistema inteligente de gerenciamento de bases de dados (IDBMS).

Algumas observações sobre a PCE e o RBC merecem melhor aprofundamento. O RBC é uma metodologia que possui etapas específicas. (Ver item anterior 2.1.1).

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

## 2.2 ENGENHARIA DE ONTOLOGIAS

No campo da Inteligência Artificial, as Ontologias são utilizadas como uma forma de representação que visa compartilhar o conhecimento de um mesmo domínio comunicado entre pessoas e sistemas. A Ontologia pode ser definida como uma linguagem formal que utiliza categorias e hierarquias para representar conceitos comuns legíveis por máquinas.

As Ontologias procuram refletir não apenas a memória léxica humana, mas também estabelecer relações conceituais baseadas no contexto do domínio trabalhado, estabelecendo uma rede constituída por conceitos unidos por diferentes relações semânticas. Elas visam os conhecimentos consensuais, desenvolvidos através de processo cooperativo, e buscam trazer um entendimento comum de determinado domínio através da relação entre palavras ou expressões indicativas que vão representar este domínio.

Nas Ciências Filosóficas, Ontologia é o ramo que estuda o ‘ser’ enquanto uma entidade que existe, e suas relações com as outras entidades; nesse sentido, tem forte comprometimento com a realidade. Esse conceito foi adaptado para as metodologias de Representação do Conhecimento com o objetivo, justamente, de focar as atividades na representação do que existe, do que é utilizado, e não do que deveria existir ou de como deveria ser. Assim, no campo da tecnologia, o conceito faz referência à formulação do esquema conceitual, dentro de um certo contexto, com a finalidade de facilitar a comparação, a classificação, a organização e o armazenamento dos textos analisados. Como a Representação do Conhecimento é um dos principais procedimentos da Inteligência Artificial, as Ontologias aparecem como um eficiente meio de efetivar tal representação.

Existem diferentes metodologias e técnicas para representação do conhecimento. No caso específico deste estudo, trabalha-se com a metodologia Representação do Conhecimento Contextualizada Dinamicamente (RC2D) (ver item 2.1.1.2); portanto, todos os conceitos utilizados nesta abordagem serão referentes a essa metodologia.

De forma simplificada, pode-se considerar a RC2D uma seqüência de ações dinâmicas de análise do contexto geral que envolve o problema enfocado. Consiste no processo de representação do conhecimento e respectivo ajuste aos objetivos do sistema, a fim de que o seu funcionamento seja potencializado.



Representar todo o conhecimento existente no mundo é tarefa infundável. Assim, para representá-lo e interpretá-lo adequadamente, é necessário restringir a atenção a uma pequena parte deste mundo, denominada ambiente de aplicação do sistema. Conhecido o ambiente, o próximo passo é a definição dos contextos de aplicação. Nesses contextos serão definidos os objetivos estratégicos do sistema, ou seja, o que se pretende atingir.

A partir da definição dos contextos e conhecidos os objetivos, são estabelecidos os domínios de aplicação do sistema. Os domínios são especificamente as áreas que detém o conhecimento que se pretende representar, recuperar e analisar. Para representar os contextos e os domínios é que são criadas as Ontologias. O ambiente de aplicação do sistema é o espaço conceitual do conhecimento no qual será desenvolvida a aplicação do sistema. Por contexto, entende-se a definição de um ambiente conceitual de aplicação do sistema. Ou seja, consiste na delimitação do escopo de atuação do sistema.

A construção de Ontologias exige a contextualização das Expressões Indicativas que venham a representar o conhecimento que se pretende recuperar. A contextualização propõe favorecer o entendimento de palavras ou Expressões Indicativas dentro de domínios específicos num contexto pré-estabelecido. Por domínio, entende-se uma subdivisão do contexto com significância efetiva para os objetivos do sistema. É o que se denomina Engenharia de Ontologias.

Para obter um conhecimento considerado crítico para a construção de um Sistema Baseado no Conhecimento que seja utilizável, o processo de Engenharia do Conhecimento segue necessariamente um círculo – através do qual se introduz, de modo incremental, um novo conhecimento na base. Infelizmente, salvo nos casos de domínios muito simples, é raro o caso em que se avalie um conhecimento por completo. Com frequência, descobrem-se novos conhecimentos durante a construção do sistema, mesmo depois de concluída a base.

Na interação dos novos conhecimentos, surge a incerteza, a ambigüidade e a incoerência que permeiam cada estágio do processo de Engenharia do Conhecimento. Os resultados são imprevisíveis – muitas vezes de elaboração intuitiva – o que tornam difíceis as suas consistências e continuidades. Para Santos (2003)<sup>43</sup>, durante a fase de aquisição do conhecimento, a dificuldade-chave é preservar a semântica da base enquanto se introduz um novo conhecimento.

---

<sup>43</sup> Santos Jr, E., Santos, E. S., Shimony, S. E. Implicitly preserving semantics during incremental knowledge base acquisition under uncertainty. *International Journal of Approximate Reasoning*. Volume 33. Number 1. April 2003. Pag. 71-94. Available at <http://www.computerscienceweb.com>.

Na abordagem desta pesquisa, durante a aquisição do conhecimento em Sistemas Baseados no Conhecimento, utilizou-se uma ferramenta da Engenharia de Ontologias que permitiu uma constante acomodação do novo conhecimento à base de conhecimento. Entretanto, para manter a consistência do conhecimento, durante o desenvolvimento das Ontologias, fez-se necessário elaborar uma metodologia que permitisse maior sincronização de conhecimento entre os engenheiros do conhecimento e os especialistas do domínio – o que será apresentado no Capítulo 3.

### 2.2.1 Relações da Engenharia de Ontologias

Estudos atuais denominados ontológicos vêm sendo retomados como substrato às pesquisas de modelagem cognitiva de conhecimentos. O trabalho de Guarino e Welty (2000) enfoca as Ontologias do ponto de vista do desenvolvimento de Sistemas Baseados em Conhecimento (SBC) (*Knowledge-Based Systems* – KBS) e discute como os princípios da Ontologia formal e Engenharia Ontológica podem ser usados na prática da Engenharia do Conhecimento.

Guarino e Welty (2000) apontam em seu trabalho uma forte conexão entre Ontologia formal e análise conceitual. Os autores afirmam que a Engenharia do Conhecimento pode contribuir para estabelecer as fundações de um campo emergente: a Engenharia Ontológica.

Da leitura de trabalhos dessa natureza, pôde-se constatar que novos campos estão surgindo, dedicados a tratar de realidades essencialmente presentes nos domínios da Ciência da Informação.

A interpretação desse fenômeno resultaria numa conclusão que aponta para a impossibilidade de hegemonia e manutenção dos domínios de velhos campos do conhecimento, desde que eles não apresentem resultados que atendam ao interesse do mercado ao qual a ciência contemporânea também se acha a serviço. Daí a urgência do desenvolvimento de pesquisas envolvendo os conhecimentos da área da Ciência da Computação.

### 2.2.1.1 Ontologias e Inteligência Artificial

Em particular, uma grande dificuldade da Inteligência Artificial (IA) é relacionada com a definição da terminologia empregada para nomear os conceitos e as relações (REZENDE, 2003)<sup>44</sup>. Pelo fato dos mecanismos utilizados pela IA terem se tornado cada vez mais padronizados, as teorias inseridas nesses mecanismos passaram a ser foco de atenção. Essas teorias, freqüentemente chamadas ‘Bases de Conhecimento’, foram construídas de maneira que refletissem o senso-comum de uma forma declarativa, tirando vantagem da potencialidade dos sistemas de raciocínio automatizados.

Foi John McCarthy (1980)<sup>45</sup> quem primeiro reconheceu a sobreposição entre o trabalho existente na Ontologia Filosófica e a atividade de construção das teorias lógicas nos Sistemas de IA. McCarthy afirmou, já em 1980, que os construtores de sistemas inteligentes precisavam primeiramente listar tudo o que existe; com isso, construir uma Ontologia do nosso mundo.

Entretanto, construir Ontologias é muito mais que um processo tecnológico. A razão pela qual as Ontologias se tornaram tão populares é devido ao que elas prometem: um entendimento compartilhado e comum de um mesmo domínio que pode ser comunicado entre pessoas e sistemas. Elas visam o domínio do conhecimento consensual, e o desenvolvimento delas é freqüentemente um processo cooperativo envolvendo diferentes pessoas em diferentes locais (DING et al, 2002)<sup>46</sup>.

Neste trabalho, desenvolveu-se uma visão diferenciada sobre a fundamentação da Ontologia na Inteligência Artificial. Buscou-se resgatar uma visão de qualidade, através da construção de relações entre palavras que revelassem um contexto consensual, uma visão compartilhada de um mesmo objeto. Assim, as pessoas, que concordam em aceitar uma Ontologia, estão dizendo que estão comprometidas com essa Ontologia. É um comprometimento de mente e alma, uma busca profunda no inconsciente e a sua linguagem. Esse é o primeiro passo para a qualidade.

---

<sup>44</sup> Rezende, Solange Oliveira (org.). *Sistemas Inteligentes, Fundamentos e aplicações*. São Paulo: Editora Manole Ltda. 2003.

<sup>45</sup> McCarthy, John. *Circumscription – A Form of Non-Monotonic Reasoning*. *Artificial Intelligence*, 5:13, 1980. p. 27-39. Também disponível em: <http://www-formal.stanford.edu/jmc/circumscription.html>. Acesso em: 10 de agosto de 2004.

<sup>46</sup> Ding, Y., Fensel, D., Klein, M., and Omelayenko, B. *The semantic web: yet another hip?* *Data & Knowledge Engineering*, 41(2/3):205-227. 2002.

### 2.2.1.2 O que são Ontologias?

Na Filosofia, a Ontologia estuda o ‘ser enquanto ser’, isto é, o ser considerado independente de suas determinações particulares, naquilo que constitui sua inteligibilidade própria. Pode-se definir Ontologia como uma teoria do ser em geral, da essência do real. Muitas vezes, o termo Ontologia é empregado como sinônimo de ‘metafísica’.

Para Martin Heidegger (2004)<sup>47</sup>, autor de *Ser e Tempo*, obra que revolucionou a Filosofia no século passado, o *Dasein* designaria o ser humano enquanto está no mundo, ou seja, o *ser inter homines*, relacionado e comunicativo, o que levou os filósofos italianos a designar o *Dasein* como vocábulo *esserci*.

Heidegger (2004) reconheceu a importância da análise ontológico-fundamental do modo de ser no mundo e do *Dasein* pré-científico e exterior à Ciência, da crítica à Ontologia cartesiana, do conceito de existência no qual está a relação de convivência, o ser-com e o ser-com-outros-*Dasein* e da noção de facticidade.

Entretanto, não se pode ver a Ciência Moderna como restrita à Ontologia da coisa e, por isso mesmo, insuficiente para dar conta do mundo, do existente humano.

Nesse contexto observa-se a definição de Ontologia de Miguel Reale (2004)<sup>48</sup>, que salienta:

Ora, o encontro em todas as civilizações de certas características comuns demonstra que, não obstante haver grandes diferenças entre elas, há também determinados valores constantes universais que resistem às mutações históricas: são as que denomino invariantes axiológicas.

Se assim é, se da análise dos valores singulares podemos atingir as invariantes axiológicas, também podemos dizer que do estudo dos entes, em geral, é possível atingir o sentido último do Ser, não como uma verdade racional comprovada, como pretendia a Metafísica clássica, mas como conjectura, ou seja, como uma conclusão razoável (destacado do original). A diferença entre “racionalidade” e “razoabilidade” é que distingue a ciência, baseada na experiência, da ciência conjectural que dela só resulta obliquamente.

Como penso ter demonstrado em meu livro *Verdade e Conjetura*, a Metafísica somente pode ser compreendida como ciência conjectural, isto é, “como se” se originasse da experiência.

<sup>47</sup> Heidegger, Martin. *Ser e Tempo*. Vol I e II. São Paulo: Editora Vozes, 2004.

<sup>48</sup> Reale, Miguel. (2004). *Perspectiva e Teoria do Ser*. Disponível em : <http://www.miguelreale.com.br/>. Acesso em 30 de maio de 2004.

A visão virtual do ser depende, em suma, da perspectiva segundo a qual é ele o observado, não se podendo, porém, afirmar, como fez Heidegger, que, desse modo, não se chega a uma “teoria do ser” (Ontologia), mas a uma “teoria dos entes” (Metafísica). A distinção não tem razão de ser, pois, em última análise, o *Dasein*, a partir do qual ele pretende fundar a Ontologia – não é senão a compreensão do ser de conformidade com uma perspectiva, a de estar ele no mundo.

Uma outra visão é entre aqueles que se podem chamar reducionistas ou adequacionistas:

- Os adequacionistas buscam uma taxonomia das entidades em todos os níveis de agregação, do microcosmo ao macrocosmo, incluindo também, no meio, o mundo mediano das entidades de escala humana.
- Os reducionistas, ao contrário, vêem a realidade nos termos de somente um único e privilegiado nível de existência, procurando estabelecer a versão final do universo pela decomposição da realidade através da sua constituição mais simples, ou procuram “reduzir” de alguma outra maneira uma aparente variedade de tipos de entidades existentes na realidade.

Isso é importante, porque as raízes da Ontologia são intercaladas com os primeiros desenvolvimentos da Filosofia e cresceram junto com ela, assim como, recentemente, a Ontologia tornou-se interconectada com o desenvolvimento da Inteligência Artificial e a Ciência dos Sistemas de Informação.

### 2.2.1.3 Ontologias e Representação do Conhecimento

A representação do conhecimento é, assim como o raciocínio automatizado, uma das questões centrais da Inteligência Artificial (VALENTE, 1995)<sup>49</sup>. Davis et al (1993)<sup>50</sup> argumenta que a representação do conhecimento é um conjunto de comprometimentos epistemológicos, mais especificamente comprometimentos ontológicos. Em outras palavras,

---

<sup>49</sup> Valente, Andre. *Legal Knowledge Engineering: A Modelling Approach*. IOS Press, (Amsterdam) and Omsa (Tokyo), 1995. Pág. 35.

<sup>50</sup> Davis, R., Shrobe, H. e Szolovits, P. What is knowledge representation? *AI Magazine*, pages 17-33. 1993.

a especificação dos termos pelos quais o mundo é visto; uma forma de encontrar a abstração ‘certa’ para que alguém fale sobre ela e as coloque para resolver problemas.

Muitas teorias da Inteligência Artificial que representam domínios específicos como Medicina, Física, ou conceitos básicos (tempo, ação) são atualmente Ontologias ou tem um forte sabor ontológico.

É importante notar que as questões ontológicas são em nível do conhecimento e não em nível simbólico, isto é, não importa se alguém use redes semânticas, ou regras de produção. A perspectiva ontológica demonstra que a estruturação do conhecimento, de um certo modo, não implica que ele deverá ser implementado de um modo similar no nível simbólico. Então, não importa qual a técnica de representação do conhecimento a ser usada – este é um dos fatores para dar atenção às Ontologias: mover a discussão para longe das questões de nível simbólico.

### 2.2.2 Modelos de Desenvolvimento de Ontologias

Para a aquisição do conhecimento na elaboração da base de conhecimento de sistemas inteligentes, optou-se pela utilização de um ambiente Web para realização de um desenvolvimento cooperativo. Existem, hoje, três grandes desenvolvimentos de abrangência mundial que utilizam a Internet para o desenvolvimento de Ontologias. São: a UNL, a *Semantic Web* e a *WordNet*.

As três iniciativas buscam facilitar o processamento automático das informações contidas em documentos, permitindo aos computadores executar operações mais inteligentes e recuperar essas informações de forma mais eficiente.

A UNL e a *WordNet* possuem o foco no relacionamento das palavras entre si e, através deste relacionamento, contextualizam-nas dentro do assunto tratado pelo documento, preferencialmente através de mecanismos automáticos.

Já a *Semantic Web* trabalha com a contextualização do documento, através da marcação de partes específicas do mesmo. Essa marcação é feita pelo autor do documento.

### 2.2.2.1 UNL – *Universal Networking Language*

É necessário que os computadores tenham uma linguagem para que possam, a partir de elementos memorizados, processar conhecimento como os humanos. A *Universal Networking Language* (UNL) surgiu como uma linguagem que serve aos computadores alcançarem o conhecimento na forma de uma linguagem natural para os computadores. O sistema *Universal Networking Language* pode ser definido como uma linguagem eletrônica que torna possível a comunicação em diferentes línguas, acelerando a eliminação de barreiras lingüísticas. A UNL é, portanto, uma linguagem para computadores expressarem informações e conhecimentos escritos em linguagem natural.

O objetivo do projeto UNL é o desenvolvimento de um sistema de comunicação multilingüe. O sistema permite a qualquer pessoa acessar os conteúdos da Internet na sua própria língua, independentemente da língua em que foram originados, assim como difundir a informação escrita em uma língua numa outra qualquer (CARDEÑOSA, 2004)<sup>51</sup>.

O *Universal Networking Language* é um sistema de informação composto por um ‘codificador’, um ‘decodificador’ e um ‘visualizador’, residente na Web, compatível com os padrões mundiais de rede. É o maior mapeamento da linguagem humana em toda a sua história, e um dos maiores projetos multidisciplinares da Engenharia do Conhecimento, ao lado dos projetos GENOMA<sup>52</sup> e SETI<sup>53</sup>.

Com o objetivo de obter um tradutor automático na Internet e contornar esse tipo de restrição, a Organização das Nações Unidas (ONU) tem promovido, por mais de cinco anos, pesquisas e estudos, responsabilizando-se por cientistas em todo o mundo. O projeto visa construir um conceito para se obter uma linguagem comum, de todos os tipos de sistemas, denominado *Universal Networking Language* (UNL). O projeto é coordenado (*worldwide*) pela fundação UNDL ([www.undl.org](http://www.undl.org)), sediada em Genebra (Suíça). O progresso obtido no desenvolvimento da linguagem UNL permite dizer com otimismo justificado que as barreiras

---

<sup>51</sup> Cardeñosa, Jesus. Internet ya tiene esperanto. Disponível em: [http://www.webzinemaker.com/admi/m6/page.php3?num\\_web=1604&rubr=4&id=26997](http://www.webzinemaker.com/admi/m6/page.php3?num_web=1604&rubr=4&id=26997).

<sup>52</sup> O projeto GENOMA (patrimônio genético que se recebe através da fecundação de cada um dos progenitores) busca a identificação de todos os aproximadamente 30.000 gens do DNA humano. Com isso, espera determinar as seqüências de 3 bilhões de conexões químicas que constituem o DNA humano.

<sup>53</sup> SETI (Search for Extraterrestrial Intelligence) é um projeto de experimento científico, desenvolvido pela Universidade de Berkeley, que usa computadores conectados à Internet, no mundo todo, na busca por inteligência extraterrestre.

de uma comunicação universal já começaram a cair, considerando o pequeno período de tempo, e as diferenças lingüísticas não deverão ser mais nenhum obstáculo para o intercâmbio eficiente entre todos os povos, de todas as línguas e regiões do mundo.

A UNL é uma linguagem artificial na forma de rede semântica, permitindo que os computadores expressem e troquem algum tipo da informação. É composto por: um vocabulário formado por Palavras Universais (UW); um jogo das relações e dos atributos, representando a sintaxe da UNL; e a base de conhecimento de UNL, que é a semântica da linguagem. A representação da informação, usando UNL, é feita sentença a sentença. A sentença é representada por um *hypergraph*, em que as UW são os nódulos, relações e atributos que constituem os arcos do gráfico. Há pelo menos uma relação binária entre a UW de uma sentença e a classificação dos assuntos e dos objetos da sentença que é expressa, respectivamente, por relações e por atributos. Desta forma, um documento UNL supõe o formato de uma lista longa das relações entre conceitos.

O sistema de UNL é composto: por uma linguagem própria de UNL, pelos usuários da língua, e por um número de ferramentas básicas para finalidades de visualização e edição. Um usuário da língua consiste em dois elementos: o conversor e o desconversor. O desconversor é um sistema de tradução que cria um texto em uma língua fornecida, baseada em um texto representado na língua de UNL. O sistema do desconversor é composto por: um *software*; uma série de regras para a geração da língua; e dicionários específicos para essa língua. O conversor é um sistema para gerar informação na linguagem UNL, de uma fonte em uma outra língua. O sistema do conversor é composto por: um *software* de conversão; uma série de regras para a análise; e dicionários.

Qualquer um que tenha acesso à Internet poderá usar o sistema UNL para converter um texto escrito em sua própria linguagem dentro da UNL; do mesmo modo, todo o texto representado em UNL pode ser convertido em sua língua nativa. A infra-estrutura da rede global de informação na internet e a UNL são formas de facilitar a comunicação entre toda essa infra-estrutura. A conclusão do projeto UNL, e de fazê-lo conhecido mundialmente, é um tema de grande importância dentro dos objetivos da ONU, ao se promover a paz e o equilíbrio entre todas as nações em torno do mundo<sup>54</sup>.

---

<sup>54</sup> Thiecher, Adilson; Hoeschl, Hugo Cesar; Zimath, Patrícia Bonina. Interlegis: virtual network of communication and information that enlarges Brazil's democracy and citizenship. In: Traumüller, Ronald. (Org.). Lecture notes in computer science 2739. Berlin Heildeberg, 2003, v. 1, p. 75-79.



Sinteticamente, o Sistema UNL pode ser assim representado (ver quadro 1):

<b>Tool UNL</b>	<ul style="list-style-type: none"> <li>- <i>Universal Word's</i> (UWs): dicionário de conceitos simples e compostos;</li> <li>- <i>Relation</i>: ligação entre duas UWs;</li> <li>- Atributos: são expressões utilizadas que descrevem a subjetividade da sentença;</li> <li>- <i>Knowledge Base (KB)</i>: conjunto que define possíveis relações binárias que darão sentido (semântico) compreensível.</li> </ul>
<b>Linguagem Server</b>	<ul style="list-style-type: none"> <li>- <i>Deconverter</i>: sistema de geração de idiomas (<i>software</i>, regras de geração e dicionário do idioma nativo);</li> <li>- <i>Encoverter</i>: sistema gerador de UNL (<i>software</i>, regras de análise e dicionários).</li> </ul>
<b>Tools for UNL</b>	<ul style="list-style-type: none"> <li>- <i>UNL Editor</i>: produz um documento em UNL utilizando o seu próprio idioma.</li> <li>- <i>UNL Proxy</i>: é um tipo especial de servidor (aquela máquina que disponibiliza informações, acesso, etc.) de http (páginas de hipertexto que se acessa na rede Internet) que funciona dentro de uma máquina <i>firewall</i> (proteção utilizada e que permite apenas acesso autorizado). Um servidor Proxy torna as páginas da Web disponíveis em todas as línguas suportadas pela UNL.</li> <li>- <i>UNL Viewer</i>: é um “visualizador”, residente na Web, compatível com os padrões mundiais de rede.</li> </ul>

Quadro 1 – UNL – *Universal Natural Language*.

No que tange ao acesso e a precisão do conhecimento, a UNL disponibilizará em língua natural uma enorme variedade e/ou diversidade de informação. Tanto o tempo quanto o custo no acesso a essas informações serão sensivelmente reduzidas, por possuir *softwares* em um idioma/língua para computadores. Contudo, este só precisará interpretar as instruções escritas no idioma para poder executar suas funções.

Neste sentido, atuando por meio de tradução, conversão e acumulação de informações, a ferramenta UNL será referência de tecnologia a serviço da comunicação, derrubando as barreiras lingüísticas e promovendo a globalização do conhecimento.

#### 2.2.2.2 *Semantic Web*

A *Semantic Web* é uma extensão da Web atual, na qual a informação possui um significado mais bem-definido, permitindo que os computadores possam processar a informação contida nas páginas, e entendê-la, executando operações que facilitem o trabalho para os usuários. Ela torna exequível o processamento automático das informações contidas

em documentos, permitindo aos computadores efetuar operações mais inteligentes e recuperar as informações de forma mais eficiente.

A *Semantic Web* abrange questões como:

- 1) Da lingüística: a semântica é sobre o significado pretendido das palavras;
- 2) Da lógica: a semântica é sobre a interpretação das linguagens;
- 3) Computacional: Ontologias são (em sua versão formal) as teorias que especificam o significado das palavras ou algum outro tipo de significado pretendido. Elas podem e devem ser representadas em alguma linguagem lógica com uma semântica explícita (por exemplo: OWL);
- 4) Podem ser vistas como um objeto em desenvolvimento, distribuído, que seja feito ‘semioticamente’ de entidades relevantes (documentos, retratos, filmes, etc.), e dos ‘tags’ associados para aquelas entidades ou parte delas. Os ‘tags’ são partes da chamada ‘assinatura’ de uma Ontologia, ou seja, são ‘definidas’ em uma Ontologia. A combinação (*match*) semântica pode significar diversas coisas, mas a idéia básica poderia ser a combinação entre duas partes, independentemente da maneira que estas partes expressam seu interesse. Por exemplo: linguagens diferentes, mídia, bases de dados, etc.

A *Semantic Web* é muito mais abrangente que uma organização de sinônimos. Um dos recursos mais importantes da *Semantic Web* são aqueles chamados *web-services*. Ou seja, páginas na Web que não só fornecem informação estática, mas permitem que se efetue alguma ação ou mudança no mundo. A *Semantic Web* permite ao usuário localizar, selecionar, desenvolver, compor e monitorar *web-services* automaticamente.

A *Semantic Web* utiliza-se da flexibilidade da estrutura *Resource Description Framework* (RDF), na qual é possível descrever o conteúdo da informação disseminada na rede, fazendo-se afirmações sobre determinado objeto e identificando suas propriedades e valores. Cada objeto ou assunto é identificado por um Identificador Universal de Registro (URI) que assegura que as palavras na Web estejam relacionadas a apenas uma definição (BERNERS-LEE et al, 2001)<sup>55</sup>.

---

<sup>55</sup> Berners-Lee, T., Hendler, J., Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, mai. 2001. Disponível em: <http://www.siam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 20 julho de 2004.

A *Semantic Web* se utiliza ainda das Ontologias para possibilitar a recuperação de conceitos. Uma Ontologia na *Semantic Web* possui uma taxonomia e um conjunto de regras de inferência. A taxonomia define as classes de objetos e as relações que se estabelecem entre eles. Forma-se assim uma estrutura em que propriedades são atribuídas, e determinadas classes e objetos, que pertençam a esta classe, herdaram suas características. Berners-Lee et al (2001) afirmam que a ambigüidade pode ser solucionada atribuindo-se diferentes URIs para cada conceito de uma palavra. Assim, os motores de busca poderão encontrar páginas que se refiram a conceitos específicos e não todas as páginas, nas quais a palavra ambígua é utilizada. A solução de ambigüidades constitui-se numa das principais preocupações dos estudos da *Semantic Web* (Bräscher, 2004)<sup>56</sup>.

### 2.2.2.3 *WordNet*

A *WordNet* é um sistema de referência léxico cujo *design* é inspirado em teorias psicolinguísticas sobre a memória léxica humana. São classificados apenas os substantivos, verbos, adjetivos e advérbios da língua inglesa, sendo organizados em conjuntos de sinônimos, cada qual representando um conceito léxico. Os conjuntos de sinônimos são ligados entre si através de outras relações diferentes.

A análise da estrutura ontológica da *WordNet* é mesmo importante, devido ser crescente o número de aplicações em que a *WordNet* está sendo utilizada – mais como Ontologia do que como um recurso lexical. Para ser usada como Ontologia, porém, algumas ligações léxicas da *WordNet* precisam ser reinterpretadas como ligações semânticas, conectando o significado pretendido junto às palavras, de acordo com a própria interpretação pessoal ou individual. Uma dessas ligações é a relação hiponímia e hipernímia, que corresponde em muitos casos à relação usual (ou IS\_A) de classificação entre os conceitos. Uma primeira tentativa para explorar os problemas semânticos e Ontológicos, que estão por trás dessa correspondência, é descrita por Guarino (1998)<sup>57</sup>.

---

<sup>56</sup> Bräscher, Marisa. A ambigüidade na Recuperação da Informação. DataGramaZero –Revista de Ciência da Informação – v. 3, n.1, fev/02. Disponível em: [http://www.dgzero.org/fev02/Art\\_05.htm](http://www.dgzero.org/fev02/Art_05.htm). Acesso em 16 de maio de 2004.

<sup>57</sup> Guarino, N.: Some Ontological Principles for Designing Upper Level lexical Resources. In: A Rubio, N. Gallardo, R. Castro and A. Tejada (eds.), Proceedings of First international Conference on Language Resources and Evaluation. ELRA – European Language Resources Association, Granada, Spain (1998) 527 –534.

### 2.3 ENGENHARIA DA MENTE: A SINCRONICIDADE DO CONHECIMENTO

“Muito embora todos concordem em que a mente tem algo que ver com o cérebro, ainda não há consenso generalizado quanto à natureza exata dessa relação”. (Texto de introdução da Antologia sobre *Consciousness in Philosophy and Cognitive Neuroscience*, citado por Frijot Capra).

O objetivo desta seção é apresentar os fundamentos que embasam o tema central desta pesquisa, isto é, a definição de uma metodologia para auxiliar o processo de construção de Sistemas Baseados em Conhecimento focado na representação do conhecimento através de Ontologias. Esse processo denomina-se *Engenharia da Mente* e engloba o estudo de pessoas, processos, tecnologias e conteúdos. É a sincronização de fatores com um único objetivo: permitir que o conhecimento ou expertise de um determinado domínio possa ser disponibilizado através de um sistema computacional, e ser utilizado na sua total compreensão por outras pessoas, ou seja, um poderoso mecanismo de transferência de tecnologia.

A *Engenharia da Mente* é um processo de sincronização do conhecimento desenvolvido com o objetivo de fixar modelos conceituais em relação ao objetivo e à aplicação do modelo de gestão baseado na Inteligência Artificial, definido pelos processos da Engenharia do Conhecimento e da Engenharia de Ontologias. A metodologia busca ainda aprimorar habilidades intelectuais para atuar no desenvolvimento de estratégias para percepção do problema de forma criativa e com qualidade.

No ambiente de desenvolvimento de sistemas, especificamente no processo denominado Engenharia do Conhecimento, não há espaço para percepções imediatas próprias da experiência do especialista, mesmo nas mais sofisticadas técnicas de Representação e Aquisição do Conhecimento da Inteligência Artificial. Em sua maioria, os métodos são organizados em módulos isolados, excluindo-se a visão do todo nas etapas, que é imprescindível para um resultado efetivo e eficaz.

Observa-se que, muito embora o ‘bom senso’ seja um parâmetro para o desenvolvimento de sistemas inteligentes, nenhuma metodologia aventurou-se em abrir espaço para as percepções provenientes das pessoas que participam da etapa de Engenharia do Conhecimento, para utilizá-las como fator de sincronização e definição de prioridades. Disciplinas como a Engenharia de *Software* existem exatamente para ‘eliminar’ este tipo de

iteração. Entretanto, o ‘bom senso’ diz que a percepção imediata (intuição) tem maior eficácia sobre a melhor solução para um problema do que a aplicação de regras como as da lógica proposicional; embora a proposta mais aceita é a de que as pessoas tentam resolver problemas dedutivos aplicando regras como as da Lógica Proposicional (ÁRIS et al, 2003)<sup>58</sup>.

Um sistema de *software* complexo se caracteriza por um conjunto de componentes abstratos de *software* (estruturas de dados e algoritmos) encapsulados na forma de procedimentos, funções, módulos, objetos ou agentes interconectados entre si, compondo a arquitetura do *software*, que deverão ser executados em sistemas computacionais (WIKIPEDIA, 2005)<sup>59</sup>.

Embora a Engenharia de *Software* tenha vindo para cuidar da criação, construção, análise, desenvolvimento e manutenção de um sistema computacional, através de uma infinidade de métodos, nenhum se volta à compreensão e à estruturação desses aspectos, e que levem em consideração questões subjetivas relativas às pessoas envolvidas no desenvolvimento do sistema, sejam os especialistas, sejam os engenheiros do conhecimento. Um dos modelos que mais se aproxima a esta abordagem foi o elaborado pelo *CommonKADs* (Schreiber, et al, 2002)<sup>60</sup>, em que se destacam modelos que fazem parte de todo o processo de Engenharia do Conhecimento envolvido na construção de SBCs.

Assim, identificar, reconhecer e aplicar os elementos próprios da percepção humana, e sua forma de iteração na solução de um problema, tornaram-se meta desta pesquisa, com o objetivo de melhorar a qualidade no processo de desenvolvimento e percepção do conhecimento dos Sistemas Baseados em Conhecimento. Com isso, surgiu a metodologia que se denominou *Engenharia da Mente*.

A *Engenharia da Mente* está mais voltada aos princípios da Engenharia de *Software* que da Ciência da Computação. Por essa razão, a metodologia envolve o uso de modelos abstratos e precisos que permitem ao engenheiro especificar, projetar, implementar e manter sistemas de *software*, avaliando e garantido suas qualidades. Em trabalhos futuros, a *Engenharia da Mente* pretende oferecer mecanismos para se planejar e gerenciar o processo de desenvolvimento, como na Engenharia de *Software*.

---

<sup>58</sup> Arís, Enrique P., González, Juan L. S., Rubio, Fernando M. Lógica Computacional. Thomson Editores Spain, 2003. Pág. 8.

<sup>59</sup> Retirado de [http://pt.wikipedia.org/wiki/Engenharia\\_de\\_software](http://pt.wikipedia.org/wiki/Engenharia_de_software).

<sup>60</sup> Schreiber, G.; Akkermans, H.; Anjewierden, A.; Hoog, R.; Shadbolt, N.; de Velde, W. V.; and Wielinga, B.. Knowledge Engineering and Management: the CommonKADS Methodology. MIT Press. Cambridge. Massachussets. 2002.

A *Engenharia da Mente* engloba a identificação da mente, ou cognição, como uma das instituições mais profundas e arcaicas da humanidade, mas também como um dos campos mais respeitados e concorridos da ciência na atualidade: o efeito da Sincronicidade na Criação do Conhecimento.

No desenvolvimento de Sistemas Baseados no Conhecimento, durante o processo de Engenharia do Conhecimento, percebeu-se que as habilidades de relacionamento e comunicação interpessoal entre os especialistas e os engenheiros do conhecimento eram fundamentais e influam na definição e organização dos dados relevantes destes sistemas.

Na implantação de um Sistema Baseado em Conhecimento (SBC), a paralisia é comum. Falta de comunicação, falta de compromisso, corporativismo e ceticismo acabam ‘matando’ os melhores planos. Apesar das excelentes intenções, a natureza humana tem suas resistências quase embutidas. Além disso, poucas pessoas parecem se posicionar contrariamente às mudanças, mas, quando se trata de mudanças de atitudes e de comportamento no nível individual, a resistência é visível.

A construção de um Sistema Inteligente não é um processo linear como desejam a maioria das metodologias. A solução foi fazer uma abordagem mais profunda e desmistificada dos valores humanos necessários para a efetivação de mudanças num ambiente normalmente em plena crise. Na implantação de qualquer programa de melhorias ou mesmo de sobrevivência, é necessário preparar a pessoa mental e emocionalmente para as implicações de trabalhar numa nova realidade de complexidade, caos e mudanças descontínuas (O’Donnel e Bancon, 1999)<sup>61</sup>.

Foi buscando um modelo para criar um ambiente ideal que se encontrou a ‘Teoria da Sincronicidade’. O inconsciente coletivo é um termo cunhado por Carl Jung (2000)<sup>62</sup> para sua teoria de que tudo no universo estava interligado por um tipo de vibração, e que duas dimensões (física e não física) estavam em algum tipo de sincronia, que fazia certos eventos isolados parecerem repetidos, em perspectivas diferentes. Tal idéia desenvolveu-se primeiramente em conversas com Albert Einstein, quando ele estava começando a desenvolver a Teoria da Relatividade. Einstein levou a idéia adiante no campo físico, e Jung, no psíquico.

---

<sup>61</sup> O’Donnel, Ken & Bancon, Brian. No Olho do furacão. Salvador,BA: Casa da Qualidade,1999.

<sup>62</sup> Jung, Carl Gustav. Os arquétipos e o inconsciente coletivo. Editora:Vozes, 2000.

A ‘Sincronicidade’ é definida como uma coincidência significativa entre eventos psíquicos e físicos. Jung (2000) postula que tais coincidências apóiam-se em organizadores que geram, por um lado, imagens psíquicas e, por outro lado, eventos físicos. As duas coisas ocorrem aproximadamente ao mesmo tempo, e a ligação entre elas não é causal.

Jung (2000) sublinha que a ‘Sincronicidade’ parece depender consideravelmente da presença de afetividade, ou seja, sensibilidade a estímulos emocionais. Segundo Jung (2000), os pensamentos vêm-nos à consciência; as intuições e pensamentos que surgem do inconsciente não são produtos de esforços deliberados para pensar, mas objetos internos, parcelas do inconsciente que pousam ocasionalmente na superfície do ego.

A matemática é um produto puro da mente, e não se mostra em parte alguma do mundo natural; no entanto, pessoas podem sentar-se em seus escritórios e gerar equações que rigorosamente predizem e captam objetos e eventos físicos. Impressionava a Jung que um produto puramente psíquico (uma fórmula matemática) pudesse ter um relacionamento tão extraordinário com o mundo físico. Por outro lado, Jung (2000) propunha que os arquétipos também servem como ligações diretas entre a psique e o mundo físico, mas não são as causas destes. Parece, sim, ligá-lo a operadores que organizam a ‘Sincronicidade’.

Para quantos que têm alguma experiência da prática da meditação, as descrições da “atitude fenomenológica” hão de parecer familiares (CAPRA, 2003)<sup>63</sup>. Com efeito, as tradições contemplativas desenvolveram, desde há muitos séculos, técnicas rigorosas para o exame e a sondagem da mente, e demonstraram que essa atividade pode ser levada a um alto grau de aperfeiçoamento no decorrer do tempo. Segundo Serpa e Silva (2004)<sup>64</sup>, há uma ‘Sincronicidade’ nas construções humanas decorrente deste fundo comum a todos os seres humanos.

As percepções de Capra (2003) e Serpa e Silva (2004) são semelhantes ao mecanismo neural específico que Maturana e Varela (2001)<sup>65</sup> propõem para explicar o surgimento de estados transitórios de consciência, chamado de “sincronização de fases”, no qual diferentes regiões do cérebro se interligam de tal modo que seus neurônios ativam-se em sincronia uns com os outros. Segundo essa hipótese, cada experiência consciente se baseia num conjunto

---

<sup>63</sup> Capra, Fritjof. *As Conexões Ocultas: Ciência para uma vida sustentável*. Ed. Cultrix. 3ª Ed. São Paulo, 2003.

<sup>64</sup> Serpa, Luiz Felipe Perret, Silva, Aderval Barros da. *A Física Quântica e a Sincronicidade*. Disponível em: [http://www.faced.ufba.br/rascunho\\_digital/textos/236.htm](http://www.faced.ufba.br/rascunho_digital/textos/236.htm). Acesso em: 19 de janeiro de 2004.

<sup>65</sup> Maturana, Humberto R. e Varela, Francisco J e *A Árvore do Conhecimento - as bases biológicas da compreensão humana*. Editora Palas Athena, 2001.

específico de células, no qual muitas atividades neurais diferentes – associadas à percepção sensorial, às emoções, à memória, aos movimentos corporais, etc – unificam-se numa totalidade transitória, mas coerente de neurônios oscilantes.

A diferença entre a “sincronização de fases” e a Sincronicidade de Jung reside no fato que a primeira oferece modelos concretos e passíveis de verificação prática para explicar a dinâmica específica do processo, e deve, assim, produzir avanços significativos na formulação de uma verdadeira Ciência da Consciência. Desse modo, pode-se dizer que o processo de sincronização inicia com a compreensão da inteligência humana, as suas manifestações inconscientes e a sua relação com as palavras.

De acordo com Lacan *apud* Miller (2002), toda investigação humana está vinculada irreversivelmente no interior do espaço criado pela linguagem. Os diferentes desdobramentos da inventividade humana, embora muito diversificados, têm uma mesma origem: a mente humana e as percepções inconscientes; daí o fato de construções distintas eventualmente conduzirem os pensamentos a uma mesma referência. Antes de poder dizer “noite e dia”, explica Lacan, a noite e o dia não existem. Não há nada além de variações de luz (LACAN *apud* MILLER, 2002)<sup>66</sup>.

Então, uma metodologia que permita que este fundo comum se manifeste, e seja reconhecida como conhecimento entre os diversos especialistas e os engenheiros do conhecimento, deve necessariamente ser uma metodologia holística. Por essa razão, buscou-se desenvolver na metodologia da *Engenharia da Mente* ações que proporcionem aflorar a percepção imediata do especialista, através da compreensão do relacionamento das palavras, definição de atributos do sistema e objetivos com qualidade e criatividade.

### 2.3.1 Bases da Sincronicidade

#### 2.3.1.1 Teorias da Linguagem

Segundo Lacan (2000), caso se considere que o inconsciente é estruturado como uma linguagem, é possível reconstruir as associações inconscientes entre as palavras, revelando,

---

<sup>66</sup> Miller, Jacques-Alain. *Percurso de Lacan: uma introdução*. 2ª Edição. Tradução de Ari Roitman. Rio de Janeiro: Jorge Zahar Editor. Pág. 136. 2002.



dessa maneira, um ‘contexto’. Ele ressalta que todas as línguas são redutíveis a um sistema de articulação, mas seu gênio, sua coerência, situa-se num jogo dessa articulação, que é única. Assim, o ponto de vista cognitivista concerne às aquisições simbólicas, aquelas que se apóiam sobre os significados cujo suporte é, geralmente, a linguagem natural, ou, às vezes, as linguagens especializadas, como as formais ou as técnicas.

O ponto de vista descrito acima supõe a aquisição dos significados elementares, mas deixa em aberto o problema da construção desses últimos e, principalmente, os processos de categorização, os quais desenvolvem um papel fundamental na construção dos significados.

A obtenção dos significados elementares numa equipe ampla requer pensamentos síncronos. Por essa razão, a representação do conhecimento baseada em Ontologias confere uma maior ‘sensação’ de inteligência ao sistema computacional desenvolvido com esta técnica. A lógica prova aquilo que a intuição descobriu. É uma métrica do inconsciente.

Diferentemente de Chomsky, Lacan (2000) sabe que não existe locutor-ouvinte ideal, isto é, alguém que é pura ficção, que pertence a uma comunidade lingüística perfeitamente homogênea, e que conhece, sem qualquer vacilação, seu idioma. Essa questão é importante para o desenvolvimento de Ontologias em ambientes cooperativos. Um locutor que nunca tem distrações, que nunca tem desvios de atenção, ou seja, um perfeito computador, é um sujeito que não existe, segundo Lacan (2000).

Na consideração da lógica da linguagem, que é muito diferente da lógica da matemática, ressalta Miller (2002), é uma tradição imaginar que a linguagem tem por função apreender, de forma exata, uma referência.

Como construir Ontologias é exatamente esse jogo de referências, não se pode imaginar, que num desenvolvimento real, que elas não se desloquem, e todo o problema de linguagem é esse: não há uma palavra adequada para dizer algo, e é sempre em relação a outros significantes que a pessoa formula alguma coisa. Por isso, a matriz mínima da linguagem, tal como Lacan (2000) a escreve, é um significante depois outro. O mínimo do significantes são dois,  $S_1$  e  $S_2$ , o que se chama na teoria dos conjuntos, ‘par ordenado’. Quando se tenta definir alguma coisa, sempre se define através de outra coisa; não se percebe que, ao dizer isso, está-se evocando precisamente o deslocamento essencial do significante; diz muito bem: a resposta geral à pergunta o que é uma A é sempre que é um B. Essa é uma espécie de argúcia do lógico, que poderia estar assinada por Lacan. A linguagem-objeto é uma ilusão. Não há linguagem, estritamente falando, que se produza sem que o efeito de sujeito não esteja

sempre já, aí. Nesse sentido, pode-se dizer que todos os significantes também são semblantes, e justamente as formalizações lógicas dão uma experiência disso. É só pôr as definições e os axiomas convenientes, e se desenvolve um sistema. Um sistema de semblantes lógicos.

A psicanálise lacaniana constitui uma teoria plausível e uma prática socialmente significativa. No entanto, ela não é uma visão de mundo totalizada ou totalizante (Lacan, 2000)<sup>67</sup>. Para Bruce Fink (1998)<sup>68</sup>, é apenas um discurso entre muitos, não o discurso final, ou essencial. A psicanálise lacaniana não é um discurso de poder. A psicanálise utiliza o poder da causa do desejo para provocar uma reconfiguração do desejo do analisando.

Na metodologia de Aquisição e Representação do Conhecimento (RC2D)<sup>69</sup>, o contexto permite a associação de expressões que não seriam possíveis dentro de uma classificação formal da linguagem. Usar a teoria lacaniana é entender o conhecimento que se manifesta no mal-entendido da comunicação. É entender que a linguagem não é um processo comunicativo perfeito.

Lacan (2000) tentou formalizar a estrutura que sustenta a fenomenologia da experiência analítica. É, evidentemente, uma estrutura complexa, pois os fenômenos que ocorrem na experiência analítica dão, à primeira vista, a impressão de que não podem estar estruturadas, mas a metáfora pode estar estruturada, a metonímia pode estar estruturada, o equívoco pode estar estruturado, a função do ‘outro’ na delimitação do sentido pode estar estruturada.

A ciência supõe a disjunção do simbólico e do imaginário, do significante e da imagem. Lacan (2000) comentou com frequência os trabalhos do importante epistemólogo francês Alexander Koyré sobre Galileu, Kepler e Newton. Sobre essa questão, Miller (2002, pág. 45) ressalta:

Depois dessas referências, em um aparte, o professor Cadenas me dizia que a ciência é algo que dá lugar, por exemplo, à equação da gravidade. É esse também o exemplo com o qual Lacan se conforma. Mas a emergência das equações decisivas da teoria da gravidade exigiu – é o que assinala Lacan, baseando-se nos estudos de Koyré –, exigiu que desaparecessem todos os valores imaginários atribuídos aos movimentos dos astros. Exigiu, segundo a expressão de Lacan, a exterminação de todo simbolismo imaginário dos céus (destacado do original).

<sup>67</sup> Lacan, Jacques. Os Seminários de Lacan. Disponível em CD ROOM, 2000. Seminário n° 11, pág.78.

<sup>68</sup> Fink, Bruce. O sujeito Lacaniano: entre a linguagem e o gozo. Rio de Janeiro: Jorge Zahar Ed., 1998. pág. 159.

<sup>69</sup> Representação do Conhecimento Contextualizada Dinamicamente.

E continua:

No fundo, retomando a expressão de Bachelard que se tornou célebre, qual era o obstáculo epistemológico? Qual era o obstáculo epistemológico que tornava impossível a formulação das equações da teoria de Newton? Em Kleper, por exemplo. Kepler continuava pensando que, dada a eminente dignidade dos astros, seu valor superior, as órbitas celestes, as órbitas dos planetas deviam ter uma forma perfeita; dada essa perfeição, o movimento dos planetas não podia ser elíptico, tinha que ser circular. Essa teoria imaginária supunha o círculo mais perfeito que a elipse, daí a exigência, poder-se-ia dizer, estética imaginária, de que o movimento dos planetas fosse circular. A esse respeito, a equação newtoniana só pôde ser formulada a partir do momento em que se renunciou a atribuir qualquer significação imaginária aos céus (destacado do original); em que não se pensou mais na dignidade dos planetas, em que se renunciou à exigência de perfeição para se contentar com esses pequenos símbolos que podem ser escritos em folhas de papel e que valem pela criação inteira.

Para Miller (2002)<sup>70</sup>, a clássica teoria do conhecimento supõe uma co-naturalidade entre o sujeito e o objeto, uma harmonia preestabelecida entre o sujeito que conhece e o objeto conhecido. A teoria do conhecimento comentou sempre o milagre da adequação do conhecimento, guardando o lugar de uma coisa em si, como diria Kant, incognoscível. Miller (2002) ressalta:

A ciência distingue-se do conhecimento, desde o começo, mesmo que seja só pelo fato de construir seu objeto. Esse princípio, notem, não é especificamente lacaniano. É também o princípio de Bachelard, por exemplo, para quem o objeto e o instrumento científico são uma teoria encarnada, essa é a sua expressão. Assinalarei, de passagem, que ocorre o mesmo com o inconsciente freudiano; na medida em que é apreendido no dispositivo novo de sua prática, também ele realiza uma teoria. Que teoria? Essa é toda a questão. Este é um primeiro ponto, muito rápido, que pode prestar-se à discussão, e direi que não é especificamente lacaniano.

A linguagem não só permite o intercâmbio de informações e de conhecimentos humanos, como também funciona como meio de controle de tais conhecimentos. Ao se fazer tal afirmativa, está-se inserindo em uma corrente ligüístico-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 linguagem (WARAT, 1995)<sup>71</sup>.

<sup>70</sup> Miller, Jacques-Alain. *Curso de Lacan: uma introdução*. 2ª Edição. Tradução de Ari Roitman. Rio de Janeiro: Jorge Zahar Editor. pág. 40.

<sup>71</sup> Warat, Luis Alberto. *O direito e a sua linguagem*. Porto Alegre: Sergio Antonio Fabris Editor, 1995. (p. 37).

### 2.3.1.2 O Método Científico e o Imaginário

A ciência supõe a disjunção do simbólico e do imaginário, do significante e da imagem. Que sabem as plantas da teoria de Fibonacci, segundo a qual certas plantas dispõem as suas folhas? Ou sobre a teoria da gravidade? Como podem obedecer estas leis? Há coisas que se constata assim, que, afinal, não há porque compreendê-las, coisas que, evidentemente, colocam Deus no horizonte da ciência.

É por esta razão que a dúvida hiperbólica de Descartes (2002)<sup>72</sup>, qual seja: o esvaziamento da esfera psíquica, o esvaziamento do universo das representações, o esvaziamento de tudo o que é imaginário – não é adequada ao processo de construção de um modelo de representação do conhecimento. O ‘cogito’, em sua identidade, só surge como resíduo ineliminável dessa operação de esvaziamento, como um simples ponto desvanescente, já que, como diz Descartes, “eu sou, eu penso”, mas por quanto tempo? Só no instante em que penso.

Embora, em certa medida, estejamos condenados a estudar a consciência por vias indiretas, essa limitação não se restringe à consciência. Aplica-se a todos os demais fenômenos cognitivos. Analogamente, eletroencefalogramas e imagens de ressonância magnética funcional retratam correlatos da mente, porém esses correlatos não são a mente. A inevitabilidade da via indireta, contudo, não implica uma eterna ignorância sobre as estruturas mentais e os mecanismos neurais básicos. O fato de as imagens mentais serem acessíveis somente ao organismo que as possui, não impede que elas sejam caracterizadas, não nega sua dependência de um substrato orgânico e não impede que gradualmente se aproxime das especificações desse substrato, isto é, esse estado de coisas não deve impedir de se tratar cientificamente os fenômenos subjetivos.

Para Damásio (2000)<sup>73</sup>, ‘todos’ os conteúdos da mente são subjetivos, e o poder da ciência provém de sua capacidade para comprovar ou refutar objetivamente a consistência de muitas subjetividades individuais.

---

<sup>72</sup> Descartes, René. Discurso do método: para bem conduzir a própria razão e procurar a verdade nas ciências. São Paulo: Ed. Paulus, 2002.

<sup>73</sup> Damásio, António. O mistério da Consciência: do corpo e das emoções ao conhecimento de si. Tradução Laura Teixeira Motta. São Paulo: Companhia das Letras, 2000. Pág. 113.

A neurociência tem se empenhado consideravelmente em compreender a base neural da representação dos objetos. Amplos estudos sobre percepção, aprendizado, memória e linguagem deram uma idéia viável de como o cérebro processa um objeto, nos aspectos sensoriais e motor, e também de como o conhecimento sobre um objeto pode ser armazenado na memória, categorizado sob os aspectos conceituais ou lingüísticos e recuperado como evocação ou reconhecimento.

Do lado do organismo, porém, a situação é outra. Para Damásio (1996), a idéia de que as representações dos objetos poderiam estar ligadas à mente e à noção do *self* têm recebido pouca atenção, embora muito se tenha descoberto sobre como o organismo é representado no cérebro. Ele aventou em seu livro *O Erro de Descartes*<sup>74</sup>:

[...] a possibilidade de que a parte da mente que denominamos *self* seja, biologicamente falando, alicerçada em um conjunto de padrões neurais inconscientes que representam a parte do organismo que chamamos de corpo propriamente dito.

A mente é tão estritamente moldada pelo corpo e destinada a servi-lo que somente uma mente poderia surgir nesse corpo. Não há mente que não tenha um corpo, não há corpo que tenha mais de uma mente. As coisas que se possui estão próximas do corpo, ou deveriam estar, para que continuem sendo suas, e isso se aplica às coisas, à pessoa amada ou às idéias. A condição de agente requer, obviamente, um corpo agindo no tempo e no espaço, e não há sentido sem esse requisito.

Não existe percepção pura de um objeto em um canal sensorial, por exemplo: a visão. As mudanças simultâneas que ocorrem durante um fato não são um acompanhamento opcional. Para perceber um objeto, visualmente ou de algum outro modo, o organismo requer tanto os sinais sensoriais especializados como os sinais provenientes do ajustamento do corpo, que são necessários para a ocorrência da percepção.

No que diz respeito às experiências subjetivas, Capra (2003)<sup>75</sup> salienta que três grandes caminhos de análise estão sendo percorridos:

- O primeiro baseia-se na introspecção, método desenvolvido bem no começo da psicologia científica.

<sup>74</sup> Damásio, António. *O Erro de Descartes: emoção, razão e o cérebro humano*. São Paulo: Companhia das Letras, 1996. Cap. 10 e Introdução.

<sup>75</sup> Capra, Fritjof. *As Conexões Ocultas: Ciência para uma vida sustentável*. São Paulo: Ed. Cultrix. 3ª Ed., 2003.

- O segundo é a abordagem fenomenológica no sentido estrito, tal como foi desenvolvida por Husserl e seus seguidores.
- O terceiro caminho baseia-se no uso dos abundantes relatos derivados da prática da meditação, especialmente na tradição budista.

Seja qual for o caminho escolhido, esses cientistas cognitivos insistem que não estão lançando um olhar casual sobre as experiências subjetivas, mas sim empregando uma metodologia rigorosa que exige uma capacidade especial e uma formação contínua, à semelhança das metodologias de outros campos de observação científica.

Na teoria de Descartes, do século dezessete, o modelo de consciência separa a nossa mente de todo o mundo em torno de nós. Sabe-se que o modelo de Descartes não serve para a estruturação deste modelo de Ontologias, nem para a organização das pessoas que irão trabalhar no desenvolvimento destas Ontologias.

Os líderes experientes sabem que as emoções, como a incerteza, o medo, a confusão e a perda de autoconfiança, são sentimentos comuns que podem preceder o surgimento de uma novidade. Para se conseguir isso é preciso ter consciência que a mente não é somente aquilo que está inserido no crânio. É necessária uma visão muito mais ampla para compreender como a mente se estende além do cérebro.

A narrativa sem palavras, proposta por Damásio (2000), baseia-se em padrões neurais que se tornam imagens, sendo as imagens a mesma moeda corrente com que é feita a descrição do objeto causador da consciência.

Para António Damásio (2000)<sup>76</sup>, a função biológica das emoções é dupla:

A primeira é a produção de uma reação específica à situação indutora. Em um animal, por exemplo, a reação pode ser correr, imobilizar-se, lutar ferozmente contra o inimigo ou iniciar um comportamento prazeroso. Nos humanos, as reações são essencialmente as mesmas, influenciadas – espera-se - pelo raciocínio e pela sabedoria. A segunda função biológica da emoção é a regulação do estado interno do organismo de modo que ele possa estar preparado para a reação específica. Por exemplo, fornecer um fluxo sanguíneo mais intenso às artérias das pernas para que os músculos recebam oxigênio e glicose adicionais, no caso de uma reação de fuga, ou alterar os ritmos cardíacos e respiratórios, no caso de imobilização.

---

<sup>76</sup> Damásio, António. Op cit. Pág. 78.

Em outras palavras, o propósito biológico das emoções é claro, e as emoções não são um luxo dispensável. As emoções são adaptações singulares que integram o mecanismo com o qual os organismos regulam sua sobrevivência.

Só se pode atingir a ‘coincidência significativa’ entre indivíduos conscientes dos seus valores, aqueles que buscam uma crescente individualização através da construção contínua da capacidade intelectual, das forças emocionais e das redes sociais. Somente assim o indivíduo terá condições de assimilar novos conhecimentos, repassar o seu conhecimento e trabalhar em sincronia com outros para atingir um objetivo comum. Esses tipos de sincronicidade provocam mais de uma sensação, por exemplo, a freqüente expressão: “Ah, agora entendi...”. Em alguns casos, o acontecimento externo ocorre primeiro e o significado subjetivo, interior, em seguida. Em outros, a coincidência significativa é entre uma imagem interior, como um sonho, e um acontecimento externo, subsequente. Em todos, os vários eventos sincronísticos, contudo, o princípio de ligação interior e exterior é o significado do acontecimento para as pessoas envolvidas. Para Jung (2000), o inconsciente coletivo é um segundo sistema psíquico da pessoa. Diferentemente da natureza pessoal da consciência, ele tem um caráter coletivo e não-pessoal.

A imaginação é aquela operação da mente que faz imagens ou retratos mentais. Estes são chamados, às vezes, também ‘pensamentos’ ou ‘idéias’. Mas, devidamente compreendido, o pensamento é um processo, isto é, um movimento da mente. O pensamento é dinâmico, mas um pensamento ou uma idéia é estático, como um retrato. A fim de que o processo de pensar possa ocorrer, deve haver alguns pensamentos ou idéias ou imagens mentais para que se trabalhe com elas, e elas serão melhores quando mais desobstruídas e fortes.

Ernest E. Wood (1936)<sup>77</sup> desenvolveu, em 1936, um estudo interessante sobre como melhorar a imaginação. Para Wood (1936), todas as pessoas são aptas a viver em um mundo mental incolor, onde se permite que as palavras substituam as idéias. Então, o que se deve remediar para que as mentes trabalhem realmente bem e para dar uma existência colorida? Primeiramente, examinar o pensamento pessoal. Nele, a atenção se move de um pensamento a outro ou, ainda, de um grupo de pensamentos a outro grupo dos pensamentos, pois a maioria das imagens pessoais é complexa.

---

<sup>77</sup> Wood, Ernest E. *Mind and Memory Training*. The Theosophical Publishing House, Ltd: (1936).

Para buscar um bom desenvolvimento mental, Wood (1974)<sup>78</sup> elaborou uma **lista das quatro estradas principais do pensamento**, que serão utilizadas na construção de Sistemas Baseados em Conhecimento, para agrupar e categorizar as informações no processo de construção de Ontologias, como metodologia para criar e visualizar as relações entre os conceitos. As leis são:

- Na PRIMEIRA LEI, a **atração entre idéias** deve ser incluída. Idéias de coisas similares aderem-se uma a outra, e sugerem-se facilmente. Denomina-se este princípio de **Lei da Classe**. Ela inclui as relações entre um objeto e a classe a que pertence, e também aquela entre objetos da mesma classe.
- A SEGUNDA LEI é a **Lei das Partes**. Quando se pensa num elefante, provavelmente dá-se forma a retratos mentais especiais de seu tronco, ou orelhas, ou pés, ou quando se pensar nas suas orelhas, pode-se também pensar em outras partes dele, tais como os olhos.
- A TERCEIRA LEI pode ser chamada de **Qualidade**. Ela expressa a relação entre um objeto e sua qualidade, e também entre os objetos que têm a mesma qualidade. Assim se pode pensar o gato como um artista, a lua como esférica, etc., ou, se além de pensar na lua, pode-se também pensar em uma moeda de prata grande, porque têm a qualidade do branco e do redondo como a aparência da terra em comum.
- A QUARTA LEI diz respeito às **experiências impressionantes e familiares referente aos indivíduos**, e tem mais conexão com a imaginação do que com a observação lógica. Para se ver ou pensar duas coisas frequentemente e fortemente juntas, a força de seu encontro na consciência tenderá a dar-lhes **uma associação permanente** na mente. Conseqüentemente intitula-se o quarto princípio de **Lei da Proximidade**. Assim, por exemplo: se se pensar em uma cama, pensa-se em dormir, não em dançar. Se se pensar no Brasil, pensa-se no café e no maravilhoso rio Amazonas, não no arroz e nas montanhas do Himalaia. Cada pessoa tem uma base independente de experiências compostas das memórias de tais relacionamentos vistos, ouvidos, ou pensados, seja ele ao vivo ou repetidamente. Dentro dessa Lei vêm também a seqüência familiar, ou a sucessão contígua, chamada popularmente de

---

<sup>78</sup> Wood, Ernest E. *Mind and Memory Training*. The Theosophical Publishing House: London, 1974. Adyar - Madras - India Wheaton - Ill. - USA. 1936. Disponível em: <http://www.led.ufsc.br/~martins/ch/mind/>. Acesso em: 13 de julho de 2004.



“causa e efeito”, como exercício e saúde, ou guerra e pobreza. Ela tem proximidade no tempo.

Em relação à Primeira Lei, deve-se mencionar um caso que é freqüentemente mal entendido - chamado contraste. Se duas coisas contrastarem, devem pertencer à mesma classe. Não se pode contrastar uma vaca com papel borrado, ou uma vara andando com a raiz quadrada de dois. Mas se pode contrastar um elefante e um rato, um papel borrado e um papel vitrificado, o sol e a lua, e outros tais pares. Assim, os contrastes pertencem à Primeira Lei.

O controle do objeto e do sentido de movimento do pensamento é chamado freqüentemente de concentração.

Embora todas as mentes trabalhem sob as mesmas leis, elas fazem isso em diferentes graus de intensidade e abrangência. Algumas mentes trabalham rapidamente, outras lentamente; algumas têm muitas idéias a oferecer, outras poucas. Algumas mentes são mais brilhantes que outras, e provavelmente todos querem ter a sua brilhante e forte. As pessoas querem pensar muitas idéias, e pensá-las bem. As pessoas querem pensar sobre toda a abrangência do assunto em estudo, não somente sobre uma parte dele, como fazem os tímidos ou maus pensadores.

O pensamento dinâmico faz uso de pensamentos estáticos, da mesma maneira como no andar há pontos da terra firme, em que os pés, alternadamente, descansam. Não se pode andar no ar. Em ambos os casos, a dinâmica necessita da estática. Suponha-se o seguinte: o gato persegue o rato, e o rato é afeiçoado ao queijo, e o queijo é obtido na leiteria, e a leiteria está entre as árvores. Não há nenhuma conexão entre o gato e as árvores, mas o pensamento moveu-se do gato para as árvores pelas palavras-chave do rato, do queijo e da leiteria. Para desenvolver o poder da imaginação, é preciso ver com clareza a distinção entre idéias e pensar, entre o pensamento estático e o dinâmico.

Para desenvolver a imaginação, Wood (1974) criou uma série de exercícios com o objetivo de treinar a mente para dar forma, com facilidade e velocidade, a retratos mentais completos e vívidos, ou idéias-imagem. Quando se conhece um objeto concreto, ele é reproduzido dentro da mente, que é o instrumento do conhecimento; e quanto mais a imagem se aproxima ao objeto, mais verdadeiro é o conhecimento que se apresenta. Na prática, tal imagem é geralmente vaga e com freqüência distorcida.

Para desenvolver a imaginação, Wood (1974) dividiu as idéias-imagens em quatro variedades: concreto simples, concreto complexo, abstrato simples e abstrato complexo.

- 1) As idéias concretas simples são reproduções mentais dos objetos pequenos ordinários da vida, como: laranja, pena, vaca, livro, chapéu, cadeira, e todas as sensações simples de som, forma, cor, peso, temperatura, gosto, cheiro e sentimento.
- 2) As idéias concretas complexas são na maior parte múltipla das simples, ou associações de uma variedade delas, como: cidade, família, jardim, formigas, areia, provisões, roupas.
- 3) As idéias abstratas simples são aquelas que pertencem a uma variedade de idéias concretas, mas não denotam nenhuma delas no detalhe, como: cor, peso, massa, temperatura, saúde, posição, valor, número.
- 4) As idéias abstratas complexas são combinações das simples, como: majestoso, esplendor, benevolência, fé.

Essas combinações de exercícios serão utilizadas em etapas específicas da aplicação da metodologia *Engenharia da Mente*, seja na etapa da Engenharia do Conhecimento, seja na etapa da Engenharia de Ontologia.

### 2.3.2 Conclusão

As teorias apresentadas nesta seção foram os argumentos base da metodologia da *Engenharia da Mente*.

Questões levantadas na psicanálise, na lingüística, na neurociência e na meditação indicam o significado dos símbolos que, supostamente, existem dentro das pessoas. Conhecer e reconhecer nas pessoas esses símbolos e compartilhar os próprios é importante para a representação do conhecimento na Inteligência Artificial e na construção de Ontologias para este tipo de aplicação.

Esta também é a conclusão de Capra (2003)<sup>79</sup>, que resume as questões levantadas neste capítulo e por vários outros cientistas:

[...] os cientistas terão de aceitar outro paradigma novo – terão de reconhecer que a análise da experiência viva, ou seja, dos fenômenos subjetivos, tem de fazer parte de qualquer ciência da consciência que mereça ser considerada como tal.

---

<sup>79</sup> Capra. Op. cit. Pág 57.

E continua:

A enorme relutância dos cientistas em se ver às voltas com os fenômenos subjetivos faz parte da nossa herança cartesiana. A divisão fundamental que Descartes operou entre a mente e a matéria, o eu e o mundo, levou-nos a crer que o mundo pudesse ser descrito objetivamente, ou seja, sem que se fizesse menção nenhuma ao observador humano.

O primeiro uso do relato imagético da relação entre organismo e objeto é informar o organismo sobre o que ele está fazendo ou, em outras palavras, responder a uma pergunta que o organismo nunca formulou: o que está acontecendo? Qual a relação entre as imagens das coisas e este corpo? O sentimento de conhecer é o começo da resposta.

Quando o cérebro de um organismo desperto, adequadamente equipado, gera consciência central<sup>80</sup>, o primeiro resultado é mais um estado de vigília e o segundo é a atenção mais focalizada no objeto causativo. Obtêm-se ambos resultados por meio de um realce dos mapas de primeira ordem que representam o objeto.

O alicerce indispensável da consciência é a consciência central, mas sua glória é a consciência ampliada. Quando se pensa na grandiosidade da consciência, o que se tem em mente é a consciência ampliada. Damásio (2000) explica que a consciência ampliada vai além do aqui e agora da consciência central, em direção tanto ao passado quanto ao futuro. É a capacidade de estar consciente de uma gama enorme de entidades e eventos, ou seja, a capacidade de gerar um senso de perspectiva individual, de propriedade, e da condição de agente sobre uma gama de conhecimentos maior do que a abrangida pela consciência central. A consciência ampliada não é o mesmo que inteligência; ela pode ser desenvolvida através de técnicas de meditação.

Estabeleceu-se, então, um quadro geral e coeso de algumas idéias a respeito de como a mente funciona e, com elas, fundamentaram-se princípios de fácil assimilação para identificar e organizar um conhecimento passível de ser reproduzido por um computador.

Na investigação, descobriu-se que não é a mente que determina o conhecimento, tampouco as sensações, mas a combinação desses dois elementos, de modo síncrono. Portanto, na *Engenharia da Mente*, o papel do especialista soma-se ao analista de sistemas e

---

<sup>80</sup> Na definição de António Damásio, consciência central, ou *self central* é inerente ao relato não verbal de segunda ordem que ocorre toda vez que um objeto modifica o proto-self. Ele pode ser acionado por qualquer objeto e sofre mudanças mínimas no decorrer de toda a vida. Somos conscientes do self central. Já o proto-self é um conjunto interligado e temporariamente coerente de padrões neurais que representam o estado do organismo, a cada momento, em vários níveis do cérebro. Não somos conscientes do proto-self.

ao engenheiro do conhecimento, não só como coadjuvantes no processo de Engenharia do Conhecimento, mas atuando no papel principal, como parte do conhecimento assimilado pelo Sistema Inteligente.

Sobre a aplicação dessas teorias, ou parte delas, será abordado no próximo do Capítulo, na qual expressões como inconsciente coletivo, semântica, a dualidade mente e corpo, psicanálise, ciência e meditação permitem uma interessante visão sobre os aspectos cognitivos que permeiam a construção dos Sistemas Baseados em Conhecimento.

## CAPÍTULO 3

### 3 AS METODOLOGIAS DA ENGENHARIA DA MENTE

Na Revisão da Literatura, salientou-se que para manter a qualidade da base de conhecimento de um Sistema Inteligente, durante a fase de Engenharia de Conhecimento e Engenharia de Ontologia, exige uma visão do contexto conjunta e síncrona dos engenheiros do conhecimento e dos especialistas. Argumentou-se que a ‘Sincronicidade’ é um elemento indispensável em ambientes que buscam a efetividade e a eficácia dos resultados e não é um elemento que deva ser isolado numa fase específica, mas permear todo o processo.

Neste Capítulo, são descritos os elementos utilizados para manter a sincronia entre os especialistas do domínio, engenheiros do conhecimento e programadores. Os especialistas do domínio devem conhecer e repassar o seu conhecimento de forma mais estruturada e coerente, criando, muitas vezes, novos caminhos para a implementação de uma solução. Por sua vez, o engenheiro do conhecimento necessita do conhecimento profundo do especialista para compreender o conteúdo e o objetivo do sistema. Para atingir esse resultado, é necessário instruir o engenheiro do conhecimento para que ele possa auxiliar o especialista na realização dos inventários de pessoas, conteúdo, processos e tecnologias. E, finalmente, o programador precisa interagir continuamente com os dois universos para compreender e criar computacionalmente algoritmos que resolvam com maior eficiência o problema do usuário.

Durante esta pesquisa, observou-se que os processos de aquisição do conhecimento, realizada pela equipe de engenheiros do conhecimento, na área de sua especialização<sup>81 82</sup>, obteve uma eficácia maior que a aquisição realizada pela mesma equipe em domínios diversos

---

<sup>81</sup> Bueno, Tânia C. D., et al. Retrieval in Jurisprudencial Text Bases using Juridical Terminology. Op. cit.

<sup>82</sup>Hoeschl, Hugo C., et. al. Olimpo: Contextual Structured Search to improve the representation of UN Security Council with Information Extraction methods. Op. cit.

de sua especialização<sup>83</sup>. No segundo caso, observou-se que foram os obstáculos de comunicação que levaram a necessidade da realização do processo de aquisição conhecimento novamente.

Por isso, sistematizou-se uma série de questões que, comprovadamente, melhoraram a velocidade e a qualidade do conhecimento representado no sistema. Assim, o compartilhamento de conhecimentos entre todos os elementos da equipe (analistas de sistemas, engenheiros do conhecimento e especialistas) tornou-se essencial na metodologia elaborada neste trabalho.

Na construção de Ontologias, mesmo sendo um trabalho de especialistas e engenheiros do conhecimento, observou-se que a construção de uma linguagem comum, feita pela equipe sobre o domínio de aplicação do sistema, influía nos resultados. Isto é, quando os programadores acompanhavam a estruturação das Ontologias, eles passaram a refletir maior conhecimento no desenvolvimento dos sistemas, e os especialistas do domínio, por sua vez, ao conhecerem detalhes da construção dos algoritmos e da linguagem de programação e representação do conhecimento, passaram a construir Ontologias de maneira mais eficiente e mais representativa do domínio.

Na elaboração da metodologia da *Engenharia da Mente*, o item ‘Compartilhamento do Conhecimento’, através da uniformização da linguagem, foi definido como um dos fatores que permitiu elevar a sincronia na equipe, que se define como primeira premissa. Essa sincronia, aliada ao conhecimento de todos elementos necessários para se atingir o objetivo para o qual o sistema esta sendo construído, permitiu a ‘Visualização’ de todos os aspectos do contexto, no qual o sistema está inserido; é o que se considera como segunda premissa.

A terceira premissa é a ‘Definição de Relevância’ – um procedimento que permite manter a qualidade no desenvolvimento dos sistemas inteligentes e das Ontologias. Por ser constante em todas as etapas, o procedimento permite uma constante reavaliação dos resultados, aumentando a atenção de todos os envolvidos e permitindo uma identificação instantânea de erros, tanto de programação quanto de concepção dos sistemas e das Ontologias.

Nas seções seguintes, descrevem-se os elementos e a relação das premissas com a construção de Sistemas Baseados no Conhecimento e de Ontologias.

---

<sup>83</sup> Ribeiro, Marcelo S.; Mattos, Eduardo da S., Bueno, Tânia C. D.; Hoeschl, Hugo C. KMAI- Knowledge Management With Artificial Intelligence. Op. cit.

### 3.1 FASES DA ENGENHARIA DA MENTE

A metodologia da *Engenharia da Mente* engloba o estudo de pessoas, processos e tecnologias através de três premissas (ver figura 4):

- 1) Compartilhamento do conhecimento;
- 2) Visualização;
- 3) Definição de relevância.

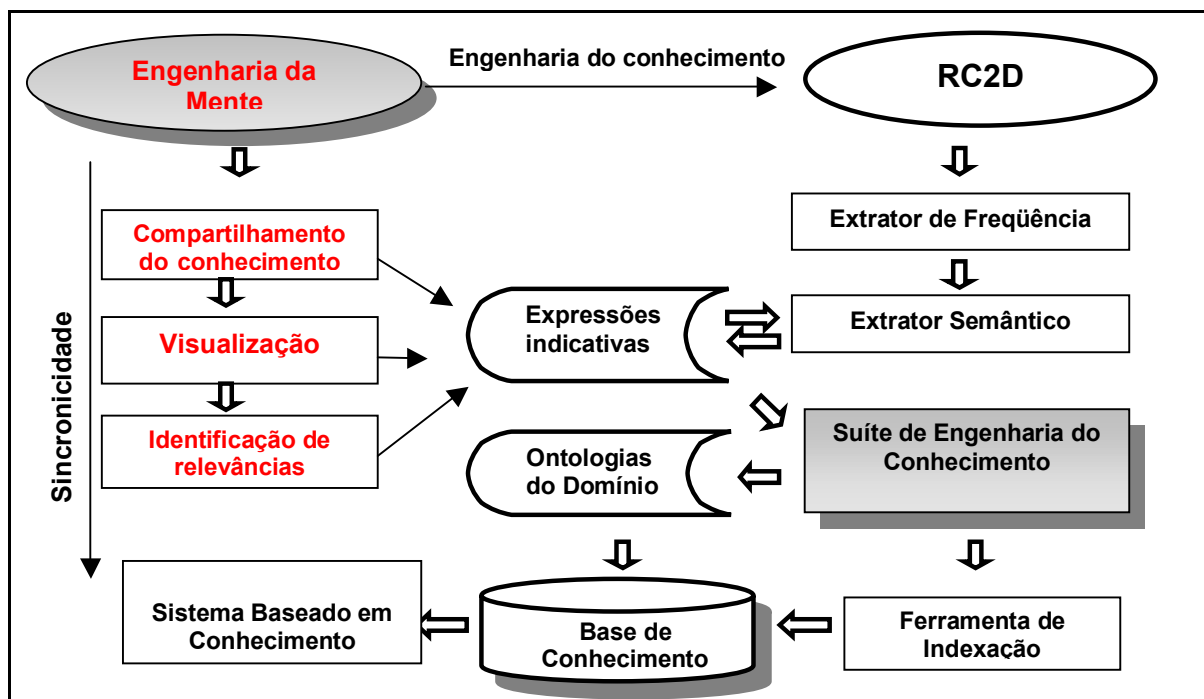


Figura 4 – Engenharia da Mente aplicada à Construção de Ontologias em Sistemas Baseados em Conhecimento.

#### 3.1.1 Fase 1: Compartilhamento do Conhecimento

Que coisas as pessoas têm que são recursos produtivos? O que têm as pessoas que se traduzem em valor para elas mesmas e para as organizações das quais fazem parte? Segundo Gratton & Ghoshal (2003)<sup>84</sup>, há três tipos de recursos que as pessoas possuem que, coletivamente, constituem o seu capital humano individual. Existem elementos, como a

<sup>84</sup> Gratton, Lynda, et.al. Managing Personal Human Capital: new ethos for the "Volunteer" Employee, The European Management Journal, vol 21, n° 1 pp1-10, February, 2003.

complexidade cognitiva e a capacidade de aprender, que fornecem os traços individuais subjacentes, nos quais o conhecimento especializado e as habilidades são baseados similarmente na sociabilidade e na confiança, que fornecem as âncoras para desenvolver e manter uma rede de relacionamentos. O cérebro só registra, aprende e ‘ramifica’ quando está aberto ao novo.

Nesta pesquisa e no desenvolvimento de protótipos, notou-se que o conhecimento dos especialistas sobre a técnica de IA, que estava sendo aplicada na modelagem do sistema, nos casos RBC e RC2D (ver Capítulo 2), permitiu uma transferência de conhecimento para a linguagem computacional de uma forma muito positiva para o escopo final dos sistemas. Para que essas transferências ocorressem de forma síncrona, sistematizaram-se três pontos principais, necessários ao desenvolvimento da primeira etapa da metodologia, o ‘Compartilhamento do Conhecimento’:

- 1) Identificação dos aspectos subjetivos para formalização da equipe;
- 2) Uniformização do vocabulário;
- 3) Inventário pessoal, de processos, de tecnologias e conteúdo.

### 3.1.2 Fase 2: Visualização

O uso de palavras como imagens, padrões neurais, representações e mapas, tão comuns na Inteligência Artificial, possuem significados variados e pouco claros. A etapa de ‘Visualização’ tem por objetivo a construção de mapas mentais que representem o domínio de aplicação, seja no desenvolvimento de sistemas baseados no conhecimento, seja na criação das Ontologias do domínio. Desenvolveu-se uma série de exercícios utilizando-se o estudo de Ernest. E. Wood<sup>85</sup> – que se fundamenta em técnicas de meditação – e nos estudos sobre a neurociência, de António Damásio, e psicanalíticos, de Lacan e Jung.

As quatro leis do pensamento mencionadas no capítulo anterior (ver item 2.3.1.2) são dadas, de maneira geral, para a finalidade atual. O pensador ressalta que algumas idéias surgem com a capacidade da mente para a comparação, isto é, através de uma faculdade lógica, enquanto outras idéias surgem simplesmente na imaginação sem nenhuma razão, a não

---

<sup>85</sup> Wood, Ernest. Op. Cit.



ser aquela que foi gravada em cima dela, em algum tempo precedente. A comparação cobre (pertence) as primeiras três leis, a imaginação somente a quarta.

Diversos estudantes podem ser convidados para escrever um ensaio sobre o assunto dos gatos. Alguns deles sentirão seus pensamentos surgir de forma abundante através dos recessos da mente, enquanto outros se sentarão e mastigarão as extremidades de suas canetas por muito tempo antes que seus pensamentos comecem a fluir.

Na *Engenharia da Mente*, a fase de ‘Visualização’ é indispensável na sincronização de conhecimentos. E, como será visto mais adiante, ela possui um resultado diferenciado inerente à área de aplicação da metodologia. É preciso notar, porém, que nem todas as imagens que o cérebro constrói se tornam conscientes (DAMÁSIO, 2000)<sup>86</sup>. Geram-se imagens demais e competição demais para a janela da mente, que é relativamente pequena, na qual as imagens podem se tornar conscientes – a janela na qual as imagens acompanham a percepção de que se está apreendendo e, em consequência, atentando devidamente para elas.

Evoca-se aqui a teoria computacional da mente (ver item 2.3.1.2), feita por Damásio (2000)<sup>87</sup>, segundo o qual as imagens que cada pessoa vê na mente não são cópias do objeto específico, mas imagens das interações entre cada pessoa e um objeto que mobilizou seu organismo, construídas na forma de padrão neural, de acordo com a estrutura do organismo. E, sob o ponto de vista biológico, entre as pessoas, as imagens podem ser suficientemente semelhantes para se construir uma imagem de uma mesma coisa. Pode-se aceitar, sem hesitar, a idéia convencional de que se forma a imagem de uma coisa específica. Mas isso não é verdade. Dessa maneira, deve-se ter cautela com o termo representação: ele facilmente evoca a metáfora do cérebro como um computador, sendo essa metáfora inadequada. O cérebro de fato executa computações, mas sua organização e seu funcionamento pouca semelhança têm com a noção comum do que seja um computador.

### 3.1.3 Fase 3: A Definição de Relevância

Esta é a fase mais importante da metodologia da *Engenharia da Mente*, e a mais complexa e difícil de se atingir, o que exige dos engenheiros do conhecimento uma dedicação

---

<sup>86</sup> Damásio, António. O mistério da Consciência: do corpo e das emoções ao conhecimento de si. Op. Cit. 404.

<sup>87</sup> Damásio, António. Ib., pág. 406.

especial à expansão das suas capacidades cognitivas. A definição do que é relevante exige sintonia com o momento, conhecimento profundo sobre o domínio e sintonia com o próprio corpo. A definição de ‘relevância’ é uma decisão emocional. É a dimensão social da consciência. É a manifestação do inconsciente coletivo.

Nesta fase faz-se a especificação do sistema e seus atributos, bem como do modelo de interface. Com base no inventário realizado e na compreensão mútua e síncrona do domínio de aplicação, é possível identificar a abrangência do Sistema Baseado em Conhecimento dentro do escopo em que ele se constrói. Perde-se em precisão, mas se ganha em rapidez. Ela se baseia em primeiras impressões e reage ao panorama global ou aos seus aspectos mais gritantes (Goleman, 1995)<sup>88</sup>, pois a mente emocional é mais rápida que a racional, agindo de maneira irrefletida, sem parar para pensar.

A rapidez com que as emoções se apossam das pessoas – como uma adaptabilidade emocional essencial – mobiliza-as para agir nas emergências, sem perda de tempo com ponderações sobre quando reagir, ou mesmo se haverá reação. Se a mente emocional segue essa lógica e suas próprias regras, com um elemento representando outro, ela precisa necessariamente definir as coisas através de sua identidade objetiva: o que importa é como são ‘percebidas’; as coisas são como parecem ser. A lembrança evocada pela percepção de alguma coisa pode ser muito mais importante do que a coisa ‘é’. Enquanto a mente racional faz conexões lógicas entre causa e efeito, a mente emocional não faz qualquer discriminação. Liga coisa com coisa que, entre si, guardam uma longínqua similaridade. Então, como controlar o incontrolável?

Na seção referente a bases da Sincronicidade da *Engenharia da Mente*, tratou-se sobre o corpo como sustentáculo do inconsciente (ver item 2.3.1.1). As pessoas tornam-se conscientes quando internamente seu organismo constrói e exhibe um tipo específico de conhecimento sem palavras – o conhecimento de que um objeto mudou o organismo – e quando esse conhecimento ocorre junto à exibição interna destacada de um objeto. Essa é a visão de Damásio (2000)<sup>89</sup>, que enumera premissas sobre o relato imaginético e não-verbal da relação entre objeto e organismo – premissas que permitem o mapeamento das imagens mentais. Através da observação da mecânica das emoções, a neurociência começa a mostrar

---

<sup>88</sup> Goleman, Daniel. *Inteligência Emocional*. Op. Cit. Pág. 306.

<sup>89</sup> Damásio, António. *Ib.* Pág. 219.

como diferentes sistemas cerebrais atuam para produzir, por exemplo: raiva, tristeza ou alegria.

Para atingir este relato imaginético e não-verbal, a metodologia utilizou-se das impressões obtidas pelo conhecimento compartilhado, pela observação dos mapas elaborados na etapa de inventário. Isso permitiu decidir sobre os elementos indispensáveis para se obter o resultado esperado pelos ‘construtores’ do Sistema Inteligente.

### 3.2 ENGENHARIA DA MENTE APLICADA À ENGENHARIA DO CONHECIMENTO

A demanda por soluções tecnológicas cada vez mais adaptada às necessidades do cliente acrescenta um ator importante no processo produtivo destes produtos: o próprio cliente. Os sistemas inteligentes são exemplos da nova era tecnológica, na qual não basta simplesmente desenvolver um aplicativo com base nas teorias computacionais tradicionais de *hardware* e *software*.

A Engenharia do Conhecimento surge como uma forma de fazer a ponte entre o que a tecnologia pode fazer e o que o usuário precisa que a tecnologia faça. Por isso é de suma importância que se tenha conhecimento de tudo aquilo que pode interferir no ambiente, onde será implantada uma solução tecnológica, sejam processos, pessoas ou outros ambientes.

No processo de Engenharia do Conhecimento, existe uma infinidade de métodos, sendo que alguns se voltam à compreensão e à estruturação do aprendizado das pessoas envolvidas, sejam os especialistas, sejam os engenheiros do conhecimento. No entanto, mesmo as metodologias que buscam a gestão dos ativos intangíveis baseados, sobretudo, na competência dos colaboradores de uma determinada organização, como o *CommonKADS*<sup>90</sup>, ela é individual. Isto é, a competência, na maioria das vezes, concentra-se no conhecimento tácito de cada indivíduo, que é pessoal.

Neste trabalho construiu-se uma metodologia para que o processo de aquisição de conhecimento permita ao especialista e ao engenheiro do conhecimento obter e compartilhar o conhecimento necessário para trabalharem em ‘sincronia’, a fim de atingirem o objetivo do sistema. O trabalho desenvolvido em sincronia possibilita a identificação mais efetiva dos atributos do sistema e a formação de uma base de conhecimento fundamentada na relação

---

<sup>90</sup> Schreiber. et al, 2002.

entre expressões relevantes de um contexto, mesmo em ambiente cooperativo de trabalho na Web.

O processo de Engenharia do Conhecimento envolve tradicionalmente a interação entre dois tipos de participantes: o engenheiro do conhecimento e o especialista, e visa a obtenção dos seguintes resultados:

- a) Inventário de pessoas, processos, tecnologias e conteúdos, visando o levantamento do fluxo de informação no que tange a aquisição, armazenamento (memória organizacional), disseminação e reutilização da informação.
- b) Requisitos e arquitetura do sistema;
- c) Modelo de documentação.

A seguir, apresenta-se a estrutura de aplicação das três premissas à Engenharia do Conhecimento.

### 3.2.1 Fase 1: O Compartilhamento do Conhecimento na Engenharia do Conhecimento

Consiste na primeira etapa da metodologia. Ao especialista, supõe-se um conhecimento profundo do domínio e da estrutura da instituição; ao engenheiro do conhecimento, supõem-se conhecimentos sobre a metodologia de aquisição do conhecimento, da estrutura de um Sistema Baseado em Conhecimento e das técnicas de Inteligência Artificial.

Os itens a seguir descrevem com detalhes os pressupostos necessários ao compartilhamento desse conhecimento, que é a compreensão da conexão do conhecimento entre o Inventário do Conhecimento Institucional e o Inventário de Pessoas.

#### 3.2.1.1 Identificando os aspectos subjetivos para formalização da equipe de EC

A prisão ao passado, a inflexibilidade e a inércia levam ao pensamento burocrático, muito comum nas instituições. Se há falha no processo de comunicação da equipe de engenheiros do conhecimento, a questão se agrava.

Para Goleman (1995)<sup>91</sup>, a solução é melhorar a capacidade imaginativa, isto é, incentivar o especialista a entender quem ele é, assim como a sua importância no ambiente de trabalho, além de incentivá-lo a ser, ou seja, exercer a sua função. Feito isso, precisa-se avaliar a fragmentação que o excesso de informações, a falta de concentração e o stress ocasionam.

Identificar e separar os condicionamentos daquilo que se define como conhecimento é essencial para a representação do conhecimento num SBC. Engenharia do Conhecimento é, sobretudo, troca de conhecimento.

Aqui, a predisposição ao aprendizado dividirá a equipe. Na definição da equipe, escolhem-se aquelas pessoas com menos resistência e mais flexibilidade e interesse em apreender (convicção, determinação e esforço).

No caso de não ser possível escolher a equipe, é importante, na fase de uniformização do vocabulário, trabalhar as resistências através da dinâmica de construção do vocabulário controlado pela equipe.

O grau de conhecimento e importância na equipe são características secundárias, mas elas influenciam, e é preciso definir novas funções de acordo com as tarefas a serem desenvolvidas.

Identificar emoções relativas à instituição e à função, que o membro da equipe exerce, é muito relevante, uma vez que elas impulsionam a busca por uma solução. Compartilhar as emoções sobre questões relacionadas ao domínio de aplicação permite um maior comprometimento no compartilhamento de informações, tão necessário à formação da base de conhecimento.

### 3.2.1.2 Uniformização do Vocabulário

A importância dos conhecimentos existentes para as novas aquisições deriva do papel fundamental que desempenham dentro da construção das representações e da idéia de que a aquisição passa, necessariamente, por essas representações.

---

<sup>91</sup> Goleman, Daniel. Inteligência Emocional. Ed. Objetiva: Rio de Janeiro, 1995.pág.

A diferença mais clara entre um novato num jogo de xadrez e um jogador experiente é a velocidade na tomada de decisão. Com a experiência, o número de elementos potencialmente relevantes que o aprendiz pode reconhecer é menor, melhorando o seu desempenho. Essa é a importância desta fase.

Na troca de conhecimentos, o especialista passa a conhecer a forma como seu conhecimento poderá se organizar, isto é, os conceitos básicos da técnica de Inteligência Artificial empregada na representação do conhecimento. Assim, ele poderá contribuir com mais efetividade e haverá maior interesse em participar do processo.

Quanto ao especialista, a troca levará a uma percepção mais imediata do escopo do sistema, e aumentará o interesse em se aprofundar no estudo do domínio. Ambos estarão preparados para lidar com a sobrecarga e para conseguir a competência necessária para planejar ou escolher uma perspectiva que determine, então, que elementos da situação devem ser tratados como importantes e quais podem ser ignorados. Ao perceber que, da vasta informação, o conhecimento se restringe somente a algumas características e aspectos possivelmente relevantes, tomar uma decisão tornar-se-á mais fácil.

Para uniformizar a linguagem que se aplicará no desenvolvimento do Sistema Baseado em Conhecimento, é preciso identificar os principais conceitos trabalhados no domínio, e a forma de como as pessoas trabalham a linguagem, de modo a auxiliar a equipe de Engenharia do Conhecimento na compreensão e determinação do contexto. Se o sistema for jurídico, por exemplo, é necessário que o engenheiro do conhecimento saiba entender o conteúdo de uma sentença ou de uma norma, assim como o seu processo e relevância. Ao especialista, não basta dominar a linguagem jurídica e os processos do domínio de aplicação: ele deverá conhecer algumas expressões comuns ao desenvolvimento de sistemas inteligentes.

### 3.2.1.3 Inventário do Conhecimento Institucional ou do Domínio de Aplicação

O inventário permite conhecer amplamente o ambiente de desenvolvimento do SBC. Especificamente, permite definir todos os elementos atuantes e também desnecessários que fazem parte da instituição ou do domínio de aplicação. Na primeira fase, não se julga o que é relevante ou não; somente procura-se ser exaustivo no inventário. Os participantes devem estar concentrados para a atividade que deverá ser concluída sem intervalos, pois a duração da atividade poderá consumir o período de algumas horas ou algumas semanas. A continuidade

implica que os engenheiros do conhecimento e os especialistas se dediquem exclusivamente à tarefa. No quadro 3, são descritos os tipos de inventários realizados dentro de uma instituição.

<b>De Processos</b>	O inventário de processos é necessário para conhecer os trâmites dos documentos e das informações dentro da instituição ou do domínio de aplicação. Normalmente esses processos são formais. No caso das instituições públicas, eles são normatizados; no caso de instituições privadas, necessita-se observar se existe confluência entre o procedimento adotado e a formalização exigida pela instituição (é muito comum a existência de processos não-implementados ou não-aceitos pelos usuários).
<b>De Tecnologias</b>	O inventário de tecnologias apresenta todos os recursos tecnológicos adotados para a atividade fim da instituição e também os recursos tecnológicos que a instituição pretende utilizar no desenvolvimento do Sistema Baseado em Conhecimento. O inventário é de sistemas computacionais ( <i>software</i> ) e também de equipamentos ( <i>hardware</i> ).
<b>De Conteúdo</b>	Este inventário consiste na descrição detalhada de toda a produção digital da instituição. Arquivos em todos os formatos, informações organizadas em diretórios ou banco de dados, enfim, toda informação disponível em formato digital.
<b>Elaboração do Mapa Conceitual Institucional</b>	Os mapas conceituais são impressões visuais coletivas sobre a associação dos diversos elementos (processos, tecnologia e conhecimento). Deve-se elaborar um mapa em equipe e de forma célere e intuitiva. Após uma exposição satisfatória desses elementos, é necessário um tempo para que os integrantes da equipe reflitam individualmente sobre o mapa elaborado.

Quadro 2 – Tipos de inventários realizados dentro de uma instituição.

#### 3.2.1.4 Inventário de Pessoas

Esta é a etapa mais importante para o processo de Engenharia do Conhecimento. Segundo Gratton e Ghoshal (2003) (ver seção 3.1), identifica-se o capital humano através de conhecimentos não-cognitivos, que são também conhecimentos importantes nas instituições e,

por essa razão, devem fazer parte do capital das organizações. Por conseguinte, procurou-se um caminho para identificá-lo e representá-lo nos Sistemas Baseados em Conhecimento.

Esta complexa rede de comunicação, entre as diversas áreas de talento, irá fornecer a necessária flexibilidade, versatilidade e adaptabilidade para as inteligências, inclusive emocional, acontecerem. (Ver quadro 4).

<p><b>Identificação das emoções e expectativas individuais</b></p>	<p>Esta etapa identifica as expectativas individuais dos especialistas em relação a si próprios e em relação ao escopo do sistema. Também se levam em consideração as expectativas dos engenheiros do conhecimento. É preciso identificar como os especialistas e os usuários do sistema se vêem no processo, e identificam as atribuições que lhes foram outorgadas, após a implementação do sistema. A importância desta etapa está no fato de que a construção de um sistema baseado no conhecimento é resultado também das expectativas individuais dos envolvidos no processo de construção da base do conhecimento. Ao se estabelecer uma equipe de especialistas, a equipe de Engenharia do Conhecimento deverá interagir harmonicamente com ela.</p>
<p><b>Elaboração do Mapa Conceitual Pessoal</b></p>	<p>Na elaboração dos mapas conceituais referentes às pessoas de uma instituição ou associadas a um domínio de conhecimento, deve-se levar em conta as impressões individuais sobre a função que a pessoa exerce e a sua relação com os elementos associados à execução do objetivo do sistema inteligente. Da mesma maneira que o mapa conceitual institucional foi elaborado, deve-se elaborar o mapa pessoal. Após uma exposição satisfatória desses elementos, é necessário um tempo para que os integrantes da equipe reflitam individualmente sobre os mapas elaborados.</p>

Quadro 3 – Tipos de Inventários de Pessoas.

Em resumo, foram identificados três pontos principais necessários ao desenvolvimento da primeira etapa da metodologia:

- 1) Conhecimento da resistência às emoções envolvidas no processo de Engenharia do Conhecimento através da identificação das expectativas individuais;
- 2) Uniformização do vocabulário;
- 3) Manutenção do compartilhamento contínuo de informações.



### 3.2.2 Fase 2: Visualização na Engenharia do Conhecimento

Permite identificar todos os atributos e valores que possam constituir o sistema, que é visualizado em forma de caso (ver item 2.2.1.1).

A ‘Visualização’ consiste em (ver item 3.1.2):

- 1) Observação rigorosa dos elementos inventariados na etapa de compartilhamento do conhecimento para acolher somente aquilo que é evidente<sup>92</sup>;
- 2) Dividir em partes para resolver o problema mediante análise;
- 3) Resolver os problemas em ordem, partindo dos mais fáceis para os mais complexos<sup>93</sup> (a diferença entre idéias simples e complexas é do grau, não do tipo, de modo que o que é simples para uma pessoa pode parecer complexo para outra);
- 4) Fazer enumerações completas, isto é, a compreensão coletiva de cada parte que compõe o sistema ou o domínio de aplicação.

### 3.2.3 Fase 3: Identificação de Relevância na Engenharia do Conhecimento

Nesta fase é feita a especificação do sistema, bem como o modelo de interface. Esta fase é baseada na visão de qualidade de Pirsig (ver item 2.3.1.1). Para Pirsig (1984), quando a qualidade entra em cena, a pré-seleção dos fatos deixa de ser arbitrária. Ela não se baseia em opiniões subjetivas e caprichosas, mas na ‘Qualidade’, na realidade em si (PIRSIG, 1984)<sup>94</sup>.

Assim, baseado no inventário realizado e na compreensão mútua e síncrona do domínio de aplicação, nesta fase é possível identificar a abrangência do Sistema Baseado em Conhecimento dentro do escopo em que ele está sendo construído.

---

<sup>92</sup> Embora, tradicionalmente, o método cartesiano exclua a emoção, por considerá-la vaga, nesta metodologia, a emoção também é considerada no inventário, ela é o centro da própria representação do conhecimento, pois não há distinção em entre corpo e mente.

<sup>93</sup> Fica claro, que o simples e complexo na ordem do conhecer, que Descartes não entende aquilo que é mais imediatamente apreensível no processo genético de aquisição de um conhecimento, e sim aquilo que é princípio de explicação de um dado, seja ele matemático, físico ou metafísico; por essa razão, adicionou-se a esses preceitos, princípios relacionados à organização da mente, baseado em recentes estudos da neurociência.

<sup>94</sup> Pirsig, Robert M. Zen e a Arte de manutenção de motocicletas: uma investigação sobre valores. Tradução de Celina Cardim Cavalcanti. Rio de Janeiro:Paz e Terra, 1984. Pág. 255.

### 3.3 ENGENHARIA DA MENTE APLICADA À ENGENHARIA DE ONTOLOGIAS

As Ontologias fornecem um vocabulário comum de uma área e define, com níveis diferentes de formalidade, o sentido dos termos e os relacionamentos entre eles. A Engenharia de Ontologias refere-se ao conjunto das atividades relacionadas ao processo do desenvolvimento da Ontologia, o ciclo de vida da Ontologia, os métodos e metodologias para a construção das Ontologias, e as ferramentas e a linguagem que as suportam (GÓMEZ-PÉREZ et al, 2003)<sup>95</sup>.

Atualmente, as Ontologias estão sendo usadas amplamente na Engenharia de Conhecimento, na Inteligência Artificial e na Ciência da Computação; nas aplicações relacionadas às áreas da Gestão do Conhecimento, Processamento da Linguagem Natural, Comércio Eletrônico, Integração Inteligente da Informação, Bio-Informática e Educação; e em muitos outros campos do conhecimento.

A metodologia apresentada nesta seção oferece as introduções principais da Engenharia de Ontologia e discorre sobre aspectos práticos de selecionar e de aplicar conhecimentos, uso de ferramentas para construir Ontologias e a própria Suíte de Engenharia do Conhecimento (ver Capítulo 2, seções 2.1 e 2.2).

#### 3.3.1 Fase 1: Compartilhamento do Conhecimento na Engenharia de Ontologias

Na Engenharia de Ontologias, o compartilhamento do conhecimento realiza-se de maneira mais simples e célere do que na Engenharia do Conhecimento. Como já se conhecem o escopo do sistema e o contexto de aplicação das Ontologias, a primeira fase tem como objetivo a elaboração de uma lista exaustiva sobre o vocabulário do domínio.

O processo de construção das Ontologias inicia-se com o compartilhamento do conhecimento entre os especialistas do domínio e o engenheiro do conhecimento, visando um trabalho sincronizado, a fim de atingir o objetivo do sistema. O especialista é o perito ou *expert*, com alto grau de conhecimento em dado domínio, que possui habilidade na

---

<sup>95</sup>Gómez-Pérez, Asunción, Fernández-López, Mariano, Corcho, Oscar. *Ontological Engineering*. Published in November 2003 by Springer Verlag as part of the *Advanced Information and Knowledge Processing* series. ISBN 1-85233-551-3

transmissão desse conhecimento. O engenheiro do conhecimento é responsável por viabilizar o compartilhamento, organização e representação do conhecimento no sistema. Para tanto, deve estimular a integração da equipe e a discussão sobre os objetivos estratégicos da organização, para identificar os assuntos relevantes a serem representados em forma de Ontologias, no contexto de trabalho definido.

O trabalho desenvolvido em sincronia possibilita a formação de uma base de conhecimento fundamentada nas relações entre as expressões relevantes de um contexto, num ambiente cooperativo de trabalho.

A coerência na construção das Ontologias somente será viável quando todos os especialistas e engenheiros do conhecimento conceberem de uma mesma forma a solução de um problema, isto é, visualizarem o objeto e os objetivos do sistema de forma síncrona e efetiva. Essas características devem ser levadas em conta no momento do compartilhamento, uma vez que aí serão definidas as diretrizes para a representação, que será realizada em seguida. (Ver seção 2.1).

#### 3.3.1.1 Identificação do Vocabulário Controlado do Domínio

O vocabulário controlado é composto pelas Expressões Indicativas extraídas dos textos identificados como fontes e das interações entre a equipe de especialistas. Esses documentos devem identificar significativamente o domínio de conhecimento que se pretende representar. As Expressões Indicativas são identificadas através do extrator de frequência. Expressões Indicativas são palavras ou composto de palavras, também denominados termos, que expressam um significado de modo a representar o domínio de aplicação do sistema.

O processo de elaboração das Expressões Indicativas tem seu início na fase de visualização do ambiente, uma vez que é na visualização que vão sendo elaboradas as relações, e essas relações também representam as ações que interagem dentro do domínio. Assim, após a visualização do ambiente, a fase seguinte é a identificação do vocabulário.

Na identificação, são considerados os termos ou palavras ou ainda as expressões utilizadas de forma corrente, isto é, utilizadas no dia a dia, além daquelas consideradas formais. Como formais, são classificados os termos técnicos ou aqueles encontrados em documentação oficial.

A distinção entre o vocabulário usual e o formal do domínio é fundamental para a elaboração do dicionário, pois, seguindo os propósitos da Ontologia, os termos e expressões elaborados devem traduzir em linguagem natural, entendida por todos, o conhecimento que está contido em cada domínio. A utilização do vocabulário usual ou formal também deve considerar o tipo de informação que consistirá em fonte para o sistema. Documentos técnicos precisam de um vocabulário formal para ser analisado, já documentos abertos, como mídia, por exemplo, precisam de um vocabulário usual.

A atividade de identificação do vocabulário conta com a participação dos especialistas do domínio, no sentido de relacionar todos os termos considerados expressivos para representar o domínio. Na identificação, todos contribuem num processo de *brainstorm*<sup>96</sup>, em que os termos vão sendo elencados, de forma visível, para que todos possam ter a noção do contexto do domínio e a pertinência de cada termo. Os membros das equipes podem também identificar se todos os termos são de conhecimento geral, se todos têm o mesmo entendimento dos conceitos, que esses termos pretendem representar, além de ser o momento de sugestões e/ou padronização de vocabulário.

O Vocabulário Controlado é uma relação de Expressões Indicativas construídas que descrevem quais os termos representam o conhecimento do domínio de aplicação do sistema, formando uma lista a ser validada pelo extrator semântico.

Para um melhor entendimento do processo, abordar-se-á a construção do Vocabulário Controlado em etapas. São elas:

- 1) Inventariar todo o domínio, isto é, catalogar todas as fontes de informação digital que servirão como base de dados do sistema. Exemplo: boletins de ocorrência.
- 2) Aplicar o extrator de frequência de palavras em cima da base de dados inventariada. Esse extrator permite explorar um grupo de documentos, analisar e organizar as palavras de acordo com a sua frequência. O extrator de frequência é um *software* de mineração de textos com abordagem estatística que permite explorar um grupo de documentos não-estruturados e retirar deles palavras organizadas por sua frequência em um documento ou distribuída por todos. Elabora grupos para visualização

---

<sup>96</sup> Significa "Tempestade Cerebral". Técnica de reunião coletiva de criação, adotada principalmente em agências de P.P. Consiste em reunir pessoas de diferentes especialidades, envolvidas na elaboração de uma campanha, para a discussão livre e descontraída, onde os participantes podem expor qualquer idéia, por mais absurda que pareça, sobre todos os aspectos relacionados à criação e ao desenvolvimento da campanha, sobre o produto, seu mercado, possibilidades, características, possíveis slogans, etc.

estatística e análise de contexto. Os resultados do extrator de frequência são avaliados pelo especialista para a construção do Vocabulário Controlado.

- 3) Comparação entre os resultados dos extratores com as necessidades dos especialistas;
- 4) Construir, junto com o especialista um Vocabulário Controlado representativo do domínio. Geralmente o domínio é organizado na forma de domínio e subdomínios.
- 5) Utilizando o Vocabulário, aplicar o extrator semântico na base de dados. O extrator semântico ajuda a encontrar outras palavras e Expressões Indicativas, construídas a partir do extrator de frequência. Essa ferramenta pode ser utilizada sempre que for necessário fazer a validação de novas expressões. O extrator semântico utiliza a base de conhecimento construída no Vocabulário Controlado e faz sua validação em comparação com o conteúdo dos documentos estudados, permitindo o refino do mapa de conhecimento textual encontrado. Trabalha, também, com a exploração dos conceitos referentes ao tema abordado, facilitando a expansão dos termos e a localização de outros, relacionados.
- 6) Avaliar o resultado com base na frequência das Expressões Indicativas encontradas e definir uma lista de palavras;

Nas seções seguintes, serão detalhadas as demais etapas da construção das Ontologias, apresentando-se o processo de Validação dos Termos e Expressões Indicativas, realizado em conjunto com as equipes de especialistas e engenheiros do conhecimento.

No quadro 5 (ver pág. 85), apresenta-se exemplos de Vocabulário Controlado no contexto da Segurança Pública – Domínio Crime Organizado: ativo ilícito.

<b>COAF</b>	<b>GAFI/FATF</b>
Combate à lavagem de dinheiro	Golpe bancário
Combate à reciclagem de dinheiro	Grupo de Ação Financeira contra
Combate à sonegação	Lavagem de Dinheiro
Conselho de Controle da Atividade Financeira	Infração bancária
Conta bancária irregular	Inteligência Financeira
Conta CC5	Lavagem de dinheiro
Conta CC-5	Lavagem de recurso
Conta de não-residente no país	Lavanderia de dinheiro
Conta em Bahamas	Ocultação de bem
Conta fantasma	Ocultação de direito
Conta laranja	Off-shores
Conta na Ilha Cayman	Omissão de bem
Conta na Ilha Virgem	Operação de lavagem
Conta na Suíça	Origem ilícita do capital
Conta no exterior	Paraíso fiscal
Conta no Uruguai	Remessa de divisa
Crime contra o sistema financeiro	Remessa de recurso
Crime de lavagem	Remessa ilegal de dinheiro
Crime organizado	Remessa ilegal de recurso
Desvio de dinheiro público	Repressão à lavagem de dinheiro
Desvio de verba pública	Repressão à reciclagem de dinheiro
Dinheiro de origem ilícita	Repressão à sonegação
Dinheiro lavado	Segredo bancário
Dinheiro obtido de origem ilícita	Sigilo bancário
Dinheiro sujo	Sigilo financeiro
Erradicação da reciclagem de dinheiro	Sonegação bancária
Esquema criminoso	Sonegação de bem
Estrutura criminosa	Sonegação de verba pública
Evasão de divisas	Sonegação na receita federal
Evasão de recurso financeiro	Transferência de fundo
Extorsão bancária	Valor de origem ilícita
	Valor objeto do crime
	Valor oriundo de crime
<b>FATF</b>	
Financial Activities Task Force	
Fraude de banco	
Fraude de verba pública	
Fraude do dinheiro público	

Quadro 4 – Vocabulário Controlado no contexto da Segurança Pública.

### 3.3.2 Fase 2: Visualização na Engenharia de Ontologias

Com base nas etapas descritas nesta metodologia e dentro do contexto de aplicação estipulado na etapa de Engenharia do Conhecimento, os engenheiros e os especialistas definem como os termos do vocabulário vão se inter-relacionar para atingir o escopo do sistema. O resultado é a criação de temas e subtemas para categorização das Ontologias.

### 3.3.2.1 Criação de Domínios e Subdomínios (Temas e Subtemas)

Definem-se as Ontologias na Suíte de Engenharia do Conhecimento através das suas relações e também pelos domínios e subdomínios. As Ontologias pertencem a um domínio ou subdomínio parcialmente ou na sua totalidade.

A definição dos domínios está essencialmente relacionada com os resultados esperados do sistema, uma vez que o sistema utilizará a estrutura de domínios para organizar e apresentar o conhecimento armazenado.

Os domínios formam uma árvore hierarquizada, agrupando os assuntos de forma a facilitar a análise das informações da base de conhecimento. Eles são formados por grupos de palavras e Expressões Indicativas conectadas entre si por meio de relações de ‘sinônimos’, ‘conexos’, ‘parte de’ e ‘tipo de’.

O limite de cada domínio será definido pela abrangência e profundidade que se pretende dar ao sistema. Por esse motivo, não é necessário seguir uma organização taxonômica ou institucional para a organização da árvore de domínios; deve-se, sim, dar preferência para uma organização que facilite a localização das informações e a verificação de destaque, para se determinar os assuntos.

### 3.3.2.2 O Desenvolvimento do Dicionário e de suas Relações

É importante salientar que, antes de se criar uma Ontologia, é imprescindível estabelecer a amplitude da mesma, pois não é possível construir uma Ontologia que abranja todos os conceitos e relacionamentos de um domínio específico. Assim, é fundamental ter clareza do que se pretende representar, porque, como já se disse, somente os especialistas da área podem definir a importância ou a relevância dos assuntos.

Para relembrar, domínio pode ser entendido como uma delimitação do escopo de atuação do sistema, isto é, expressa o vocabulário relacionado a um contexto, no caso em questão: Segurança Pública. O entendimento comum do conceito de domínios vai permitir uma recuperação com melhor desempenho.

Existem diversas metodologias para construção de Ontologias, mas para cada caso devem ser consideradas a aplicação e suas possíveis extensões. Neste Capítulo são estudadas as etapas de finalização do processo de construção de Ontologias para utilização no sistema com base no Vocabulário Controlado. A partir das Expressões Indicativas, o Vocabulário é expandido por meio de relações técnicas e usuais e da constante validação realizada pelo extrator semântico, que tem a função de definir o valor do índice mais útil para a recuperação de um documento.

A construção das Ontologias é baseada nas relações de sinônimo, tipo de (hipernímia e hiponímia), parte (meronímia e holonímia) e conexo, cujos conceitos foram definidos a partir do estudo de outras teorias e metodologias de aquisição e estruturação de conhecimento em outros domínios de aplicação que também utilizam as Ontologias para representar o conhecimento.

A definição das relações (sinônimos, homônimos e hiperônimos) tem como base o vocabulário usual e formal sobre o assunto, construídos na etapa anterior, conforme a necessidade do sistema, que serão utilizados na recuperação das informações, como por exemplo: os boletins de ocorrência.

As relações de sinônimo, tipo de (hipernímia e hiponímia), parte de (meronímia e holonímia) e relação de conexo são utilizadas no Sistema KMAI para expressar uma relação matemática de proximidade e frequência entre os termos. Isto significa que não se tem somente uma definição taxonômica a ser elaborada. Por exemplo: se, no domínio mamífero, a relação 'tipo de' for definida com todos os tipos de mamíferos existentes, isso pode não servir para o propósito do sistema. Isto é, será necessário investigar na base de informações (Internet ou Banco de Dados) se existe uma relação de frequência e proximidade entre os termos associados. Então, evita-se fazer uma listagem 'de tipos' ou 'partes'. É o que se chama de 'contextualização'. A definição das relações está mais próxima da sistemática do que da taxonomia.

Além disso, trabalha-se com o que se denomina 'Expressões Indicativas'. Expressões Indicativas procuram indicar que um determinado 'tipo' ou 'parte de' está no texto do documento. Por exemplo: tribo indígena e seus sinônimos (aldeia de índios, povo indígena, etc.) estão presentes em mais de 90% dos documentos não estruturados ou semi-estruturados (notícias, textos, relatórios, etc.) que expressam o nome das tribos (v.g., Guarani, Kaxinawá, etc.). São mais de 300 tipos de tribos no Brasil; se contadas as denominações de suas variações, seriam em torno de mil (1000) termos para definir esta Ontologia. Se utilizadas as



Expressões Indicativas, os documentos podem ser recuperados com a mesma precisão, utilizando em torno de 23 termos.

Para a elaboração das Ontologias, de modo específico, utilizam-se as Leis do pensamento de Wood<sup>97</sup>. Observa-se que algumas idéias surgem através da capacidade mental de comparação – isto é, através de uma faculdade lógica – enquanto outras surgem simplesmente na imaginação sem nenhuma razão, a não ser aquela que foi gravada em cima dela em algum tempo precedente. A comparação pertence às primeiras três leis; a imaginação, somente à quarta. Ver quadro 5 para entender melhor o conceito dessas relações:

<b>Relações existentes na Suíte de Engenharia do Conhecimento</b>	<b>Descrição</b>	<b>Como se define</b>
<b>Sinônimos</b>	É uma relação existente entre expressões com o mesmo significado, dentro de um mesmo domínio, ou seja, as expressões de uma mesma relação podem ser substituídas sem alterar a compreensão do texto. Ex.: tráfico de drogas é <u>sinônimo</u> de narcotráfico.	Lei da Qualidade. Ela expressa a relação entre um objeto e sua qualidade, e também entre os objetos que têm a mesma qualidade.
<b>Conexos</b>	Relação Conexo: representa a conexão existente entre termos fortemente relacionados que não se enquadra em nenhum outro tipo de relação. Ex.: trânsito na fronteira é <u>conexo</u> de controle migratório.	Lei da Proximidade. Ela diz respeito às experiências impressionantes e familiares referente a nós próprios, e tem mais conexão com a imaginação do que a observação lógica. Se ver ou pensar duas coisas freqüentemente e fortemente juntas, a força de seu encontro na minha consciência tenderá a dar-lhes uma associação permanente em minha mente.
<b>Isso é um tipo de - (hipernímia) Isso é um tipo disso (hiponímia)</b>	É a relação existente entre expressões da qual se depreende uma relação de categoria e classe, ou gênero e espécie. É a relação existente entre um lexema mais específico ou subordinado e um lexema mais geral, ou superordenado. Ex.: cocaína é um <u>tipo de</u> substância entorpecente.	Lei da atração entre idéias, Idéias de coisas similares aderem-se uma a outra, e sugerem-se facilmente. Nós chamaremos este primeiro princípio de lei da classe. Ela inclui as relações entre um objeto e a classe a que pertence, e também aquela entre objetos da mesma classe. Chega-se a ela pela observação, de modo momentâneo, das suas semelhanças e diferenças.
<b>Isso é parte de (meronímia) É parte disso (holonímia)</b>	É aquela relação que determina uma idéia de fração e todo, sendo mais comumente encontrada na estrutura de organizações (ex.: Secretaria é <u>parte de</u> Ministério) ou nos casos em que a descrição das partes é tão relevante quanto a previsão do todo.	Lei das Partes. Quando você pensa sobre um elefante você provavelmente dará forma a retratos mentais especiais do seu tronco, ou orelhas, ou pés, ou quando você pensar nas suas orelha, você pode também pensar em outras partes dele, tais como os olhos.

Quadro 5 – Leis da Percepção aplicada à construção de Ontologias Suíte de Engenharia do Conhecimento.

<sup>97</sup> Wood, Ernest E., Mind and Memory Training. Op. Cit. pág.11.

### 3.3.3 Fase 3: Identificação de Relevâncias na Engenharia de Ontologias

Nesta fase, identifica-se a profundidade da representação da Ontologia, ou seja, definem-se os seus limites em relação ao contexto. Podem-se criar Ontologias indefinidamente. A estrutura da Suíte de Engenharia do Conhecimento permite uma constante manutenção e atualização com a base de conhecimento do sistema inteligente. Então, estabelecer a extensão das Ontologias para que o sistema inteligente possa atingir o escopo, para o qual se construiu, é ponto fundamental para a viabilidade desse sistema.

No processo de construção das Ontologias, é imprescindível a uniformização do vocabulário para a criação de um entendimento compartilhado do domínio de aplicação do sistema. Através de processo colaborativo entre o engenheiro do conhecimento e a equipe de especialistas, primeiramente é realizada a uniformização do conhecimento, que tem por objetivo o entendimento comum do contexto da aplicação.

A construção dos vocabulários está diretamente ligada à representação do conhecimento de um domínio. Assim, o conhecimento desse domínio precisa ser comum a todos os membros da equipe. A participação do especialista é de fundamental importância, uma vez que, além do conhecimento das estratégias e procedimentos usuais, também será o especialista que fornecerá informações sobre o que deve ser recuperado com eficiência pelo usuário e quais as particularidades de cada conteúdo.

Para que seja possível uma perfeita recuperação das informações, as linguagens utilizadas pelos usuários precisam ser ‘entendidas’ pelo sistema de recuperação empregado. A utilização das Ontologias vai favorecer esse entendimento comum que tem seu ponto de partida na sincronização das equipes, na percepção e na elaboração dos conceitos.

Na sincronização, a visualização do ambiente é fundamental, em que todos devem ter o mesmo entendimento do que vai ser representado e do que se pretende como recuperação. Na visualização, as dinâmicas, como a percepção do ambiente, elementos envolvidos, rotinas, materiais utilizados, delimitação física, são alguns dos componentes utilizados para que todos os membros da equipe estabeleçam a mesma definição, tenham o mesmo conceito sobre o que vai ser representado. Esse processo de integração das equipes envolvidas no compartilhamento das informações determina a definição das Expressões Indicativas e a construção dos relacionamentos necessários para a representação do conhecimento,

identificando as fontes básicas de informação que servirão como insumo para a construção das Ontologias.

A partir do entendimento comum do contexto do sistema, é realizada a definição dos domínios de aplicação. Na identificação de um Domínio, são necessários levantar os tipos de informação, as fontes, o tipo de linguagem e a especificação exata do tipo de problema que o sistema se propõe a resolver, bem como estudar a documentação (bibliografia) sobre o assunto. Depois da definição dos domínios de aplicação, é elaborada uma lista de palavras e Expressões Indicativas para cada domínio, denominada Vocabulário Controlado. Com base no vocabulário, elaboram-se as Ontologias e, expandindo-se esses conceitos, criam-se novas Expressões Indicativas relacionadas para representar o contexto da aplicação.

Da mesma forma que a definição dos domínios tem grande influência nos resultados apresentados pelo sistema, a criação das Expressões Indicativas e suas relações também deve considerar alguns parâmetros.

- a) Aos objetivos do sistema: deve-se observar em que partes do sistema as Ontologias serão utilizadas e de que maneira, para que a representação seja útil para todas elas, evitando apresentar inconsistências nos resultados dos diferentes módulos.
- b) Na tipificação dos documentos utilizados para a construção das Expressões Indicativas, é possível utilizar documentos técnicos, legislação e até mesmo documentos internos da organização, como as atas de reunião. No entanto, é de suma importância que a representação seja focada nos documentos que servirão como fontes de informação para o sistema.
- c) Na tipificação dos documentos das fontes de informação do sistema, a representação do conhecimento deve ter como foco os documentos que irão compor a base do sistema, e que serão recuperados e apresentados nas interfaces de análise. Portanto, não é eficiente uma representação que contenha somente termos técnicos, quando o objetivo do sistema for a recuperação de notícias da mídia.

Assim, a partir de uma definição do termo de maior abrangência, os demais vão sendo construídos conforme o seu relacionamento, formando as teias que, no sistema, estabelecem os relacionamentos semânticos. O processo de elaboração do vocabulário, que vai compor o dicionário, é desenvolvido em etapas passíveis de avaliações constantes.

Quando se elabora um vocabulário, todos contribuem com seu conhecimento para que os termos representem com clareza todos os tópicos ou ações que se pretende representar.

Essa contribuição vai além da sugestão, pois tanto especialistas quanto engenheiros do conhecimento extraem dos termos sugeridos aqueles de maior relevância. A identificação da relevância dos termos vem da noção, mesmo que inconsciente, do conceito e das suas características, além de um conhecimento prévio dos documentos. Outra contribuição, quando da utilização de fontes de mídia, é identificar a qualidade e a quantidade dos itens recuperados com o termo.

Assim, o termo mais abrangente e com maiores características vai ser o ponto inicial das futuras relações. Da mesma forma, a relevância está na pertinência e na abrangência do termo para a representação do domínio. A relevância é determinada pela propriedade do termo em representar as informações necessárias, e deve ir de encontro ao objetivo da busca. O termo será relevante quando concentrar características que determinem o seu grau de importância, isto é, aquilo que ele representa engloba diversos determinantes, que são imprescindíveis para a representação do conhecimento.

O especialista atua de forma efetiva na determinação de relevância, contribuindo com seu conhecimento prévio para a identificação dos termos relevantes. A identificação de relevância dos termos determina que o dicionário final tenha objetividade, clareza, abrangência e pertinência.

O objetivo do dicionário, que faz parte do Editor de Ontologias, apresentado no item 4.1.4.5, é concentrar os termos que melhor representem o domínio, dentro do contexto pretendido. Assim, o usuário poderá se valer dele para recuperar informações valiosas para a gestão das suas atividades. Além disso, a identificação da relevância dos termos favorece a concentração das informações, além de impedir a construção de listas extensivas de termos, que não tenham representatividade efetiva.

### 3.3.3.1 Validação dos Termos

Como visto anteriormente, o processo de elaboração do vocabulário constitui-se em várias etapas. A validação, teoricamente, está classificada como etapa de finalização do processo de construção do vocabulário. A validação pode representar uma nova construção, razão pela qual não se pode afirmar categoricamente que seja uma etapa de finalização. As avaliações são constantes, caracterizando a dinamicidade da Ontologia como linguagem de representação do conhecimento.

Nesta etapa de validação, os engenheiros do conhecimento, especialistas no sistema que participaram do processo de Engenharia do Conhecimento e elaboração do vocabulário, analisam a construção dos termos e expressões. São consideradas, aqui, as relações construídas, a representatividade dos termos, as construções de sinônimos, as construções partitivas (tipo de, parte de) e as relações conexas. Atendendo às particularidades do sistema, essas relações são uma forma de classificação voltadas ao processo de recuperação.

Assim, os termos são avaliados pela frequência que aparecem e pela qualidade das informações que recuperam. Baseados nos termos do vocabulário, as Ontologias são construídas com o objetivo de expansão das formas de recuperação. A validação dos termos visa, então, essa expansão, na qual as Ontologias devem representar uma recuperação contextualizada do primeiro ao último texto.

### 3.4 CONCLUSÃO

Após a delimitação do escopo do aplicativo e da representação do conhecimento pelas equipes de Engenharia do Conhecimento e dos especialistas, a fase de aquisição baseada em Ontologias consiste na etapa mais importante do processo, uma vez que, por meio dela, verificou-se a adequação do aplicativo na resolução do problema. Justamente, por esse motivo, que a integração constante das duas equipes é de fundamental importância para a construção do Dicionário de Ontologias.

O trabalho entre a equipe necessita não somente o conhecimento profundo sobre o domínio de aplicação do sistema, mas também sobre a organização da sua base de conhecimento. A criação de um ambiente computacional na Web permitiu um maior compartilhamento de informações e resultado entre as equipes.

## CAPÍTULO 4

### 4 A METODOLOGIA APLICADA EM SISTEMAS

#### 4.1 SISTEMAS BASEADOS EM CONHECIMENTO DESENVOLVIDOS COM A TÉCNICA RC2D E PCE

##### 4.1.1 JurisConsulta®

O JurisConsulta®<sup>98</sup> é um sistema de recuperação de jurisprudência, desenvolvido com técnicas de Inteligência Artificial, que permite:

- Facilitar o acesso do profissional do direito e também do cidadão a estas informações armazenadas em banco de dados disponíveis na Internet (Figura 5, pág. 94);
- Atualizar automaticamente a base de conhecimento do sistema, permitindo que novos casos sejam disponibilizados para o usuário, sem a intervenção do mesmo. (Figura 6, pág. 94).

O JurisConsulta utiliza como técnica de Inteligência Artificial o Raciocínio Baseado em Casos (RBC). O RBC usa experiências anteriores e semelhantes para a solução de um problema novo. O RBC é baseado no princípio de analogia, assumindo que problemas semelhantes tem soluções semelhantes (AMONDT, 1994)<sup>99</sup>. Por essa razão, o RBC é uma técnica muito adequada ao domínio jurídico, pois emprega o mesmo tipo de raciocínio utilizado pelos juristas na solução de um problema.

---

<sup>98</sup> Todos os direitos reservados da marca e uso da aplicação reservados para o IJURIS – Instituto de Governo Eletrônico, Inteligência e Sistemas. Copyright 2002.

<sup>99</sup> Amondt, A., Plaza, E., “Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches”. AI Communications, 17(1), 1994.

**JurisConsulta - Elaboração**

## Elaboração da Consulta

Alto (84%)

**Descreva o seu problema:**

O meu cliente foi preso em flagrante por tráfico de drogas. Estou pedindo a desclassificação para o delito de uso, embora ele portasse uma razoável quantidade de maconha, o objetivo não era a comercialização da mesma. A droga era para distribuição gratuita entre seus amigos numa festa. E há ausência de provas suficientes que indiquem o tráfico de entorpecentes.

### Delimitação da Pesquisa (Opcional)

Habilita

Por tipo de recurso:  Por tipificação:

Por data de publicação: De:  Até:  Por relator:  Por local (origem do processo):

Por resultado:  Concedido  Denegado

Número de casos : 2247  
Nº de casos delimitados: 101

**Dispara o processo de busca pelos casos mais similares ao elaborado.**

Consultar   
 Limpar   
 Voltar

©1999 - PPGE

Figura 5 – Interface do Sistema JurisConsulta.

Fonte: JurisConsulta

**JurisConsulta - Atualização**

## Atualização da Consulta

Arquivos de origem	Posição
26419AC.TXT	1º
26423AC.TXT	2º
26424AC.TXT	3º
26426AC.TXT	4º
26427AC.TXT	6º
26428AC.TXT	7º

Arquivos convertidos	Situação
26426ACB.TXT	OK

Acórdão :  Recurso :  Relator :

Local:  Data :  Resultado :

DJ: 8.260 DATA: 28/05/91 PAG: 08 Apelação criminal n. 26.926, de Canoinhas. Relator: Des. José Roberge. FURTO - Pretendida desclassificação para a espécie privilegiada do parágrafo 2o. do artigo 155 do CP. Não cabimento. A ausência de prejuízo da vítima não basta para caracterizar a forma privilegiada, que exige seja de pequeno valor a coisa furtada. Sursis. Não impede a concessão o fato de o agente responder outro processo. Concedido a favor em ambos, circunstância perfeitamente possível, será o mesmo cassado se as duas condenações transitarem em julgado. Vistos, relatados e discutidos estes autos de apelação criminal n. 26.926, da comarca de Canoinhas (1a. Vara), em que é apelante Mauri Pinheiro, sendo apelada a Justiça, por seu Promotor: ACORDAM, em Segunda Câmara Criminal, por votação unânime, dar provimento parcial ao recurso para conceder o sursis. Custas de lei. 1 - Na comarca de Canoinhas, perante o MM. Juiz de Direito da 1a.

**Selecione um ou mais Acórdãos que deseja converter em Casos.**

Editar   
 Confirmar   
 Voltar

©1999 - PPGE

Figura 6 – Módulo de Indexação Automática do Sistema JurisConsulta.

Fonte: JurisConsulta

Para possibilitar o processo de conhecimento textual não estruturado, ‘um Vocabulário Controlado e um Dicionário de Termos Normativos’ – baseado na terminologia jurídica usual e na teoria jurídica – são integrados na recuperação e no processo de extração do conhecimento. Assim, o usuário pode descrever uma questão em linguagem natural, e o ajuste da situação é iniciado pela comparação 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, ele será ordenado de forma decrescente, de acordo com os valores obtidos com a métrica de similaridade; e os casos mais úteis, ou seja, aqueles com valores de similaridade mais altos, serão apresentados como o resultado da consulta. O usuário poderá, ainda, selecionar na íntegra o caso mais apropriado para solução do seu problema jurídico. (Ver figura 7).

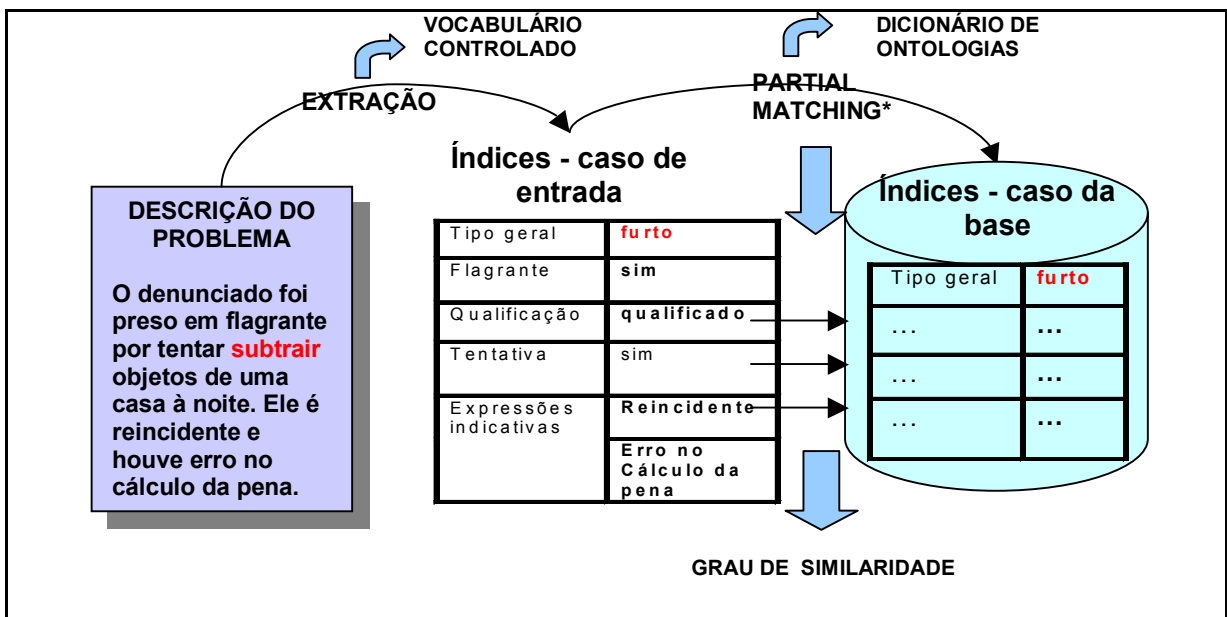


Figura 7 – Processo de Recuperação Baseado em Casos do Sistema JurisConsulta.

As principais partes do sistema JurisConsulta® são:

- 1) Estrutura de representação de casos legais e identificação de informação relevante para a recuperação de jurisprudência apropriada.
- 2) Extração automática da informação do texto legal para inclusão de novos casos jurídicos na base de caso.



A representação e recuperação de casos no sistema JurisConsulta considera que um termo normativo pode ter várias interpretações nas decisões jurídicas e são eles que indicam quais são os termos relevantes para serem empregados na construção de argumentos persuasivos na solução de problemas jurídicos. Por essa razão, 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 desses termos (com base na terminologia usual dos Tribunais) para representar a eficiência do conhecimento jurídico presente em suas decisões. Assim, a forma de representação terá maior relevância jurídica que a extração de conhecimento feita exclusivamente baseada na informação contida no próprio texto jurisprudencial.

Os argumentos são persuasivos na 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 medida (WARAT, 1994)<sup>100</sup>. Deste modo, para apoiar o processo de recuperação e a extração automática dos índices relativos ao conteúdo jurídico dos textos da jurisprudência criminal, um Vocabulário Controlado e um Dicionário de Termos Normativos foram desenvolvidos. (Ver figura 8).

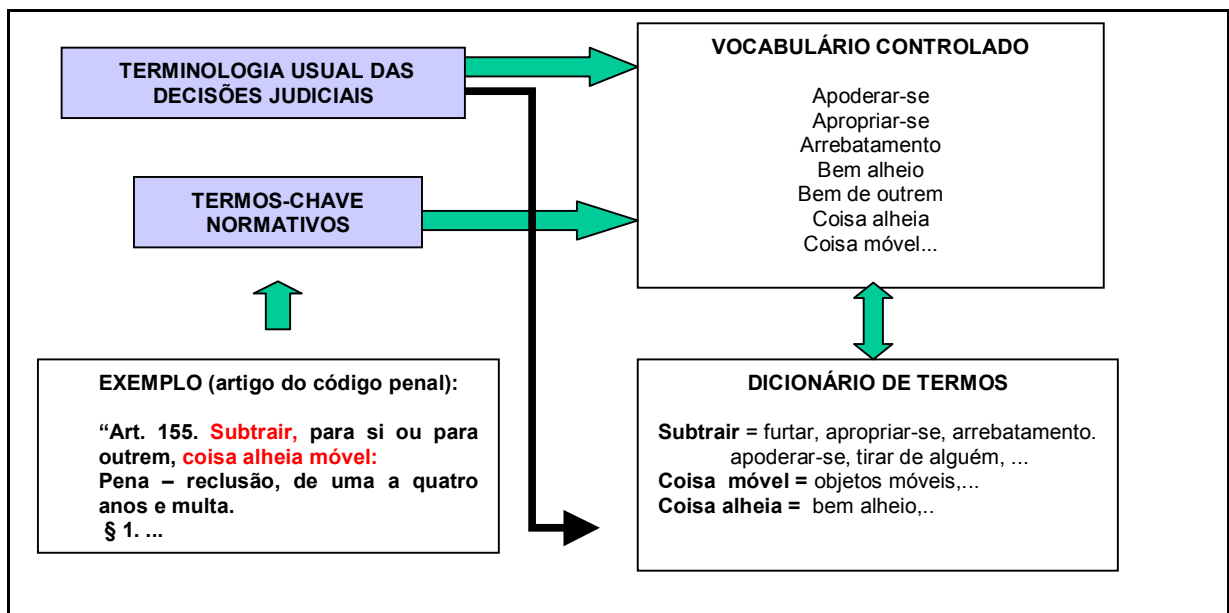


Figura 8 – Definição das Expressões Indicativas no Sistema JurisConsulta baseada na Teoria da Argumentação Jurídica.

<sup>100</sup> Warat, Luiz. Alberto. Interpretação da Lei: temas para uma reformulação. Porto Alegre: Sergio Antonio Fabris Editor, 1994.

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

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

Com base no ajuste da situação fornecida, os casos 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’.

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

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

Em que,

$S$  = 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. 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 três expressões, não existindo número

limitado de valores), enquanto que o índice ‘Tipificação’ e seus subíndices possuem um único valor. Assim, os valores dos subíndices da ‘Tipificação’ será sempre 1/5 do valor dos subíndices da ‘Expressão Indicativa’. No exemplo dado na Figura 9, a similaridade global terá o valor de 65% (sessenta e cinco por cento).

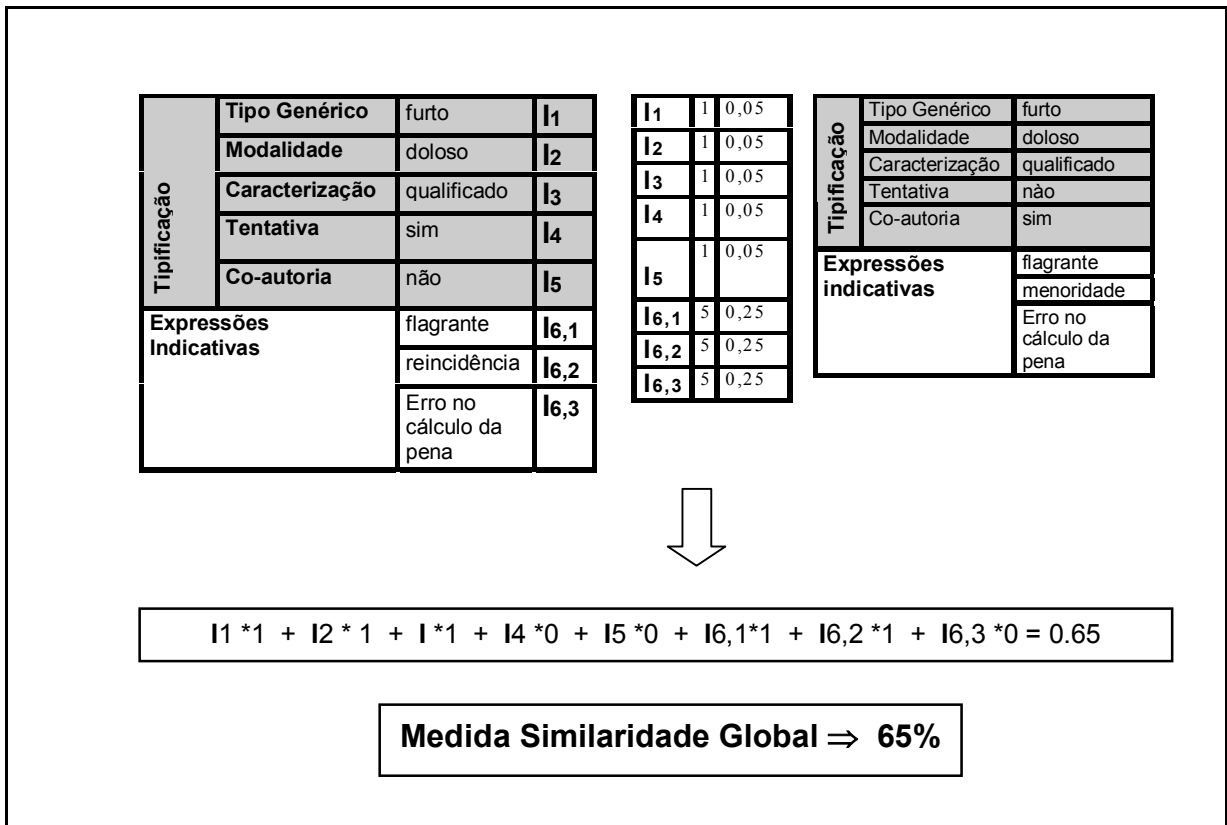


Figura 9 – Medida da Similaridade Global do Sistema JurisConsulto

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 essa razão, o valor da similaridade local deste subíndice é um (1), multiplicado por um fator de normalização. Ao contrário, no subíndice tentativa, que possui valores diferenciados, o valor da similaridade local é zero, multiplicado por um fator de normalização.

Além disso, no subíndice tipo genérico ( $f_1$ ), a determinação do valor poderá ser auxiliada pelo Dicionário de Termos Normativos. Por exemplo: se na descrição do ajuste da situação, o tipo geral seja descrito como ‘assassinato’ e o tipo geral declarado no caso é ‘homicídio’ (um sinônimo para assassinato), então, pelo uso do Dicionário de Termos

Normativos, essas 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:

$$f_{7,1}(gt_s, gt_{ci}) = \begin{cases} 1 & (gt_s = gt_{ci}) \vee (gt_s \text{ n } gt_{ci}) \\ 0 & \textit{else} \end{cases}$$

Em que,

$gt_s$  = tipo genérico do ajuste da situação

$gt_{ci}$  = tipo genérico do caso  $i$

$n$  = sinônimo do dicionário

O índice ‘Expressões Indicativas’ é representado por um conjunto de valores, ou seja, todos os valores encontrados com base no Vocabulário Controlado serão representados. Aqui, a similaridade local é calculada pela comparação de cada valor fornecido do caso da entrada com cada valor do caso da base de casos; desse 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 do índice. Pela integração do Dicionário de Termos Normativos, como acontece no índice  $f_I$ , é possível considerar ainda os sinônimos no cálculo da similaridade local.

Assim, no cálculo da similaridade local do índice ‘Expressões Indicativas’ ( $f_6$ ), em que há multiplicidade de valores, a fórmula fica assim definida:

$$\begin{array}{c}
 f_6(\{ie_{s1}, \dots, ie_{sn}\}, \{ie_{ci1}, \dots, ie_{cin}\}) \\
 = \sum_k \max\{n(ie_{sk}, ie_{cil}) \mid ie_{cil} \in \{ie_{ci1}, \dots, ie_{cin}\}\} \\
 \hline
 |\{ie_{s1}, \dots, ie_{sn}\}|
 \end{array}$$

#### 4.1.2 Olimpo®

O Sistema Olimpo®<sup>101 102</sup> tem a sua performance centrada na combinação de aspectos derivados do RBC e da recuperação de informações textuais, somados a uma adequada organização do conhecimento referente ao assunto focado (no caso específico, as resoluções do Conselho de Segurança da ONU) (HOESCHL, 2001).

O uso simultâneo de técnicas de Recuperação de Informação e Raciocínio Baseado em Casos possibilita uma entrada textual extensiva. Isso torna o sistema diferenciado em termos de performance de recuperação de informação em documentos textuais.

De um modo geral, o funcionamento do sistema apresenta um fluxo similar a outros sistemas baseados em conhecimento, segundo o qual, uma entrada manual sofre um processo de ajuste e, após, é submetida à base de documentos, dentre os quais são escolhidos os melhores, semelhante ao descrito no sistema JurisConsulta. Esse processo faz com que o espaço de busca seja reduzido de forma mais eficiente, não se tratando de uma simples seleção de documentos de acordo com as suas características superficiais, mas de uma comparação preliminar orientada pelo contexto relacionado à questão de entrada. O cálculo da métrica de similaridade do Sistema Olimpo® é semelhante ao sistema JurisConsulta, descrito no item anterior. O que os diferencia é o processo de representação e aquisição do conhecimento, denominado RC2D. Esses são os processos básicos de consulta realizada pelo sistema Olimpo. (ver item 2.1.1.2).

<sup>101</sup> Hoeschl, Hugo César. Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU. Tese de Doutorado. Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina. Florianópolis: 2001.

<sup>102</sup> Todos os direitos reservados da marca e uso da aplicação reservados para a WBSA – Sistemas Inteligentes S.A. Copyright 2001.

A pesquisa é considerada ‘contextual’ e ‘estruturada’ pelas seguintes razões:

- a) É levado em consideração o contexto dos documentos armazenados quando da formação da estrutura retórica do sistema;
- b) O contexto norteia o processo de ajuste da entrada bem como da comparação e seleção dos documentos;
- c) Quando da elaboração da consulta, a entrada não está limitada a um conjunto de palavras, ou à indicação de atributos, podendo assumir o formato de uma questão estruturada pelo conjunto de um longo texto somado à possibilidade de acionamento de atributos específicos, que funcionam como ‘filtros’ que fazem uma seleção preliminar dos documentos a serem analisados.

No Sistema Olimpo®, a informação contida nos documentos é representada em forma de casos, que consiste no conteúdo do documento original somado a um conjunto de oito (8) índices em forma de par atributo-valor: *subject, date, number of the Resolution, meeting, country, acronyms, decisions, and indicative expressions*. Esses índices fazem parte da interface do sistema. (Ver figura 10, pág. 102).

O processo de filtragem reduz o espaço de análise do sistema, atuando de forma tal que não seja um simples particionamento da base de dados, pois trata de questões de ‘mérito’, e não de ‘forma’, ensejando, inclusive, uma visualização preliminar do universo da pesquisa, antes mesmo da consulta ser deflagrada.

Além disso, o controle do nível de profundidade da pesquisa (*search level*) permite uma seleção de documentos de acordo com a maior ou menor ocorrência de Expressões Indicativas dentro de cada resolução, antes de se efetuar a comparação.

Um exemplo de sua eficiência foi constatado com a expressão *stability pact*. Segundo os métodos tradicionais, uma busca textual – estruturada com pesos e atributos – somente registraria a presença dessa expressão se ela constasse da mesma forma. Este fato geralmente é verificável mediante uma regra de produção do tipo: ‘se...’, ‘então...’, em que ‘se’ *stability pact* está presente no texto, ‘então’ lance a pontuação ‘x’ no cálculo de similaridade.

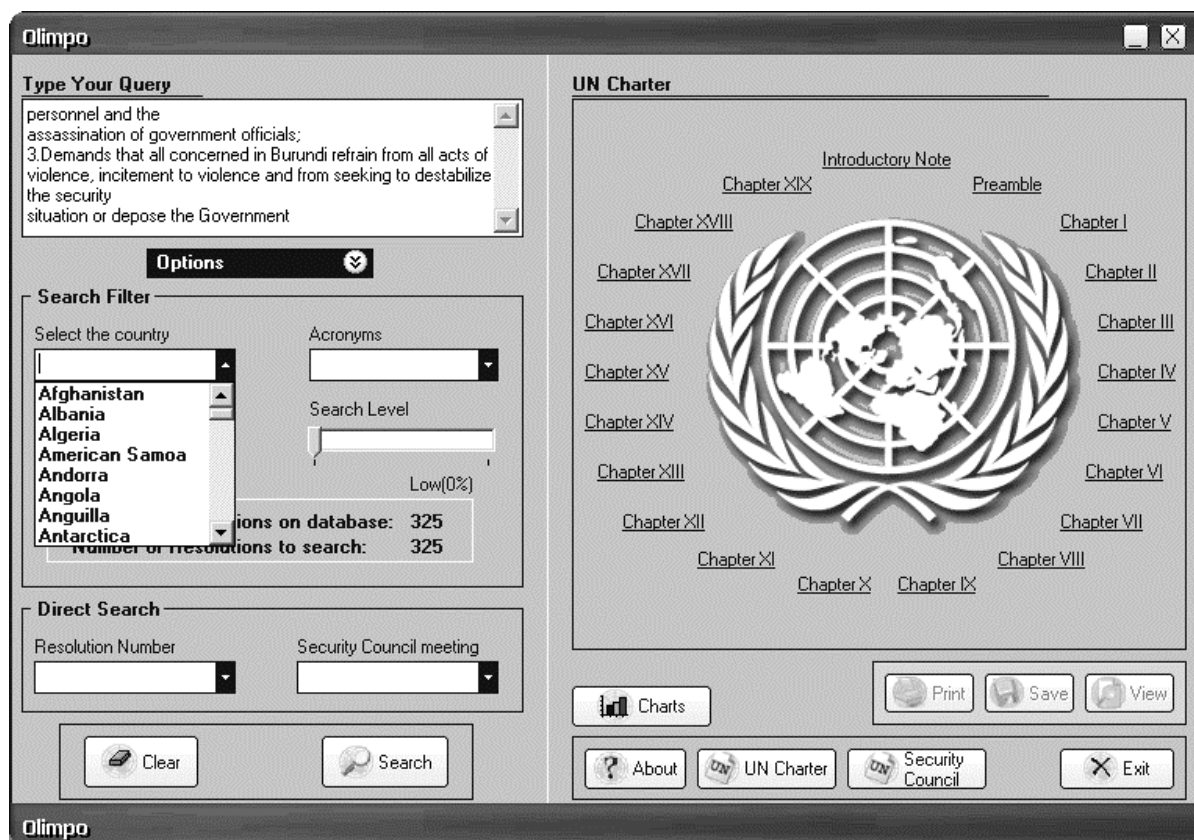


Figura 10 – Interface do Sistema Olimpo®.

Fonte: Olimpo

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

Tabela 1 – Exemplo de Resultado do Cálculo de Similaridade do Sistema Olimpo®

Classificação	Número da Resolução	Percentual de similaridade
1º.	1.244	100%
2º.	1.048	66,70%
3º.	1.159	66,70%
4º.	1.230	66,70%
5º.	1.265	66,70%
6º.	1.007	33,30%
7º.	1.014	33,30%
8º.	1.029	33,30%
9º.	1.054	33,30%
10º.	1.063	33,30%

#### 4.1.3 *Alpha Themis*®

O *AlphaThemis*®<sup>103 104</sup> consiste em um sistema inteligente para a extração do conhecimento contido nas súmulas dos Tribunais nacionais, configurando-se em uma potente ferramenta de mineração de textos em documentos norteadores de decisões judiciais. A técnica responsável pela recuperação de súmulas empregada no aplicativo é derivada do desejo de se mesclar uma pesquisa estruturada em texto e atributos que pudessem ensejar a formação de um contexto (próprio para o sucesso da busca). A modelagem conferida ao *software* tem como idéia básica tratar o conhecimento contido nas súmulas, e seus respectivos enunciados, de forma tal que a interface do sistema permita ao usuário formular uma consulta na qual seja possível mesclar atributos fixos com texto livre (BUENO et al, 2003).

O uso de pesos dinâmicos nos filtros de busca torna o *AlphaThemis*® ainda mais flexível na formulação da pesquisa. Trata-se da possibilidade de alteração dinâmica da relevância dos atributos dentro do contexto pesquisado, calibrando-se, individualmente, os pesos de cada um dos atributos, dentro de uma escala de 0 a 100. (Ver figura 11, pág. 104).

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

O sistema *AlphaThemis*® conta também com um módulo estatístico e um módulo específico, em que cada um dos enunciados das súmulas pode ser visualizado diretamente.

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

---

<sup>103</sup> Todos os direitos reservados da marca e uso da aplicação reservados para a WBSA – Sistemas Inteligentes S.A. Copyright 2001.

<sup>104</sup> Bueno et al.. Analyzing the use of dynamic weights in legal case based system. In: NINTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 2003, Edimburgo. Proceedings of the Conference. New York: ACM, 2003. v. 1, p. 136-141.



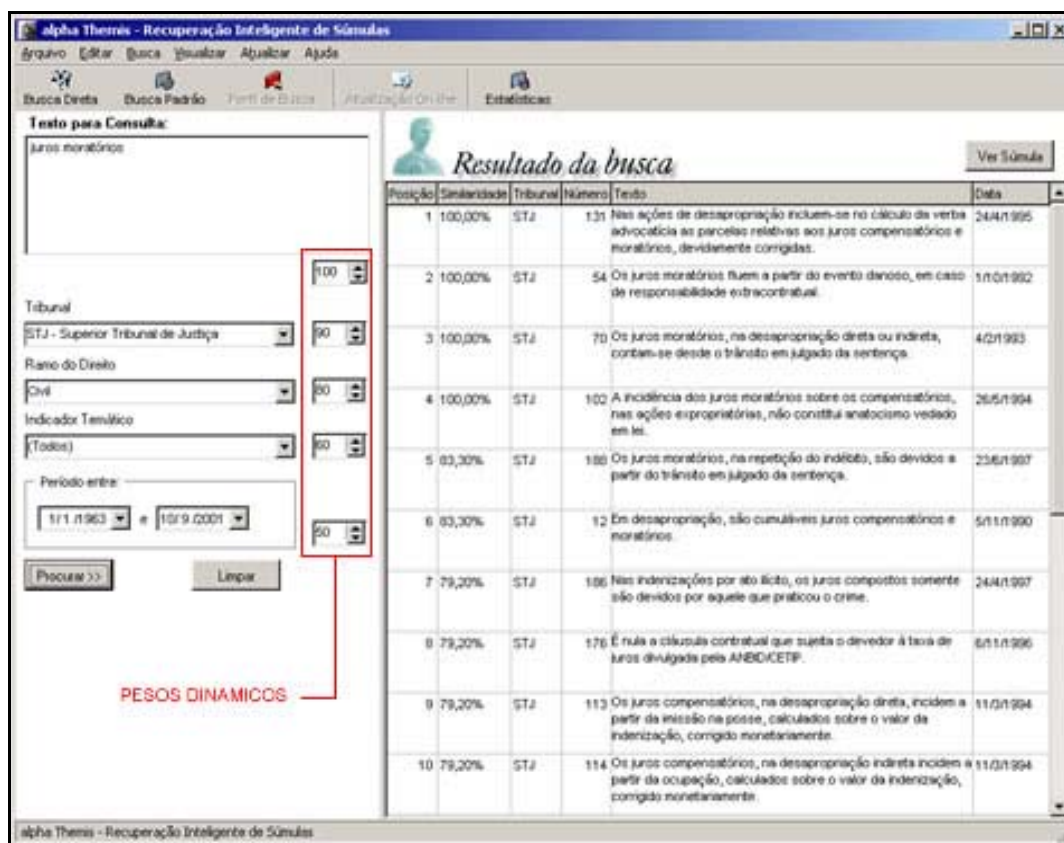


Figura 11 – Interface AlphaThemis®, mostrando a saída dos dados e ressaltando o uso de pesos dinâmicos

Fonte: AlphaThemis

Semelhante ao JurisConsulta® e ao Olimpo®, como técnica de extração de informações, o *AlphaThemis*® identifica os valores retirados pela localização dos conceitos e palavras no texto, mediante a comparação com o vocabulário previamente construído pelos especialistas do domínio, bem como pela inferência de informações via contexto da súmula. Ressalta-se que um dos pontos fortes dessa argumentação é a tarefa executada pelo programa que permite extrair conceitos subjetivos não diretamente explicitados no texto.

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

- 1) Tribunal (lista fechada, 7 itens, valor: 0,5 ou 0);
- 2) Ano (lista fechada, 34 itens, valor: 0,5 ou 0);
- 3) Ramo do Direito (lista fechada, 35 itens, valor: 1 ou 0);
- 4) Ramo secundário (lista fechada, 35 itens, valor: 1 ou 0);
- 5) Indicador temático central (lista fechada, valor: 1 ou 0);

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

Cálculo da métrica de similaridade:

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

Segundo nível ( $S_n$ ): leva em consideração as **palavras informadas** na entrada, operando sobre os campos '7' a '9', que equivalem aos outros 50% da métrica total. Cada palavra equivale a 1,0. Exemplo: se sete (7) palavras foram informadas,  $7 = 50\%$ . Na busca, duas (2) não foram encontradas; uma (1) foi encontrada no campo nove (9) (recebe o valor 0,5), duas (2) foram encontradas no campo oito (8) (recebem o valor 0,7 cada uma delas), e as demais foram encontradas no campo sete (7) (recebe o valor 1,0 cada uma delas). O valor total é  $0,5 + 1,4 + 2,0 = 3,9$  que corresponde a 27,85% da similaridade total (100%), tendo em vista que  $7 = 50\%$ . O valor final da **similaridade** é  $P_n + S_n$ , o que, no exemplo citado, corresponderia a  $30 + 27,5 = 57,85\%$ .

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

Além disso, o resultado apresenta também um somatório visual das informações solicitadas, isto é, todos os mecanismos que o usuário ativou na realização da consulta, bem como os valores respectivamente atribuídos, estão ali consignados. Isto significa que a Medida da Similaridade Global inclui todos os atributos, inclusive quando o usuário não combina alguns parâmetros da busca (e.g. Tribunal, Ramo do Direito, etc.) com o texto da busca.

Posteriormente foram feitas modificações no valor de alguns atributos, bem como realizada a exclusão de outros. Para adequar essas modificações, um novo cálculo da similaridade foi realizado, conforme descrito a seguir:

$$sim_c(S_a, S_b) = \sum_{i=1}^n (sim_l(c_{ia}, c_{ib}) * p_c), \text{ Onde,}$$

$sim_c$  é a similaridade do conceito (termo);  $sim_d$ , a similaridade da data;  $sim_{cr}$ , a similaridade do Tribunal;  $sim_a$ , a similaridade do ramo do direito;  $sim_t$ , a similaridade do tema;  $n_c$ , o número de conceitos  $S_a$ ;  $p_c$ , o peso dinâmico do conceito;  $p_d$ , o peso dinâmico da data;  $p_{cr}$ , o peso dinâmico do Tribunal;  $p_a$ , o peso dinâmico do ramo do direito; and  $p_t$ , o peso dinâmico do ramo do direito, onde  $c_{ia}$  é o número de conceitos  $i$  proveniente de  $S_a$ ,  $c_{ib}$ , o número de conceitos  $i$  proveniente de  $S_b$ ,  $sim_l(c_{ia}, c_{ib})$ .

Cada similaridade local tem a sua própria fórmula. A similaridade do conceito (termo)  $sim_c(S_a, S_b)$  é calculada através da seguinte fórmula:

A similaridade local entre  $c_{ia}$ ,  $c_{ib}$ , e  $p_c$ , o peso dinâmico do conceito (termo).

A similaridade de  $sim_l(c_{ia}, c_{ib})$  é calculada através:

$$sim_l(c_{ia}, c_{ib}) = \begin{cases} 1 & \text{if } c_{ia} = c_{ib} \\ x & \text{if } \exists r \in Rc \quad c_{ia} = r \\ 0 & \text{if } c_{ia} \neq c_{ib} \wedge \forall r \in Rc \quad c_{ia} \neq r \end{cases}$$

Em que,  $Rc$  representa o conjunto dos termos conexos e  $r$  representa um termo conexo.

Se um termo conexo proveniente da *súmula* é encontrado, o valor da similaridade é  $x$ , o qual representa a relação entre os termos.

A similaridade de data, Tribunal e o ramo do Direito são calculados da mesma maneira. Se a data da *súmula* for a mesma, a similaridade é 1, multiplicado pelo respectivo peso dinâmico definido pelo usuário. Se não for, a similaridade é 0.

A similaridade do tema é calculada através da seguinte fórmula:

$$sim_t(S_a, S_b) = \begin{cases} 1 * p_t & \text{if } t_{S_a} = t_{S_b} \\ y * p_t & \text{if } t_{S_a} = st_{S_b} \\ 0 & \text{if } t_{S_a} \neq t_{S_b} \wedge t_{S_a} \neq st_{S_b} \end{cases}$$

Em que,  $t$  é o tema primário,  $st$  é o tema secundário,  $p_t$  o peso dinâmico, e  $y$  é o valor que representa a relação entre o tema primário e o tema secundário.

Nesse sistema foi implementada a primeira versão da Suíte de Engenharia do Conhecimento (ver figura 12), principal objeto de estudo desta tese.

Figura 12 – Primeira versão da Suíte de Engenharia do Conhecimento.

No sistema *AlphaThemis*®, a primeira versão do módulo de Engenharia do Conhecimento permitiu uma aquisição do conhecimento integrada. E, embora a Suíte não tivesse quase nenhum recurso tecnológico para verificação de inconsistências, essa versão melhorou o tempo e a qualidade do trabalho dos engenheiros do conhecimento.

#### 4.1.4 Sistema KMAI®

O Sistema KMAI® é um instrumento de Gestão do Conhecimento em instituições onde as atividades de inteligência possuam importância destacada, enfocando a disponibilização e a fácil recuperação de informações estratégicas que contribuam para a dinamização do processo de tomada de decisões. O sistema é uma evolução dos sistemas anteriormente descritos (Jurisconsulto, Olimpo e *AlphaThemis*). Ele utiliza as tecnologias do PCE e RC2D (HOESCHL, 2001).

O Sistema KMAI apresenta estrutura modular, sendo dividido conforme descrito a seguir:

- a) Módulo Básico: Contemplam a análise textual e gráfica, notas informativas, fontes de informação e ambiente de administração.
- b) Módulo Básico com Editor de Ontologias: Além do módulo básico também contempla o editor de ontologias. O editor de ontologias permite a expansão dos domínios da base de conhecimento do sistema, ferramenta importante para a evolução e expansão das buscas e análises no sistema.
- c) Módulo Básico com SC-Info: Contempla módulo básico com o SC-Info. Esse módulo permite que o cliente construa novos observadores para as fontes de informação.
- d) Módulo Completo: Esse módulo contempla o módulo básico com editor de ontologias e SC-Info.

Concebido como modelo conceitual, o Sistema KMAI® possui estrutura física/virtual e uma estrutura metodológica. A estrutura física compreende o *software* para coleta, armazenamento e tratamento de informações com alto valor agregado, que visa subsidiar o processo de tomada de decisão. A estrutura metodológica é composta pela Engenharia do Conhecimento que contempla a personalização do sistema, e a Engenharia de Ontologias para a construção das expressões que farão parte da rede de Ontologias, insumo do sistema, e treinamento no Editor de Ontologias.

O sistema para a Gestão do Conhecimento Estratégico, para a produção e compartilhamento de grandes volumes de informações e de conhecimentos para organizações públicas e privadas, permite a coleta (fontes digitais de informações), processamento,

monitoramento, pré-análise e difusão das informações; confecção de dossiês e relatórios; agregar valor às pré-análises; e descobrir conhecimentos ocultos.

O KMAI® possibilita a busca em linguagem aberta e resgata os documentos mais semelhantes com o texto de entrada. Utiliza técnicas de Inteligência Artificial e metodologias de recuperação da informação desenvolvidas no Brasil; possui capacidade de reunião de informações digitais, advindas da Internet ou de arquivos locais em uma base de conhecimento classificada e indexada de forma automática; indexa novos documentos à base de dados; permite a definição dos conceitos e contextos, que serão utilizados pelas rotinas de inteligência artificial para entenderem o universo dos assuntos; e proporciona a visualização gráfica de todos os dados armazenados pelo sistema, classificados pela definição ontológica, termos, fonte, período ou combinação destes atributos.

A figura 13 ilustra as etapas do ciclo de produção de informações estratégicas:

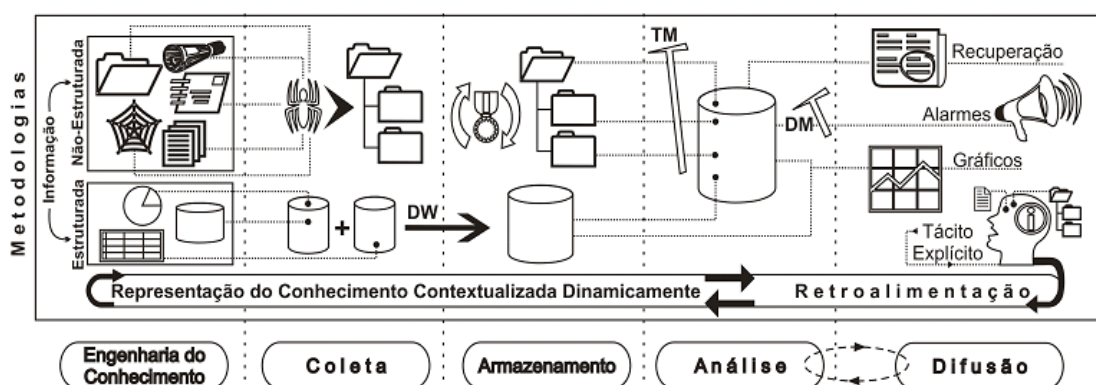


Figura 13 – Processos do Sistema KMAI®.

Fonte: KMAI

No módulo de monitoramento de fontes digitais abertas, os agentes inteligentes de coleta monitoram alvos digitais específicos, que serão definidos na etapa de Engenharia do Conhecimento. Os agentes automáticos de coleta organizam as informações colhidas, contextualizando-as em relação ao objetivo da pesquisa.

A indexação automática dos documentos acontece num ambiente em que é possível ao analista adicionar Notas Informativas, em texto aberto, aos relatórios, às informações quantitativas sobre notícias colhidas e às fontes visitadas.

No módulo de gerenciamento, o sistema conta com uma interface que indica o funcionamento do sistema de Coleta de Informações (SC-Info), trazendo informações

quantitativas sobre fontes visitadas, notícias colhidas e relatórios sobre eventuais problemas que possam ter ocorrido no processo de coleta periódico.

A plataforma contempla também o conhecimento tácito produzido constantemente pelos analistas, inserido de forma explícita através de Notas Informativas. As informações são indexadas através da Estrutura de Ontologias definidas pelos especialistas do domínio de aplicação do sistema. (Ver figura 14).

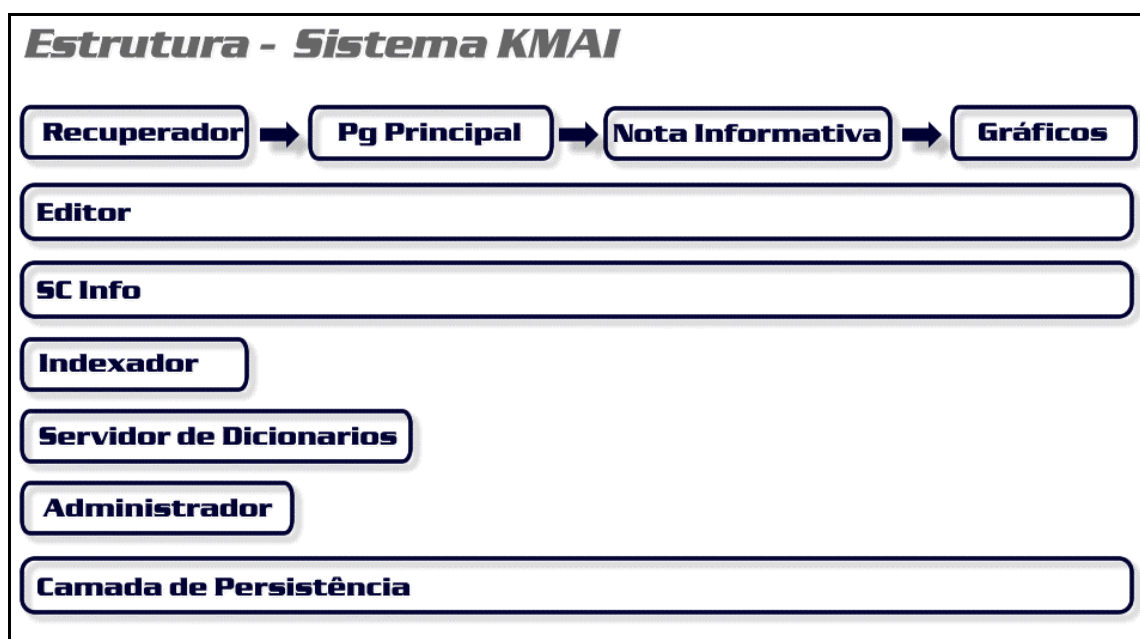


Figura 14 – Estrutura do Sistema KMAI.

Fonte: KMAI

O sistema KMAI®, no plano de armazenamento, utiliza-se de uma estrutura aberta baseada em XML, em contato direto com um banco de dados indexado inteligente e automaticamente.

O sistema KMAI® tem por objetivo apoiar investigações e operações de inteligência na coleta e análise de informações relevantes provenientes de fontes de mídia. Disponibiliza ferramentas para aumentar a capacidade de geração de cenários e caminhos para a investigação com base em fontes secundárias.

A seguir, algumas interfaces do Sistema KMAI®.

#### 4.1.4.1 Análise Textual

Permite uma busca inteligente em linguagem aberta e recupera os documentos mais semelhantes com o texto de entrada, armazenados na base de conhecimento, advindas de fontes elencadas pelo cliente (Figura 15).

Figura 15 – Análise Textual do Sistema KMAI.

Fonte: KMAI

O critério de organização das respostas apresentadas é, primeiramente, a similaridade dos documentos encontrados com o texto de entrada, que é identificada pela cor da legenda apresentada ao lado esquerdo de cada registro. O sistema considera o período de busca selecionado pelo usuário e, dentre os documentos compreendidos nesse período, organiza-os por ordem de similaridade. Num segundo momento, caso haja empate no cálculo da similaridade, é utilizada a data, sendo apresentados primeiramente os documentos mais recentes.

Outra opção de resposta disponível no sistema é a geração de um gráfico temporal, indicando a evolução do assunto descrito para pesquisa no período selecionado (Figura 16, pág. 112). O gráfico pode indicar tendências de aumento ou diminuição de repercussão de



determinada matéria que está sendo acompanhada, além de apresentar um histórico de como cada fonte trata as matérias pesquisadas no tempo.

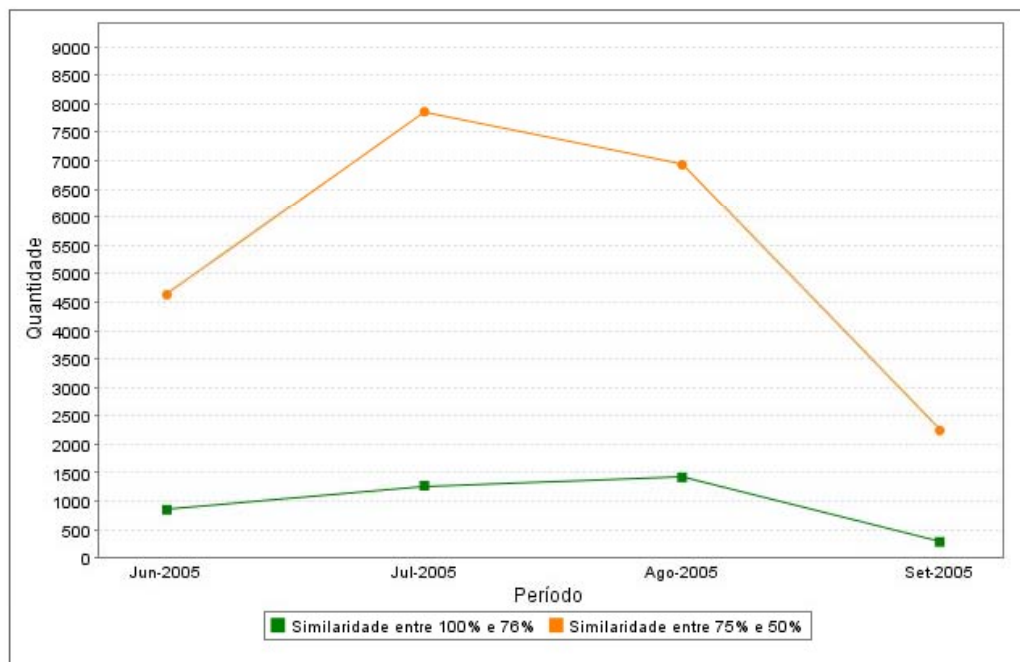


Figura 16 – Gráfico dos Resultados.

Fonte: KMAI

#### 4.1.4.2 Análise Gráfica

O diferencial dos gráficos gerados pelo Sistema KMAI<sup>®</sup> é que eles possibilitam agrupar os registros da base de conhecimento por assunto, considerando o contexto dos documentos, conforme a organização efetivada nas Ontologias (ver figura 17, pág. 113). Isto porque as Ontologias permitem que se organize o conhecimento de forma que o sistema seja capaz de identificar, nos documentos indexados, a qual assunto ele se refere e de que forma ele deve ser armazenado para posterior recuperação.

Permite a geração de gráficos quantitativos (barras) das informações contidas na base de conhecimento do sistema, considerando os filtros disponíveis na interface como fonte, data e domínio. Proporciona também a geração de gráficos temporais (de linha) que viabilizam o acompanhamento da evolução temporal de um determinado assunto pesquisado. (Ver figura 18, pág. 113).

Fonte: Todas

Domínio:

- Conceitos de Metrologia
- Grandezas e Unidades
- Resultados de Medição
- Características dos Instrumentos de Medição
- Padrões
- Medições
- instrumentos de medição

Classificar por:

Fonte  Sub-domínio  Termo

Ordem:

Quantidade  Alfabética

Período: 23/09/2005 até 23/09/2005 (dd/mm/aaaa)

**GERAR GRÁFICO**

Figura 17 – Apresentação dos Filtros para Geração dos Gráficos.

Fonte: KMAI

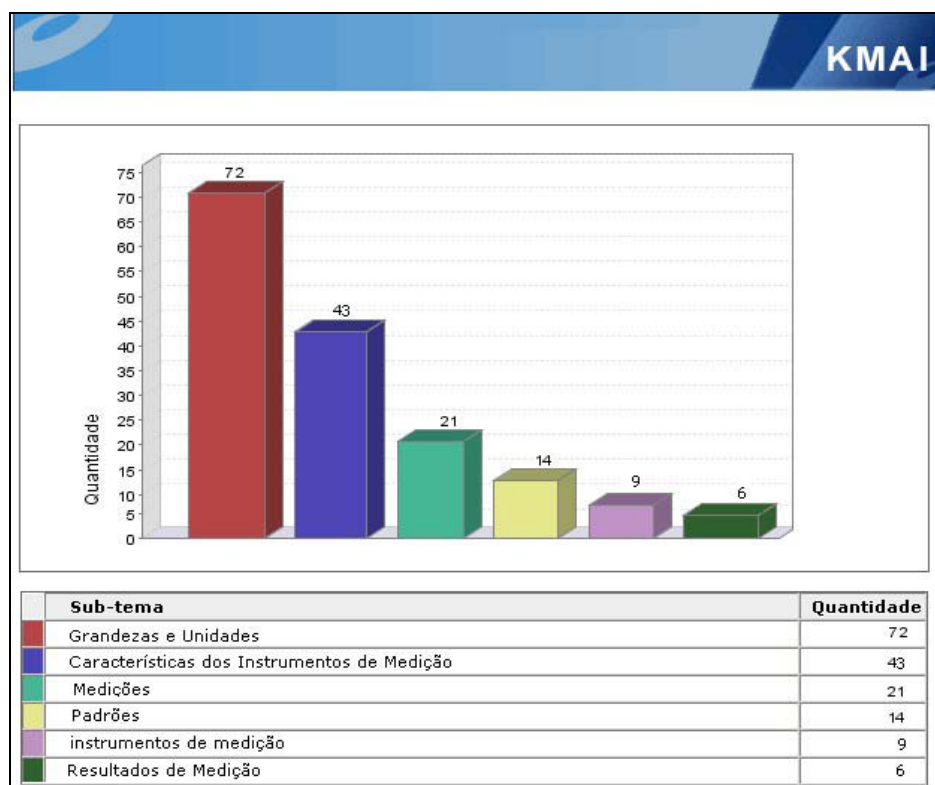


Figura 18 – Geração de Gráficos por Assunto.

Fonte: KMAI

#### 4.1.4.3 Nota Informativa

Permite ao analista agregar o seu conhecimento ao sistema, possibilitando a sua difusão entre os outros interessados, e armazena aos estudos uma memória, dados, informações ou tendências percebidas individualmente por cada um, que podem ser de grande utilidade no futuro. Essa funcionalidade também permite a inserção de arquivos, os quais farão parte da base de conhecimento.

#### 4.1.4.4 Ambiente de Administração

É o ambiente onde o administrador do sistema concederá os privilégios de cada usuário. Indicará quais funcionalidades cada colaborador terá acesso e em que nível, bem como a configuração de outras características do aplicativo parametrizadas com base no perfil do usuário (Figura 19).

The screenshot displays the 'Administração de Usuários' (User Administration) interface. At the top, there is a navigation bar with 'Página Inicial', 'Análise', 'Notas Informativas', 'Ontologias', and 'Administração'. The main content area is titled 'Usuários Cadastrados' and shows a table with 7 users. Below the table is a 'Cadastro de Usuários' (User Registration) form with fields for Name, Login, Password, User Type, Permissions, Security, and Unit.

Nome Completo	Unidade Policial	Tipo de Usuário	Ações
Raimundo Cristiano Melarica	1o DP Central de Polícia	Escrivão	View, Edit, Delete
Edmarcin Marcelin Di Calvalcanti	17o DP Hominidine	Investigador	View, Edit, Delete
Ortolânio Alcico	27o DP Furtos e Roubos	Delegado	View, Edit, Delete
Adailson Melandri	12o DP - Proteção a Criança e ao Adolescente	Operador	View, Edit, Delete
Raimunda Aparecida Pitenga	1o DP Delegacia da Mulher	Delegada	View, Edit, Delete
Elence Eunice	34o DP - Narcóticos	Operador	View, Edit, Delete
Eustáquio Medroso	1o DP - Central de Polícia	Escrivão	View, Edit, Delete

**Legenda de Ações:**  Visualizar  Editar  Excluir

**Cadastro de Usuários**

Nome Completo:

Login:

Senha:

Repita a Senha:

Tipo de Usuário:

Permissões:

- Cadastro de Pessoas Desaparecidas
- Cadastro de Veículos Roubados/Furtados
- Autoria de Exames de Procedência
- Administrador do Sistema
- Cadastro de Avisos

Segurança:

- Alterar senha no próximo login
- Bloquear Usuário
- Ativo

Unidade Policial de Matrícula:

Área reservada para as mensagens de controle desta tela

WBSA

Figura 19 – Cadastro de Usuários

Fonte: KMAI

#### 4.1.4.5 Editor de Ontologias

O Editor de Ontologias é um sistema que foi concebido para auxiliar o trabalho da equipe de especialistas na construção de Ontologias. Consiste em uma estrutura que relaciona termos complexos, considerando seus conceitos no domínio de conhecimento específico do aplicativo, permitindo que este reconheça o contexto dos documentos em análise. O aplicativo será descrito detalhadamente nas seções 4.2 e 4.3 deste trabalho.

#### 4.1.4.6 SC – INFO

O Sistema KMAI<sup>®</sup> realiza a coleta de informações das fontes selecionadas pelos especialistas. Assim, a instituição pode monitorar as fontes que melhor lhe convir.

O SC - Info (Sistema de Coleta de Informações) é o módulo responsável pela coleta de documentos diretamente das suas fontes. Para cada fonte, é criado um observador, que é configurado através do preenchimento de um arquivo XML. Depois de configurado o observador, pode-se testá-lo e, em seguida, inseri-lo no sistema. A lista de fontes, que estão sendo coletadas, é disponibilizada para que o usuário realize a sua manutenção.

Este módulo permite a construção de novas fontes pelos próprios usuários do sistema. (Figura 20).



The screenshot shows the KMAI SC - Info module interface. At the top, there is a navigation menu with options: Página Inicial, Análise, Notas Informativas, Gráficos, Ontologias, Administração, and Log Out. The main content area is titled 'Agenda de Coletas:' and contains a table with the following data:

Fonte	Próxima Coleta	Ações
Associação Brasileira de Normas Técnicas	30/09/2005 05:30:00	[Edit] [Delete]
Instituto Nacional de Metrologia	30/09/2005 05:30:00	[Edit] [Delete]
National Physical Laboratory	29/09/2005 16:34:18	[Edit] [Delete]

Below the table, there is a 'Legenda de Ações:' section with two icons: a document icon for 'Editar' and a trash can icon for 'Excluir'.

Figura 20 – Módulo SC - Info.

Fonte: KMAI

## 4.2 A CONSTRUÇÃO DE ONTOLOGIAS NO SISTEMA KMAI

A organização das informações no Sistema KMAI® é realizada através das Ontologias (Ver figura 21). As Ontologias compõem o coração do sistema, pois é baseada nas mesmas Ontologias em que são feitos todo o processamento, o armazenamento das informações coletadas e a organização da base de conhecimento, além de interferir de forma decisiva na qualidade da resposta apresentada ao usuário.

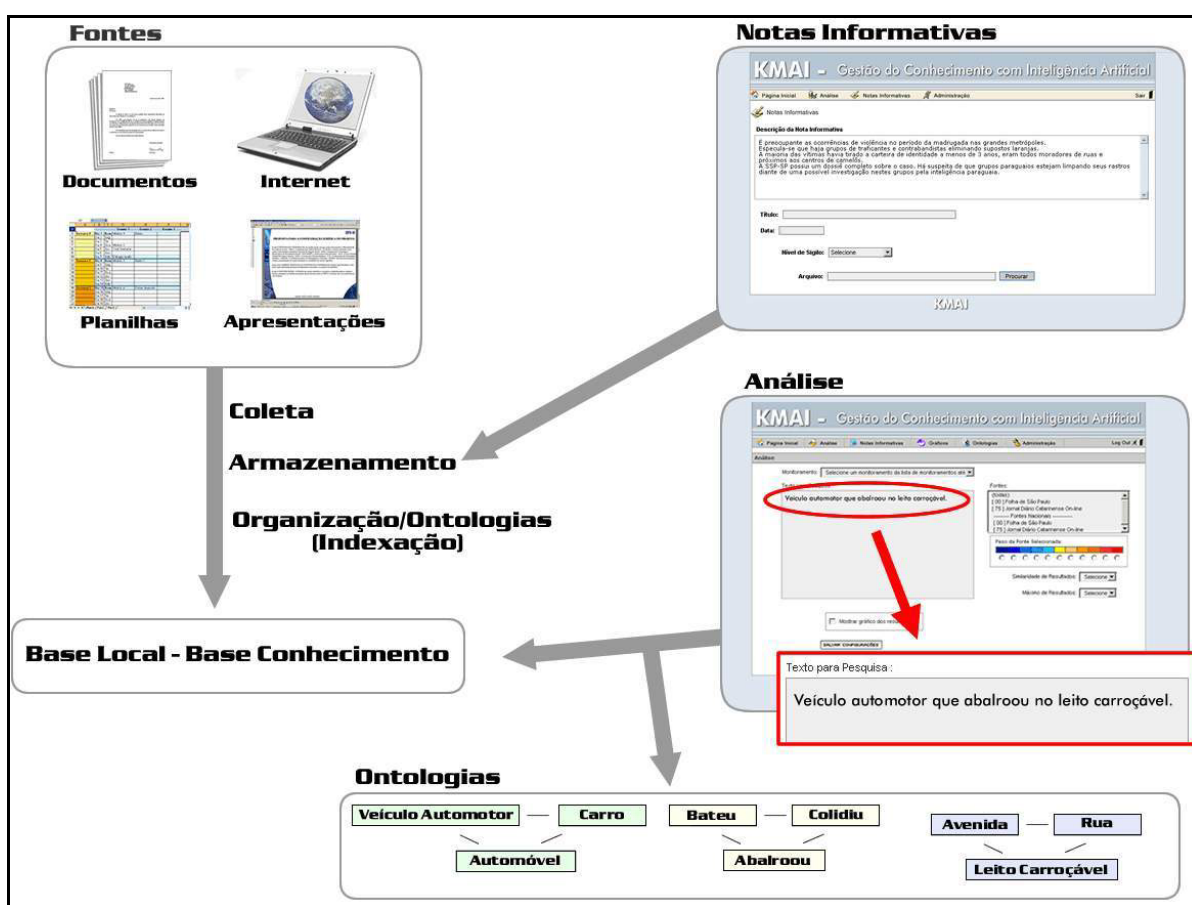


Figura 21 – Ontologias no Sistema KMAI.

Comportando sua aplicação para diferentes escopos, contempla a recuperação de informações em documentos estruturados, semi-estruturados e não estruturados. O formato dos documentos, os objetivos estratégicos do sistema e o conteúdo das fontes de informação influenciam diretamente na metodologia utilizada para a construção das Ontologias, como descrito a seguir.

A participação das Ontologias no sistema se dá, basicamente, em três momentos distintos, conforme segue:

Primeiro, na fase chamada de Coleta, o sistema extrai as informações das diversas bases previamente selecionadas pelo usuário. Cada um dos documentos é, então, indexado com base nas Ontologias construídas pelos especialistas e engenheiros do conhecimento, durante a fase de Engenharia do Conhecimento. O sistema, então, irá identificar os documentos com todas as Expressões Indicativas que forem localizadas no seu texto, armazenando-os de forma organizada na Base de Conhecimento. Assim, é possível fazer uma pré-classificação dos documentos de acordo com o que foi previsto na organização do conhecimento promovida pelas Ontologias.

Num segundo momento, as Ontologias terão fundamental importância nas interfaces de análise disponibilizadas ao usuário final do sistema. O processo inicia no instante em que o usuário digita um texto para pesquisa. Nesse momento são identificadas as Expressões Indicativas descritas pelo usuário que coincidem com aquelas previstas nas Ontologias. As Expressões Indicativas que foram identificadas vão então determinar o fluxo da cadeia de relações. O fluxo de relações se estabelece de acordo com a classificação das Ontologias em: conexos; parte de; parte disso; tipo de; e tipo disso. Isso significa que existe uma relação dinâmica entre a forma de como o usuário entra com uma Expressão Indicativa na interface de recuperação e a forma de como são definidas as relações no Editor de Ontologias para essa expressão.

A dinamicidade da utilização das Ontologias reside no fato de que quem determina a importância das Expressões Indicativas a serem consideradas é o próprio usuário, no momento da descrição do texto para pesquisa. Daí em diante, o sistema irá priorizar as Expressões Indicativas e buscar as derivações interessantes para cada caso, conforme previsto na base de conhecimento. Percebe-se que, *a priori*, não há hierarquia na organização das Ontologias na base de conhecimento. O peso das relações somente será considerado com base no que for requisitado pela pesquisa, que determinará o contexto buscado pelo usuário.

O terceiro momento, em que as Ontologias têm forte participação, é na utilização do Editor de Ontologias. O editor fica disponível no sistema, integrado a sua arquitetura, e permite a atualização constante da base de conhecimento com novas expressões não previstas anteriormente. A cada nova alteração realizada nas Ontologias, o sistema re-indexa todos os textos da base para garantir que as interfaces de análise possam trabalhar com todos os casos constantes da base de conhecimento, mesmo aqueles armazenados antes de realizada a

alteração. Isso permite a verificação da pré-existência de documentos que tratem de um assunto, mesmo que ele normalmente seja pouco monitorado, e só passe a ter importância a partir de um determinado momento. É possível, nesse caso, construir um histórico do assunto tratado, localizando seu ponto de início.

No desenvolvimento da Suíte de Engenharia do Conhecimento, foram utilizados, como referência, desenvolvimentos na área de Engenharia de Ontologia de abrangência mundial como: *WordNet*, *Semantic Web* e a *Universal Networking Language* (UNL), cujo escopo é conectar pessoas e computadores em diferentes lugares do mundo, tendo a Internet como meio e fim. Além disso, os desenvolvimentos possuem em comum, com a presente abordagem, a estruturação de uma base de conhecimento através da construção de Ontologias, com a utilização de ferramentas de apoio *on line*.

A Suíte de Engenharia do Conhecimento foi desenvolvida para ser utilizada junto com a metodologia denominada Representação do Conhecimento Contextualizada Dinamicamente (RC2D), (ver item 2.1.1.2), que consiste num processo dinâmico de análise do contexto geral que envolve uma temática enfocada. A metodologia é apoiada por ferramentas como o Extrator de Frequência, Extrator Semântico e a própria estrutura tecnológica da Suíte de Engenharia do conhecimento para auxiliar na tarefa.

A metodologia consiste nos seguintes passos:

- 1) Inventariar todo o domínio, isto é, catalogar todas as fontes de informação digital que servirão como base de dados do sistema;
- 2) Aplicar o Extrator de Frequência de Palavras em cima da base de dados inventariada;
- 3) Avaliar a informação constante na base de dados com as necessidades do especialista;
- 4) Construir, junto com o especialista, um Vocabulário Controlado representativo do domínio;
- 5) Utilizando esse vocabulário, aplicar o Extrator Semântico na base de dados;
- 6) Avaliar o resultado com base na frequência das Expressões Indicativas encontradas e definir uma lista de palavras;
- 7) Construir Ontologias para utilização no sistema com base no Vocabulário Controlado;
- 8) Utilizar a Suíte de Engenharia do Conhecimento para estabelecer relações entre as palavras e aumentar o vocabulário;

- 9) Definir os termos conexos com base no conhecimento especialista;
- 10) Definir sinônimos, homônimos e hiperônimos com base doutrinária e principalmente considerando a legislação sobre o assunto.

A criação de uma infra-estrutura na Web para construção e criação de Ontologias provou ser ferramenta eficaz e efetiva para a aquisição do conhecimento em sistemas inteligentes.

#### 4.2.1 Ontologias e Cálculo da Similaridade do Sistema KMAI

O Dicionário de Ontologias fornece uma estrutura de organização do conhecimento que permite interpretar e analisar o conteúdo dos documentos armazenados na Base de Dados. Baseado nessa estrutura teórica, o Dicionário de Ontologias fornece índices e expressões para a valoração do grau de similaridade entre o texto de entrada e o texto na base de dados.

Após cada caso da base serem valorados, eles são 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, e apresentados como o resultado da consulta. Os mais semelhantes são apresentados ao usuário, que seleciona na íntegra o mais apropriado para a melhor solução do seu problema. O processo de recuperação é dividido em:

- 1) Ajuste da Situação;
- 2) Cálculo da Similaridade;
- 3) Seleção do melhor caso.

O sistema, então, extrairá as informações relevantes automaticamente da descrição em linguagem natural. Isto é realizado pela integração de um Vocabulário Controlado, constituído pela totalidade dos termos que integram o Dicionário de Ontologias, elaborado para um domínio particularizado. Com base nessa descrição, uma descrição formal é gerada pela definição dos índices para um caso específico. A 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 Ontologias é usado para indicar a similaridade entre os termos individuais e cada relação entre as Expressões Indicativas, que têm um valor determinado, conforme tabela 2 (ver pág. 120).



Tabela 2 – Valor dos Pesos das relações no Cálculo da Similaridade

Relação da Suíte de Engenharia do Conhecimento	Pesos para o Cálculo da Similaridade
Sinônimos	0,99
Tipo de (hipernímia e hiponímia)	0,4
Parte de (meronímia e holonímia)	0,3
Termos Conexos	0,7

O próximo passo é definir quais são os casos mais úteis, dentre esses relevantes identificados, no ajuste da situação, através do cálculo da métrica da Similaridade.

$$Sim(A, B) = \frac{\left( \sum_{i=1}^n Sim1_L(Ta_i) \right) * P_d + \left( \sum_{j=1}^m Sim2_L(Pa_j) \right) * P_p}{n * P_d + m * P_p}$$

$$Sim1_L(Ta_i) = \begin{cases} 1,0 & se \ \exists Tb_j \in B \mid Ta_i = Tb_j \\ 0,99 & se \ \exists Tb_j \in B \mid Ta_i \text{ (sinônimo) } Tb_j \\ 0,7 & se \ \exists Tb_j \in B \mid Ta_i \text{ (conexo) } Tb_j \\ 0,3 & se \ \exists Tb_j \in B \mid Ta_i \text{ (parte de) } Tb_j \\ 0,3 & se \ \exists Tb_j \in B \mid Tb_j \text{ (parte de) } Ta_i \\ 0,4 & se \ \exists Tb_j \in B \mid Ta_i \text{ (tipo de) } Tb_j \\ 0,4 & se \ \exists Tb_j \in B \mid Tb_j \text{ (tipo de) } Ta_i \\ 0 & ao \ \text{contrário} \end{cases}$$

$$Sim2_L(Pa_i) = \begin{cases} 1,0 & se \ \exists Pb_j \in B \mid Pa_i = Pb_j \\ 0 & ao \ \text{contrário} \end{cases}$$

Observações:

**Sim1** corresponde à similaridade entre os termos da consulta e os termos dos documentos obtidos através do dicionário.

**Sim2** corresponde à similaridade por palavra-chave.

A métrica de similaridade leva em consideração tanto o termo indexado pelo dicionário quanto os demais termos que não estão no dicionário. Pode-se determinar uma relação de importância entre os dois conjuntos de termos através dos pesos  $P_d$  e  $P_p$ .

É necessário determinar o peso para as relações ‘parte de’ e ‘tipo de’.

Exemplo:

Dicionário = {"reforma da previdência"}

Texto Consulta = {"O presidente Lula enviou a reforma da previdência"}

Termos indexados: consulta = {"reforma da previdência"}

Palavras-chave: consulta = {"presidente", "Lula", "enviar"}

Termos indexados documento 1 = {"reforma da previdência"}

Palavras-chaves documento 1 = {"presidente", "congresso"}

Termos indexados documento 2 = { }

Palavras-chave documento 2 = {"Lula"}

Supondo  $P_d = 2$  e  $P_p = 1$ :

$$Sim(C, D1) = \frac{(1) * 2 + (1) * 1}{1 * 2 + 3 * 1} = \frac{2 + 1}{5} = \frac{3}{5} = 0,6 = 60\%$$

$$Sim(C, D2) = \frac{(0) * 2 + (1) * 1}{1 * 2 + 3 * 1} = \frac{0 + 1}{5} = \frac{1}{5} = 0,2 = 20\%$$

Supondo  $P_d = 1$  e  $P_p = 1$ :

$$Sim(C, D1) = \frac{(1) * 1 + (1) * 1}{1 * 1 + 3 * 1} = \frac{1 + 1}{4} = \frac{2}{4} = 0,5 = 50\%$$

$$Sim(C, D2) = \frac{(0) * 1 + (1) * 1}{1 * 1 + 3 * 1} = \frac{0 + 1}{4} = \frac{1}{4} = 0,25 = 25\%$$

Supondo  $P_d = 3$  e  $P_p = 1$ :

$$Sim(C, D1) = \frac{(1) * 3 + (1) * 1}{1 * 3 + 3 * 1} = \frac{3 + 1}{6} = \frac{4}{6} = 0,66 = 66\%$$

$$Sim(C, D2) = \frac{(0) * 3 + (1) * 1}{1 * 3 + 3 * 1} = \frac{0 + 1}{6} = \frac{1}{6} = 0,1667 = 16,67\%$$

Supondo  $P_d = 1$  e  $P_p = 2$ :

$$Sim(C, D1) = \frac{(1) * 1 + (1) * 2}{1 * 1 + 3 * 2} = \frac{1 + 2}{7} = \frac{3}{7} = 0,4286 = 42,86\%$$

$$Sim(C, D2) = \frac{(0) * 1 + (1) * 2}{1 * 1 + 3 * 2} = \frac{0 + 2}{7} = \frac{2}{7} = 0,2857 = 28,57\%$$

Assim, através desse resultado, pôde-se verificar que as Ontologias permitem similaridade maior entre a questão de entrada e o texto da base de dados.

Por exemplo, se num texto de aproximadamente trinta (30) linhas, existem somente duas (2) Expressões Indicativas constantes do texto de entrada (questão), mas, no Dicionário, essas palavras possuem uma infinidade de relações que também constam do texto, as outras relações também participarão do cálculo da similaridade. Se, ao contrário, nesse mesmo texto de trinta (30) linhas, somente duas (2) palavras-chave são encontradas, somente essas participarão do cálculo da similaridade, resultando num valor de similaridade inferior àquele elaborado através das Ontologias.

#### 4.3 SUÍTE DE ENGENHARIA DO CONHECIMENTO

A Suíte de Engenharia do Conhecimento é uma ferramenta de Engenharia de Ontologias. Ela é uma estrutura computacional independente para o desenvolvimento, criação e edição de Ontologias e bases de conhecimento para ser utilizada na Web. É uma ferramenta concebida para auxiliar o trabalho da equipe de engenheiros do conhecimento e especialistas

na construção dos Dicionários de Ontologias. Consiste em uma estrutura que relaciona termos complexos, considerando seus conceitos no domínio de conhecimento específico do aplicativo, permitindo que este reconheça o contexto dos documentos em análise, isto é, a sua aplicação é contextualizada.

A Suíte de Engenharia do Conhecimento é um módulo do Sistema KMAI® (ver item 4.1.4). O módulo apresenta características configuráveis do sistema que permite a atualização constante de sua inteligência, preservando a capacidade de compreensão evolutiva da linguagem. Essa ferramenta torna a aplicação capaz de entender a relação existente em contextos específicos, como por exemplo: faca com arma branca; fibras artificiais como fibras sintéticas; podendo também se adaptar a novas expressões denominadas indicativas. Isso se deve à construção das Ontologias.

O desenvolvimento de Ontologias procura refletir as relações conceituais baseadas no contexto do domínio trabalhado e tem por objetivo facilitar o acesso de usuários não conhecedores da linguagem técnica, estabelecendo uma rede constituída por conceitos unidos por diferentes relações semânticas.

Existe uma variedade de usos possíveis para as bases de conhecimento e Ontologias dos dados disponíveis em um domínio de aplicação. Com base nos dados fornecidos, vários sistemas poderão ser construídos, tais como: sistemas de recuperação de informação, sistemas de gestão do conhecimento e sistemas estatísticos.

A Suíte de Engenharia do Conhecimento tem as seguintes características:

- A edição de Ontologias é orientada ao usuário (amigável);
- A linguagem de criação de Ontologias é a natural;
- Permite o compartilhamento de Ontologias por classe de usuário;
- Sistema de busca de palavras com Inteligência Artificial;
- Relatório de atividades por usuário;
- Ambiente de administração para aprovação, exclusão e inclusão de Ontologias;
- Ferramenta compatível com linguagens abertas.

#### 4.3.1 Módulos da Suíte de Engenharia do Conhecimento

A Suíte é um Editor de Ontologias, estruturada de forma a permitir um trabalho cooperativo na Web entre a equipe de engenharia de conhecimento e a equipe de especialistas. Ela é composta basicamente de quatro módulos:

- 1) Cadastro de novas ontologias;
- 2) Sistema de busca por similaridade;
- 3) Editor de relações;
- 4) Ambiente de Administração. (Ver quadro 6).

<b>MÓDULO</b>	<b>DESCRIÇÃO</b>	<b>IMPORTÂNCIA</b>
<b>Cadastro de domínios (temas e subtemas)</b>	Neste ambiente, o usuário define o tema e subtema no qual ele irá inserir uma nova expressão indicativa. Um domínio pode ser categorizado em inúmeros temas e subtemas.	Permite a elaboração de um dicionário contextualizado, pela seleção de temas e subtemas para a classificação das expressões indicativas.
<b>Busca</b>	É um sistema de busca baseado na similaridade. Fornece uma lista de expressões indicativas semelhantes presentes na base de conhecimento em ordem alfabética após consulta feita pelo usuário. Ele é utilizado no cadastro, na edição e no módulo administração das Ontologias.	Informa a respeito de outros termos já cadastrados na base que tenham alguma semelhança fonética com o termo digitado. Essa ferramenta permite a verificação de possíveis erros de digitação, além de impedir o cadastramento de um mesmo termo mais de uma vez.
<b>Editor</b>	São apresentados os campos com todas as relações possíveis de serem formadas, são elas: 1) Sinônimos; 2) Conexos; 3) Isso é tipo de; 4) É um tipo disso; 5) Isso é parte de; 6) É parte disso. O Editor apresenta as relações já cadastradas e permite excluir relações. Cada relação possui um peso em relação à expressão indicativa definida na busca pelo usuário.	Através desse cadastro vai se formando a árvore de relacionamentos, sempre considerando a semelhança de todos os termos cadastrados com os já existentes na base. Estas relações permitem ao sistema expandir o contexto da busca. A organização da árvore permite a definição dinâmica dos pesos das expressões indicativas de acordo com a entrada do usuário.
<b>Administração</b>	Nível alto: permite inserir temas e subtemas, validar exclusões, incluir e excluir usuários, verificar produtividade de cada usuário e verificar históricos dos dicionários, temas, subtemas e expressões indicativa. Nível médio: permite verificar produtividade e históricos. Nível baixo: permite verificar históricos.	Integração do conhecimento e validação, relação entre as palavras de acordo com o contexto definido entre o escopo do dicionário, os temas e subtemas.

Quadro 6 – Módulos da Suíte de Engenharia do Conhecimento.

A metodologia de construção de Ontologias da Suíte de Engenharia do Conhecimento trabalha com extratores automáticos de padrões (de Frequência e Semântico) em colaboração com engenheiros do conhecimento e especialistas do domínio abordado, conforme especificações encontradas na metodologia Representação do Conhecimento Contextualizado Dinamicamente (RC2D).

#### 4.3.1.1 Módulo de Criação de Novas Ontologias

O termo Ontologia, no campo da tecnologia, faz referência à formulação de esquema conceitual, dentro de um certo domínio, com a finalidade de facilitar a comparação, a classificação, a organização e o armazenamento dos textos analisados. Assim, as Ontologias permitem que o sistema computacional realize raciocínios indutivos, semelhantes ao do especialista, quando analisa e armazena os documentos. Como a representação do conhecimento é um dos principais procedimentos da Inteligência Artificial, as Ontologias aparecem como um eficiente meio de efetivar tal representação.

O Dicionário de Ontologias procura refletir não apenas a memória léxica humana, mas também estabelecer relações conceituais baseadas no contexto do domínio trabalhado, estabelecendo uma rede constituída por conceitos unidos por diferentes relações semânticas. (Ver figura 22, pág. 126).

Os conceitos descritos a seguir foram definidos a partir do estudo de outras teorias e metodologias de aquisição e estruturação de conhecimento em outros domínios de aplicação que também utilizam as Ontologias para representar o conhecimento.

- 1) Relação **Sinônimos**: É a relação existente entre expressões com o mesmo significado, dentro de um mesmo domínio, ou seja, as expressões de uma mesma relação podem ser substituídas sem alterar a compreensão do texto. Ex.: tráfico de drogas é sinônimo de narcotráfico.
- 2) Relação **Tipo de (hipernímia e hiponímia)**: É a conexão existente entre expressões da qual se depreende uma relação de categoria e classe, ou gênero e espécie. É a relação existente entre um lexema mais específico ou subordinado e um lexema mais geral ou superordenado. Ex.: cocaína é um tipo de substância entorpecente.

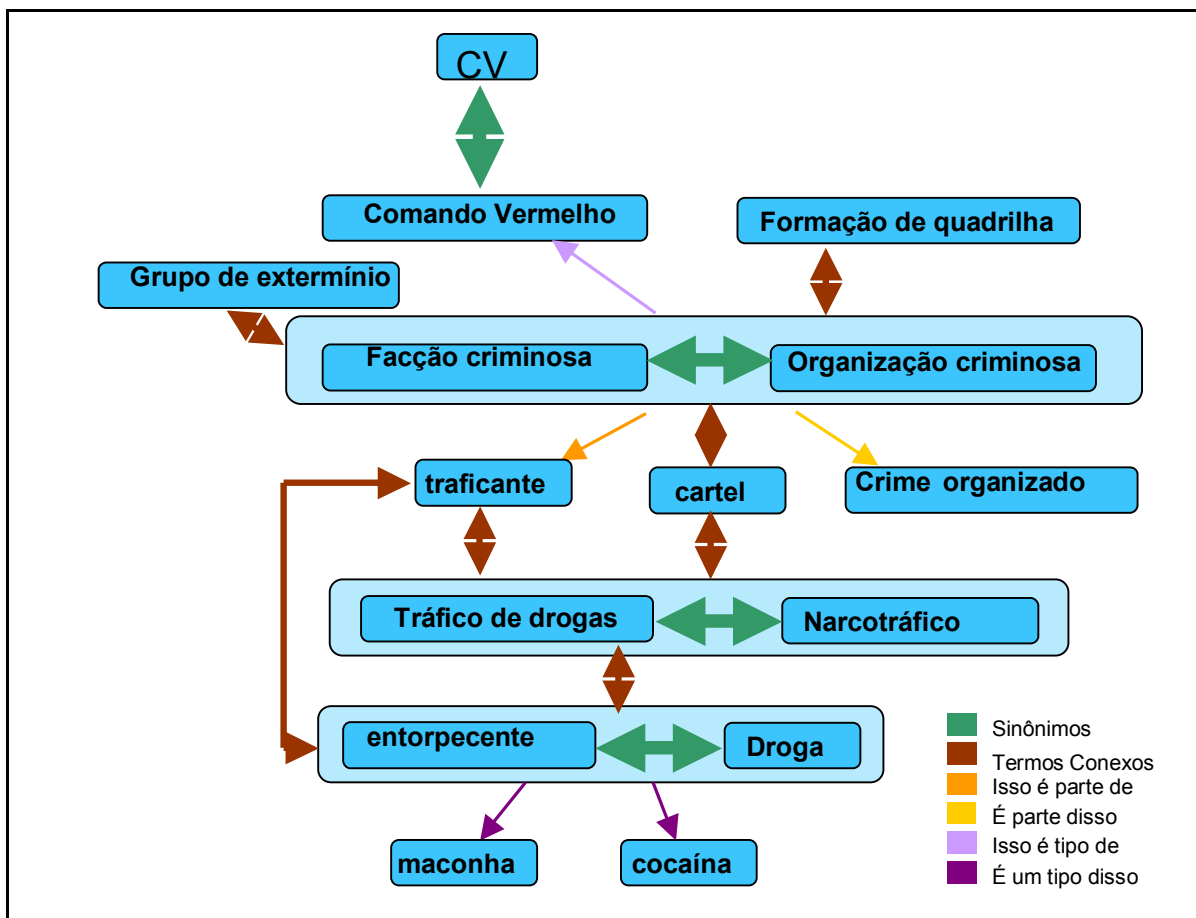


Figura 22 – Exemplo de uma Ontologia e suas relações na Suíte de Engenharia do Conhecimento.

- 3) Relação **Parte de (meronímia e holonímia)**: É aquela relação que determina uma idéia de fração e o todo, sendo mais comumente encontrada na estrutura de organizações (ex.: Secretaria do Ministério do Planejamento é parte de Ministério do Planejamento) ou nos casos em que a descrição das partes é tão relevante quanto à previsão do todo.
- 4) Relação **Termos Conexos**: Representa a conexão existente entre termos fortemente relacionados que não se enquadram em nenhum outro tipo de relação. Ex.: trânsito na fronteira é conexo de controle migratório.

A interface apresentada na figura 23 (ver pág. 127) é a que permite a construção efetiva das relações de sinônimo, tipo, parte e conexo. No canto superior esquerdo, é apresentado um histórico do cadastro, com a seção em que se encontra a Suíte (Inserir), o Tema e Subtema e o termo em evidência para cadastro.

Figura 23 – Módulo de Cadastro de Novas Ontologias.

Fonte: BRASIL (2006)

Assim, como em todas as outras interfaces da Suíte, a expressão ‘isso’ ou ‘disso’ no tipo de relação, ‘Isso é, Tipo de’, sempre diz respeito ao termo em evidência; neste caso, ‘remessa de dinheiro’. Pode-se perceber que as relações de ‘tipo’ e ‘parte’ se repetem, pois podem ser feitas tanto do macro para o micro como do micro para o macro.

Para efetivar o cadastro das relações, basta digitar os termos que se pretende relacionar com o termo em evidência, no campo da relação desejada, apresentado ao lado esquerdo da interface e, ao final, clicar no botão “Relacionar”, que aparece abaixo. No caso de não haver necessidade de relacionar o termo em evidência com qualquer outro termo, basta simplesmente clicar no botão “Relacionar”, que o sistema cadastra o termo isoladamente.

No lado direito da interface, o sistema apresenta as relações referentes ao termo já cadastrado no Banco de Dados, inserido pelo usuário, dentro do quadro “Relações existentes no Dicionário”. Como o mesmo termo pode ser cadastrado em mais de um Tema ou Subtema, também é apresentado um histórico, mostrando onde esse termo está cadastrado e as relações



que possui. No canto inferior esquerdo da interface, o botão “Remover Seleccionadas” permite que os termos selecionados, dentro do quadro, sejam removidos, excluindo-se, assim, a relação pré-existente. Esse procedimento ainda precisa ser aprovado pelo administrador do sistema para produzir efeitos.

O módulo de inclusão permite a elaboração de um dicionário contextualizado, em que o usuário define o Tema e Subtema no qual será inserido um novo termo. (Ver figura 24). A escolha do Tema e Subtema fazem parte da Engenharia do Conhecimento do Sistema Baseado no Conhecimento. Neste exemplo, os Temas e Subtemas foram definidos para aplicação num sistema de Gestão do Conhecimento para o Monitoramento de Crises do Governo Brasileiro.

The image shows a web application interface for 'Gabinete de Segurança Institucional' (GSI) under the 'Gestão Saei' system. The main navigation bar includes links for 'Página Inicial', 'Análise', 'Notas Informativas', 'Gráficos', 'Ontologias', 'Administração', and 'Log Out'. The current page is titled 'Dicionário de Ontologias - Inserir' and is for 'Inclusão de novas relações'. The form contains three main sections: 'Tema' with a dropdown menu set to 'Crime Organizado', 'Subtema' with a dropdown menu set to 'Lavagem de dinheiro', and 'Digite a Palavra' with a text input field containing 'remessa de dinheiro' and a 'Verificar' button. At the bottom, there is a logo and the text 'SAEI / GSI / PR'.

Figura 24 – Escolha de Tema e Subtema para inserção

Fonte: BRASIL (2006)

#### 4.3.1.2 Módulo de Exclusão de Ontologias

Esta seção permite a exclusão de termos ou a exclusão de determinadas relações do termo em evidência dentro do Subtema.

Na figura 25 (ver pág. 129), verifica-se a possibilidade de exclusão de um termo do dicionário como um todo, ou seja, ao clicar no botão “Excluir”, todas as relações de termo digitado, em todos os Subtemas em que ele se encontra, serão excluídas. Esse termo passará a não mais produzir efeitos no sistema.

Figura 25 – Exclusão de termo do Dicionário

Fonte: BRASIL (2006)

Também é possível excluir todas as relações de um termo dentro de um mesmo Tema. Nesse caso, o termo digitado será excluído de todos os Subtemas do Tema selecionado que o apresentarem, conforme figura 26.

Figura 26 – Exclusão de termo do Tema

Fonte: BRASIL (2006)

A última opção é excluir as relações de um termo somente num Subtema específico, conforme mostra a parte superior da figura 27. (Ver pág. 130). No caso de excluir apenas alguma(s) relação(ões) específica(s) do termo em evidência, o procedimento concentra-se na parte inferior da interface, não havendo a necessidade de digitar o termo na caixa superior.

A segunda grande caixa da interface apresenta todas as relações de todos os termos do Subtema selecionado, com a data de inserção da relação, o usuário que a criou, o tipo de

relação e uma caixa de seleção ao final da linha. Essa última serve para que seja selecionada a relação que se deseja excluir, podendo ser selecionada mais de uma relação por vez.

Se a intenção for excluir uma relação de sinônimos, é preciso que seja marcado qual dos termos está saindo da relação, ou seja, qual deles não fará mais parte do grupo de sinônimos existente. Nesse caso, o sinônimo excluído perde a relação com os demais sinônimos, mas mantém todas as relações de outro tipo (conexo, parte e tipo), permanecendo no Subtema, porém num grupo separado de relações.

Gabinete de Segurança Institucional  
Gestão Saei

Página Inicial | Análise | Notas Informativas | Gráficos | Ontologias | Administração | Log Out

**Dicionário de Ontologias - Excluir**

Exclusão de relações

**Tema**  
Crime Organizado

**Subtema**  
Lavagem de dinheiro

**Digite a Palavra**  
transferência de fundo [Excluir]

**Excluir:**  do Dicionário  do Tema  do SubTema

Palavra 1	Palavra 2	Data	Usuário	Relação
<input checked="" type="radio"/> remessa ilegal de dinheiro	<input type="radio"/> remessa ilegal de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input type="radio"/> remessa ilegal de dinheiro	<input checked="" type="radio"/> transferência de fundo	02/04/2004	valter	Sinonimo <input checked="" type="checkbox"/>
<input checked="" type="radio"/> remessa ilegal de dinheiro	<input type="radio"/> evasão de recurso financeiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input checked="" type="radio"/> repressão à lavagem de dinheiro	<input type="radio"/> repressão à lavagem de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input checked="" type="radio"/> repressão à lavagem de dinheiro	<input type="radio"/> combate à reciclagem de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input checked="" type="radio"/> repressão à lavagem de dinheiro	<input type="radio"/> repressão à reciclagem de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input checked="" type="radio"/> repressão à lavagem de dinheiro	<input type="radio"/> erradicação da reciclagem de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>
<input checked="" type="radio"/> repressão de dinheiro	<input type="radio"/> remessa ilegal de dinheiro	02/04/2004	valter	Sinonimo <input type="checkbox"/>

<< Voltar Excluir

SAEI / GSI / PR

Figura 27 – Exclusão de termo do Subtema.

Fonte: BRASIL (2006)

#### 4.3.1.3 Módulo de Busca e Edição

Esta seção agrega duas funções principais: busca e edição de termos. Por meio da busca, é possível localizar na Base de Dados termos foneticamente semelhantes ao digitado,

indicando a sua colocação no Dicionário. Assim, pode ser evitado o re-cadastro de um mesmo termo e possíveis erros de digitação, além de serem apresentadas às relações que o termo selecionado possui.

Após o termo ser pesquisado, o sistema apresenta uma lista com informações sobre as relações do termo pesquisado. (Ver figura 28). A primeira coluna traz o termo que se relaciona com o pesquisado; a segunda, o tipo de relação; a terceira, o Subtema no qual encontra-se aquela relação; a quarta, a data de criação da relação e, por último, um botão “Editar” em cada linha. Na coluna “Relação”, ressalta-se que as palavras, ‘isso’ e ‘disso,’ referem-se sempre ao termo pesquisado.



The screenshot shows the SAEI system interface. At the top, there is a navigation bar with links for 'Página Inicial', 'Análise', 'Notas Informativas', 'Gráficos', 'Ontologias', 'Administração', and 'Log Out'. Below this is a search bar with the text 'Dicionário de Ontologias - Buscar / Editar' and a search input field containing 'transgênico'. The main content area displays a table with the following data:

Palavra/Termo	Relação	Subtema	Data Criação	
Biodiversidade e Transgênicos	Isso é Parte de	Biodiversidade e Transgênicos	02/04/2004	Editar
material transgênico	Conexo	Biodiversidade e Transgênicos	02/04/2004	Editar
indivíduo geneticamente modificado	Conexo	Biodiversidade e Transgênicos	02/04/2004	Editar
biotecnologia	Isso é Parte de	Biodiversidade e Transgênicos	02/04/2004	Editar
manipulação genética	Isso é Tipo de	Biodiversidade e Transgênicos	02/04/2004	Editar
soja transgênica	Tipo disso	Biodiversidade e Transgênicos	02/04/2004	Editar
OGM	Sinonimo	Biodiversidade e Transgênicos	02/04/2004	Editar
planta bioreatora	Tipo disso	Biodiversidade e Transgênicos	02/04/2004	Editar
organismo geneticamente modificado	Sinonimo	Biodiversidade e Transgênicos	23/06/2004	Editar
...	..	Rindiversidade e	..	..

Below the table, there is a red message: '\* Relações em vermelho já foram alteradas e estão aguardando aprovação.' At the bottom right, there is a button labeled '<< Voltar'.

Figura 28 – Visualização das relações e possibilidade de edição.

Fonte: BRASIL (2006)

O botão “Editar”, ao final de cada linha, permite que sejam alterados a grafia dos termos ou o tipo de relação destes com o termo pesquisado. (Ver figura 28). Depois de efetivada a alteração, a linha correspondente ao termo que foi alterado aparecerá destacada em vermelho, conforme lembrete no final da página, e aguardará a aprovação do administrador do sistema para produzir efeitos. A última linha apresenta o botão “Voltar”, que, ativado, retorna à página anterior.

No caso de alteração do tipo de relação entre sinônimos, deverá ser escolhido qual deles deixará de ser sinônimo e se haverá outro tipo de relação. Nesse caso, o termo que deixar de ser sinônimo perderá também todas as suas outras relações de tipo, parte e conexo. (Ver figura 29).

Gabinete de Segurança Institucional

Gestão Saei

Página Inicial Análise Notas Informativas Gráficos Ontologias Administração Log Out

Dicionário de Ontologias - Buscar / Editar

Pesquisa e Alteração de relações

Termo 1 : OGM

Termo 2 : transgênico

Relação:  Sinonimo  Isso é Tipo de  Tipo disso  
 Conexo  Isso é Parte de  Parte disso

<< Voltar

SAEI / GSI / PR

Figura 29 – Interface de alteração de relação de sinônimos.

Fonte: BRASIL (2006)

#### 4.3.1.4 Módulo Relatório

Para visualizar a listagem de todos os termos cadastrados em determinado Subtema e suas respectivas relações, deve ser utilizada a seção “Relatório”. Aqui, todos os termos podem ser vistos em evidência, o que gera uma proposital repetição das relações, modificando-se apenas o ponto de partida.

Para visualizar um relatório, deve-se acessar a seção específica da Suíte, selecionar um Tema e um Subtema, e clicar no botão “Verificar”. (Ver figura 30, pág. 133).

Gabinete de Segurança Institucional

Gestão Saei

Página Inicial Análise Notas Informativas Gráficos Ontologias Administração Log Out

Dicionário de Ontologias - Relatório

Visualização das relações

Tema  
Meio Ambiente

Subtema  
Biodiversidade e Transgênicos Verificar

bioprospecção

- fim econômico da biodiversidade - Sinônimo
- bioprospecção - Sinônimo
- bionegócio - Sinônimo
- negócio sustentável - Conexo
- Biodiversidade e Transgênicos - Isso é parte de

bionegócio

- fim econômico da biodiversidade - Sinônimo
- bioprospecção - Sinônimo
- bionegócio - Sinônimo
- negócio sustentável - Conexo
- Biodiversidade e Transgênicos - Isso é parte de

OGM

- OGM - Sinônimo
- transgênico - Sinônimo
- organismo geneticamente modificado - Sinônimo
- patrimônio genético - É parte disso
- indivíduo geneticamente modificado - Conexo
- material transgênico - Conexo
- algodão transgênico - É um tipo disso
- milho transgênico - É um tipo disso
- soja transgênica - É um tipo disso
- biodiversidade - Isso é um tipo de
- planta bioreatora - É um tipo disso
- Biodiversidade e Transgênicos - Isso é parte de

pirataria biológica

- biopirataria - Sinônimo
- pirataria biológica - Sinônimo
- tráfico de espécie - É um tipo disso
- comércio ilícito da fauna - É um tipo disso
- comércio legal da fauna - É um tipo disso
- mercado paralelo da fauna - É um tipo disso
- contrabando da fauna - É um tipo disso
- tráfico de planta - É um tipo disso
- comércio ilegal da flora - É um tipo disso
- mercado paralelo da flora - É um tipo disso
- mercado negro da flora - É um tipo disso
- contrabando da flora - É um tipo disso
- tráfico da flora - É um tipo disso
- Biodiversidade e Transgênicos - Isso é parte de

Figura 30 – Relatório dos Subtemas.

Fonte: BRASIL 2006

#### 4.3.1.5 Módulo de Administração

A administração do sistema é parte extremamente relevante no desempenho da Suíte de Engenharia do Conhecimento. As seções descritas a seguir apenas estão disponíveis para o administrador do sistema, que é o responsável pela aprovação, ou não, dos procedimentos efetuados pela equipe de analistas.

As alterações sugeridas pelos especialistas passam a produzir efeito, e essa decisão é irrevogável. Uma vez aprovada, não há como cancelar a aprovação posteriormente, seja ela na inclusão, edição ou exclusão de Ontologias. As atividades de aprovação são: Confirmar Novas Relações; Confirmar Alterações; Confirmar Exclusão.

O ambiente de administração, além da aprovação das alterações efetuadas na Suíte, é responsável pela criação dos Temas e Subtemas que definem o contexto, no qual as Ontologias serão trabalhadas pelos especialistas ou engenheiros do conhecimento.

Esta seção é a que permite a inclusão, exclusão ou alteração de temas na Suíte de Engenharia do Conhecimento. A interface inicial apresenta uma lista com todos os Temas já existentes. (Ver figura 31).

Gabinete de Segurança Institucional

Gestão Saei

Página Inicial | Análise | Notas Informativas | Gráficos | Ontologias | Administração | Log Out

**Dicionário de Ontologias - Manter Tema**

Visualização das inclusão de Temas

Tema	Editar	Excluir
Governabilidade/Estabilidade Institucional	Editar	Excluir
Conceito Estratégico Nacional	Editar	Excluir
Políticas Públicas	Editar	Excluir
Operações de Paz	Editar	Excluir
Assuntos Internacionais	Editar	Excluir
Defesa e Integração Nacional	Editar	Excluir
Inteligência	Editar	Excluir
Lista de Órgãos	Editar	Excluir

**Cadastro de Temas**

Nome do Tema:

Nome do Subtema:

Salvar

SAEI / GSI / PR

Figura 31 – Cadastro de Novo Tema.

Fonte: BRASIL (2006)

Para criar um novo Tema, é preciso digitar seu título na caixa disponível na metade inferior da tela, ao lado de “Nome do Tema”, e também digitar o título do Subtema, na caixa ao lado de “Nome do Subtema”. Não é possível criar um Tema sem pelo menos um Subtema, podendo este último ser a repetição do primeiro, sendo isso um pré-requisito do sistema.

O Dicionário considera o nome do Tema e Subtema como um termo de sua base. Por esse motivo, verifica sua semelhança fonética com outros termos já cadastrados no sistema. Ao ser efetivado o cadastro, o nome do novo Tema já aparece na lista apresentada na interface.

Dentro de um Tema já existente, ou de um recém criado, é possível incluir novos Subtemas. O procedimento é basicamente o mesmo da inclusão de Temas. A diferença é que, antes de cadastrar o nome do novo Subtema na caixa ao lado de “Nome do Subtema”, encontrada na parte inferior da interface, é preciso selecionar, na caixa “Nome do Tema”, o Tema no qual se deseja incluir o novo Subtema. (Ver figura 32).

The screenshot shows the 'Gabinete de Segurança Institucional' interface for 'Gestão Saei'. The main menu includes 'Página Inicial', 'Análise', 'Notas Informativas', 'Gráficos', 'Ontologias', 'Administração', and 'Log Out'. The current page is 'Dicionário de Ontologias - Manter Subtema', showing 'Visualização das inclusão de Subtemas'. A table lists a subtopic: 'Novo subtema do Novo TEMA' with 'Editar' and 'Excluir' actions. Below is the 'Cadastro de Subtemas' form, where 'Nome do Tema' is 'Novíssimo TEMA' and 'Nome do Subtema' is 'Mais novo do Novíssimo TEMA'. A 'Salvar' button is present at the bottom.

Figura 32 – Cadastro de novo Subtema.

Fonte: BRASIL (2006)

Assim, como acontece na seção descrita anteriormente, o sistema localiza termos foneticamente semelhantes ao nome do Subtema sugerido, para evitar expressões repetidas na base. Se não houver mais a necessidade de manter algum Subtema na base, o sistema também permite a exclusão de Subtemas. Nesse caso, todos os termos cadastrados dentro desse Subtema deixarão de existir, caso não estejam cadastrados em outro Subtema. Como esta já é uma tarefa do administrador, uma vez excluído, não há como recuperar o Subtema posteriormente.



### 4.3.2 Avaliação da Suíte de Engenharia do Conhecimento

A estrutura de avaliação descrita neste tópico serve para avaliar as diferentes ferramentas de construção de Ontologia. Ela foi proposta por Duineveld et al (1999)<sup>105</sup> e chama-se: *WONDERTOOLS (Web-based ONtology DEscriptions and Research of its TOOLS)*<sup>106</sup>. Os autores especificaram em sua pesquisa um número de critérios relevantes (ver lista abaixo, item 2). Esta pesquisa utilizou inicialmente uma estrutura de avaliação de sistemas de *software* proposta por Kabel (1997)<sup>107</sup>, e acrescentou critérios específicos para aplicação em ferramentas de Engenharia de Ontologias.

Neste trabalho foram avaliadas as seguintes ferramentas: *Ontolingua*<sup>108</sup>, *WebOnto*<sup>109</sup>, *ProtégéWin*<sup>110</sup>, *OntoSaurus*<sup>111</sup> e ODE.

As ferramentas foram avaliadas em três dimensões.

- 1) **Dimensão geral:** Refere-se aos aspectos das ferramentas que também podem ser encontradas em outros tipos de programas. Atribuí-se esta dimensão à informação da relação usuário-interface e as diferentes ações que o usuário pode executar. Para o exemplo, a primeira pergunta do item 1: Geral: avalia a clareza da interface.
- 2) **Segunda dimensão:** A dimensão das Ontologias refere-se às questões relacionadas às Ontologias encontradas nas ferramentas, tais como: a quantidade de ajuda na construção de Ontologias e o ‘nível’ de premissas fornecido. Para o exemplo, pode-se encontrar informação sobre o que são as Ontologias e quando é útil fazer uma? Há Ontologias de exemplo?
- 3) **Última dimensão:** É aquela da cooperação, que é usada para avaliar a capacidade de suporte da ferramenta para construir uma Ontologia por diversas pessoas em locais diferentes. Algumas ferramentas fornecem uma funcionalidade extra para suportar

<sup>105</sup>Duineveld, A. J. et al, 1999. WonderTools? A comparative study of ontological engineering tools. *Twelfth Workshop on Knowledge Acquisition, Modeling and Management*. Voyager Inn, Banff, Alberta, Canada.

<sup>106</sup> WonderTools. Disponível em: <http://www.swi.psy.uva.nl/wondertools/>. Acesso em: 20 nov. 2004.

<sup>107</sup> Kabel, S.C.(1997). “System evaluation framework”. Master’s thesis, University of Amsterdam.

<sup>108</sup> Ontolingua. Disponível em: <http://ontolingua.nici.kun.nl>. Acesso em: 29 de novembro de 2004. Disponível em: <http://www-ksl-svc.stanford.edu>. Acesso em: 28 de novembro de 2004.

<sup>109</sup> WebOnto. Disponível em: <http://webonto.open.ac.uk/>. Acesso em: 17 nov. 2004.

<sup>110</sup> ProtégéWin. Disponível em: <http://smi-web.stanford.edu/projects/prot-nt/>. Acesso em: 18 de nov. 2004.

<sup>111</sup> OntoSaurus: Disponível em: <http://sevak.isi.edu:8300/loom/shuttle.html> (username and password required).

(auxiliar) essa dimensão, enquanto, em outras, somente é possível salvar a Ontologia em um determinado formato e exportar esse formato usando o e-mail ou o FTP.

Para avaliar as ferramentas nessas dimensões, empregou-se uma lista de verificação, a qual fornece pontos para se ter em mente, quando as ferramentas são usadas para construir uma Ontologia. A lista de verificação consiste em um número de perguntas para cada dimensão. A maioria das perguntas pode ser respondida dando uma pontuação para um determinado aspecto da ferramenta. As pontuações usadas são: de 1 a 10, sendo que 1 significa extremamente ruins e 10 representa os que são excelentes.

Entretanto, o maior interesse na avaliação está na explicação do por quê que uma determinada pontuação foi fornecida, do que na pontuação *per se* e, conseqüentemente, não foi dado demasiada significação para pontuações específicas. Ao invés disso, usou-se bom (+), razoável (0) e ruim (-). Outras questões perguntam apenas se alguma opção está disponível, que pode ser respondida com sim ou não.

Usando essas avaliações, foi possível comparar determinados aspectos de diferentes ferramentas. Na metodologia, foram avaliados os seguintes itens:

1) Geral:

- Avaliação da clareza da Interface (acessibilidade);
- Avaliação da consistência da Interface;
- Avaliação da velocidade de atualização, depois que novos dados são introduzidos;
- Há uma boa visão geral da Ontologia?
- O significado dos comandos é claro?
- As mudanças são identificáveis por um determinado comando claro ao usuário?
- Avaliação da estabilidade da ferramenta;
- A ferramenta requer uma instalação local?
- Avaliação do sistema de ajuda.

2) Ontologia:

- É possível usar heranças múltiplas?
- É possível criar decomposições exaustivas e/ou desarticuladas? (+ fácil fazer isto)

- A ferramenta faz a verificação dos dados novos para a sua consistência com a Ontologia? Em que nível?
- Há exemplos de uma Ontologia disponível na ferramenta?
- A ferramenta fornece bibliotecas das Ontologias que podem ser reutilizadas? Através de qual operação (inclusão, união, etc)?
- Há níveis altos de premissas (*high-level primitives*)?
- Existe informação sobre os termos usados na construção de uma Ontologia no sistema de ajuda?

### 3) Cooperação:

- A ferramenta permite a edição síncrona da mesma Ontologia por usuários diferentes?
- Há maneiras de travar a Ontologia (lock)?
- É possível navegar numa Ontologia, se ela for travada (locked)?
- As mudanças feitas por outros usuários são fáceis de reconhecer?
- É possível exportar o código das Ontologias em vários formatos?
- É possível importar uma descrição de Ontologia de uma outra ferramenta?

Os resultados da avaliação realizada por Duineveld et al (1999)<sup>112</sup>, bem como as ferramentas avaliadas, estão resumidas no gráfico 1:

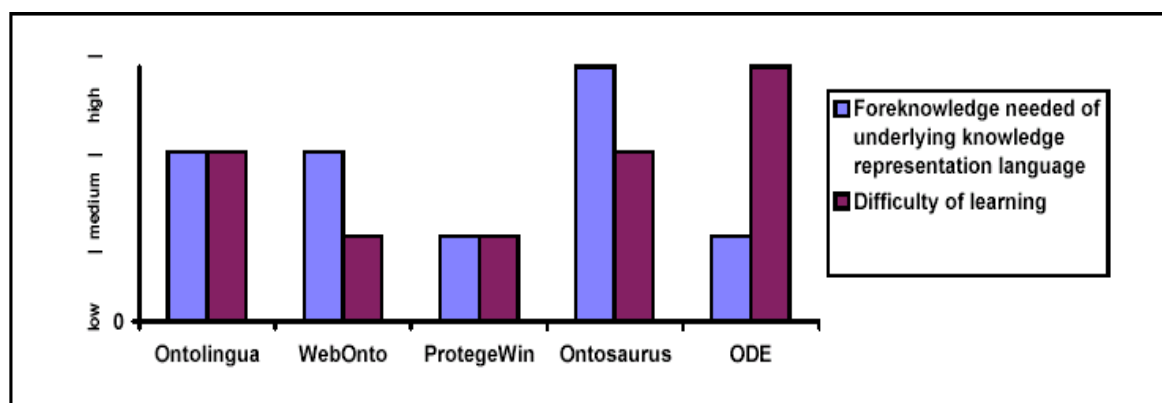


Gráfico 1 – Avaliação das Ferramentas de Engenharia de Ontologia.

Fonte: Duineveld et al (1999)

<sup>112</sup>Duineveld, A. J. et al, 1999. WonderTools? A comparative study of ontological engineering tools. *Twelfth Workshop on Knowledge Acquisition, Modeling and Management*. Voyager Inn, Banff, Alberta, Canada.

*ProtégéWin* e o ODE não requerem muito conhecimento da linguagem subjacente de representação e, conseqüentemente, são visados por usuários não potenciais. A aprendizagem no *ProtégéWin* é fácil, pois sua interface é simples/direta. Já o ODE é mais difícil de aprender, porque usa um formato de tabela. A penalidade que o *ProtégéWin* paga pela simplicidade é que ele fornece primitivos de menor *high-level*, e não permite a modelagem de axiomas. O *Ontolingua*, o *OntoSaurus* e o *WebOnto*, todos, usam a estrutura de uma linguagem subjacente na representação do conhecimento. Isso permite uma modelagem mais complexa, mas a desvantagem é que os usuários necessitam saber as linguagens antes de poderem construir as Ontologias (é especialmente o caso do *OntoSaurus*, que usa o ‘*Loom*’). O *OntoSaurus* foi desenvolvido principalmente como uma ferramenta de navegação.

No restante desta discussão, as ferramentas foram comparadas de acordo com as três dimensões definidas na estrutura de avaliação: Geral (interface); Ontologias; e Cooperação. A tabela 03 fornece um resumo dos resultados, em que foi compilada a escala 0-10 em uma escala de três níveis {+, 0, -}. A escala mais (+) significa positiva, por exemplo: a característica está disponível ou corretamente implementada. A escala zero (0) significa razoável, por exemplo: a característica está disponível, mas é difícil usá-la. A escala menos (-) é negativa, por exemplo: a característica não é suportada e não é executada corretamente. 'NA' significa não aplicável. (Ver Tabela 03: os resultados da pesquisa realizada por Duineveld et al (1999)):

Tabela 03 – Resumo dos Resultados da Avaliação de Ferramentas de Engenharia de Ontologias.

CRITÉRIO	ONTO LINGUA	WEB ONTO	PROTÉGE WIN	ONTO SAURUS	ODE
<b>GERAL</b>					
1.1 Clareza da Interface	-	+	+	-	-
1.2 Consistência da Interface	+	+	+	+	+
1.3 Velocidade de Atualização	-	0	+	-	+
1.4 Visão Geral	0	+	+	+	-
1.5 Compreensão dos Comandos	+	+	+	+	0
1.6 Identificação das Mudanças	0	0	0	0	0
1.7 Estabilidade do Sistema	+	+	+	+	-
1.8 Instalação Local	Não	Não	Sim	Sim/Não	Sim
1.9 Sistema de Ajuda	+	-	+	+	-
<b>ONTOLOGIAS</b>					
2.1 Heranças Múltiplas	Sim	Sim	Sim	Sim	Sim
2.2 Tipos Decompostos	+	+	-	+	+
2.3 Verificação de Consistência	+	+	+	+	+

Fonte: Duineveld et al (1999)

#### 4.3.2.1 Resultado da Avaliação da Suíte EC

Na tabela 04, estão os resultados da avaliação da Suíte de Engenharia do Conhecimento. Os critérios foram estabelecidos conforme o descrito no item 4.3.2, baseados na metodologia proposta por Duineveld et al (1999)<sup>113</sup>. Participaram da avaliação 6 (seis) usuários (as) da ferramenta.

Para avaliação foi compilada a escala 0-10 em uma escala de três níveis {+, 0, -}. A escala mais (+) significa positiva, por exemplo, quando a característica está disponível ou corretamente implementada. O zero (0) significa razoável, por exemplo, quando a característica está disponível, mas é difícil usá-la. A escala menos (-) é negativa, por exemplo, quando a característica não é suportada, nem é executada corretamente. 'NA' significa não aplicável. (Ver tabela 04).

Tabela 04 – Resultado da Avaliação da Suíte de Engenharia do Conhecimento.

CRITÉRIOS	SUÍTE DE ENGENHARIA DO CONHECIMENTO
<b>GERAL</b>	
1.1 Clareza da Interface	+
1.2 Consistência da Interface	+
1.3 Velocidade de atualização	+
1.4 Visão Geral	0
1.5 Compreensão dos comandos	+
1.6. Identificação das Mudanças	+
1.7 Estabilidade do Sistema	+
1.8 Instalação local	NA
1.9 Sistema de Ajuda	-
<b>ONTOLOGIA</b>	
2.1 Herança Múltipla	NA
2.2 Tipos decompostos	NA
2.3.1 Verificação de consistência	-
2.3.2 Nível de verificação	-
2.4 Exemplos de Ontologias	-
2.5 Reutilização de Ontologias	+
2.6 Nível elevado de Premissas	NA
2.7 Ajuda das Ontologias	-
<b>COOPERAÇÃO</b>	
3.1 Sincronia da Edição	+
3.2 Bloqueamento das Ontologias	-
3.3 Navegação enquanto Bloqueadas	NA
3.4 Reconhecimento das Mudanças	+
3.5 Facilidade de exportação	-
3.6 Facilidade de importação	-

<sup>113</sup> WonderTools. Disponível em: <http://www.swi.psy.uva.nl/wondertools/>. Acesso em: 20 nov. 2004.

Com o resultado da avaliação, observou-se que a ferramenta Suíte de Engenharia do Conhecimento possui excelente nível de usabilidade, não exigindo do usuário o conhecimento de uma linguagem específica, pois a especificação é feita em linguagem natural, além de possuir uma interface intuitiva para o usuário.

Especificamente no tocante à fundamentação das Ontologias, a ferramenta possui ainda poucos recursos, fato compensado pelo uso da metodologia RC2D e outras ferramentas, como os Extratores de Frequência e Semântico.

Quanto ao critério de cooperação, a ferramenta possui as principais funções; no entanto, ela precisa ser aprimorada tecnologicamente para facilitar o seu uso pelo especialista.

#### 4.3.3 CONCLUSÃO

A Suíte de Engenharia do Conhecimento foi desenvolvida de forma a permitir um trabalho cooperativo entre pessoas em diferentes lugares, estruturando uma base de conhecimento contínua e de fácil visualização (árvore de conhecimento), através de redes de relacionamento.

Os testes, na construção das Ontologias, ficaram voltados aos conceitos que as Ontologias pretendem. Assim, quando se fez os testes, foram considerados os objetivos do sistema, quais sejam: para que o sistema se destina ou qual o tipo de resposta que os usuários pretendem e em qual contexto. Esta clareza, sobre os objetivos do sistema, foi fundamental, pois determinou que a arquitetura de construção das Ontologias seguisse sempre os mesmos critérios, em todos os domínios.

Da mesma forma, a utilização das palavras-chaves também foram consideradas. As palavras-chaves, apesar de recuperarem as informações, não representaram os únicos componentes da Ontologia, inclusive, em alguns casos, até não estavam presentes na Ontologia. As Expressões Indicativas, formadas pelos termos, representaram o contexto, fazendo com que o sistema recuperasse as informações, apresentadas de outras formas, ou seja, as relações partitivas ou de pertinência como ‘tipo de’ e ‘parte de’.

Considerando que, no Sistema KMAI, as relações de sinônimo, tipo de (hipernímia e hiponímia), parte de (meronímia e holonímia) e relação de conexo foram utilizadas para expressar uma relação matemática de proximidade e frequência entre os termos, nos testes,

verificou-se que as relações construídas não se restringiram apenas a uma taxonomia. Isso significou que, para cada contexto, somente foram pertinentes os termos que tiveram grande relação entre si, dentro desse universo. Assim, observou-se que, para o sistema, é possível não interessar a relação de todos os tipos de armas ou todos os tipos de crimes; irá depender da definição do contexto, do propósito do sistema. Nos testes e avaliação de Ontologias, recomenda-se que isso deve ser considerado.

A Suíte de Engenharia do Conhecimento revelou-se adequada aos melhores parâmetros de avaliação existentes na área de ferramentas de Engenharia de Ontologias. A Suíte permite não somente a comunicação com o especialista, como também a integração da equipe de Engenharia do Conhecimento, que precisa estar em total sintonia no momento do processo de produção. O ambiente facilita a concentração na tarefa, possibilita a manifestação de soluções criativas e estimula a colaboração mútua. Mas o mais importante é a linguagem natural utilizada para representar as Ontologias na Suíte, isto é, o usuário prescinde de qualquer conhecimento em programação computacional. Isso torna a Suíte de Engenharia do Conhecimento uma ferramenta única do gênero.

Portanto, pôde-se verificar a eficiência e eficácia deste ambiente na qualidade da representação do conhecimento. Além disso, o constante acompanhamento dos especialistas e o seu conhecimento sobre os métodos de representação permitiram que eles interferissem positivamente na construção das Ontologias a qualquer momento e em qualquer lugar com acesso Web.

## RESULTADOS

A metodologia da *Engenharia da Mente* foi elaborada a partir de observações dos elementos utilizados pela equipe de pesquisadores para o desenvolvimento de Sistemas Baseados em Conhecimentos (SBC) antes e após a implantação do processo de Engenharia de *Software*. Os melhores resultados foram alcançados quando se aplicou somente a metodologia de Engenharia do Conhecimento para a construção dos Sistemas Inteligentes. No entanto, mesmo a Engenharia do Conhecimento não apresentou elementos suficientes para que a equipe alcançasse os resultados de eficiência e eficácia, pois a metodologia era, ainda, experimental.

O número de Ontologias cresceu significativamente, quando foram identificados os sinônimos e as relações, razão pela qual, num teste direto na Web, o retorno de documentos apresentou excesso de informação ou informação não relevante. Porém, a partir da aplicação da ferramenta Suíte de Engenharia do Conhecimento, nos testes da ferramenta, os mesmos sinônimos e as relações reduziram o número de documentos recuperados, porém representaram todas as informações contextualizadas.

Na avaliação e teste, utilizaram-se as Ontologias como recurso de busca. Os documentos recuperados foram analisados um a um, verificando se a recuperação contemplou os sinônimos, as relações partitivas (tipo de, parte de) ou os conexos e, ainda, se do primeiro ao último documento, todo o assunto foi contextualizado. Como o sistema diferencia os termos por cores, a avaliação tornou-se bastante otimizada.

Quando a recuperação apresentou algum texto que não tinha pertinência ao assunto, procedeu-se a uma avaliação das relações construídas a partir do termo. Observou-se que, quando não houve um perfeito entendimento do contexto e nem uma visualização das ações dentro dele, muitas relações acabaram sendo equivocadamente construídas. Após a avaliação, todas as relações foram reconstruídas para uma nova série de testes. Assim, individualmente, todos os termos e Expressões Indicativas foram testados na ferramenta. Como as fontes foram definidas na Engenharia do Conhecimento, e tiveram como critério para essa seleção uma



importante base de consulta, as Ontologias apresentaram a mesma linguagem encontrada nos textos.

Ao associar ferramentas e procedimentos de Engenharia de *Software* à Engenharia do Conhecimento, alguns elementos foram criados para se obter os resultados esperados. Por essa razão, os resultados a seguir apresentados foram obtidos pela equipe durante a fase da sistematização da metodologia da *Engenharia da Mente*. No quadro 7, é possível visualizar um resumo do processo e resultados:

<b>Área de aplicação da Engenharia da Mente</b>	<b>Importância</b>	<b>Visualização e definição de relevância</b>	<b>Resultado</b>
Engenharia do Conhecimento	<p>Maior efetividade na representação e aquisição do conhecimento em sistemas baseados no conhecimento.</p> <p>Permite criar novos algoritmos e soluções tecnológicas para o sistema que está sendo construído;</p>	<p>Atributos e valores do sistema;</p> <p>Interface do sistema;</p> <p>Mapa conceitual em relação ao domínio de aplicação.</p>	<p>1) Celeridade na implantação do sistema;</p> <p>2) Sincronização da equipe evitando o retrabalho;</p> <p>3) Uniformização do Vocabulário;</p> <p>4) Compartilhamento contínuo de Visões.</p>
Engenharia de Ontologias	<p>Reutilização do conhecimento organizado, Estudos da Ontologia Aplicada.</p>	<p>Vocabulário Contextualizado;</p>	<p>1) Maior coerência entre as relações das palavras;</p> <p>2) Maior efetividade das Ontologias na aplicação em Sistemas Baseados no Conhecimento.</p>

Quadro 7 – Objetivos da Aplicação da Engenharia da Mente.

Contudo, o resultado mais importante foi que, no decorrer desta pesquisa, a metodologia da *Engenharia da Mente* pôde ser aplicada na prática. Durante dois anos, entre 2002 e 2004, na Secretaria de Acompanhamento e Estudos Institucionais (SAEI), do Gabinete de Segurança Institucional da Presidência da República Federativa do Brasil, a Suíte de Engenharia do Conhecimento foi implementada, conforme os critérios descritos neste estudo, numa primeira fase de forma experimental. Nos dias atuais, o sistema funciona atrelado ao fluxograma do órgão governamental, apresentando, segundo fontes oficiais, excelente desempenho, proporcionando grande produtividade nas atividades de busca e recuperação de informações, sobre o qual o sistema foi projetado.

## CONSIDERAÇÕES FINAIS E DESENVOLVIMENTOS FUTUROS

As ferramentas de Ontologia filosófica têm sido aplicadas para resolver problemas práticos, num ramo chamado ‘Ontologia Aplicada’<sup>114</sup>. Observa-se que essas teorias estão adentrando em novos domínios, como a ciência do Direito, do Comércio, da Medicina, da Geografia, entre outras.

Entretanto, esta pesquisa demonstrou que o crescimento de aplicações da Engenharia de Ontologia deve ser observado sob uma visão mais ampla, que reflita os esforços de cientistas da computação e da informação que olham por de trás do vasto mundo que os artefatos de computação e informação relatam. Em outras palavras, na atualidade, os sistemas de informação baseados em Ontologia devem ser considerados como um novo ramo de aplicações práticas que está requerendo ser explorado por métodos com rigor filosófico.

A sistematização e organização das equipes de especialistas do domínio junto com a equipe de Engenharia do Conhecimento se mostraram como o grande desafio no desenvolvimento de sistemas de gestão do conhecimento. As metodologias de Engenharia do Conhecimento, na sua quase totalidade, baseiam-se em entrevistas com o especialista para realizar aquisição do conhecimento. Porém, no início desta pesquisa, observou-se que as entrevistas não eram suficientemente representativas. O primeiro passo foi explicar as técnicas de IA e a forma de representação do conhecimento, fator que melhorou e muito o conhecimento repassado. Por essa razão, na metodologia da *Engenharia da Mente*, o especialista tem um papel ativo no desenvolvimento dos Sistemas Inteligentes; mas não se resume a isso.

A metodologia para a sincronização de conhecimentos descrita neste trabalho produziu uma excepcional coerência entre as relações semânticas das chamadas Expressões Indicativas, principalmente pelo apoio de toda a estrutura computacional durante o processo. Isso

---

<sup>114</sup> Koepsell, David R. *The Ontology of Cyberspace: Law, Philosophy, and the Future of Intellectual Property*, Chicago: Open Court. 2000.

possibilitou ao engenheiro do conhecimento e ao especialista desenvolverem muito mais que o conhecimento do domínio, produziu competências como consciência própria, disciplina, persistência e empatia.

É importante ressaltar que a elaboração de Ontologias, de forma contextualizada, permitiu uma indexação automática de informações pelo sistema. Porém, revelou também que uma aquisição do conhecimento pouco abrangente levará a um sistema pouco inteligente.

Portanto, pode-se concluir que a metodologia proposta tem por finalidade atuar na gestão da mudança, em que Sistemas Inteligentes devem ser implantados em ambientes onde as pessoas estão motivadas para modificar o procedimento de tomada de decisão. Para isso, as pessoas precisam estar preparadas. Se for para repetir o mesmo procedimento, a metodologia da Fábrica de *Software* é mais adequada.

Sob essa perspectiva, a *Engenharia da Mente* é um novo modelo de Representação e Aquisição do Conhecimento, em que a compreensão das emoções dos especialistas e a organização de sua rede de relacionamentos fazem parte da base de conhecimento do Sistema Inteligente, influenciando na sua modelagem e na construção de Ontologias. Através da participação constante do especialista no processo de aquisição do conhecimento, e não somente o seu conhecimento técnico do domínio, é possível obter um Sistema Inteligente com maior eficácia.

O cérebro humano necessita de um bom treinamento tanto quanto um computador necessita de uma boa programação; é a nossa métrica da inteligência.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AMONDT, A., Plaza, E., **Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches**. AI Communications, 17(1), 1994.
- ARÍS, Enrique P.; GONZÁLEZ, Juan L. S.; RUBIO, Fernando M. **Lógica Computacional**. Thomson Editores Spain: Madrid, 2003.
- BAESA-YATES, R. and RIBEIRO NETO B., **Modern Information Retrieval**. Addison Wesley, Reading, Mass, 1999.
- BARABÁS, Albert-László. **Linked: How everything Is connected to Everything Else and What It means for Business, Science, and Everyday Life**. Plume: New York, 2003.
- BENJAMINS, V.R., 1998. **The ontological engineering initiative (KA)<sup>2</sup>**. Formal Ontology in Information systems. IOS Press, Amsterdam.
- BERNERS-LEE, T.; HENDLER, J., Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, mai. 2001. Disponível em: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 20 julho de 2004.
- BITTENCOURT, Guilherme. **Inteligência artificial – Ferramentas e teorias**. Florianópolis: Editora da UFSC, 1998.
- BRABO, L. (2002). Filosofia Sistemática. <http://www.filosofia.catolico.org.br>
- BRÄSCHER, Marisa. A ambiguidade na Recuperação da Informação. DataGramaZero. **Revista de Ciência da Informação** – v. 3, n.1, fev/02. Disponível em: [http://www.dgzero.org/fev02/Art\\_05.htm](http://www.dgzero.org/fev02/Art_05.htm). Acesso em 16 de maio de 2004.
- BRASIL. Presidência da República. Gabinete de Segurança Institucional. Disponível em <http://www.presidencia.gov.br/gsi/>. 2006
- BUENO, Tania C. D.; HOESCHL, Hugo C.; BORTOLON, Andre; BARCIA, Ricardo. Engineering of Minds: The Synchronicity Between Artificial Intelligence and the Management of Knowledge in Collaborative Networked Organizations. Proceedings IADIS International Conference www/internet 2004. Madrid, Espanha.
- BUENO, Tânia C. D.; HOESCHL, Hugo C.; BORTOLON, Andre; MATTOS, Eduardo da Silva; RIBEIRO, Marcelo Stopanovski. Analyzing the use of dynamic weights in legal case based system. In: Ninth International Conference On Artificial Intelligence And Law, 2003, Edimburgo. Proceedings of the Conference. New York: ACM, 2003. v. 1, p. 136-141.
- BUENO, Tânia C. D. **O Uso da Teoria Jurídica para Recuperação em Amplas Bases de Textos Jurídicos**. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 1999.
- BUENO, Tânia C. D.; WANGENHEIN, Christiane Gresse Von; HOESCHL, Hugo Cesar; MATTOS, Eduardo da Silva; BARCIA, Ricardo Miranda. Retrieval in Jurisprudential Text Bases using Juridical Terminology. Proceedings of 7th International Conference in Intelligence Artificial And Law - ICAIL, 1999, Oslo. ACM, 1999.

CAMERON, Julia. **Guia Prático para a Criatividade: o caminho do Artista**. Rio de Janeiro: Ediouro, 1996.

CAMPOS, Liliana S. L. Q; LIMA, Martha M. C. de Andrade. **Mapeamento dos Capitais Intelectuais Necessários ao Monitoramento de Temas pela Secretaria de Acompanhamento e Estudos Institucionais da Presidência da República**. Projeto Final. Pós Graduação Latu Sensus, Inteligência Empresarial e Gestão do Conhecimento. COPPE/UFRJ, 2002.

CAPRA, Fritjof. **As Conexões Ocultas: Ciência para uma vida sustentável**. Editora Cultrix. 3ª Ed. São Paulo, 2003.

CARDEÑOSA, Jesus. Internet ya tiene esperanto. Disponível em: [http://www.webzinemaker.com/admi/m6/page.php3?num\\_web=1604&rubr=4&id=26997](http://www.webzinemaker.com/admi/m6/page.php3?num_web=1604&rubr=4&id=26997). Acesso em 30 de novembro de 2004.

CASTORIADIS, Cornelius. **A instituição imaginária da sociedade**. Tradução de Guy Reynaud. 3ª Edição. Rio de Janeiro: Paz e Terra, 1982.

\_\_\_\_\_. **As encruzilhadas do labirinto, III: o mundo fragmentado**. Tradução de Rosa Maria Boaventura. Rio de Janeiro: Paz e Terra, 1992.

CURY, Augusto Jorge. **Inteligência Multifocal: Análise da construção dos pensamentos e da formação de pensadores**. Editora Cultrix: São Paulo, 1999.

DAMÁSIO, António. **O mistério da Consciência: do corpo e das emoções ao conhecimento de si**. Tradução Laura Teixeira Motta. 5ª edição. São Paulo: Companhia das Letras, 2000.

DAVIS, R.; SHOROB, H.; SZOLOVITS, P. What is knowledge representation? **AI Magazine**, pages 17-33. 1993.

DING, Y., Fensel, D., KLEIN, M.; OMELAYENKO, B. **The semantic web: yet another hip?** *Data & Knowledge Engineering*, 41(2/3):205-227. (2002)

DESCARTES, René. **Discurso do método: para bem conduzir a própria razão e procurar a verdade nas ciências**. São Paulo: Ed. Paulus, 2002.

DUINEVELD, A. J.; STOTER, R.; WEIDEN, M. R.; KENEP, B.; BENJAMINS, V. R.: Wondertools? A comparative study of ontological engineering tools. In the **International Journal of Human-Computer Studies**, July 2000.

DREYFUS, Hubert L. Intelligence Without Representation. Disponível em: <http://www.hfac.uh.edu/cogsci/dreyfus.html>. Acesso em: 19 jan. 2004.

ERIKSONN, H. et al.. **Automatic Generation of Ontology Editors**. Twelfth Workshop on Knowledge Acquisition, Modeling and Management. Voyager Inn, Banff, Alberta, Canada 1999.

GELB, Michael G. **Aprenda a pensar com Leonardo da Vinci**. Editora Ática: São Paulo, 2003.

GOLEMAN, Daniel. **Inteligência Emocional**. Ed. Objetiva: Rio de Janeiro, 1995.

GOLEMAN, David. Emotional Intelligence: What it is and Why it Matters. 2000. Disponível em: [http://www.eiconsortium.org/research/what\\_is\\_emotional\\_intelligence.htm](http://www.eiconsortium.org/research/what_is_emotional_intelligence.htm). Acesso em: 19 jan. 2004.

- GRATTON, Lynda; GHOSHAL, Sumantra. Managing Personal Human Capital: new ethos for the "Volunteer" Employee, **The European Management Journal**, vol 21, n° 1 pp1-10, February, 2003.
- GUARINO, Nicola; WELTY, Chris. A Formal Ontology of Properties. In: Dieng, R., and Corby, O., eds, Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management. Spring-Verlag LNCS Vol. 1937:97-112. October, 2000.
- GUARINO, N. Some Ontological Principles for Designing Upper Level lexical Resources. In: A Rubio, N. Gallardo, R. Castro and A. Tejada (eds.), Proceedings of First international Conference on Language Resources and Evaluation. ELRA – European Language Resources Association, Granada, Spain (1998) 527 –534.
- HARRIS, Judith. **Jung e a Yoga**. – A ligação Corpo-Mente. Editora Claridade, São Paulo, 2004.
- HOESCHL, Hugo C. Sistema Olimpo: Tecnologia da Informação Jurídica para o Conselho de Segurança da ONU. Tese de Doutorado. Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina. Florianópolis:2001.
- HOESCHL, Hugo. C.; BUENO, Tania. C. D.; BARCIA, Ricardo. M.; BORTOLON, Andre; MATTOS, Eduardo da Silva. Olimpo: Contextual Structured Search to improve the representation of UN Security Council with Information Extraction methods. Proceedings da 8a International Conference on Artificial Inteligence and Law, ICAIL - 2001, St. Louis. New York: ACM SIGART, 2001. p.217 – 218.
- HOESCHL, Hugo. C. et al, 2003. Structured Contextual Search For The Un Security Council. Proceedings of the fifth International Conference On Enterprise Information Systems. Anger, France, v.2. p.100 - 107
- HOSFSTEDE, Geert H. **Cultures and organizations: software of the mind**. New York:MacGraw-Hill, 1991.
- HOPCKE, Robert H. **Sincronicidade – Ou Por Que Nada É Por Acaso**. Editora Record. 2000.
- JUNG, Carl Gustav. **Os arquétipos e o inconsciente coletivo**. Editora:Vozes, 2000.
- KOLODNER, Janet L. **Case-based reasoning**. Morgan Kaufmann Publishers, Inc. 1993. San Marco, CA94403.
- LACAN, Jacques. **Os Seminários de Lacan**. Disponível em CD ROOM, 2000.
- MARQUES NETO, Agostinho Ramalho. Subsídios para pensar a possibilidade de articular direito e psicanálise. Texto disponível em papel.
- MATURANA, Humberto R.; VARELA, Francisco J. **A Árvore do Conhecimento - as bases biológicas da compreensão humana**. Editora Palas Athena, 2001.
- LIPP, Marilda E. Novaes [organizadora]. **Mecanismos Neuropsicofisiológicos do Stress: Teoria e Aplicações Clínicas**. São Paulo: Casa do Psicólogo, 2003.
- MENEZES, Luíz César de Moura. **Gestão de Projetos**. São Paulo:Atlas, 2001.
- MILLER, Jacques-Alain. **Perspectiva do Seminário 5 de Lacan: as formações do Inconsciente**. Tradução: Maria Josefina S. Fuentes. Rio de Janeiro: Jorge Zahar Ed., 1999.
- MILLER, Jacques-Alain. **Percursos de Lacan: uma introdução**. 2ª Edição. Tradução de Ari Roitman. Rio de Janeiro: Jorge Zahar Editor.

O'DONNELL, Ken; BANCON, Brian. **No Olho do furacão**. Salvador,BA: Casa da Qualidade,1999.

OJEDA, Sergio R.; HILL, Jennifer; HILL Diane F.; COSTA, Maria E.; TAPIA, Veronica; CÓRNEA, Anda; MA, Ying J. The Oct-2 POU Domain Gene in the Neuroendocrine Brain: A Transcriptional Regulator of Mammalian Puberty. *Endocrinology* Vol. 140, No. 8 3774-3789. 1999:The Endocrine Society. Disponível na Internet em <http://endo.endojournals.org/cgi/content/full/140/8/3774>. Acesso em 07 de julho de 2004.

Ontolingua. Disponível em: <http://www-ksl-svc.stanford.edu>. Acesso em: 28 de novembro de 2004.

Ontolingua. Disponível em: <http://ontolingua.nici.kun.nl>. Acesso em: 29 de novembro de 2004.

OntoSaurus. Disponível em: <http://sevak.isi.edu:8300/loom/shuttle.html> (username and password required).

PINKER, Steven. **Como a mente funciona**. São Paulo: Companhia das Letras,1998.

PIRSIG, Robert M. **Zen e a Arte de manutenção de motocicletas: uma investigação sobre valores**. Tradução de Celina Cardim Cavalcanti. Rio de Janeiro:Paz e Terra, 1984.

PRESSMAN, Roger. **Engenharia de Software**. São Paulo: Makron Books, 1995.

ProtégéWin. Disponível em: <http://smi-web.stanford.edu/projects/prot-nt/>. Acesso em: 18 de nov. 2004.

REALE, Miguel (2004). *Perspectiva e Teoria do Ser*. Disponível em : <http://www.miguelreale.com.br/>. Acesso em 30 de maio de 2004.

REZENDE, Solange O. (org.). **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri-SP: Manole, 2003.

RIBEIRO, Jorge Ponciano. **Gestalt-terapia: o processo grupal: uma abordagem fenomenológica da teoria do campo e holística**. São Paulo:Summus, 1994.

RIBEIRO, Marcelo Stopanovski. **KMAI, da RC<sup>2</sup>D à PCE. Gestão do conhecimento com inteligência artificial, da representação do conhecimento contextualizado dinamicamente à pesquisa contextual estruturada**. [2004]. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

RIBEIRO, Marcelo Stopanovski; MATTOS, Eduardo da Silva; BUENO, Tânia C. D.; HOESCHL, Hugo Cesar. KMAI- Knowledge Management With Artificial Intelligence. The Symposium on Professional Practice in AI in the First IFIP International Conference On Artificial Intelligence Application And Innovations. Toulouse, 2004.

SANTOS JÚNIOR, Eugene; SANTOS Eugene S.; SHIMONY, Solomon Eyal. Implicitly preserving semantics during incremental knowledge base acquisition under uncertainty. *International Journal of Approximate Reasoning*. Volume 33, p. 71-94. 2003.

Science does it with feeling. *The economist*. July 20th 1996,. p.71 a 73

Semantic Web. Disponível em: <http://www.w3.org/2001/sw/>. Acesso em: 19 jan. 2004.

SERPA, Luiz Felipe Perret; SILVA, Aderval Barros da. *A Física Quântica e a Sincronicidade*. Disponível em: [http://www.faced.ufba.br/rascunho\\_digital/textos/236.htm](http://www.faced.ufba.br/rascunho_digital/textos/236.htm). Acesso em: 19 de janeiro de 2004.

SHELDRAKE, Rupert. Sociedade, espírito e ritual: ressonância mórfica e inconsciente coletivo. 1997. Disponível em: <http://ligiacabus.sites.uol.com.br/traducoes/sheldrake01.htm>. Acesso em: 23 de agosto de 2004.

SCHANCK, R. **Dynamic Memory: A theory of learning in computers and people**. New York, Cambridge University Press, 1982.

THIECHER, Adilson; HOESCHL, Hugo Cesar; ZIMATH, Patrícia Bonina. Interlegis: virtual network of communication and information that enlarges Brazil's democracy and citizenship. In: Traumüller, Ronald. (Org.). Lecture notes in computer science 2739. Berlin Heildeberg, 2003, v. 1, p. 75-79.

TREFILI, James. **Somos diferentes? Um cientista explora a inteligência ímpar da mente humana**. Rio de Janeiro:Rocco, 1999.

UNL – Universal Networking Language. Disponível em: <http://www.unl.ias.unu.edu/unlsys/index.html>. Acesso em: 19 jan. 2004.

VALENTE, Andre. **Legal Knowledge Engineering: A Modelling Approach**. IOS Press, (Amsterdam) and Omsha (Tokyo), 1995.

WARAT, Luis Aberto. **Introdução Geral ao Direito: O Direito não estudado pela teoria jurídica moderna**. Sergio Antonio Fabris Editor: Porto Alegre, 1997.

WARAT, Luis Alberto. **O direito e a sua linguagem**. 2ª Versão. Sergio Antonio Fabris Editor: Porto Alegre, 1995.

WARAT, Luiz. Alberto. **Interpretação da Lei: temas para uma reformulação**. Porto Alegre: Sergio Antonio Fabris Editor, 1994.

WebOnto. Disponível em: <http://webonto.open.ac.uk/>. Acesso em: 17 nov. 2004.

WOOD, Ernest E. **Mind and Memory Training**. The Theosophical Publishing House:London, 1974.

WonderTools. Disponível em: <http://www.swi.psy.uva.nl/wondertools/>. Acesso em: 20 nov. 2004.

Wordnet. Disponível em: <http://www.cogsci.princeton.edu/~wn/>. Acesso em: 19 jan. 2004.



## **ANEXOS**

**Anexo I – Paper aprovado no DEXA 2005 - Database and Expert Systems Applications, 16 Conferência Internacional, Copenhagen, Dinamarca**

**Knowledge Engineering Suite: A Tool to Create Ontologies for Automatic Knowledge Representation in Knowledge-Based Systems**

# Knowledge Engineering Suite: a Tool to Create Ontologies for Automatic Knowledge Representation in Knowledge-based Systems

- Tania Cristina D'Agostini Bueno<sup>1</sup>, Hugo Cesar Hoeschl<sup>2</sup>, Andre Bortolon<sup>2</sup>, Eduardo Mattos<sup>1</sup>, Cristina Souza Santos<sup>1</sup>

<sup>1</sup>WBSA Sistemas Inteligentes SA, Parque Tecnológico Alfa, Centro de Tecnologia IlhaSoft , SC 401 Km 1 - Módulo 10 - Térreo B - João Paulo - 88030-000 - Florianópolis, SC – Brasil  
{tania,mattos,cristina}@wbsa.com.br  
<http://www.wbsa.com.br>

<sup>2</sup>Instituto de Governo Eletrônico, Inteligência Jurídica e Sistemas – IJURIS, Rua Lauro Linhares, 728 – sala 105 – Trindade - 88036-0002 - Florianópolis – SC – Brasil  
{hugo, andre}@ijuris.org  
<http://www.ijuris.org>

**Abstract.** This paper is focused on the process of systematic knowledge acquisition to be used in knowledge-based systems. The result is a computational structure that can be used inside the organization (Intranet) as well as outside (Internet). This structure is the Knowledge Engineering Suite, an ontological engineering tool to support the construction of ontologies in a collaborative environment and is based on observations from the Semantic Web, UNL (Universal Networking Language) and WordNet. We use both a knowledge representation technique called DCKR to organize knowledge, and psychoanalytic studies, focused mainly on Lacan and his language theory to develop a methodology called Mind Engineering to improve the synchronicity between knowledge engineers and specialists on a particular domain.

## 1 Introduction

The importance of knowledge-based systems is that they provide some particular characteristics of human intelligence to the computer, such as the capacity to understanding natural language and simulate reasoning under uncertainty conditions. Definition of the relevant information to be inserted into a knowledge-based system is a major problem in the construction of such systems, mainly because the process is basically experimental and depends mostly on the ability of the knowledge engineer. In particular, there is a high difficulty related to the definition of the terminology used to nominate the concepts and the relations. [1] Despite the high number of methods to perform the knowledge acquisition process, there is no one that deals with the understanding and learning of people involved in the process, both specialists and knowledge engineers.

More recently, the notion of ontology has become popular in fields such as intelligent information integration, information retrieval on the Internet, and knowledge management. The reason is partly due to what they promise: a shared and common understanding of some domain that can be communicated through people and computers [2]. Cooperative work has been used by different development teams worldwide, with reference to WordNet, Semantic Web and UNL (Universal Networking Language) through the construction of ontologies using collaborative tools. The use of ontological engineering tools, or metatools, to support the Knowledge Engineering process enables the process of organizing a knowledge base established on the relationship between relevant expressions within a context. Ontologies, as a basis for automatic generation of knowledge acquisition tools, simplify the system specification phase by taking advantage of ontologies defined during the Knowledge Engineering process [3]. Nevertheless, experience shows that often the bottleneck of building sharable ontologies lies more in the social process than in the technology itself [4]. Therefore, a methodology for the process of knowledge acquisition was developed, so that the specialist and the knowledge engineer can work in synchronicity, in cooperative networked organizations. We call this methodology Mind Engineering. This synchronization process begins with the understanding of human intelligence, its unconscious manifestations

and its relationship with words, since, according to Lacan, every human investigation is linked irreversibly to the inner space created by language.

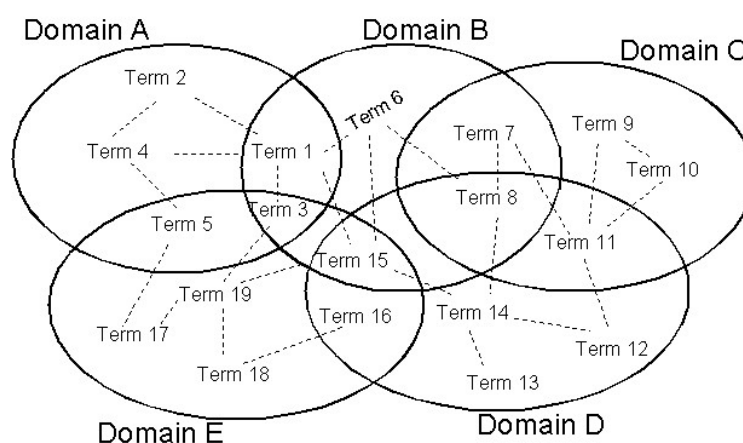
In the present development, a tool was created to support the Knowledge Engineering process by assisting developers in the design and implementation of ontologies on a specific domain.

In earlier works, we used a methodology called DCKR (Dynamically Contextualized Knowledge Representation) [5]. DCKR allows to build a knowledge base, improving the construction of the ontology of the domain and the automatic representation of cases in knowledge-based systems, either in the legal area [6] or any other knowledge management domain [7].

It follows a description of the methodology for knowledge synchronization. This methodology allowed an exceptional coherence among the semantic relations of what are called 'indicative expressions', mainly by the support of all this computational structure during the process. This allowed the knowledge engineer and the specialist to develop, more than the knowledge representation of the domain, abilities such as an inherent conscience, discipline, persistence, and empathy.

## 2 Knowledge Representation in Knowledge-based Systems

We use a special process to extract and represent knowledge in the process of developing knowledge-based systems. The main purpose is to allow an automatic process of text indexing, on the basis of a controlled vocabulary and a dictionary of normative terms, constructed persuasively through the relevance of pre-defined terms, called key-normative terms [8]. Given the need to turn the acquisition process faster, it was necessary to evolve the process using IR (Information Retrieval) techniques to associate the relevance of the terms with the frequency of the words added to the controlled vocabulary and the dictionary of normative terms; this approach resulted in a methodology of knowledge representation called DCKR - Dynamically Contextualized Knowledge Representation [9]. DCKR is a methodology of knowledge representation centered on a dynamic process of acquisition of knowledge from texts, defined through the elaboration of a controlled vocabulary and a dictionary of terms, associated to an analysis of frequency of the words and indicative expressions of the specific context (see figure 1).



**Fig. 1.** The semantic relations of indicative expressions

### 1.1 UNL, Semantic Web and WordNet

In the process of knowledge acquisition for the preparation of a knowledge base of intelligent systems, methodologies that use web environments and cooperative development

have to be used. Nowadays, there are three main solutions worldwide that use the Internet for the development of ontologies: UNL, Semantic Web and WordNet.

UNL (Universal Networking Language) [10] is a language for computers to share information through a network. It is meant for representing the natural language, so that computers can process the text and represent it in different languages.

WordNet [11] is a lexicon reference system inspired in psycholinguistic theories on the human lexical memory. The nouns, verbs, adjectives and adverbs of the English language are, organized in sets of synonyms, each one representing a lexical concept. Different semantic relations link the sets to each other.

The Semantic Web [12] is an extension of the current Web, in which the information has a very well defined meaning, allowing computers to process the information contained in web pages and to understand it, performing operations that facilitate the users' work.

The three initiatives are meant to facilitate the automatic processing of information contained in documents, allowing computers to perform more intelligent operations and to retrieve information in a more efficient way.

## 1.2 The Use of Ontologies in the System

The ontologies structure is the heart of a knowledge-based system that uses DCKR methodology. The reason for that is because all processing and storage of gathered information and knowledge base organization is done using this structure. It also plays an important role in the quality of the results presented to the user.

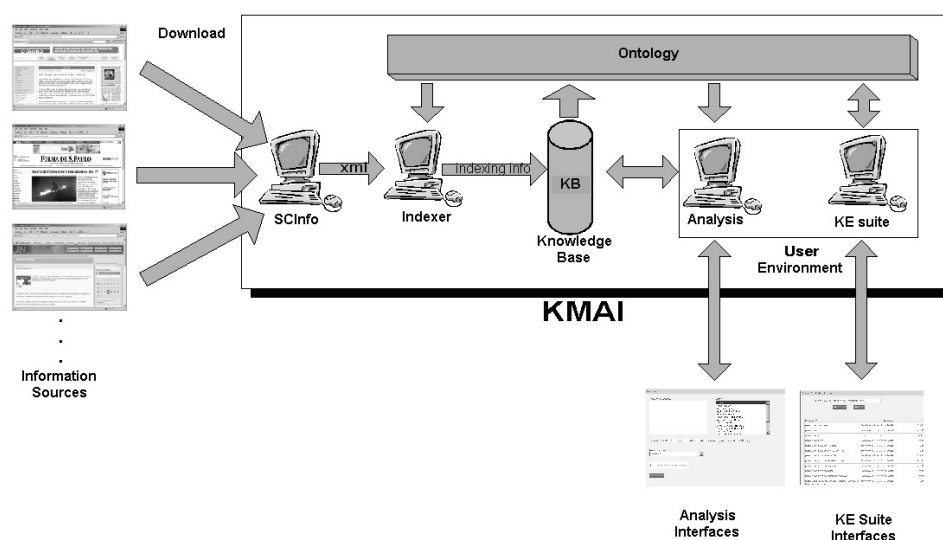


Fig. 2. Ontologies in the system

The participation of the ontology structure in the system occurs in three moments (see figure 2). At the first moment, the system extracts information from different previously selected sources. Each one of these documents is indexed based in the ontologies defined by the specialists and knowledge engineers during the knowledge engineering process. It means that the system will mark the documents with all indicative expressions found in the text, storing them in an organized way in the knowledge base. Thus, it is possible to make a pre-classification of the cases in the base according to what was defined in the knowledge organization promoted by the ontologies.

In a second moment, the ontologies are important in the analysis interface available to the user. The process begins at the moment in which the user types the input text for the search. At this point, the indicative expressions defined by the user that coincide with the ones presented in the ontology are identified. These expressions identified in the entry case determine the stream of relations that will be used by the system. It means that there is a dynamic relation between the way the user enters the indicative expression in the analysis

interface and the way the relations in the Knowledge Engineering Suite are defined for this expression.

The first versions of the Knowledge Engineering Suite worked with key expressions, an approach that resulted in some rigidity in the ontology organization, for the weight of the information that was typed by the user in the search text was not considered. This rigidity is acceptable in cases in which the content of the documents stored in the system is standardized, with a small degree of variation. But in cases with broader domains and with different sources of information with no control over its contents, this approach was not efficient.

For this reason, it was decided to apply an approach that turns the use of ontology more dynamic in the analysis process. In this new approach, the importance of the indicative expressions to be considered is defined by the user. The system gives priority to the expressions and search for the corresponding derivations for each case, according to the knowledge base. A priori, there is no hierarchy in the organization of the ontology in knowledge base. The weight of the relations will be based only in what is required by the search, where the context intended by the user is defined.

The third moment where the ontology takes part is in the Knowledge Engineering Suite, available in the system and integrated in its architecture. Through the Knowledge Engineering Suite the user is able to update the knowledge base with new expressions. At each new update in the ontology, the system re-indexes all the texts stored in the knowledge base, so the users may use this new ontology organization to search for documents previously indexed. It allows the verification of old documents that are related to a context that is important at the present moment. This way, it is possible to define a dateline about a subject, locating its start point.

### **1.3 The Knowledge Engineering Suite**

The Knowledge Engineering Suite is an Ontological Engineering Tool for collaborative work on the Web, aiming to ease the sharing of knowledge between the Knowledge Engineering team and the specialist team. The Suite allows to build relationships between complex terms, considering its concept in the specific domain of application. These relationships are based on AI (Artificial Intelligence) techniques [13], theories of language, Semantic Web, WordNet, and UNL.

The creation of an infrastructure for the acquisition of knowledge for cooperative work on the Web is an efficient and effective tool of knowledge acquisition for intelligent systems. Many different techniques of Knowledge Acquisition exist; but Knowledge Engineering Suite (see figure 3) is integrated as part of DCKR methodology. Here, tools such as the Frequency Extractor, the Semantic Extractor and the Knowledge Engineering Suite have been associated with the methodology to help in the task of knowledge acquisition.

**Theme:** Meio Ambiente  
**Sub Theme:** Biodiversidade  
**Word:** diversidade biológica  
**Synonyms**  
diversidade genética

**This is a type of**

**It belongs to this type**  
material genético

**This is a part of**  
biossegurança  
biotecnologia

**It is part of this**  
bioprospecção

**Related terms**

**File**

**Relationships already filed to this word:**

**Theme:** Meio Ambiente  
**Sub Theme:** Biodiversidade  
**It is a type of biodiversidade**

soja transgênica  
 espécime de fauna  
 espécime de flora

**biodiversidade is part of**  
 Biodiversidade e Transgênicos

**Related terms of biodiversidade**  
 megadiversidade

Fig. 3. Editing Module - Ontology construction (insertion and consistency checking)

This application works with extractors of automatic standards in conjunction with knowledge engineers and domain specialists as according to specifications found in the methodology DCKR, which consists of a dynamic process of analysis of the general context involving the theme to be focused on. The Suite is an editor of ontologies structured in a form to allow a cooperative work on the Web between the Knowledge Engineering team and the specialists team.

This computational environment of shared access has two main objectives: organization and representation of knowledge, and updating of the knowledge base. It is basically composed by four modules, which are:

1. **Filing.** It allows to set up a contextualized dictionary, by selecting themes and sub-themes for the classification of indicative expressions. In this environment the user defines the theme and sub-theme under which new indicative expressions will be inserted. A domain can be categorized in various themes and sub-themes;

2. **Search.** It informs about other terms already filed on the base, which have some phonetic similarity with the term typed. This tool allows the verification of possible typing errors, besides preventing duplicated filing of the same term. It is a search system based on phonetic similarity. It supplies the user with a list of similar indicative expressions found in the knowledge base, in alphabetical order, when a query is typed by the user. The search module is used in the filing, edition and administration modules;

3. **Relationship Editor.** Allows the building of the relationship tree. This Module works with the similarity between all the terms filed and the ones already existing on the base. These relationships allow the system to expand the search context. The organization of the tree allows the dynamic definition of the weights of the indicative expressions according to the query of the user. The fields with all available relationships are presented. They are the following: -synonyms; -related terms; “this is a type of”; “it belongs to this type”; “this is a part of”; “it is part of this”. The editor presents the existing relationships and allows the inclusion of them (see figure 3). Each relationship has a weight related to the defined indicative expression in the query by the user.

4. Administration Environment. The knowledge integration and the validation between words are made in accordance with the context of themes and sub-themes. The environment is organized in three levels: High Level - allows inserting themes and sub-themes, to validate exclusions, to include and to exclude users, to check productivity of each user and to check descriptions of the dictionaries, themes, sub-themes and indicative expressions; - Medium level- allows checking productivity and historical data; and, Low level- allows checking descriptions.

The definition of related concepts implies research work or help from a knowledge specialist on the matter. They are terms that can be considered as synonyms of themes and secondary themes, as well as close to the application context. An identifiable limit does not exist for the number of related concepts. Therefore it is important to observe the application of the terms in real cases. The specialists are helped in this task by a technological structure.

The module of related concepts is used by the domain specialists. They can work in their office, and then the contents are integrated into the knowledge base through the knowledge acquisition module (see figure 3). In order to enable the specialists work, a methodology based on the Theory of Juridical Argumentation [2] and Extensive Interpretation is used.

All the concepts, linked each other, generate a semantic-like network. This network improves the system capacity to recognize concepts, independently of finding it or not in the text. The network is organized into levels, indicating the “distance” between two concepts. These levels are used later on in the similarity measure.

However, all this structure and methodology was not enough to turn the cooperative work efficient and effective. A more holistic approach was necessary, which allows a greater coherence between the relations of the expressions, mainly in the definition of the related terms where the participation of the specialist is almost exclusive. It is important to highlight that this structure of contextualized ontologies allows automatic information indexing by the system and a knowledge acquisition that gives more qualitative answers in the retrieval process.

### **3 Enabling the Synchronicity in a Collaborative Networked Organization**

The different unfolding of the human inventivity, although it is so diversified, has the same origin, the unconscious mind and the human perceptions. This is because distinct constructions eventually lead the mind to the same reference. Therefore we created a methodology that allows the immediate perception of the specialist to arise, without the pretension to reach all the knowledge, but with clear objectives, for example, to eliminate the common resistance of people to technological innovations, standing out the importance of management of human capital. [14].

During the development of tasks of Knowledge Engineering, it was observed that the efficiency of the acquisition process had a direct relation with good relationship between the knowledge engineer and the domain specialist, no matter what the quality or content of the interviews were, or the efficient application of the support tools. Thus, keeping this relationship in perfect synchrony is a key factor for the success of the system and a challenge for which the stages defined in the present work serve as a model of relative success.

Common sense tells us that immediate perception (intuition) has greater effectiveness on the best solution for a problem than the application of rules of the propositional logic. However, the most accepted proposal is people trying to solve deductive problems applying rules such as those of the propositional logic. According to Lacan [15], if we consider that the unconscious is structured as a language, it is possible to reconstruct the unconscious associations between the words, thus disclosing a context.

There are elements, like the cognitive complexity and the capacity to learn, that supply the underlying individual traces on which the specialized knowledge and abilities are based, and similarly, sociability and confidence supply the anchors to develop and to keep a net of relationships. Thus, identifying that non-cognitive knowledge is also important knowledge of the institutions and, for this reason, they must be part of the capital of these organizations, it is necessary to look for a way to identify it and to represent it in the knowledge based



systems. Therefore, this complex net of communications between the diverse areas of talent will provide the necessary flexibility, versatility and adaptability intelligences.

All the languages are structured as an articulating system. But their character and coherence is a unique articulated system. Thus the cognitive point of view concerning the symbolic acquisitions has as foundation the meanings generally supported by natural language or specialized languages such as the formal ones. To have these elementary meanings present in the work of a team requires synchronous thinking.

This synchronization process starts with the understanding of human intelligence, its unconscious manifestations and its relationship with words. Therefore, in accordance with Lacan [16], every human investigation is tied irreversibly in the interior of the space created by the language. But, for the success of this dynamics of cerebral exercise, it is essential the person to be in a positive attitude. The brain registers, learns and builds ramifications only when it is open to what is new.

### **3. Mind Engineering Methodology**

There are many different techniques of Knowledge Acquisition. We created Mind Engineering (see figure 3) to help developing the following process (DCKR methodology): (1) Inventory of the entire domain (classification of all sources of digital information that will be in the system database); (2) Application of the word frequency extractor based on the database inventoried; (3) Comparison between extractor results with the specialist needs; (4) Construction of a representative vocabulary of the domain by the specialist and knowledge engineers; (5) Application of the semantic extractor on the database using the representative vocabulary (indicative expressions); (6) Definition of a list of words based on the evaluation of the results of the frequency of the indicative expressions found in the inventory (7) Construction of the ontologies in the Knowledge Engineering Suite based on this controlled vocabulary (8) Definition of synonyms, related terms, homonyms, hyponyms, hypernyms and meronyms.

The acquisition of knowledge carried out by the team of Knowledge Engineers had a bigger effectiveness in the area of its specialization [5] [6] than the acquisition performed by the same team in domains different from its specialization [7], where some obstacle of communication caused the need of a new acquisition process to be implemented.

Not having synchronization problems, the deep knowledge of the specialists on the AI technique applied in the system modeling (e.g., Case-Based Reasoning) allowed the transference of knowledge into the computational language in a very positive way for the final target of the system.

It was observing the elements presented in the two processes that we were able to systematize a series of questions, improving the speed and quality of knowledge represented in the system.

Additionally, uncommon procedures of knowledge acquisition were adopted, such as neurolinguistics and meditation techniques, to defragment the emotional memory of the specialist and to facilitate the learning process (see figure 4). This happened due to the following problems: (1) Resistance against the system; (2) Difficulty to reproduce the process of decision-making; (3) Low quality of the knowledge handled.

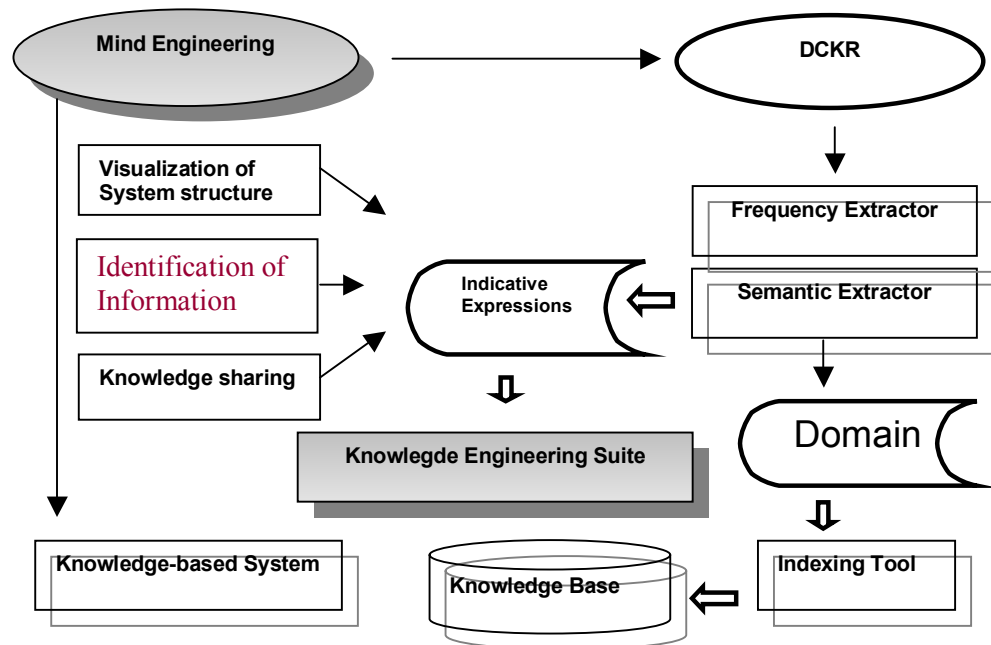


Fig. 4. Mind Engineering Methodology applied to the construction of ontologies in Knowledge-based Systems

However, the focus object is not the area of application of the system (domain), but the work of the specialist and the knowledge engineer to define the target of the system and create the knowledge base of this system. To identify and to classify knowledge levels is essential, therefore both (specialists and engineers) have to be trained on the learning process; that requires them to overcome the comfort zone. Knowledge Engineering is mostly a process of knowledge exchange.

The importance of existing knowledge for new acquisitions comes from the basic role they play inside the construction of the representation and from the idea given by that acquisition process to those representations. The importance of this phase is the exchange of knowledge; the specialist gets to know how his knowledge could be organized, that is, the basic concepts of the technique of Artificial Intelligence used in the representation of knowledge. Thus he will be able to contribute with more effectiveness and will have a greater interest in participating in the process. For the knowledge engineer, the exchange will lead to a more immediate perception of the target of the system and will increase the interest in going deeper in the study of the domain. Both will be prepared to deal with this overload and to obtain the ability necessary to plan or to choose a perspective that determines which elements of the situation must be treated as important elements and which can be ignored. By perceiving that the vast information or knowledge is reduced to a number of characteristics and relevant aspects, the decision making process becomes easier.

The continuous sharing of the established visions makes the specialists and engineers to work in better cooperation in the construction of the ontologies of the domain. This productive process is continuous and can lead to changes in the system implementation phase.

#### 4 Conclusion

The systematization and organization of domain specialist teams together with the team of Knowledge Engineering became a big challenge in the development of knowledge management systems. The cooperative work between the teams does not only require the deep knowledge on the application domain, but also on the organization of its knowledge base. The creation of a computational environment on the web allowed a greater sharing of information and better results of the teams in the construction of knowledge-based systems.

The Knowledge Engineering Suite enables a cooperative work among people in different places, structuring a continuous knowledge base and easy visualization (knowledge tree) through relationship networks and supplies an exceptional coherence among the semantic

relations of the indicative expressions, mainly by the support of all this computational structure during the process. This allowed the knowledge engineer and the specialist to develop much more than the knowledge of the domain, but abilities such as conscience itself, discipline, persistence, and empathy.

## References

1. Resende, Solange Oliveira. *Sistemas Inteligentes: fundamentos e aplicações*. Barueri, SP: Manole, 2003.
2. Duineveld, A. J. et al, 1999. WonderTools? A comparative study of ontological engineering tools. Twelfth Workshop on Knowledge Acquisition, Modeling and Management. Voyager Inn, Banff, Alberta, Canada.
3. Eriksson, H. et al, 1999. Automatic Generation of Ontology Editors. Twelfth Workshop on Knowledge Acquisition, Modeling and Management. Voyager Inn, Banff, Alberta, Canada.
4. Benjamins, V.R., 1998. The ontological engineering initiative (KA)<sup>2</sup>, Formal Ontology in Information systems. IOS Press, Amsterdam.
5. Hoeschl, Hugo. C. Bueno, Tania. C. D., Barcia, Ricardo. M., Bortolon, Andre., Mattos, Eduardo Da Silva. Olimpo: Contextual structured search you improve the representation council of UN security with information extraction methods In: *Artificial International conference on intelligence and law, 2001*, St. Louis. ICAIL 2001 Proceedings. New York: ACM SIGART, 2001, p.217 – 218.
6. Bueno, Tânia Cristina D'Agostini. *O Uso da Teoria Jurídica para Recuperação em Amplas Bases de Textos Jurídicos*. 1999. 94 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 1999.
7. Ribeiro, Marcelo Stopanovski. *KMAI, da RC<sup>2</sup>D à PCE. Gestão do conhecimento com inteligência artificial, da representação do conhecimento contextualizado dinamicamente à pesquisa contextual estruturada*. [2004]. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2003.
8. Bueno, Tânia C. D. et al, 1999. JurisConsulta: Retrieval in Jurisprudencial Text Bases using Juridical Terminology. Proceedings of the Seventh International Conference On Artificial Intelligence And Law. ACM, New York.
9. Hoeschl, Hugo. C. et al, 2003. Structured Contextual Search For The Un Security Council. Proceedings of the fifth International Conference On Enterprise Information Systems. Anger, France, v.2. p.100 – 107.
10. UNL. Universal Networking Language. Available at: <http://www.unl.ias.unu.edu/unlsys/index.html>. Access on: 19 jan. 2004.
11. WORDNET. Available at: <http://www.cogsci.princeton.edu/~wn/>. Access on: 19 jan. 2004.
12. Semantic Web. Available at: <http://www.w3.org/2001/sw/>. Access on: 19 jan. 2004.
13. Kolodner, J. *Case-Based Reasoning*. Morgan Kaufmann, Los High, CA. 1993.
14. Gratton, Lynda, Ghoshal, Sumantra. Managing Personal Capital Human: new ethos will be the "Volunteer" Employee, *The European Management Journal*, vol 21, n° 1 pp1-10, February, 2003.
15. Lacan, Jacques. *Os seminários de Lacan*. Disponível em CD Rom, 2000.
16. Miller Jacques-Alain, 1988. *Curso de Lacan: uma introdução*. Jorge Zahar Editor Ltda, 2a edição, Rio de Janeiro.

**Anexo II - Paper aprovado no IADIS – Internacional Conference  
www/internet 2004, Madrid, Espanha.**

**Engineering of Minds: The Synchronicity Between Artificial  
Intelligence and the Management of Knowledge in Collaborative  
Networked Organizations**

# ENGINEERING OF MINDS: THE SYNCHRONICITY BETWEEN ARTIFICIAL INTELLIGENCE AND THE MANAGEMENT OF KNOWLEDGE IN COLLABORATIVE NETWORKED ORGANIZATIONS

Tania C. D. Bueno, Hugo C. Hoeschl

*IJURIS – Instituto de Governo Eletrônico e Inteligência Jurídica  
Rua Lauro Linhares, 728, sala 105, Florianópolis BRASIL – CEP 88036-002  
{tania,hugo}ijuris.org*

Andre Bortolon, Ricardo M. Barcia

*UFSC – Universidade Federal de Santa Catarina \*  
Florianópolis-BRASIL \*  
{bortolon,rbarcia}@eps.ufsc.br*

## ABSTRACT

The present work is focused on the systematisation of a process of knowledge acquisition for its use in intelligent management systems. The result was the construction of a computational structure for use inside the institutions (Intranet) as well as out of them (Internet). This structure was called Knowledge Engineering Suite an ontological engineering tool to support the construction of ontologies in a collaborative setting and was based on observations made at Semantic Web, UNL (Universal Networking Language) and WorldNet. We use a knowledge representation technique called DCKR to organize knowledge and psychoanalytic studies, focused mainly on Lacan and his language theory to develop a methodology called Engineering of Minds to improve the synchronicity between knowledge engineers and specialist in a particular knowledge domain.

## KEYWORDS

Ontological Engineering Tool, knowledge representation, Artificial Intelligence, Ontology, Lacan, Intelligent Management Systems.

## INTRODUCTION

The importance of the Knowledge Based Systems is in the fact that they provide the computer with some peculiar characteristics of human intelligence, such as the capacity to understand natural language and simulate reasoning in uncertainty conditions. Defining the relevant information to be inserted into a Knowledge Based Systems is the great problem in the development of intelligent systems, mainly because the process is basically experimental and depends greatly on the ability of the knowledge engineer. In particular, a great difficulty is related to the definition of the terminology used to nominate the concepts and the relations (Rezende, 2003). In the knowledge acquisition, there are a great number of methods, but none of them deals with the understanding and learning of the people involved, either the specialists, or the knowledge engineers.

More recently, the notion of an ontology is being so popular in fields such intelligent information integration, information retrieval on the Internet, and knowledge management. The reason is in part due to what they promise: a shared and common understanding of some domain that can be communicated across people and computers (Duineveld et al, 1999). Different developments of a worldwide range have a reference in cooperative work as a WordNet, Semantic Web and UNL (Universal Networking Language) through the construction of ontologies using collaborative settings. The use of ontological engineering tools or metatools to support the knowledge engineering process allows the organization of a knowledge base established on the relationship between relevant expressions of a context. The use of ontologies as a basis for automatic generation of knowledge acquisition tools simplifies the tool-specification process by taking advantage of ontologies defined as part of the knowledge engineering process (Eriksson et al, 1999). Nevertheless, experience shows that often the bottleneck of building sharable ontologies lies more in the social process than in the technology (Benjamins, 1998). For this reason, we develop a methodology to the process of knowledge acquisition to allow the specialist and the knowledge engineer to work in synchronicity, in cooperative networked organizations. We call this methodology engineering of minds. This synchronization process initiates with the understanding of human intelligence, its unconscious manifestations and its relationship with the words, since, in accordance with Lacan, every human investigation is linked irreversibly in the inner space created by language. In the present

development, we create a tool to support the knowledge engineering process by assisting developers in the design and implementation of ontologies in a specific domain.

In earlier works, we use a methodology called DCKR (Dynamically Contextualized Knowledge Representation (Hoeschl, 2001). DCKR allows the construction of a knowledge base, improving and the construction of the domain ontology process, and the automatic representation of cases in knowledge-based systems, either in the legal area (Bueno et al, 2003), or in the area of knowledge management (Ribeiro, 2003).

In the next section, the methodology for the knowledge synchronization is described. This methodology this work allowed an exceptional coherence among the semantic relations of what is called 'indicative expressions', mainly by the support of all this computational structure during the process. This allowed the knowledge engineer and the specialist to develop much more the knowledge representation of the domain, but abilities as inherent conscience, discipline, persistence and empathy.

## ENGINEERING OF MINDS

During the development of works of knowledge engineering, it was observed that the efficiency of the knowledge acquisition process had a direct relation with the good relationship between the knowledge engineer and the domain specialists. The quality or the content of the interviews, as well as the efficiency of the support tools, had been secondary factors in the final result. Then, in the process of knowledge engineering, it was noticed that the abilities of relating to others and interpersonal communication were essential and that they had an influence in the performance of the knowledge based system.

The Knowledge Engineering Suite was developed for its use with the methodology of knowledge representation called Dynamically Contextualized Knowledge Representation - DCKR (Hoeschl, 2001). Tools such as the frequency Extractor, Semantic Extractor and the Engineering Suite of the knowledge to assist in the task.

There are many different techniques of Knowledge Acquisition. We created the Engineering of Minds (see figure 1 in the item 2.1) to help developing the following process: 1. Inventory of the entire domain, classifying all the sources of digital information that will be the system database. 2. Application of the word frequency extractor based on the database inventoried; 3. Comparison between extractor results with the specialist's needs. 4. Construction of a representative vocabulary of the domain, by the specialist and knowledge engineers. 5. Application of the semantic extractor on the database; using the representative vocabulary (indicative expressions). 6. Definition of a list of words based on evaluation of the result of the frequency of the indicative expressions found in the inventory. 7. Construction of the ontologies in the Knowledge Engineering Suite based on this controlled vocabulary. 8. Definition of synonyms, homonyms and hypernyms.

### 2.1 The Unconscious and the Word

The *Common-sense* tells us that the immediate perception (intuition) has greater effectiveness on the best solution for a problem than the application of rules of the propositional logic. Although the most accepted proposal is that people try to solve deductive problems applying rules such as of the propositional logic. According to Lacan, if we consider that the unconscious is structured as a language, it is possible to reconstruct the unconscious associations between the words, thus disclosing, a context.

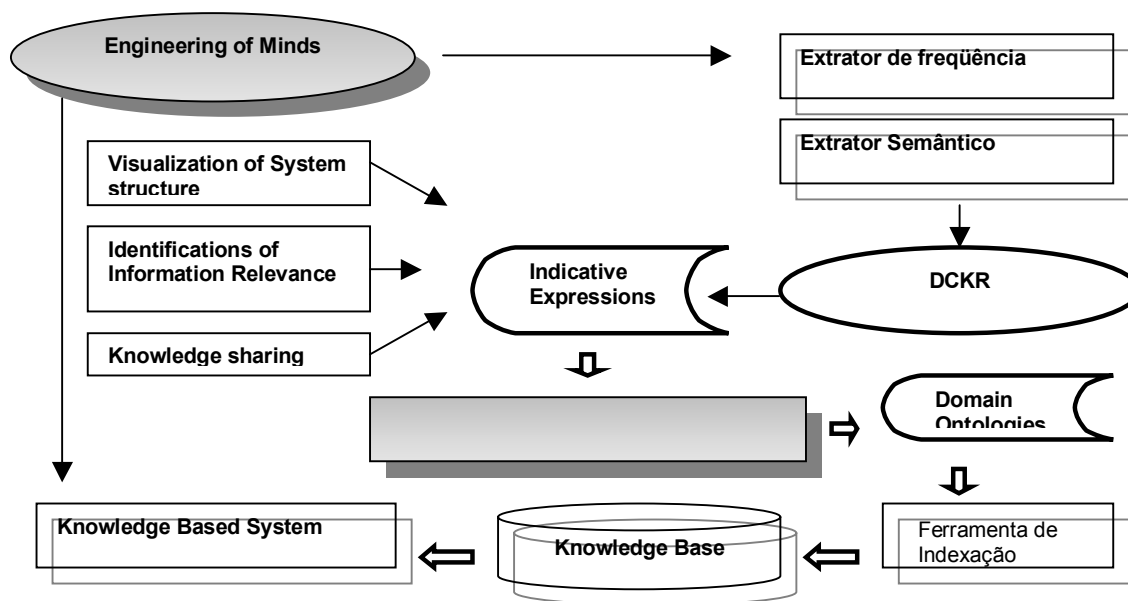


Figure 1. Engineering of Minds Methodology

All the languages are structurable as an articulating system. But their character, their coherence is in an articulation system, which is unique. Thus, the cognitive point of view concerns the symbolic acquisitions, those that have as a foundation the meanings whose support is, generally, natural language, or, at times, specialized languages, as the formal or the technical ones. The attainment of these elementary meanings in a wide team requires synchronous thoughts. Logics prove what intuition discovered. It is a metrics of the unconscious.

This synchronization process initiates with the understanding of human intelligence, its unconscious manifestations and its relationship with the words, since, in accordance with Lacan (Miller, 1988), every human investigation is linked irreversibly in the inner space created by language. The different outspread ways of human creativity, even being so very diversified, have the same origin, human mind and unconscious perceptions; deriving from this is the fact that distinct constructions eventually lead to thoughts to the same reference. Therefore, it will be this complex net of communication between the diverse areas of talent that will to supply the necessary flexibility, versatility and adaptability so that intelligences, including emotional intelligence, may happen. But, for the victory of this dynamics of cerebral gymnastics, it is primordial that the person is in a positive attitude. The brain only registers, learns and ramifies when it is open to what is new.

## KNOWLEDGE ENGINEERING SUITE

The Suite is an editor of ontologies structured in a way to develop a cooperative work on the Web between the knowledge engineering team and the specialists' team, allowing the relationship of complex terms, considering its concepts in the specific domain of the application. This Computational environment of shared access has two main objectives: organization and representation of the knowledge and update of the Knowledge Base.

Four modules compose it basically, they are: **1.Register**. It allows the elaboration of a contextualized dictionary, for the selection of topics and sub-topics for the classification of the indicative expressions. In this environment the user defines the topic and sub-topic in which it will insert a new indicative expression. A domain can be categorized in innumerable topics and sub-topics; **2. Search**. It informs about other terms already registered in the base, which have some phonetic similarity with the term typed. This tool allows the verification of possible typing errors, besides preventing the registration of the same term more than once. It is a search system based on similarity. It supplies the user with a list of similar indicative expressions present in the knowledge base in alphabetical order after consultation made by the user. It is used in registers, in the edition and the administration and the module of the ontologies. **3. Editor of relationship**. The relationships tree starts being formed through this register, always considering the similarity between all the terms registered and the ones already existing in the base. These relationships allow the system to expand the search context. The organization of the tree allows the dynamic definition of the loads of the indicative expressions according to the entrance of the user. The fields with all the relationships available to be formed are presented. They are the following: *-synonyms; -Connected terms; -This is type of; - It is a type of this; - This is part of; - It is part of this.*

The editor presents the registered relationships and allows excluding them. Each relationship possessed one load related to the defined indicative expression in the search by the user. **4. Administration Environment.** The knowledge integration and the validation between words is made in accordance with the context the topics and sub-topics. This topic is organized in three levels: - High Level - it allows to insert topics and sub-topics, to validate exclusions, to include and to exclude users, to verify productivity of each user and to verify descriptions of the dictionaries, topics, sub-topics and indicative expressions; - Medium level- it allows to verify productivity and historical data; and, Low level- it allows to verify descriptions.

## CONCLUSION

The creation of a computational environment on the Web enabled a greater sharing of information and better results among the teams in the construction of knowledge based systems. The Knowledge Engineering Suite was developed to enable a cooperative work among people in different places, structuring a continuous knowledge base and easy visualization (knowledge tree), through relationship nets. The methodology for the knowledge synchronization described in this work allowed an exceptional coherence among the semantic relations of what is called 'indicative expressions', mainly by the support of all this computational structure during the process.

## ACKNOWLEDGEMENT

Our thanks to researches and students of the Juridical Institute of Intelligence and Systems - IJURIS: Cristina Souza Santos and Aline Junckes.

## REFERENCES

- Benjamins, V.R., 1998. The ontological engineering initiative (KA)<sup>2</sup>, *Formal Ontology in Information systems*. IOS Press, Amsterdam.
- Bueno, Tânia C. D. et al, 1999. JurisConsulta: Retrieval in Jurisprudencial Text Bases using Juridical Terminology. *Proceedings of the Seventh International Conference On Artificial Intelligence And Law*. ACM, New York.
- Hoeschl, Hugo. C. et al, 2003. Structured Contextual Search For The Un Security Council. *Proceedings of the fifth International Conference On Enterprise Information Systems*. Anger, France, v.2. p.100 - 107
- Duineveld, A. J. et al, 1999. WonderTools? A comparative study of ontological engineering tools. *Twelfth Workshop on Knowledge Acquisition, Modeling and Management*. Voyager Inn, Banff, Alberta, Canada.
- Eriksson, H. et al, 1999. Automatic Generation of Ontology Editors. *Twelfth Workshop on Knowledge Acquisition, Modeling and Management*. Voyager Inn, Banff, Alberta, Canada.
- Miller Jacques-Alain, 1988. *Percurso de Lacan: uma introdução*. Jorge Zahar Editor Ltda, 2ª edição, Rio de Janeiro.
- Rezende, S., 2003. *Sistemas Inteligentes: fundamentos e aplicações*. Manole, Barueri, SP.
- Ribeiro, Marcelo S., 2003. KMAI, da RC<sup>2</sup>D à PCE. Gestão do conhecimento com inteligência artificial, da representação do conhecimento contextualizado dinamicamente à pesquisa contextual estruturada. *Dissertação (Mestrado em Engenharia de Produção)* – Universidade Federal de Santa Catarina, Florianópolis.



**Anexo III - Paper aprovado na ICAIL – International Conference on Artificial Intelligence and Law 2003, Edimburgh, Scotland**

**Analyzing the use of Dynamic Weights in Legal Case Based System**

# Analyzing the use of Dynamic Weights in Legal Case Based System

Tania C. D. Bueno, Andre Bortolon  
Federal University of Santa Catarina  
Florianópolis, SC, Brazil  
tel. +55 48 3025-6702  
{bueno, bortolon}@eps.ufsc.br

Hugo C. Hoeschl  
Ijuris  
Florianópolis, SC, Brazil  
tel. +55 48 3025-6609  
digesto@digesto.net

Eduardo S. Mattos, Marcelo S.  
Ribeiro  
Web Intelligence Systems S/A.  
Florianópolis, SC, Brazil  
Tel. +55 48 234-5434  
{mattos, marcelo}@wbsa.com.br

## ABSTRACT

This paper presents a case-based system to retrieve legal statements of Brazilian High Courts that emphasizes the use of theoretical structures to represent cases. AlphaThemis system has the characteristics from other case-based systems. The major innovation is to allow the user modify the weights used to determine the importance of each feature in similarity measure. This innovation is called dynamic weights. Their objective is to improve the retrieval accuracy in systems that have a human expert supporting the case base creation. In this paper, we analyze the recall and usefulness of dynamic weights, suggesting a new calibration of attribute relevance to achieve their total efficiency.

### Categories and Subject Descriptors

Legal Knowledge-Based Systems: Case based Reasoning, text mining, and knowledge representation.

### General Terms

Documentation, Performance, Design, Standardization, Languages, Theory, Legal Aspects and Verification.

### Keywords

Dynamic Weights, Case Based Reasoning, knowledge representation, case law analysis.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '00, Month 1-2, 2000, City, State.

Copyright 2000 ACM 1-58113-000-0/00/0000. \$5.00.

## 1. INTRODUCTION

It is common in CBR systems the use of weights to determine the importance of each feature in similarity measure [1] [9]. The AlphaThemis system works in a similar way to other case-based systems [2] [5] [8], but the user can modify the weights used to determine the importance of each feature in similarity measure. These weights improve the retrieval accuracy in systems that have a human expert supporting the case base creation. One of the proposals of this paper is to present the use of weights calibrated by the user when s/he is doing the search. This hypothesis is possible since AlphaThemis' users have enough knowledge on the domain, making them able to choose which attribute has more importance in the context. This AlphaThemis' feature is called dynamic weights. Each dynamic weight is linked to one of the attributes available in the user interface. Users can set the value between 0 and 100. This value represents the importance the respective attribute has in this specific query. In addition, the values are used in the normalization process.

In AlphaThemis design, the Dynamically Contextualized Knowledge Representation (DCKR) [6] methodology enables, on

the developed application, characteristics of the human analogical reasoning, as a knowledge engineering technique to represent and retrieve the legal texts. To perform the information extraction task and the addition of new documents in the case base, we developed a Knowledge Engineering Module, facilitating the description of concepts and attributes of the cases in a semi-automatic manner.

## 1.1 Application Area

The Brazilian Courts, when necessary, issues *Súmulas*, texts with around three lines, to consolidate juridical interpretations about some legal decisions in disconformities. However, the language of the documents is very formal and not systematized, becoming its application unusual. Despite the availability of those documents on Internet and commercial systems, the access difficulty remains. Nevertheless, the development of efficient and effective system to retrieve the adequate *Súmula* really yields significant benefits for the solution of legal problems.

The *Súmula* is a short document, but the same statement can be applied to resolve several law cases. The case base of the AlphaThemis system is constituted by *súmulas* of all Brazilian High Courts (STF, STJ, TRF1, TRF2, TRF3, TRF4 e TRF5)<sup>15</sup>, totaling 1037 documents.

In the next section, related works are compared. The case representation and the knowledge acquisition module are described in Section 3. Section 4 has the system description itself. In the Section 5 is described the recall test. The system evaluation is shown in section 6. Future Works appear in Section 7, finishing with the Conclusions.

## 2. RELATED WORK

*JurisConsulta* [2] and *Olimpo* [5] are case-based approaches to retrieve information from textual documents, demonstrating how a theoretical-juridical structure enables the automatic extraction of document's indexes. In *JurisConsulta*, a controlled vocabulary and a dictionary of normative terms support the knowledge representation. *Olimpo* uses Structured Contextual Search (SCS), a generic methodology that emphasizes the use of information retrieval methods combined with Case-Based Reasoning (CBR). In that methodology, a list of relevant expressions is subject to statistic analysis in order to improve the automatic case representation.

## 3. KNOWLEDGE REPRESENTATION

A structured representation maps the textual description of the document, supporting the knowledge representation represented in the module of Knowledge Engineering. (See figure 1).

<sup>15</sup> Available on Internet ([www.stf.gov.br](http://www.stf.gov.br), [www.stj.gov.br](http://www.stj.gov.br); [www.cjf.gov.br](http://www.cjf.gov.br)).

**Número**

**Tribunal**

**Data**

**Texto**

A restituição da importância adiantada, à conta de contrato de câmbio, independe de ter sido a antecipação efetuada nos quinze dias anteriores ao requerimento da concordata.

**Referências**

Referências Normativas:

Lei n. 4.728/65;  
Decreto-Lei n. 7.661/45;  
Súmula 495 do STF;  
Súmula 36 do STJ;

Fontes:

RSTJ VOL.:00080 PG:00017

**Características**

**Termos Conexos**

Antecipação efetuada; Importância adiantada; adiantamento; Falências; quebra; massa falida; crime falimentar; empresa comercial; sociedade comercial; Lei

**Ramos do Direito**

Direito Civil; Direito Falimentar; Direito Comercial; Direito Financeiro;

**Indicador Temático Central**

Restituição em Contrato de Câmbio;

**Indicador Temático Secundário**

Requerimento de Concordata; Antecipação de Valor em Concordata; Restituição de Importância Adiantada;

Figure 1. Knowledge Acquisition Module

### 3.1 Legal Domain Knowledge

In the present system, we use the extensive interpretation as the technique to improve the knowledge representation process of the law statements. This interpretation technique extends the range of the purely literal terms of the norm, enclosing cases that, although possible in their context, are found outside of their verbal expression. Several criteria were elaborated to justify the distinction between analogy and extensive interpretation. We believe that the only acceptable criterion is one that looks for the difference regarding to the several effects of the analogical extension and the extensive interpretation: the effect of the former is the creation of a new juridical norm; the effect of latter is the application of a norm in unforeseen cases.

It is true that the law rarely provide models, which are so clear-cut and formally refine that they can be directly transferred in to a computer representation, but the importance of the fact that the law can provide workable solutions to coordination problems, tested both in the doctrinal dialectics and in the legal practice, could hardly be underestimated [4]. Thus, the knowledge representation in the AlphaThemis system is a semi-automatic process (see figure 2), in which specialists worked on an extensive interpretation of the *súmulas* content structuring the dictionary of terms. This process allows the determination of the local similarity among values of an index.

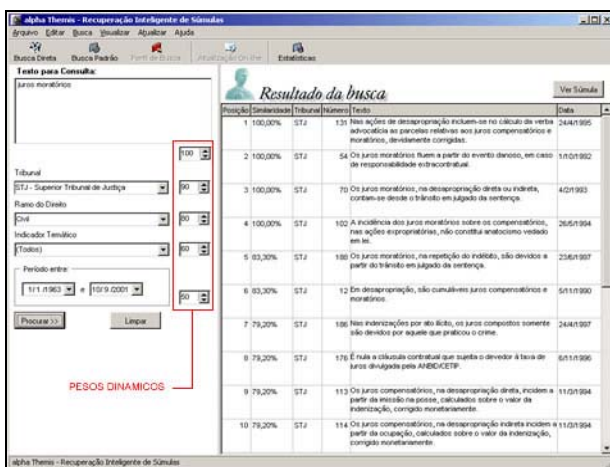


Figure 2. AlphaThemis® interface highlighting the use of dynamic weights

### 3.2 Case Representation

AlphaThemis retrieves *súmulas* allowing the user to formulate a query mixing fixed attributes (parameters) and free text. The present system was developed using as base the CBR methodology that allows the representation of a juridical text in case form, through the use of indexes. Each case is structured as a set of attribute-value pairs, as shown in table 1. Also, it was defined indexes, representing characteristics of the cases that indicate their usefulness in a specific situation. The group of the attributes and indexes were determined by law domain and Artificial Intelligence specialists.

Therefore, a set of indexes has been determined to facilitate the retrieval process (see table 1 and 2), which consist in form of nine attribute-value pairs: Court, *Súmula* Number, Concepts, Date, Law Area, Theme, Secondary Theme, Related Concepts and Normative References.

### 3.3 Information Extraction

The analysis of the document's structure and analysis is done through a knowledge engineering process.

The system comprises a methodology and a technological infrastructure for the integration of *Súmulas* in the database, as shown in figure 1. The case attributes are manually instantiated through the data extraction from the *súmula* texts and values provided by the knowledge specialist.

The *súmulas* are semi-structured texts. They follow a format defined by the law. Some relevant information can be identified through their location in the document. For instance, in STJ – Superior Tribunal de Justiça, the publication date is always described preceded for the keyword "DJ DATA".

For the information that demands theoretical support to be obtained and is not explicitly stated in the document, the values need to be inferred. For instance, using their experience and knowledge, the domain specialists identify which Law Area the matter of the *súmula* belongs, even though defining the attribute Law Area. So, we can define two extraction forms: by location and by inference.

#### 3.3.1 Extraction by Location

In the AlphaThemis system, some values can be identified through the position of the keyword in the text. All the represented information in the attributes Court, *Súmula* Number, Concepts and Date (see Table 1) can be extracted by their location in the text. They are called direct values.

Indirect values are those that can be identified in a certain location. They are identified according to the text structure, but such values need a grammatical identification. This methodology is used to find the theme and secondary theme attributes. For instance, the value for the Theme Attribute in the most of the cases is the subject of the phrase.

### 3.3.2 Extraction by Inference

In the knowledge acquisition of the AlphaThemis system, bibliographical sources were used to help in the task. In the example below, the domain specialist counting on his own professional experience defines the law area attribute. If the specialist has any difficulty to identify the Law Area, he can look it up in bibliographical sources, available on Internet or in the indexed books. There are two attributes identified using this methodology:

**Table 1. Description of the Indexes and respective weights**

Attributes	Descriptions	Weights
1- Court	The name of the responsible Court for the elaboration of the decision.	0.5
2- <i>Súmula</i> Number	The number supplied to order the legal decisions for the Court. Used only for identifying the statement.	No weight
3- Concepts	It is the body of <i>súmula</i> . Text with approximately 3 lines, constituted, usually, by only one sentence. In this body the relevant information are identified.	1 by concept found
4- Date - DJ	It is the publication date in the Official Diary.	0.5
5- Law Area	It is the dogmatic subdivision of the Brazilian Right (e.g., Penal Right, Constitutional Right, Civil Rights, etc.). The totality of the branches in the law domain is around 40.	1
6- Theme	It indicates the main subject of the <i>súmula</i> . In the recovery, the index best defines the research context.	1
7 -Secondary Theme	This attribute includes secondary subjects that, in some way, were presented in the considerations involving the decision. It can have multiple values and, according to the importance of the subjects treated, can be as useful as the central theme in the retrieval process.	0.9
8- Related Concepts	They indicate relative subjects approached in the <i>súmula</i> . A <i>súmula</i> can be indexed by n values of related concepts, according to your importance and inclusion. Consequently, this factor increases the possibility of the <i>súmula</i> to be recovered, although with a smaller similarity percentile.	0.7 by related concept found
9 -Normative References	It indicates norms, codes and other relative sources to the subject of the <i>súmula</i> . Indicates the precedents that gave origin to the <i>súmula</i>	No weight

**Law Area.** The specialist observes that the matter in the STF's *súmula* 585, belongs to tributary area. Therefore, the subject is classified as Law Area of the TRIBUTARY LAW. However, it also involves subjects of international law. So, the subject is classified as belonging to the INTERNATIONAL LAW.

**Related Concepts.** The definition of the related concepts implies in research or great knowledge specialist on the matter. They are juridical terms that can be considered as synonyms of themes and secondary themes, as well as close to the *súmula* application context. For instance, one *súmula* that disposes about Paternity Suit has as related themes: filiation recognition and DNA exam.

An identifiable limit doesn't exist for this attribute. Then it is important to observe the application of the terms in concrete cases. The specialists are helped in this task by a technological structure. The module of related concepts is for the use of domain specialists. They can work in their office, and then the contents are integrated to the knowledge base through the knowledge acquisition module (see figure 1). In order to enable the specialists' work, a methodology based on the Theory of Juridical Argumentation [2] and Extensive Interpretation is used.

All the concepts, linked each other, generate a semantic-like network. This network improves the AlphaThemis capacity to recognize concepts independently of location or not in text. The network is organized by levels, indicating the "distance" between two concepts. These levels are used later in the similarity measure.

## 3.4 Dynamic Weights

The Dynamic Weights can change the relevance of the search parameters, turning the search results more similar to the user's request. Each parameter weight can be calibrated on a 0 to 100 scale (see figure 2). For example, the theme weight is fixed on 100, the Court weight on 50 and the other parameters with 0. The most similar cases retrieved by the system will have the Court and the Theme fixed by the user, but not necessarily; if in the case base an adequate case cannot be found the system returns records related to other Courts documents with that fixed theme.

## 4. ALPHATHEMIS SYSTEM

AlphaThemis retrieves *súmulas* based on a formulated query mixing fixed attributes (parameters) and free text. Using a similarity measure and according to certain information required by the user, the system performs a classification of the documents in the database. All the documents are searched and there is no threshold. In the interface, user can select some parameters: Court, Law Area, Theme, and Date. Also, he can set the dynamic weights, providing to the system what are the parameters that have more importance in this specific query (see figure 3).

The Secondary Theme and Related Concepts are available for similarity calculation, but not as parameters in the search context. The concepts are extracted from the text of legal documents and the text formulated by the user. The extracted concepts, as well from the situation presented by the user as from the document, are used to compute a similarity value.

### 4.1 Retrieval of *súmulas*

The system gets the user's query and generates another document having the same attributes of a *súmula* from the case base. This generated case is compared with the entire case base, providing the answer to the user ranked by similarity. *Súmulas* that have the same similarity are untied according to the date.

$$sim(S_a, S_b) = \frac{sim_c(S_a, S_b) + sim_d(S_a, S_b) + sim_{cr}(S_a, S_b) + sim_a(S_a, S_b) + sim_t(S_a, S_b)}{n_c(S_a) * p_c + p_d + p_{cr} + p_a + p_t}$$

Figure 3. Similarity measure formula

#### 4.1.1 Similarity measure

The global similarity measure includes all attributes, even when the user doesn't combine search parameters (e.g. courts, law area, etc.) with the textual query. The global similarity measure is calculated using the formula showed in figure 3. There,  $sim_c$  is the concept similarity;  $sim_d$ , the date similarity;  $sim_{cr}$ , the court similarity;  $sim_a$ , the law area similarity;  $sim_t$ , the theme similarity;  $n_c$ , the number of concepts of  $S_a$ ;  $p_c$ , the concept dynamic weight;  $p_d$ , the date dynamic weight;  $p_{cr}$ , the court dynamic weight;  $p_a$ , the law area dynamic weight; and  $p_t$ , the theme dynamic weight.

Each local similarity has an own formula. The concept similarity  $sim_c(S_a, S_b)$  is calculated by:

$$sim_c(S_a, S_b) = \sum_{i=1}^n (sim_l(c_{ia}, c_{ib}) * p_c), \text{ where } c_{ia} \text{ is the } i\text{th}$$

concept from  $S_a$ ,  $c_{ib}$ , the  $i$ th concept from  $S_b$ ,  $sim_l(c_{ia}, c_{ib})$ . The local similarity between  $c_{ia}$ ,  $c_{ib}$ , and  $p_c$  the concept dynamic weight.  $sim_l(c_{ia}, c_{ib})$  is calculated by:

$$sim_l(c_{ia}, c_{ib}) = \begin{cases} 1 & \text{if } c_{ia} = c_{ib} \\ x & \text{if } \exists r \in Rc \quad c_{ia} = r \\ 0 & \text{if } c_{ia} \neq c_{ib} \wedge \forall r \in Rc \quad c_{ia} \neq r \end{cases}$$

$Rc$  represents the related concepts set and  $r$  represents one related concept. If any related concept from a *súmula* is found, the similarity value is  $x$ , which represents the relation between the concepts.

The date, court, and law area similarities are calculated through the same way. If both *súmula* data are the same, the similarity is 1 multiplied by the respective dynamic weight. Otherwise, the similarity is 0.

The theme similarity is calculated by:

$$sim_t(S_a, S_b) = \begin{cases} 1 * p_t & \text{if } t_{S_a} = t_{S_b} \\ y * p_t & \text{if } t_{S_a} = st_{S_b} \\ 0 & \text{if } t_{S_a} \neq t_{S_b} \wedge t_{S_a} \neq st_{S_b} \end{cases},$$

where,  $t$  is the primary theme,  $st$  is the secondary theme,  $p_t$  the theme dynamic weight, and  $y$  is the value representing the relation between the secondary theme and the primary theme.

## 5. RECALL TEST

The adopted procedure to do the recall test in AlphaThemis was to invite some users who are experts in the domain, but work in different areas, to make some questions. It was chosen experts in *súmulas* because lawyers that don't work with them could make questions on subjects that don't have a *súmula* on.

For each question, it was elaborated a complete query, with a law area and a theme. Then, the query was applied without dynamic weights and with dynamic weights highlighting the text, law area, or theme (each one separated). Each user made an own evaluation according to their specific work area.

Besides the complete query, it was selected from the case base the set of documents that answer the question. Since AlphaThemis measures the similarity with all *súmulas* and, due to conceptual premises, any *súmula* that has similarity superior to 0 would be returned, the recall test has a particular aspect. Only the set of *súmulas* that contains the ones who were the query's answer were considered to calculate the recall. For instance, supposing that a question has 20 *súmulas* as its answer. But, AlphaThemis returned

the 20th *súmula* at the 23rd position, although the system returned around 80 *súmulas* in total due to other attributes. Therefore, it was only considered the set of 23 *súmulas* to the recall test. This characteristic became the recall test very similar to the precision test.

Although the recall almost has never been 100%, always the document returned in both first and second position was an answer to the question.

## 6. EVALUATION

It is a characteristic of AlphaThemis to provide a better recall as the number of words of the query increases. The association with other attributes helps the system to better determine the scope and the context to be searched. Our preliminary study had identified that the weights could help specifying the context and could also give a meaningful improvement to the recall.

Evaluating the results, it was observed that the huge amount of words used as a query in AlphaThemis (it is possible to use more than 200) became the influence of dynamic weights almost irrelevant, since the weight of other attributes is minimized when there is more than the double of words than the number of attributes selected. The recall increased from 70% to 85% when were used few words and dynamic weights. When more words were used together with dynamic weights, the recall just was improved from 85% to 88%. Despite of these values, it is relevant to say that in 100% of the cases, the first document returned really is a solution to the user's problem.

It is necessary to make clear that there are two different types of query in AlphaThemis. The first one only has words, without any other attribute. The second one is more contextualized and uses the attributes. The dynamic weights only belong to this latter.

The usage of dynamic weights in AlphaThemis brought results that were different from the expected ones. According to preliminary studies, it was supposed that the recall should be increased by around 20%. However, the implementation just got an improvement of 5%. As this objective was not accomplished, the group is supposing to review the relevance of each attribute in the similarity measure. The main hypothesis is to restrict the maximum weight of the query's text to 50%.

## 7. FUTURE WORKS

Since the tests haven't achieved the expected result and according to the study done after them, the group decided to review the similarity measure, aiming at a total efficiency of dynamic weights.

The proposal is to divide the attributes in two groups. First group contains the fixed attributes, while the second has the variable ones. The new similarity measure will have the following principles:

1. the sum of the fixed attributes (court, date, law area, theme and secondary theme) corresponds to 50% of the total. This value keeps valid even if the user does not select one or more of the attributes. If none of them is selected, the value is 0. The normalization is just done within this group.
2. the sum of variable attributes (concepts and Related Concepts) is equal to the other 50%. It must exist at least one concept, and the local similarities are calculated as now.

For instance, with the current similarity measure, if there are four similar concepts in two *súmulas*, from a group of seven, and both have the same law area and theme, the similarity is 66.67%. With the new measure, the similarity is 78.57%.

This division in two groups with the same weight intends to avoid the excessive weight given to the text when the user types more the double of words than there are other attributes.

## 8. CONCLUSION

The adopted techniques, highlighting Case-Based Reasoning, provide a powerful tool to retrieve legal documents, due to its similar characteristics to legal reasoning. Legal knowledge engineering provides knowledge cumulated by the specialist through the technological development environment. All these techniques together get an improvement with the use of dynamic weights.

The similarity measure using two attribute groups balances the powerful AlphaThemis' capacity to manage high amounts of words and that knowledge provided by the specialist.

AlphaThemis represents an innovative solution since it is a text mining and search tool tailored to work with juridical knowledge, but with a support technology able to be applied to any other field.

## 9. ACKNOWLEDGMENTS

Our thanks to lawyers and Law students of the Juridical Institute of Intelligence and Systems - IJURIS: Cristina Souza Santos, Aline Junckes and Thais Helena B. Garcia, law students who has been working in knowledge acquisition module. Filipe Costa and Erica Ribeiro, Lawyers who helped us in the system evaluation.

## 10. REFERENCES

- [1] Amond, A., Plaza, E. Case-Based Reasoning: Fundamental Issues, Methodological Variations, and System Approaches. *AI Communications*, 17(1), 1994.
- [2] Bueno, Tânia Cristina D'Agostini. O uso da teoria jurídica para Recuperação em Amplas Bases de Textos Jurídicos. *Dissertação de Mestrado, PPGE/UFSC. Florianópolis: 1999.*
- [3] Bueno, Tania Cristina D'Agostini, Hoeschl, Hugo Cesar, Mattos, Eduardo da Silva, Barcia, Ricardo Miranda, Wangenheim, Christiane Gresse Von. *JurisConsulta: Retrieval in Jurisprudencial Text Bases using Juridical Terminology*. In: *The Seventh International Conference On Artificial Intelligence And Law, 1999, Oslo*. Proceedings of the Conference. New York: ACM, 1999. v.1. p.147-155.
- [4] Conte, R., Falcone, R., Sartor, G. *Introduction: Agents and Norms: How to fill the gap? Artificial Intelligence and Law*. Vol. 7, nº 1, march 1999. Kluwer Academic Publishers. Netherlands, 1999.
- [5] Hoeschl, Hugo Cesar, Barcia, Ricardo Miranda, Bueno, Tânia Cristina D'Agostini, Mattos, Eduardo da Silva, Bortolon, André, Donatti, Fabrício Tadeu. *Sistema Olimpo*. Florianópolis, 2000. Software registrado.
- [6] Hoeschl, Hugo Cesar, Bueno, Tânia Cristina D'Agostini, Mattos, Eduardo da Silva, Bortolon, André, Barcia, Ricardo Miranda. *Olimpo: Contextual Structured Search to improve the representation of UN Security Council Resolutions with information extraction methods*. In: *The 8th International Conference On Artificial Intelligence And Law, 2001, St.Louis, MO, USA*. Proceedings of the Conference. New York: ACM, 2001.
- [7] HOESCHL, H. C., BUENO, T. C. D., BORTOLON, A., MATTOS, E. S., RIBEIRO, M. S. *AlphaThemis - from Text into Knowledge* In: *1st Workshop on Automatic Deduction and Artificial Intelligence (IDEIA), in the 8th Iberoamerican Conference on Artificial Intelligence (IBERAMIA), 2002, Sevilha*. Proceedings of the IDEIA. , 2002. v.1. p.91 - 100
- [8] Lenz, M.; Burkhard, H.-D. *CBR for Document Retrieval: The FAI/Q Project*. *Lecture Notes in Artificial Intelligence: 2nd Int. Conference on CBR, ICCBR97*. D. Leake, E. Plaza (ed.)-Berlin: Springer Verlag, 1997.
- [9] Wheeler, Richard. *I KNEW THAT: An Introduction to Case-based Reasoning*. *Artificial Intelligence Applications Institute (AIAl)*, University of Edinburgh. <http://users.pandora.be/richard.wheeler1/pdfs/CBR1.pdf>