

EUNICE PALMEIRA DA SILVA

**CLASSIFICAÇÃO DE INFORMAÇÃO USANDO  
ONTOLOGIAS**

Maceió  
Setembro de 2006

UNIVERSIDADE FEDERAL DE ALAGOAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM  
COMPUTACIONAL DO CONHECIMENTO

## CLASSIFICAÇÃO DE INFORMAÇÃO USANDO ONTOLOGIAS

Dissertação apresentada ao Curso de Pós-Graduação em Modelagem Computacional do Conhecimento da Universidade Federal de Alagoas como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional do Conhecimento.

EUNICE PALMEIRA DA SILVA

Maceió  
Setembro de 2006

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
**Bibliotecária Responsável: Helena Cristina Pimentel do Vale**

S586c Silva, Eunice Palmeira da.  
Classificação de informação usando ontologias / Eunice Palmeira da Silva.  
– Maceió, 2006.  
xi, 116f. : il.

Orientador: Luiz Gonçalves de Freitas.  
Dissertação (mestrado em Modelagem Computacional de Conhecimento) –  
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2006.

Bibliografia: f. 108-116.

1. Inteligência artificial. 2. Sistemas de recuperação da informação –  
Classificação. 3. Ontologia. 4. Sistemas multiagentes. I. Título.

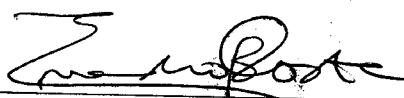
CDU: 004.8

Dissertação apresentada como requisito parcial para obtenção do grau de Mestra pelo Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina:



---

Frederico Luiz Gonçalves de Freitas – Orientador  
Centro de Informática  
Universidade Federal de Pernambuco



---

Evandro de Barros Costa - Examinador  
Instituto de Computação  
Universidade Federal de Alagoas



---

Guilherme Bittencourt - Examinador  
Departamento de Automação e Sistemas  
Universidade Federal de Santa Catarina

Maceió, Setembro 2006

# Resumo

Apesar dos aspectos positivos que a Internet possui e do potencial que permite, existe a problemática, que consiste em encontrar a informação necessária em meio a uma enorme quantidade de documentos disponíveis na rede. Faltam, ainda, ferramentas capazes de tratar semanticamente a informação contida em documentos que seguem uma estrutura preocupada apenas com a exibição dos seus dados. O sistema MASTER-Web, resolve o problema da extração integrada de páginas-conteúdo pertencentes às classes que integram um grupo (*cluster*). Neste contexto propomos a extensão dessa ferramenta para a classificação de artigos científicos baseada em ontologias. Para isso foi construída uma ontologia do domínio de Inteligência Artificial e adotadas estratégias de classificação utilizando sistemas de regras. A abordagem apresentada aqui, emprega esta ontologia e técnicas de classificação textual para extrair dos artigos informações úteis, e daí inferir sobre os temas tratados nestes artigos. Essa combinação conduziu a resultados bastante significativos: por exemplo, o sistema é capaz de identificar no texto as subáreas de IA que ele aborda e deriva conclusões, distinguindo os assuntos tratados pelo artigo daqueles que são brevemente citados no texto. A aplicação de técnicas simples e uma ontologia bem formada levam a resultados de classificação promissores, independentemente da estrutura do documento, propondo uma solução eficiente e plausível.

# Abstract

Although the positive aspects that Internet possesses and the potential it permits, there is a problematic that consists on finding needed pieces of information among the deluge of available documents on the web. Tools that are able to semantically treat the information contained in the documents which follows a structure only focused on data presentation are still lacking. The MASTER-Web system solves the problem of integrated extraction of content-pages that belong to classes which form a cluster. In this context, we propose the extension of this tool to the scientific articles classification based on ontologies. To achieve this goal, an ontology for the Artificial Intelligence domain was constructed and rule-based classification strategies were adopted. The approach presented here employs this ontology and textual classification techniques to extract useful pieces of information from the articles in order to infer to which themes it is about. This combination led to significative results: e.g. in the texts, the system is able to identify the specific subdivisions of AI and entails conclusions, distinguishing correctly the themes of the articles from the ones that are briefly mentioned in the texts. The application of simple techniques and a detailed ontology lead to promising classification results, independently of the document structure, proposing an efficient and plausible solution.

*Dedico este trabalho a*

*Meu pai, **Sebastião Palmeira**, cujo maior legado não foi bens materiais, nem mesmo os estudos, mas os valores que me transmitiu durante sua passagem nessa vida,*

***Fred Freitas**, que dedicou, além da orientação, uma generosa amizade, confiança, incentivo e indicação de rumos que se tornaram imprescindíveis à realização deste trabalho.*

# Agradecimentos

A Deus, por iluminar o meu caminho e permitir alcançar mais um objetivo.

A minha mãe Joselita, por ter plantado em mim a semente da educação e que hoje colhe junto comigo mais um bom fruto. A ela não só o meu agradecimento mas também os meus parabéns.

Ao meu irmão Eduardo Palmeira, por se fazer presente, apesar da distância física, torcendo e acreditando em mim.

A amiga Raquel Cabral, que me acolheu em sua casa e tornou-se parceira nestes anos de luta, angústia, mas também de alegrias e realizações.

A Klebson, que em nossas conversas, muitas vezes online, falava sobre sua experiência no mestrado e me encorajava.

A Ribinha, por ser uma excelente figura humana, presenteou-me com sua alegria e ajudou-me a superar um dos momentos mais difíceis de minha vida.

A Arandir, pelas palavras de apoio e por sua impagável solidariedade.

A Tércio, por dispor tempo para tirar dúvidas.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Internet e seus problemas de busca . . . . .	1
1.2	Contextualização do problema . . . . .	3
1.2.1	Componentes e números . . . . .	3
1.2.2	Problemas de classificação . . . . .	4
1.2.3	Classificação textual . . . . .	5
1.3	Estado da Arte . . . . .	7
1.4	Contribuições . . . . .	9
1.5	Organização do texto . . . . .	10
<b>2</b>	<b>Ontologias</b>	<b>12</b>
2.1	Conceito de ontologia . . . . .	12
2.2	Componentes das ontologias . . . . .	14
2.3	Tipos de Ontologias . . . . .	15
2.4	Áreas que usam ontologias . . . . .	16
2.5	Engenharia da Ontologia . . . . .	18
2.5.1	Critérios para projeto de Ontologias . . . . .	18
2.5.2	Metodologia . . . . .	21
2.6	Ontologia para o domínio de Inteligência Artificial . . . . .	23
2.6.1	Propósito . . . . .	24
2.6.2	Construção . . . . .	25

2.6.3	Pluralidade de visões e conceitos de IA . . . . .	27
2.6.4	Descrição da ontologia de IA . . . . .	28
<b>3</b>	<b>Uso do MASTER-Web para classificação</b>	<b>31</b>
3.1	MASTER-Web . . . . .	31
3.1.1	Tarefas dos agentes . . . . .	33
3.1.2	Conhecimento dos agentes . . . . .	35
3.2	MASTER-Web para classificação . . . . .	37
3.2.1	Alterações no código . . . . .	37
3.2.2	Contribuições na base de conhecimento . . . . .	39
3.2.3	O Motor de Inferência Jess . . . . .	49
3.2.4	Considerações . . . . .	50
<b>4</b>	<b>Experimentos</b>	<b>52</b>
4.1	Metodologia . . . . .	52
4.1.1	Base experimental . . . . .	53
4.2	Estudo de caso e resultados . . . . .	55
4.2.1	Estudo de caso . . . . .	55
4.2.2	Resultados . . . . .	55
4.3	Discussão dos Resultados . . . . .	59
<b>5</b>	<b>Trabalhos Relacionados</b>	<b>61</b>
5.1	Técnicas de Classificação . . . . .	61
5.1.1	Técnicas baseadas em conhecimento . . . . .	61
5.1.2	Técnicas baseadas em aprendizado de máquina . . . . .	62
5.2	Recuperação de Informação . . . . .	64
5.2.1	Técnicas de Recuperação de Informação . . . . .	64
5.3	Sistemas Similares . . . . .	68
5.3.1	OntoSeek . . . . .	68
5.3.2	SARI . . . . .	69

5.4	Considerações . . . . .	71
<b>6</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>72</b>
6.1	Contribuições . . . . .	72
6.2	Trabalhos Futuros . . . . .	75
<b>A</b>	<b>Exemplos de Processamento</b>	<b>77</b>
A.1	Reconhecimento de seções e classes principais . . . . .	77
A.2	Classificação de artigos . . . . .	85
<b>B</b>	<b>Base Experimental</b>	<b>92</b>
B.1	Base de Teste . . . . .	92
	<b>Referências Bibliográficas</b>	<b>108</b>

# Lista de Figuras

2.1	Cyc: Ontologia de alto-nível (adptada de (LG90)). . . . .	15
2.2	Tipos de ontologias quanto ao nível de dependência (adaptada de (Gua98)).	16
2.3	Primeira definição de Viagem na linguagem KIF (GPFLC04). . . . .	19
2.4	Segunda definição de Viagem na linguagem KIF (GPFLC04). . . . .	20
2.5	Parte da ontologia de IA, destacando diferentes níveis. . . . .	22
2.6	Parte dos atributos da Classe <i>Feed Forward ANN</i> , ressaltando os relaciona- mentos. . . . .	23
2.7	Ambiente do Protégé com a descrição de parte da ontologia de IA. . . . .	24
2.8	Parte dos atributos da classe <i>Brute-force</i> , ressaltando o atributo <i>SYNONYMS</i> .	26
2.9	Algumas classes e seus relacionamentos no domínio de Redes Neurais. . . .	29
2.10	Algumas classes e seus relacionamentos no domínio de Busca. . . . .	29
2.11	Algumas classes e relacionamentos no domínio de Representação do Conhe- cimento. . . . .	30
2.12	Algumas classes e seus relacionamentos no domínio de Aprendizado de Má- quina. . . . .	30
3.1	Visão geral da arquitetura de sistemas multiagentes cognitivo para extração integrada de dados da Internet (Fre02). . . . .	32
3.2	Detalhe de um agente, evidenciando as tarefas a serem realizadas (Fre02)..	34
3.3	Classe <i>Web-Page</i> e alguns de seus atributos e facetas (Fre02). . . . .	35
3.4	Classe <i>ArticleWebPage</i> e alguns de seus atributos e facetas. . . . .	38
3.5	Visão geral da classificação de artigos. . . . .	40

3.6	Trecho do código HTML, destacando as seções <i>Abstract</i> e <i>Conclusion</i> do artigo <i>Neural Networks</i> . . . . .	41
3.7	Classes principais definidas explicitamente para o domínio de IA. . . . .	42
3.8	Resultado do reconhecimento direto de classes principais para o artigo <i>Neural Networks</i> . . . . .	42
3.9	Atributos que identificam relações fortes entre as classes para o domínio de IA. . . . .	43
3.10	Resultado do reconhecimento de classes principais através de atributos no artigo <i>Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models</i> . . . . .	44
3.11	Relacionamento indireto entre as classes <i>Knowledge Level</i> e <i>Knowledge Representation Formalisms</i> . . . . .	45
3.12	Resultado do reconhecimento de classes principais através de relação indireta no artigo <i>OWL Web Ontology Language Overview</i> . . . . .	45
3.13	Níveis de importância das categorias para ordenação dos resultados da classificação. . . . .	48
4.1	Trecho do código HTML, destacando as marcações usadas para destacar as seções <i>Abstract</i> e <i>Contents</i> do artigo <i>Neural Networks</i> . . . . .	54
4.2	Trecho do código HTML, destacando as marcações usadas para destacar as seções <i>Abstract</i> e <i>Introduction</i> do artigo <i>Analysis of Three-Dimensional Protein Images</i> . . . . .	55
4.3	Resultado da classificação do artigo <i>Neural Networks</i> . . . . .	56
4.4	Resultado da classificação do artigo <i>Advances in Distributed Security</i> . . . . .	57
4.5	Resultado da classificação do artigo <i>Analysis of Three-Dimensional Protein Images</i> . . . . .	57
4.6	Resultado da classificação do artigo <i>AI Planning Resources on the Web</i> . . . . .	58
5.1	Exemplo de Remoção de <i>Stopwords</i> (HLM <sup>+</sup> 04). . . . .	65

5.2	Exemplo de Aplicação de <i>Stemming</i> (HLM <sup>+</sup> 04). . . . .	66
5.3	Exemplo de uma consulta em um sistema que utiliza indexação (Cha03). . . . .	67
5.4	Arquitetura do SARI (Tav98). . . . .	70
A.1	Parte do artigo <i>Neural Networks</i> . . . . .	81
A.2	Parte do artigo <i>Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models</i> . . . . .	82
A.3	Parte do artigo <i>OWL Web Ontology Language Overview</i> . . . . .	84
A.4	Parte do artigo <i>AI planning resources on the Web</i> . . . . .	86
A.5	Parte do artigo <i>Analysis of Three-Dimensional Protein Images</i> . . . . .	87
A.6	Parte do artigo <i>Advances in Distributed Security</i> . . . . .	88
A.7	Parte do artigo <i>Information Retrieval Issues</i> . . . . .	90
A.8	Parte do artigo <i>Learning Sets of Related Concepts: A Shared Task Model</i> . . . . .	91

# Lista de Tabelas

3.1	Exemplo da pontuação do item sinônimo referente a classe principal <i>Knowledge Representation Formalisms</i> . . . . .	46
3.2	Exemplo da pontuação do item tipos de seção referente a classe principal <i>Knowledge Representation Formalisms</i> . . . . .	48
3.3	Exemplo do resultado da ordenação das classes principais reconhecidas em um artigo. . . . .	48
4.1	Distribuição dos artigos por domínio. . . . .	53
4.2	Percentuais de acerto quanto à classificação de artigos por área de reconhecimento. . . . .	56

# Capítulo 1

## Introdução

### 1.1 Internet e seus problemas de busca

Até há pouco tempo a informação científica e tecnológica era disseminada apenas em formatos impressos, disponíveis em unidades especializadas de informação. Atualmente encontra-se disponível também em formatos eletrônicos. A revolução tecnológica, inegavelmente, elevou o volume de informação e a tornou mais acessível, sobretudo aquela que se encontra na Internet. O número de redes conectadas à Internet cresce a cada dia, disponibilizando serviços, softwares, documentos e possibilitando acesso e recuperação de informação de maneira rápida e fácil. Essa fonte abundante de informação e pesquisa rompe as fronteiras geográficas, tornando possível o intercâmbio global de conhecimentos e oferecendo heterogeneidade, não apenas no que diz respeito aos assuntos abordados pelos documentos disponíveis mas também nas diferentes formas de apresentação da informação e formatos destes documentos. Porém, essa heterogeneidade constitui um fator complicador, uma vez que exige esforços para pesquisar a informação em meio a essa diversidade.

Apesar dos aspectos positivos que a Internet possui e do potencial que permite, existe uma problemática, que consiste em encontrar a informação necessária em meio a tanta informação. Deste novo problema, surge a necessidade de ferramentas eficazes na pesquisa de informação. Os engenhos de busca, por exemplo, se apresentam como



ferramentas que tentam percorrer toda a Internet em busca da informação, sob a forma de documentos ou endereços de páginas Web. Embora alguns documentos dispostos na Internet disponham de requisitos que assegurem uma busca eficaz por parte de alguns softwares, muita informação pode não ser encontrada por falta de ferramentas capazes de tratar semanticamente a informação contida em documentos que seguem uma estrutura preocupada apenas com a exibição dos seus dados. Outro problema decorrente do crescente e contínuo volume de informação é a lista de endereços retornada pelas ferramentas de busca, apresentando um grande número de páginas irrelevantes, e que leva o usuário a realizar uma verdadeira 'garimpagem' para encontrar informação útil.

Um problema típico enfrentado pelos usuários em suas buscas por informação está relacionado à semântica das palavras, ou seja, ao seu significado. Segundo (Nav04) elementos da língua como a polissemia, propriedade que uma palavra tem de apresentar vários significados, afeta a precisão <sup>1</sup> das pesquisas, retornando muitos documentos que não têm relação com a pesquisa desejada. Isto revela que selecionar material textual relevante a partir de algumas palavras-chave, sem levar em consideração a semântica destas palavras pode acarretar uma série de resultados indesejáveis, longe do verdadeiro sentido que pretendia o usuário.

Por outro lado, o comportamento e padrões de perguntas dos usuários de sistemas de busca são um fator importante a ser considerado para o sucesso dos resultados apresentados nesse tipo de sistema. Segundo os estudos de (SX00), os usuários perdem muito tempo utilizando serviços de busca na Web, embora façam consultas simples para encontrar a informação desejada; apresentando comportamento como segue:

- formulam consultas com poucos termos;
- poucos são os que acrescentam ou removem termos ao tentar novamente uma consulta, em geral, apenas trocam um termo por outro;
- checam poucas páginas de resultados;

---

<sup>1</sup>Número de documentos relevantes encontrados / número total de documentos encontrados.

- e poucos utilizam técnicas avançadas de busca, operadores booleanos (emprego de E e do OU) e o *feedback* de relevância, que permite que o usuário refine progressivamente sua pesquisa.

## 1.2 Contextualização do problema

A Web pode ser vista como uma vasta, distribuída e heterogênea coleção de documentos, onde informações são consultadas constantemente. As seções seguintes, descrevem alguns conceitos do ambiente Web e apresentam números relacionados aos domínios de informação existentes, bem como problemas e técnicas de classificação textual.

### 1.2.1 Componentes e números

A Web (World Wide Web - WWW) é formada basicamente por quatro componentes, o HTTP (Hyper Text Transfer Protocol), o HTML (Hyper Text Markup Language), o servidor para a Web e o browser. O HTTP define a estrutura das mensagens e o modo como o servidor e o cliente trocam as mensagens. As páginas disponíveis na Web são construídas no formato HTML, podendo conter diversos objetos como imagem, applet Java, um clipe de áudio, etc. O servidor Web é uma aplicação que disponibiliza as páginas Web. Um browser é um agente usuário, que apresenta a página Web solicitada e fornece numerosas características de navegação e configuração (KR03).

Em ciência da informação, o hipertexto é, antes de tudo, um complexo sistema de estruturação e recuperação da informação em forma multissensorial, dinâmica e interativa (Par99). Sua característica de interatividade em virtude da conexão propiciada pelos links (ligações para outras páginas), conduz o leitor a navegar em diversos mundos de conhecimentos, o que reflete também a temporalidade do hipertexto. Além da heterogeneidade que aciona não apenas o sentido visual, mas também sonoro, revelando o poder de hipermídia.

Em se tratando do volume de informação posto na Internet, o número de domínios

registrados no Brasil em 2005, segundo (Fap05), é de 762.520 domínios oficialmente cadastrados. Além dos domínios registrados no Brasil, há, ainda, um número não contabilizado, e difícil de se determinar, de domínios que estão sob a responsabilidade de brasileiros residentes no Brasil, mas que são registrados em outros países. Qualquer tentativa de contagem é mais lenta que o número de páginas que entra na Internet a todo o momento. O crescimento do volume de informação na Internet se dá tanto pela criação de novos canais de correio eletrônico, grupos de discussão, conferências eletrônicas, quanto pelo surgimento de documentos eletrônicos como periódicos, livros, teses, dicionários, enciclopédias, almanaques, catálogos on-line, transpostos de ou gerados a partir de fontes impressas.

É incontestável que o funcionamento da Web aumentou o volume de informação e levou a Internet para os lares e empresas. Porém, esse ambiente não estruturado com elevado volume de informação impõe aos usuários dificuldade e consumo de tempo para encontrar documentos relevantes em meio a tanta informação.

### **1.2.2 Problemas de classificação**

A classificação e mapeamento da Web, como uma tentativa de capturar toda e qualquer informação e inferir sobre sua relevância e utilidade, é inviabilizada basicamente pelo grande volume de documentos em formatos heterogêneos, e pela ausência de estrutura da informação. Logo, essa tentativa conduz à insatisfação do usuário quanto aos resultados obtidos, constituindo restrições práticas à utilização do potencial oferecido pela Internet. Além disso, impõe ao usuário uma classificação manual sobre a lista de páginas apontadas, ou ainda pelo refinamento de sua consulta utilizando palavras-chave mais apropriadas, na tentativa de adequar as fontes fornecidas às suas reais necessidades.

Aumentar a utilização dos recursos disponíveis, através de uma recuperação que atenda as reais necessidades dos usuários, constitui um problema que vem sendo abordado. Porém os mecanismos de busca convencionais não são suficientes para atender

as requisições pretendidas. Apenas a indexação de palavras de um texto não é suficiente para melhorar o aproveitamento do seu potencial, uma vez que essa técnica possui várias limitações. Para (Nav04) apesar dos mecanismos de busca representarem uma fonte preciosa de informação, um problema típico enfrentado pelos usuários que utilizam esse tipo de sistema está relacionado à semântica das palavras, ou seja, ao seu significado. De acordo com (RL94), uma bem conhecida limitação das técnicas baseadas em palavras é o sinônimo, diferentes palavras e frases podem expressar diferentes conceitos. Por exemplo, as palavras 'faz', 'manufatura' e 'produz' todas se referem ao conceito de produção, o que pode dificultar a descoberta de documentos relevantes. Para (XC96), esse é um problema fundamental visto que os autores nem sempre usam as mesmas palavras que os usuários para descrever o mesmo conceito. (Car00) ressalta que o processo de especificação da consulta geralmente é uma tarefa difícil. Há freqüentemente uma distância semântica entre a real necessidade do usuário e o que ele expressa na consulta formulada.

Outros problemas poderão surgir durante esse processo que busca responder, da melhor forma, consultas de informações feitas por usuários. Problemas que certamente precisarão ser analisados e suas soluções apresentadas, estão relacionados à relevância dos textos e posteriormente de suas sentenças. A seção seguinte trata do processo de classificação de texto e de tarefas de classificação usadas por algumas aplicações.

### 1.2.3 Classificação textual

Classificação textual consiste no processo de selecionar textos dentro de uma ou mais categorias. Com a Internet fazendo parte da rotina de muitas pessoas em todo mundo, surgiram muitas aplicações que realizam classificação textual. Algumas aplicações apresentam soluções utilizando tarefas de classificação como as descritas abaixo segundo (JM02):

- **Roteamento** - aplicações enviam automaticamente documentos, que possuem informações específicas, aos seus usuários. A técnica aqui utilizada é tipicamente

a inserção, por parte do usuário, de uma busca pré-definida que indica o tipo de informação desejada.

- **Indexação** - cada documento é descrito por um conjunto de palavras-chave, chamadas termos de indexação, onde coleções de documentos são separadas por categorias de acordo com os termos indexados.
- **Filtragem** - os documentos são classificados em duas classes: os considerados irrelevantes são desprezados e os relevantes são publicados, indexados ou encaminhados aos usuários.

Tais tarefas podem ser analisadas em relação aos dados contidos nos documentos. O entendimento dos dados (conteúdo, palavras-chave, título, etc.) é um fator importante e revela o grau de complexidade associado ao mesmo. Segundo (JM02), dentre as propriedades inerentes aos dados a serem levadas em consideração previamente figuram:

- **Granularidade:** refere-se ao nível de detalhe dos dados existentes em um documento. Quanto mais detalhada for a informação, menor será o nível de granularidade;
- **Dimensionamento:** diz respeito a quantidade de traços classificatórios. Um mesmo documento pode conter informações sobre mais de uma área, dessa forma pode ser classificado em mais de uma categoria ;
- **Exclusividade:** se o mesmo documento pode pertencer a uma ou mais de uma categoria;
- **Topicalidade:** se o documento pode conter um só ou vários tópicos, ou seja, um documento trata especificamente sobre um tema ou pode conter múltiplos tópicos.

Como descrito no capítulo 5, existe uma variedade de formas abordadas para o problema de classificação textual, onde várias áreas de pesquisa estão envolvidas nesse

processo, como por exemplo, Recuperação de Informação, Inteligência Artificial e Processamento de Linguagem Natural.

### 1.3 Estado da Arte

Muito se tem pesquisado na tentativa de tornar a recuperação da informação mais eficiente. Segundo (LL03), um dos grandes desafios está na forma como os documentos poderão ser organizados de modo a facilitar o acesso a informações relevantes.

É preciso disponibilizar softwares que retornem a informação de forma satisfatória em relação às solicitações dos usuários, independentemente da estruturação do documento. É observada uma tendência para o desenvolvimento de programas capazes de executar automaticamente tarefas múltiplas, tais como selecionar documentos de uma base com enfoque no seu conteúdo, agrupá-los em categorias ou classes e deles extrair determinados conjuntos de informações (JM02). Contudo, algumas destas ferramentas atuam sobre domínios muito restritos, estando vinculadas a bases específicas, não possuindo portanto flexibilidade para recuperação, classificação e extração de dados em diferentes campos.

Para dar suporte a buscas mais eficientes que retornem o que realmente o usuário espera encontrar têm sido feitos vários estudos. Uma nova geração da Web está sendo desenvolvida, denominada Web Semântica (BLHL01), em que a Web atual é estendida com estruturas de meta-informação que organizam e dão significados bem definidos aos recursos da Web. Estas estruturas definem formalmente relacionamentos entre termos, sendo formada pelas chamadas ontologias.

Existe uma variedade de ferramentas de busca. Algumas mais conhecidas e genéricas como o Yahoo!<sup>2</sup>, Google<sup>3</sup> e AltaVista<sup>4</sup>, que utilizam técnicas desenvolvidas pela área de Recuperação de Informação, e outras mais especializadas cujo objetivo

---

<sup>2</sup><http://www.yahoo.com.br>

<sup>3</sup><http://www.google.com>

<sup>4</sup><http://www.altavista.com>

é atender a requisitos mais específicos como WebKB (CDF<sup>+</sup>98), CiteSeer (BLG98) e DEADLINER (KGC<sup>+</sup>00). Apesar da ampla cobertura das primeiras ferramentas, os resultados muitas vezes apresentam uma grande quantidade de endereços de páginas inúteis ou irrelevantes. Segundo (Fre02), isso ocorre porque estes mecanismos possuem capacidade de representar as páginas com análises baseadas apenas no nível léxico, uma vez que utilizam algoritmos matemáticos para atribuir relevância às páginas.

Por outro lado, as ferramentas especializadas, lidam com menos informação que as genéricas, concentrando-se em uma ou mais categorias. Porém, tal informação é muito mais específica, de qualidade superior, não apenas pela restrição do universo, mas principalmente pela capacidade, por parte das aplicações, em dotar de semântica a busca. Classificam documentos de uma base com enfoque no seu conteúdo e, utilizam técnicas de diversas áreas de pesquisa, como Recuperação de Informação, Inteligência Artificial e Processamento de Linguagem Natural.

O MASTER-Web (*Multi-Agent System for Text Extraction, Classification and Retrieval over the Web*) é uma arquitetura de Sistemas Multiagentes Cognitivos. Resolve o problema da extração integrada de páginas-conteúdo pertencentes às classes que integram um grupo (*cluster*), como por exemplo o meio científico. Um grupo se caracteriza por possuir um conjunto de entidades inter-relacionadas, onde cada instância de uma entidade é representada por uma ou mais páginas. O meio científico, por exemplo, possui entidades como eventos, pesquisadores, artigos, etc. Cada agente responsabiliza-se por uma entidade, e eles colaboram entre si, aproveitando o relacionamento existente entre as classes, e, por conseguinte, entre as páginas. Por exemplo, num evento (representado por uma página de chamada de trabalhos ou "*Call for Papers*- CFP), existem âncoras para páginas de pesquisadores, institutos de pesquisa e vice-versa.

Para representar o conhecimento dos agentes foram empregadas ontologias, que segundo (Fre02) desempenham um papel fundamental na arquitetura do MASTER-Web. Servindo não só como vocabulário de comunicação entre agentes, como também na definição e organização apropriadas de conceitos, relações, e restrições. As ontologias

integram a arquitetura do sistema, contudo, não fazem parte do programa. Deste contexto, decorre o fato de que as ontologias possam ser trocadas, reusadas, refinadas e, uma vez que constituem o conhecimento sobre determinado assunto, proporcionem flexibilidade no aprendizado. O domínio de aplicação do MASTER-Web não está restrito a páginas ou sítios, mas a grupo de classes da Web, permitindo a aplicação dessa ferramenta em outros domínios da Web, além do científico. O que se pretende é facilitar o acesso à informação veiculada na Internet sob diferentes áreas através da classificação de informações contidas nos próprios documentos.

As tecnologias utilizadas pelas ferramentas citadas deixam clara a evolução da pesquisa nessa área em busca de resultados mais eficientes e automatizados. Os pesquisadores incorporaram noções semânticas nos sistemas, onde documentos são analisados levando-se em conta os termos e suas relações dentro do contexto em que estão inseridos. Ao contrário dos processos tradicionais, onde cada documento é descrito por um conjunto de palavras-chave, chamadas termos de indexação.

## 1.4 Contribuições

A classificação textual tem despertado a atenção de muitos pesquisadores, o que justifica o grande volume de trabalho destinado à construção de sistemas eficientes que apresentem desempenho adequado e satisfatório. Por outro lado, muitos trabalhos na área de Inteligência Artificial vêm explorando o uso de ontologias (vide capítulo 2) como uma maneira formal para especificar informações conceituais de um domínio, compartilhar e reusar conhecimento. Neste contexto, ontologia serve como base para sistemas de classificação por possuir granularidade variável. Assim, o presente trabalho tem como um primeiro objetivo, construir uma ontologia sobre o domínio de Inteligência Artificial (IA). Essa ontologia, inicialmente, abrange algumas das subáreas, teorias, técnicas, métodos e aplicações desse domínio, podendo ser compartilhada e expandida pela inclusão de novas subáreas, conceitos e relações.

Um segundo objetivo é propor uma modelagem para classificação de informação,



utilizando artigos científicos e aplicando a ontologia de IA, como uma abordagem declarativa para a construção de uma base de conhecimento que descreva o domínio de Inteligência Artificial e possibilite derivar interpretações a cerca dos assuntos tratados nos artigos. Essa modelagem é aplicada à ferramenta MASTER-Web, de modo a entender esta ferramenta e provar que a mesma tem suporte a classificação automática em diferentes áreas, possibilitando que pesquisas estruturadas possam ser realizadas independente de domínio.

## 1.5 Organização do texto

Este trabalho está estruturado da seguinte forma:

- O capítulo 2 apresenta algumas definições de ontologias e componentes usados em sua construção, bem como os tipos e áreas que as utilizam. Logo após, discorre-se sobre Engenharia da Ontologia, abordando os critérios para projeto e metodologia de construção. Ao fim do capítulo, é apresentada a ontologia para o domínio de Inteligência Artificial.
- O capítulo 3 discorre sobre o uso do MASTER-Web para classificação textual. Inicialmente, examina-se a arquitetura desta ferramenta. Em seguida, são apresentadas as alterações no código do sistema, estendendo-o para classificação de informação em artigos científicos. Por fim, é feita uma abordagem geral sobre a base de conhecimento adicionada a ferramenta para classificação de artigos.
- O capítulo 4 descreve a metodologia usada para realizar os experimentos e os resultados obtidos. A princípio, discorre-se sobre a base de artigos científicos utilizada nos experimentos e em seguida é analisado o estudo de caso. A partir daí os resultados atingidos são apresentados, avaliados e discutidos.
- No capítulo 5 é feita uma discussão sobre os trabalhos relacionados à classificação textual. Inicialmente são relatadas técnicas aplicadas em problemas de

---

classificação de texto e recuperação de informação. Além do relato dos trabalhos prévios, este capítulo fala resumidamente sobre sistemas de classificação textual similares.

- No capítulo 6 são revisadas as contribuições, feitas as considerações finais e relatados os trabalhos futuros.

# Capítulo 2

## Ontologias

Esse capítulo é dedicado a explicar ontologias como um formalismo de representação de conhecimento explícito. As próximas seções descrevem conceitos de vários autores sobre o termo, assim como seus componentes, tipos e algumas áreas que utilizam ontologias para representação e reuso do conhecimento. A área de Engenharia da Ontologia também é investigada, ressaltando os critérios para projeto e metodologia de construção de ontologias. Por fim, é apresentada a ontologia para o domínio de Inteligência Artificial, destacando seu propósito, construção, a pluralidade de visões no domínio de IA e as principais classes e relacionamentos dentro do escopo dessa ontologia.

### 2.1 Conceito de ontologia

O termo ontologia tem origem no grego '*ontos*', ser, e '*logos*', palavra. Foi introduzido no século 19 por filósofos alemães. Segundo (Gua98), a ontologia como um ramo da Filosofia se refere a um sistema de categorização para a organização da realidade. Neste contexto a ontologia é uma descrição sistemática da existência que busca prover uma definitiva e exaustiva classificação de entidades em todas as suas esferas.

Na Ciência da Computação, ontologias foram desenvolvidas pela Inteligência Artificial visando criar representações que fossem além da descrição de simples instâncias do domínio considerado (MG03). Sua criação veio facilitar o compartilhamento e reuti-

lização de informação (Fen01). Atualmente, o uso de ontologias vem se difundindo em diversas áreas que buscam desenvolver um vocabulário contendo os conceitos relativos ao domínio de aplicação.

São apresentadas muitas definições de ontologia na literatura científica, a definição mais famosa e referenciada pela literatura foi definida por (Gru93b):

*"Uma ontologia é uma especificação explícita de uma conceitualização".*

(SBF98) por sua vez propõe a seguinte definição:

*"Uma ontologia é uma especificação explícita e formal de uma conceitualização compartilhada".*

(SBF98) explica essa definição como segue:

- Uma **conceitualização** refere-se a um modelo abstrato de algum fenômeno no mundo que identifica conceitos relevantes desse fenômeno.
- No que diz respeito a **explícito**, os tipos de conceitos usados e suas restrições estão definidos de forma clara, explicitamente, como conceitos, propriedades, relações, funções, restrições, axiomas e terminologia.
- **Formal** refere-se ao fato de que a ontologia é compreensível para agentes e sistemas.
- **Compartilhada** reflete o fato de que uma ontologia captura conhecimento consensual, isto é, ela não é privada a um indivíduo mas aceita por um grupo.

Muitas definições são propostas na literatura sob diferentes pontos de vistas, apresentando elementos comuns e complementares dentro da mesma realidade. A Inteligência Artificial, por sua vez tem explorado as ontologias para especificação de termos de domínios promovendo seu compartilhamento e reuso entre entidades de software.

## 2.2 Componentes das ontologias

Ontologia define termos e relações compondo um vocabulário de uma área através de um formalismo que utiliza cinco componentes básicos: conceitos, relações, funções, axiomas e instâncias, os quais estão conceituados abaixo segundo (Gru93b):

- **Conceitos:** são uma representação usada em amplo sentido. Um conceito pode ser abstrato ou concreto, elementar ou composto, real ou fictício. Algo como uma tarefa, uma função, uma ação, uma estratégia, etc.
- **Relações:** representam um tipo de associação entre conceitos do domínio. Exemplo de relacionamento entre os conceitos *Autor* e *Livro* é o relacionamento *autor-de*.
- **Funções:** caso especial de relações, onde um conjunto de elementos tem uma relação única com um outro elemento. Um exemplo é *Mãe-de*.
- **Axiomas:** servem para modelar sentenças que são sempre verdadeiras. Um axioma no domínio de viagem seria afirmar que não é possível viajar de trem dos Estados Unidos para a Europa.
- **Instâncias:** são usadas para representar elementos em uma ontologia. Maceió pode ser uma instância para o conceito cidade.

Os elementos apresentados formam uma linguagem expressiva proporcionando uma hierarquia de categorias ontológicas através dos tipos de conceitos e suas relações. Assim, a hierarquia de categorias proporciona herança múltipla, onde os tipos de conceitos e relações podem possuir axiomas que determinam suas propriedades, propagando os seus subtipos em forma de herança. As relações definidas podem, ainda, restringir as referências dos tipos de conceitos.

Tendo sido apresentados os componentes de uma ontologia, a próxima seção propõe expor os tipos mais comuns de ontologias e fornecer alguns exemplos classificados dentro desses tipos.

## 2.3 Tipos de Ontologias

Ontologias podem ser classificadas em diferentes tipos de acordo com sua generalidade. (Gua98) identifica quatro tipos de ontologias, cujas dependências estão ilustradas na figura 2.2.

- **Ontologias de alto-nível:** descrevem conceitos muito gerais como espaço, tempo, matéria, objeto, evento, ação, etc., os quais são independentes de um problema ou domínio particular. Sendo assim, é bem razoável ter-se uma ontologia de alto-nível compartilhada por grandes comunidades de usuários. Um exemplo de ontologia de alto-nível é a ontologia Cyc, representada pela figura 2.1:

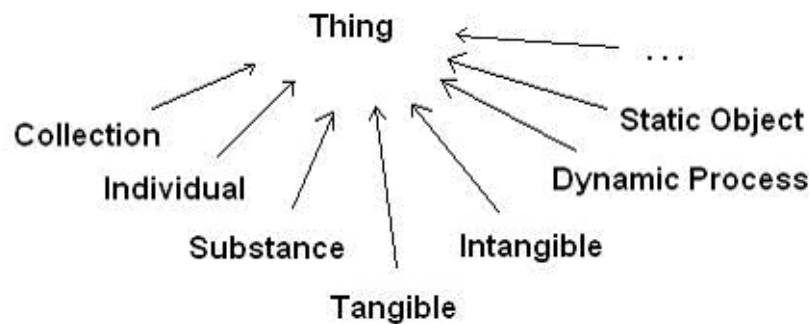


Figura 2.1: Cyc: Ontologia de alto-nível (adptada de (LG90)).

- **Ontologias de domínio:** descrevem o vocabulário relacionado a um domínio específico através da especialização de conceitos introduzidos nas ontologias de alto-nível. Por exemplo, medicina, direito e indústria de computadores.
- **Ontologias de tarefa:** descrevem um vocabulário relacionado a uma tarefa genérica, ou atividade através da especialização de conceitos presentes na ontologia de alto-nível. Por exemplo, análise de requisitos de software.
- **Ontologias de aplicação:** descrevem conceitos tanto das ontologias de domínio quanto das de tarefas, são freqüentemente especializações de ambas as ontologias. São as ontologias mais específicas por serem utilizadas dentro das aplicações.

Por exemplo, agências de viagens da Espanha especializadas em destinos Norte Americanos (GPFLC04).



Figura 2.2: Tipos de ontologias quanto ao nível de dependência (adaptada de (Gua98)).

A classificação dos tipos de ontologias apresentados nesta seção utiliza a generalidade dos assuntos como critério principal. Outros tipos de classificação podem ser encontrados na literatura, como em (MVI95) que define tipos de ontologias relacionando-as à sua função. Em (UG96), a classificação está relacionada ao grau de formalismo do vocabulário e, em (UJ99), à sua aplicação.

## 2.4 Áreas que usam ontologias

O uso de ontologias vem se disseminando em diversas áreas. Vários campos de pesquisa têm reconhecido a importância das ontologias para representação e reuso do conhecimento, bem como características que as qualificam para promover interoperabilidade e facilidade de integração entre unidades e fontes de informação.

Uma nova geração da Web denominada Web Semântica (BLHL01), visa estender a Web atual com estruturas de meta-informação que organizam e dão significados bem definidos aos recursos da Web. A Web Semântica usa ontologias que definem formalmente conceitos e relacionamentos entre termos.

Já na Biologia, é possível encontrar uma variedade de vocabulários que especificam o domínio biológico. O projeto Gene Ontology (GO) (Con01) fornece um conjunto

padronizado e estruturado de vocabulários específicos da Biologia que pode ser usado para descrever os genes de qualquer organismo.

Na área de Direito ontologias têm sido aplicadas em domínios como direito penal, processual e civil (Kra97), sendo largamente usadas para recuperação de informação em documentos normativos como proposto por (Lam00). (Lam00) propõe o projeto de uma ontologia a partir de textos publicados pelo *Journal Officiel de la République Française* para recuperação de informação e formalização da organização conceitual de um sistema de informação baseados nestes textos.

No comércio eletrônico, vários padrões são usados para classificar produtos e serviços. (MR05) apresenta uma ontologia de comércio eletrônico, chamada SNAP, desenvolvida para recomendar produtos e serviços aos consumidores. Esta ontologia é utilizada em uma variedade de domínios incluindo comércio bancário, planejamento financeiro e telefonia.

A área de segurança computacional também reconhece a importância das ontologias como um mecanismo para auxiliar na tarefa de detecção de intrusos e manutenção da segurança. (BMM04) propôs e desenvolveu uma ontologia capaz de representar informações semânticas a respeito das vulnerabilidades atualmente conhecidas, com o objetivo de utilizá-la para classificação de vulnerabilidades em sistemas computacionais.

Diversas áreas têm feito uso de ontologias como um instrumento estratégico para representação e gerenciamento de conhecimento. De acordo com (Rio05), existem muitos projetos sendo desenvolvidos, e alguns já finalizados, com objetivos de criar ontologias genéricas que possam ser utilizadas por diversos sistemas em todo o mundo. Um exemplo disso é o projeto On-To-Knowledge<sup>1</sup>, que aplica ontologias para disponibilizar informações eletronicamente, com objetivo de melhorar a qualidade da gerência do conhecimento em grandes e distribuídas organizações. O foco é criar facilidades para aquisição, manutenção e acesso on-line a informações vindas de diversas fontes, promovendo a sua reutilização.

---

<sup>1</sup>[www.ontoknowledge.org](http://www.ontoknowledge.org).



## 2.5 Engenharia da Ontologia

Agora que temos uma visão geral de ontologias, tendo visto conceitos, componentes, tipos e o uso de ontologias em diferentes áreas de conhecimento, vamos discorrer sobre o processo de construção de ontologias. Veremos, a seguir, princípios para construção de ontologias e uma metodologia para a construção de ontologias.

### 2.5.1 Critérios para projeto de Ontologias

Uma ontologia quando escolhida para representar conceitos de um domínio é resultado de um projeto, logo critérios de avaliação para guiar o processo de projeto são necessários. Abaixo são listados os critérios preliminares de projeto de uma ontologia, propostos por (Gru93a), onde foram assumidos o propósito do compartilhamento de conhecimento e a interoperabilidade entre programas:

- **Clareza:** definições devem ser objetivas. Sempre que for possível, uma definição deve ser declarada através de axiomas lógicos, sendo preferível que seja completa, com condições necessárias e suficientes, melhor que uma definição parcial, definida apenas por condições necessárias ou suficientes. Todas as definições devem ser documentadas com linguagem natural.
- **Coerência:** se uma sentença inferida a partir de axiomas contradiz uma definição ou exemplo dado informalmente, então a ontologia é incoerente.
- **Extensibilidade:** uma ontologia deve permitir que novos termos possam ser definidos para usos especiais baseados no vocabulário existente, de maneira que não seja requerida a revisão das definições previamente existentes.
- **Mínimo compromisso com implementação:** a conceituação deve ser especificada no nível do conhecimento, isto é, sem depender de uma codificação particular no nível simbólico ou de codificação.

- **Mínimo compromisso ontológico:** uma ontologia deve requerer o compromisso ontológico mínimo suficiente para dar suporte às atividades de compartilhamento de conhecimento desejadas.

Os princípios acima serão evidenciados usando exemplos do domínio de viagem, ilustrados nas figuras 2.3 e 2.4, cujas definições estão escritas na linguagem KIF<sup>2</sup> (GPFLC04).

```
(define-class Travel(?travel)
  "A journey from place to place"
:axiom-def
  (and (Superclass-Of Travel Flight)
        (Subclass-Of Travel Thing)
        (Template-Facet-Value Cardinality)
          arrivalDate Travel 1)
        (Template-Facet-Value Cardinality)
          departureDate Travel 1)
        (Template-Facet-Value Maximum-Cardinality)
          singleFare Travel 1))
:def
  (and (arrivalDate ?travel Date)
        (departureDate ?travel Date)
        (singleFare ?travel Number)
        (companyName ?travel String)))
```

Figura 2.3: Primeira definição de Viagem na linguagem KIF (GPFLC04).

Nesses exemplos, a sentença KIF após a palavra-chave *:def* estabelece restrições lógicas sobre os argumentos. Comparando esses exemplos, percebemos que o segundo exemplo possui definições mais completas e precisas que o primeiro. Na segunda definição não há restrições desnecessárias, ou seja, especificou-se que para ocorrer uma viagem são condições necessárias e suficientes a existência de uma data de partida e uma data de chegada, e apenas como condições necessárias o nome da empresa e valor da passagem. Assim, a segunda definição satisfaz o requisito *clareza*.

Por outro lado, esta definição não obedece ao princípio de *comprometimento mínimo com a implementação*, dado que o atributo valor da passagem (*singleFare*) pode não

<sup>2</sup> KIF (GF92), Knowledge Interchange Format, é uma linguagem de ontologia baseada em lógica de primeira ordem.

```

(define-class Travel( ?travel)
  "A journey from place to place"
:axiom-def
  (and (Superclass-Of Travel Flight)
        (Subclass-Of Travel Thing)
        (Template-Facet-Value Cardinality)
          arrivalDate Travel 1)
        (Template-Facet-Value Cardinality)
          departureDate Travel 1)
        (Template-Facet-Value Maximum-Cardinality)
          singleFare Travel 1))
:iff-def
  (and (arrivalDate ?travel Date)
        (departureDate ?travel Date))
:def
  (and (singleFare ?travel Number)
        (companyName ?travel String)))

```

Figura 2.4: Segunda definição de Viagem na linguagem KIF (GPFLC04).

ser modelado em outra linguagem, caso essa nova linguagem não possua o tipo número (*Number*). Para minimizar esse problema, (Gru93b) propõe ontologias para modelar quantidades físicas, compostas por classes que agrupam, por exemplo, unidades de medida. Para nosso exemplo, uma classe para agrupar moedas correntes (Real, Euro, Dólar, etc.) poderia substituir o tipo *Number*. A inclusão dessa classe nas ontologias que modelam quantidades físicas esclarece o conceito de *extensibilidade*.

Para exemplificar *coerência*, vamos usar um outro exemplo de (GPFLC04) para o domínio da ontologia de viagem. Supondo que nesse domínio exista um axioma formal estabelecendo que é impossível viajar de trem entre Estados Unidos e Europa, esse mesmo axioma proibirá a existência de instâncias de viagens de trem entre Madri e Nova York, sendo estas instâncias das cidades da Europa e Estado Unidos, respectivamente.

Por fim, vamos esclarecer o critério *mínimo compromisso ontológico* usando o atributo data de chegada (*arrivalDate*), cujo tipo é *Date*. Nesse caso, é preciso especificar o tipo de data, por exemplo, se é absoluta (dia, mês e ano) ou relativa. Enfim, ter o mínimo de compromisso ontológico é estabelecer acordos quanto ao vocabulário compartilhado, a fim de garantir a consistência quanto às definições na ontologia.

## 2.5.2 Metodologia

Para (GP99) o processo de construção de ontologias assemelha-se mais a uma arte do que a uma atividade de engenharia. Dado que, durante o processo de desenvolvimento, muitas das decisões tomadas são conduzidas por preceitos e critérios adotados por desenvolvedores, que muito embora possam fazer uso de uma metodologia, não seguem rigorosamente todas as fases pré-estabelecidas.

Para a construção da ontologia apresentada neste trabalho, foi utilizada a metodologia proposta por (NM01). Esta abordagem metodológica inclui a identificação do propósito e escopo da ontologia, o nível de especialização ou generalização, quais tarefas serão realizadas (especificação, extensão, manutenção) e revisão da ontologia, implicando ciclos de vida formando um processo iterativo. É proposto, portanto, um possível processo de desenvolvimento de ontologia, onde é especificada uma seqüência de passos a serem seguidos para realização das tarefas.

Abaixo são descritos os passos percorridos de acordo com a metodologia citada:

- **Enumerar os termos do domínio.** Consiste em criar uma lista com os termos do domínio. Este passo constitui um ponto oneroso para o desenvolvimento da ontologia, por requerer aquisição de uma quantidade significativa de conhecimento do domínio abordado, explorando a capacidade de capturar conhecimento por parte do desenvolvedor. Para extrair o vocabulário utilizado pela comunidade de Inteligência Artificial, foram utilizadas técnicas como análise de sumários, pesquisas em diversas bibliografias, entre livros, artigos, sites especializados e entrevista com especialista.
- **Definir as classes e a hierarquia de classes.** Neste passo os termos do domínio são colocados em uma hierarquia, podendo seguir basicamente três abordagens: *top-down*, *botton-up* e *middle-out*. No processo *top-down*, os conceitos mais genéricos são inicialmente definidos e seqüencialmente especializados. No processo *botton-up*, as classes mais especializadas são definidas com subsequente

agrupamento em classes mais gerais. Na estratégia *middle-out*, conceitos mais centrais são definidos e depois são generalizados ou especializados apropriadamente. A ontologia apresentada neste trabalho utilizou por vezes a abordagem *top-down* e outras vezes a *botton-up* e *middle-out*, para desenvolver a hierarquia de classes. Assim, a figura 2.5 exemplifica os diferentes níveis de generalização, onde o conceito *Knowledge Representation Formalisms* foi identificado em um nível mais alto e a partir dele chegou no conceito *Production Systems*, enquanto que o conceito *Scripts* foi definido e depois relacionado ao conceito *Language-oriented Formalism*. Quanto às classes *Higher Order Logic* e *Semantic Networks*, essas foram definidas e posteriormente especializadas e generalizadas, adequando-se dentro da ontologia proposta.

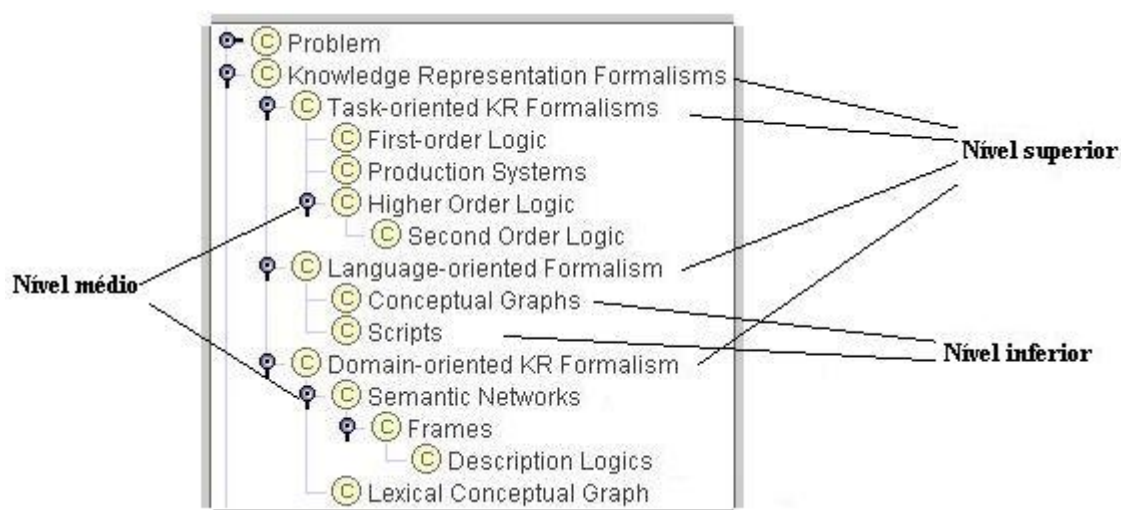


Figura 2.5: Parte da ontologia de IA, destacando diferentes níveis.

- Definição das propriedades.** As classes normalmente possuem atributos que armazenam valores descritivos de suas características. Nesse passo, os atributos são definidos podendo ter valores do tipo numérico, lógicos e cadeias de caracteres. As classes podem, ainda, ter outras classes como valores de atributos, dando origem ao formato de rede e caracterizando a multiplicidade hierárquica das classes. Por exemplo, a classe *Feed Forward ANN* possui os atributos *Propagation-Scheme* e *Input Layer*, que têm como possíveis valores instâncias da classe *Propagation*

*Rule* e *Input Layer*, respectivamente, como mostra a figura 2.6. Esses atributos que unem duas classes são conhecidos como relacionamentos.

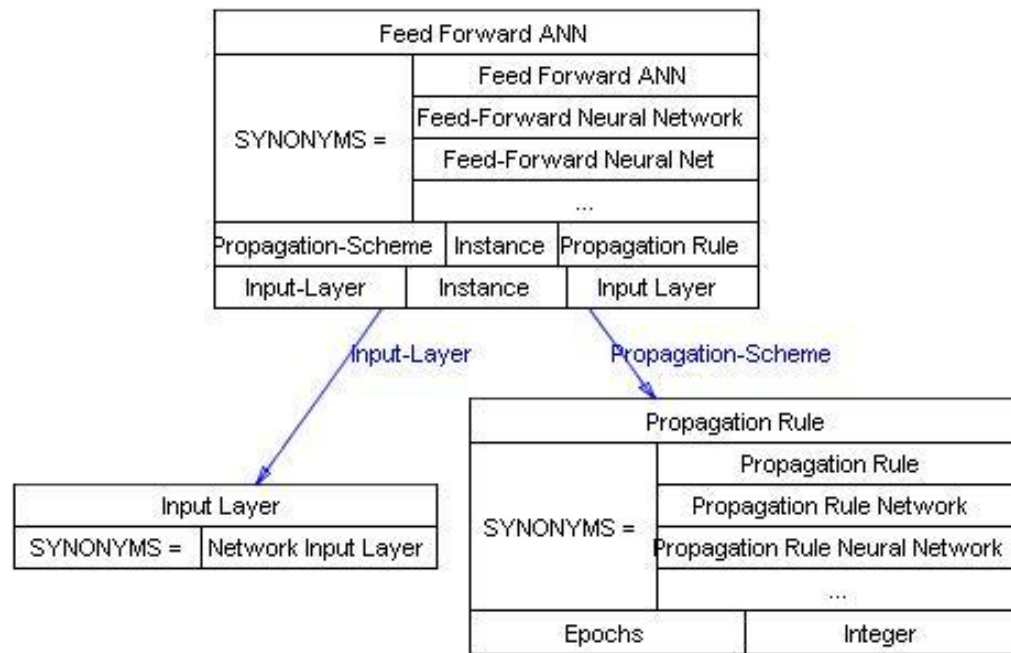


Figura 2.6: Parte dos atributos da Classe *Feed Forward ANN*, ressaltando os relacionamentos.

- **Implementação da ontologia.** A partir do modelo conceitual, a ontologia foi implementada através do ambiente de edição de ontologia Protégé-2000 (PRO00). Neste ambiente foram descritos os conceitos pertencentes à ontologia, juntamente com seus atributos e relacionamentos como ilustra a figura 2.7.

## 2.6 Ontologia para o domínio de Inteligência

### Artificial

Neste trabalho apresentamos uma ontologia que contém informações sobre os conceitos investigados no domínio de Inteligência Artificial. Seus ramos, subáreas, teorias, técnicas, métodos e aplicações são identificados privilegiando a classificação dos conceitos e informações que dizem respeito aos seus significados, ao mesmo tempo são inseridas relações semânticas entre esses conceitos estabelecendo uma hierarquia conceitual.

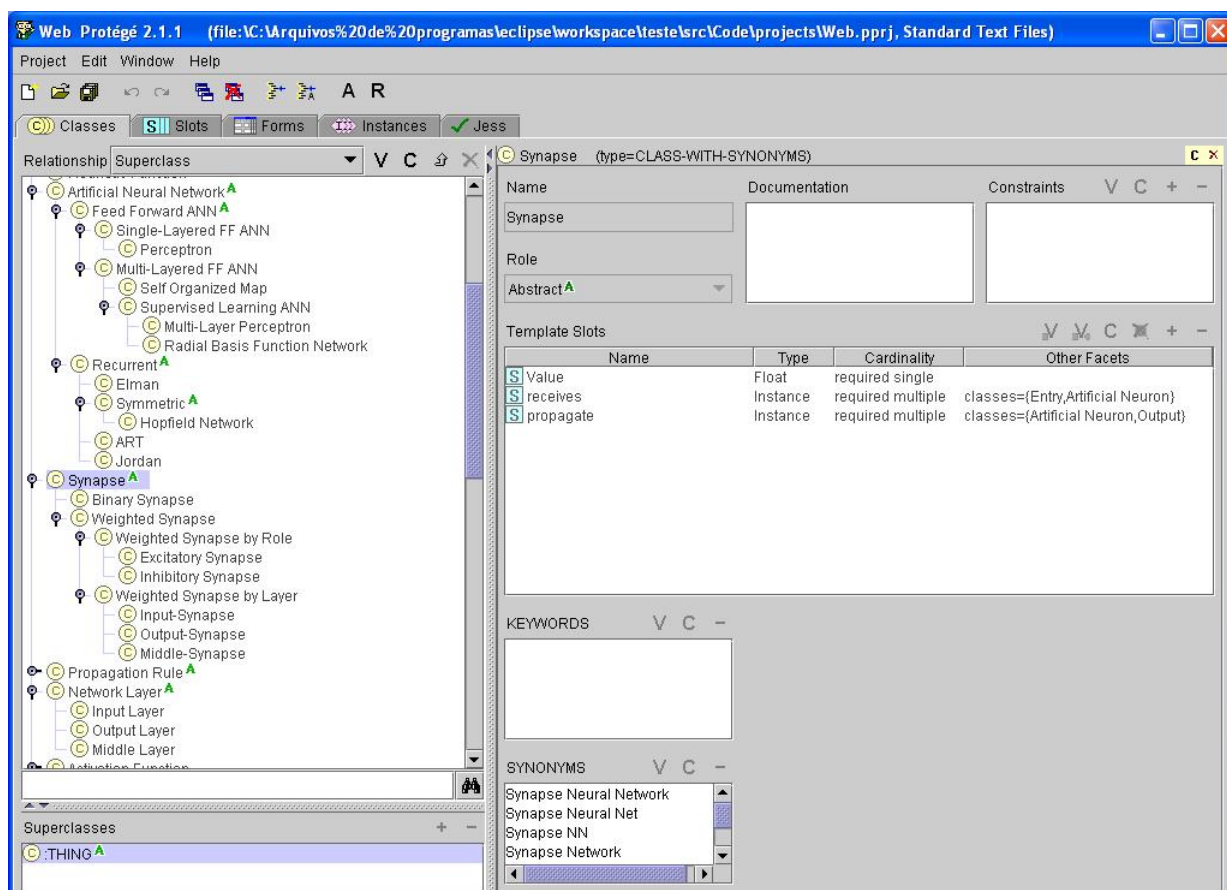


Figura 2.7: Ambiente do Protégé com a descrição de parte da ontologia de IA.

Dessa forma, tipos de conhecimento envolvendo a área de IA foram categorizados e suas dependências especificadas.

### 2.6.1 Propósito

O propósito geral desta ontologia é servir como repositório semântico para a representação do significado, formando um vocabulário estruturado e organizado para ser utilizado em conformidade com uma conceitualização de um domínio por determinada comunidade. Podendo ser empregada em sistemas de conhecimento, como sistemas extratores ou classificadores de informação, compondo uma base de conhecimento e servindo como recurso de apoio à informação.

Por permitir compartilhamento e reuso do conhecimento do domínio, sua utilização não está restrita ao contexto do desenvolvimento de software, mas poderá ser utilizada,

também, em outros contextos. A ontologia serve como documentação, fornecendo vocabulário sistematizado de termos, relações e regras, permitindo análise do conhecimento do domínio e o entendimento da estrutura da informação.

### 2.6.2 Construção

No processo de construção da ontologia foram encontradas várias dificuldades, principalmente no que diz respeito aos termos usados no domínio em questão. Alguns conceitos lingüísticos, como sinonímia e polissemia, foram identificados e soluções foram aplicadas para que a representação do conhecimento se tornasse mais completa e informações relevantes fossem encontradas, principalmente, quando da utilização da ontologia por sistemas extratores ou classificadores de informação.

- **Sinonímia** - Embora exista um vocabulário típico neste domínio, é encontrado na literatura o uso de muitos termos distintos para um mesmo conceito, denotando a preferência de um termo a outro por pesquisadores e autores, como é o caso do método de busca sem informação, onde alguns autores usam *Brute-force Method* enquanto outros preferem *Blind Search Method*.

Para esse problema, a exemplo de outros sistemas existentes, adotamos a inclusão de sinônimos na metaclasses. Essa nova propriedade, assemelha-se a um dicionário de sinônimos permitindo a expansão semântica da ontologia. A figura 2.8, mostra alguns dos sinônimos adicionados ao atributo *SYNONYMS* da classe *Brute-force*.

- **Polissemia** - Apesar de típico, o vocabulário não é exclusivo, isso pode significar que termos usados em IA também são usados em outros domínios, assim os termos podem ter múltiplos significados. Como o objetivo dos sistemas de IA é tentar imitar o comportamento humano, isso implica a realização de estudos dos mais variados aspectos do comportamento, utilizando-se da interdisciplinaridade para melhor conhecer a cognição humana. Essa abordagem interdisciplinar, segundo (Has03), envolve pesquisas no âmbito da Filosofia, Psicologia, Neurociência, entre



Brute-force		
SYNONYMS =	Brute-force Search	
	Blind Search Method	
	Weak Search Method	
	...	
Optimal	Symbol	YES
		NO
		DEPENDS
Branching factor		Integer
Complete	Symbol	YES
		NO
		DEPENDS
Time Complexity	Instance	Complexity
...		

Figura 2.8: Parte dos atributos da classe *Brute-force*, ressaltando o atributo *SYNONYMS*.

outras. Assim, termos como *Grammar* e *Negotiation* são originalmente utilizados em outras áreas, em IA termos como esses são empregados e adaptados a modelos computacionais para melhor simular a capacidade de aprendizagem do ser humano e sua habilidade em realizar funções complexas.

Nesse caso, muitos sistemas aderem a criação de uma lista de termos genéricos que devem ser ignorados em processos como classificação e extração de informação, por se tratar de termos ambíguos que podem comprometer a precisão dos resultados. Ao contrário desta técnica, optamos por uma metodologia que gerencia a ambigüidade pela criação de um vocabulário controlado formado por termos considerados não-genéricos. Assim, o atributo *SYNONYMS*, apresentado anteriormente, representa essa lista de termos para cada classe, onde o nome da classe deve ser incluído na lista, se considerado um termo não-genérico, como mostra a figura 2.8, onde o termo *Brute-force* não é incluído na lista de sinônimos.

Outra dificuldade encontrada, foi em relação a multiplicidade de visões na área de IA, que por sua especificidade será abordada na próxima seção.

### 2.6.3 Pluralidade de visões e conceitos de IA

A tarefa de realizar uma classificação que retrate de forma mais abrangente possível toda extensão do assunto, estabelecendo um caráter ontológico preciso e bem estruturado sobre o domínio de Inteligência Artificial, foi dificultada pela dimensão deste campo e pela falta de consenso em algumas subáreas. Segundo (Bit98), devido ao fato da IA ser uma ciência relativamente nova, a comunidade de pesquisadores da área está longe de um consenso sobre seus fundamentos. Entretanto, (Bit98) reconhece que uma posição aceita por grande parte da comunidade da IA é que uma teoria em IA consiste na especificação do conhecimento necessário a uma capacidade cognitiva, capacidade essa que inclui atividades tais como solução de problemas, uso de linguagem, tomada de decisões, percepção e alguns elementos de controle motor, principalmente no que diz respeito ao planejamento de trajetórias.

A IA abrange uma enorme variedade de subcampos, desde áreas de uso geral, como aprendizado e percepção, até tarefas específicas como jogos de xadrez, demonstração de teoremas matemáticos e diagnósticos de doenças (RN03). A própria dimensão e as diversas relações existentes entre os problemas, métodos, técnicas, etc. contribuem para estabelecer a falta de concordância da comunidade de pesquisa científica quanto às várias ramificações da área, considerando as sutis interseções envolvidas.

Alguns autores apresentam discordâncias quanto à classificação dos ramos, em algumas ramificações é estabelecido um certo consenso, entretanto em outras a classificação é desigual. Para (MDSC02a), algumas subáreas estão estruturadas e solidificadas, como é o caso da Simulação Baseada em Multiagentes. Porém não há discussões para definir conjuntamente uma classificação que retrate de forma mais precisa o atual estágio de pesquisa e desenvolvimento.

Assim, a qualidade da ontologia depende não só das definições oriundas de levantamento bibliográfico e consultas a especialistas, mas também da capacidade em identificar, especificar e posteriormente executar as transformações necessárias entre entidades ontológicas. É necessário, portanto, salientar que o processo de construção

de ontologia deve apresentar capacidade de evolução e adaptação às diferentes necessidades impostas pela interligação entre as ontologias e pelas transformações sofridas pelo domínio.

O desenvolvimento de ontologias é um processo moroso, complexo e subjetivo, exigindo considerável esforço manual e habilidade para distinguir não apenas conceitos e suas relações, mas também a consistência do domínio.

#### 2.6.4 Descrição da ontologia de IA

É apresentada nesta seção uma modelagem ontológica para o domínio de IA como uma proposta para especificar informações conceituais que promova compartilhamento e reuso de conhecimento, bem como permita análise do conhecimento do domínio e o entendimento da estrutura da informação.

Para a construção da ontologia, optou-se por desenvolver, separadamente, ontologias por subáreas e aplicações que posteriormente são agrupadas para compor a ontologia em questão. O critério para escolha das subáreas foi a existência de especialistas em IA disponíveis para auxiliar na construção da ontologia, assim o escopo dessa ontologia compreende as áreas de Busca, Representação do Conhecimento, Redes Neurais e Aprendizado de Máquina. Dessa forma, ontologias de outras subáreas de IA poderão ser incorporadas estendendo a ontologia existente.

Na figura 2.9, destacamos alguns relacionamentos entre as principais classes definidas na ontologia de Redes Neurais. Dois tipos de Redes Neurais, *Recurrent* e *Feed Forward ANN* e a relação *Inputs*<sup>\*3</sup>, entre as classes *Artificial Neural Network* e *Entry*, representando as múltiplas entradas nos neurônios artificiais, são apresentadas nessa figura.

Já a figura 2.10, apresenta alguns relacionamentos entre as principais classes definidas na ontologia de Busca.

---

<sup>3</sup>Na figura 2.9, o uso do asterisco (\*) após a definição do tipo de um atributo assinala que ele pode assumir múltiplos valores, e não um só como os outros atributos.

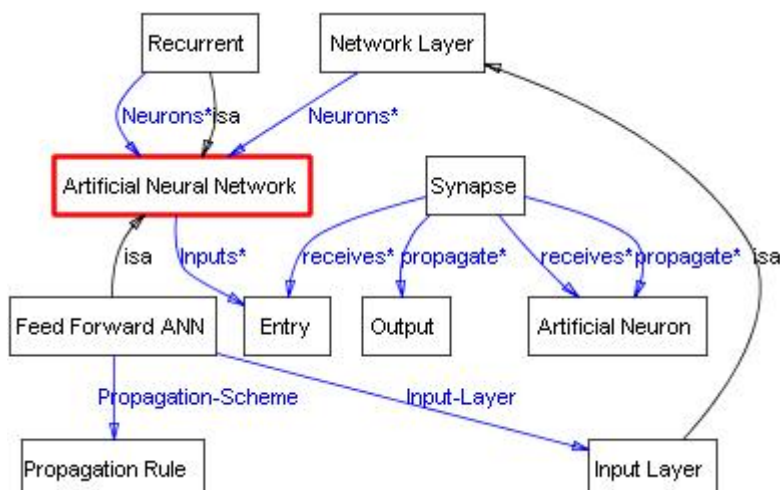


Figura 2.9: Algumas classes e seus relacionamentos no domínio de Redes Neurais.

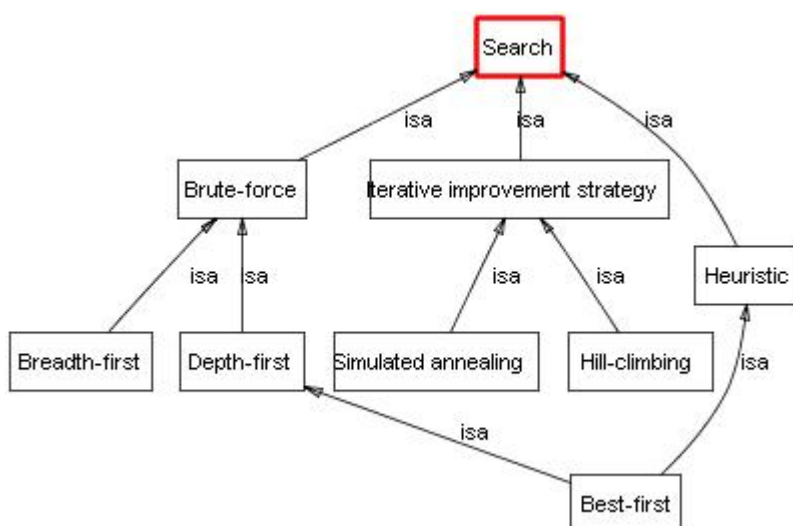


Figura 2.10: Algumas classes e seus relacionamentos no domínio de Busca.

Na figura 2.11, estão representados relacionamentos entre as principais classes definidas na ontologia de Representação do Conhecimento.

Na figura 2.12 estão presentes as subclasses da classe *Machine Learning* na ontologia de Aprendizado de Máquina.

A motivação para o emprego de ontologias é beneficiar-se dos relacionamentos entre as classes, permitidos por seus atributos (*slots*). Esses relacionamentos são explicitamente nomeados e desenvolvidos com especificação de regras, tornando a semântica mais precisa e sem ambigüidade, permitindo diferentes níveis de detalhe, dependendo

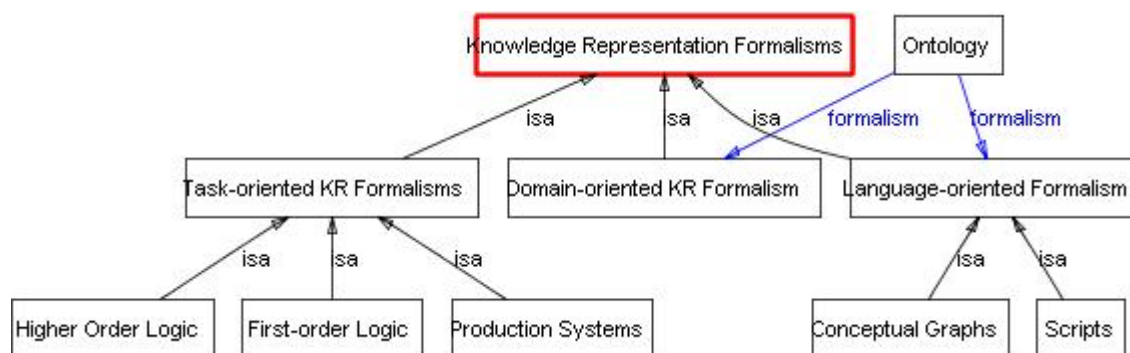


Figura 2.11: Algumas classes e relacionamentos no domínio de Representação do Conhecimento.

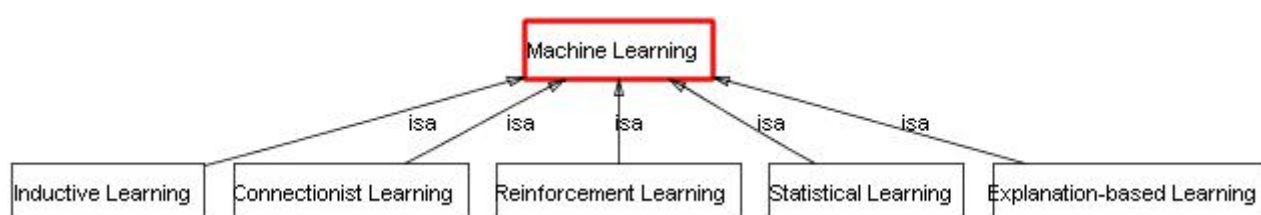


Figura 2.12: Algumas classes e seus relacionamentos no domínio de Aprendizado de Máquina.

do grau de granularidade da árvore e do propósito da taxonomia. Além disso, ontologias favorecem a inferência pela representação explícita do conhecimento e propiciam compreensão mais apurada do domínio abordado. Para (MDSC02b), ontologias possibilitam o compartilhamento e reuso do conhecimento, troca de informações e suporte à interoperabilidade entre sistemas computacionais.

Apesar dos aspectos positivos das ontologias, algumas limitações podem estar presentes em função, por exemplo, das características do domínio. No caso de IA, a dinamicidade da área impõe alterações nas entidades ontológicas que acompanhem o desenvolvimento e transformações sofridas pelo domínio, refletindo de forma atualizada os conceitos da área.

# Capítulo 3

## Uso do MASTER-Web para classificação

Um dos principais objetivos deste trabalho é propor uma modelagem para classificação de informação em artigos científicos, utilizando a ontologia descrita no capítulo anterior. Essa modelagem é aplicada à ferramenta MASTER-Web (*Multi-Agent System for Text Extraction, classification and Retrieval over the Web*), de modo que este capítulo é dedicado a explicar seu uso para classificação textual, destacando a arquitetura dessa ferramenta. Neste mesmo capítulo, são apresentadas as alterações no sistema, para atingir o objetivo desejado, incluindo uma abordagem geral sobre base de conhecimento.

### 3.1 MASTER-Web

É uma arquitetura de Sistemas Multiagentes Cognitivos para resolver o problema da extração integrada de entidades pertencentes às classes que integram um grupo de páginas ou (*cluster*) (Fre02). A motivação principal para o emprego de sistemas multiagentes é beneficiar-se dos relacionamentos entre as classes. A visão geral da arquitetura está ilustrada na figura 3.1.

Cada agente, representado como um círculo na figura 3.1, é um especialista no reconhecimento de páginas que correspondem a instâncias da classe que ele processa

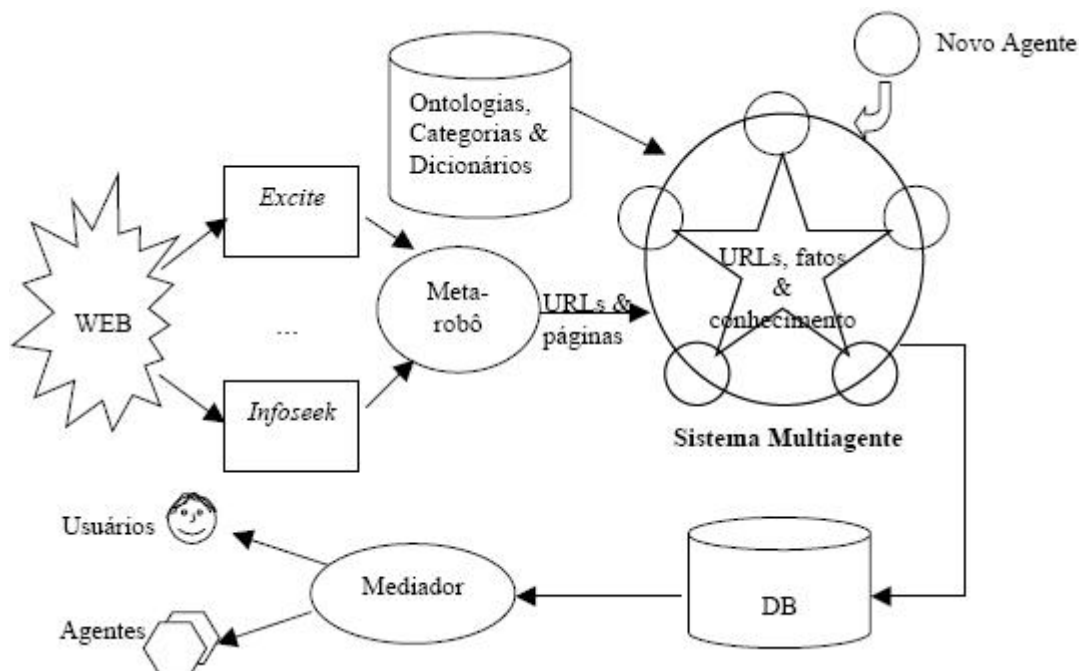


Figura 3.1: Visão geral da arquitetura de sistemas multiagentes cognitivo para extração integrada de dados da Internet (Fre02).

(por exemplo, páginas de pesquisadores, chamadas de eventos científicos - '*Call for Papers*' -, artigos e outras do grupo científico). Além de reconhecer as páginas, ele filtra e classifica, extraindo também seus atributos (por exemplo, áreas de pesquisa e instituição dos pesquisadores). A estrela indica troca de mensagens contendo regras de reconhecimento e fatos (conhecimento dos agentes), além das URLs sugeridas entre os agentes.

Cada agente possui um meta-robô, que se conecta a múltiplos mecanismos de busca - como *Altavista*, *Excite*, *Infoseek* e outros. Ele consulta os mecanismos de busca com palavras-chave que garantem cobertura<sup>1</sup> em relação à classe de páginas processada pelo agente. (e.g., os termos '*call for papers*' e '*call for participation*' para o agente CFP). Devido à falta de precisão, o conjunto de páginas resultante das consultas recai em vários grupos funcionais além de páginas-conteúdo, apresentando muitas listas, mensagens, páginas-conteúdo de outras classes, e lixo. As URLs são dispostas numa

<sup>1</sup>Cobertura (recall) significa o quociente entre o total de documentos relevantes recuperados sobre o total de documentos relevantes.

fila de URLs de baixa prioridade. Cada agente continuamente acessará, além desta fila, outra de alta prioridade, que armazena URLs sugeridas por outros agentes ou presentes em páginas da categoria funcional listas. Afinal, estes endereços são obtidos dentro de um contexto mais confiável e com maior probabilidade de ser relevantes do que as listas de resultados dos mecanismos de buscas.

Um mediador tem a função de ajudar as consultas aos dados, provendo visões não-normalizadas - mais simples - da base de dados, e permitindo a qualquer usuário ou agente beneficiar-se do acesso aos dados extraídos.

Ao entrar no sistema, os agentes registram-se e anunciam-se aos outros agentes, mandando fatos e regras de reconhecimento de páginas e ponteiros úteis a si próprio, que serão empregadas pelos outros para lhe indicarem sugestões de páginas. O novo agente receberá, reciprocamente, regras úteis aos outros agentes. Assim, quando um agente acha informação que dispara alguma das regras referentes aos outros, este agente repassa a informação (ponteiro ou página) ao agente que lhe enviou a regra disparada.

### 3.1.1 Tarefas dos agentes

Como pode ser observado na figura 3.2, cada agente desempenha quatro tarefas consecutivas no processamento de cada página (Fre02): validação, pré-processamento, reconhecimento e extração.

- **Validação:** Nesta fase são eliminadas páginas inacessíveis, já existentes no BD e em formatos que os agentes não possam processar. Mesmo as páginas inválidas ficam armazenadas no banco de dados, porque o robô freqüentemente depara-se com ponteiros repetidos, só as recuperando novamente se sua data de última modificação foi alterada. A validação acelera todo o processamento, além de evitar trabalho redundante, economizam banda passante e tempo de processamento, gastos com a recuperação de páginas inúteis.
- **Pré-processamento:** Esta fase tem por meta representar as páginas de várias maneiras, tais como conteúdo com e sem HTML, palavras-chave e freqüências, e-



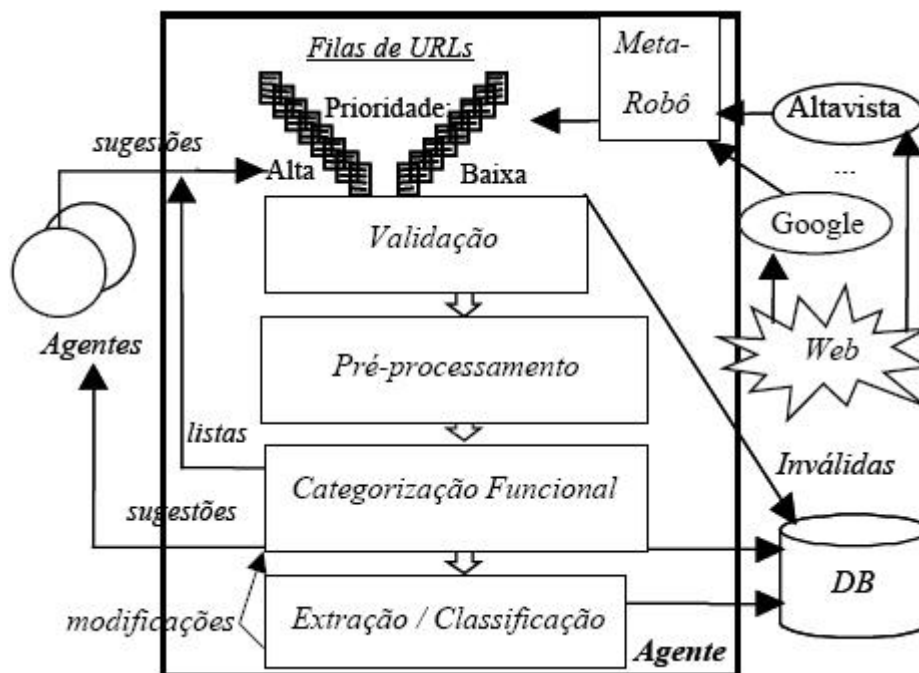


Figura 3.2: Detalhe de um agente, evidenciando as tarefas a serem realizadas (Fre02).

mails, e outros, com dados extraídos delas, aplicando, se necessário, recuperação de informação e processamento de linguagem natural (PLN). Os dados passam ao motor de inferência. Esses dados estão reproduzidos na figura 3.3, que apresenta a classe *Web-Page*, definida na ontologia Web<sup>2</sup>.

- **Categorização Funcional:** Durante esta fase, um agente classifica as páginas em grupos funcionais, como lista, mensagem, lixo ou reconhecida como membro da classe processada pelo agente ou da classe de outro agente do sistema. As páginas ou links a serem sugeridas aos outros agentes são encontradas e enviadas, quando uma das regras enviadas por eles é disparada pelo motor de inferência. Por exemplo, uma âncora com a palavra '*conference*' é útil para o agente CFP.
- **Extração e Classificação:** São extraídos os atributos, armazenados na base de dados, ou, pela inconsistência ou inexistência destes, a classificação da página é corrigida em relação aos grupos funcionais. Por exemplo, a presença de datas

<sup>2</sup>Contém definições de *hyperlink*, termo e frequência, e de página da Web em suas várias representações e atributos.

com mais de um ano de intervalo numa página de CFP, denunciam uma página erroneamente classificada como chamada de trabalhos.

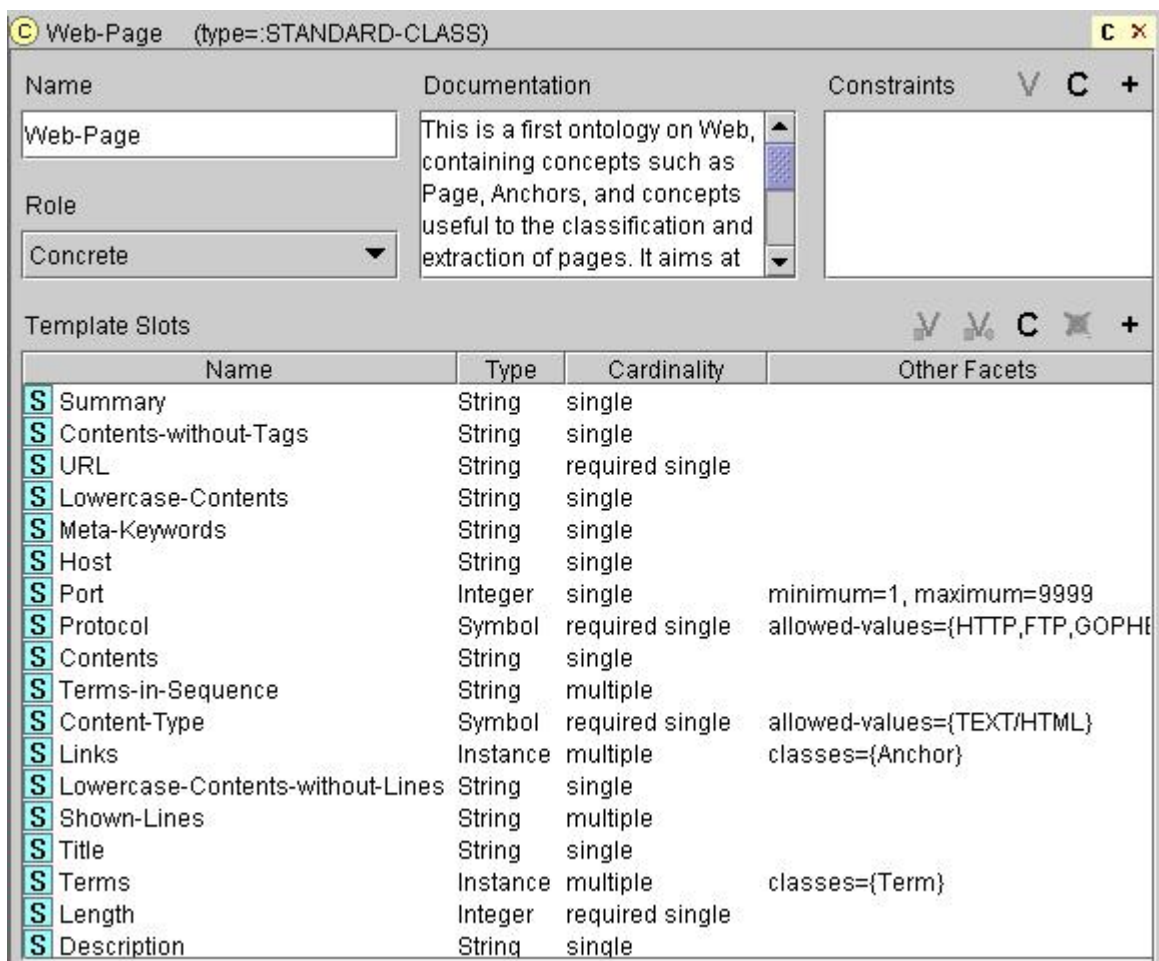


Figura 3.3: Classe *Web-Page* e alguns de seus atributos e facetas (Fre02).

### 3.1.2 Conhecimento dos agentes

Para representar o conhecimento dos agentes são usadas ontologias. As ontologias são utilizadas como vocabulário de comunicação entre agentes e, também para definir e organizar apropriadamente, conceitos, relações, e restrições. Portanto, permitem que o sistema seja aplicado em domínios diferentes. O que evidencia a portabilidade do sistema. Quatro ontologias se fazem necessárias:

- **Ontologia do domínio:** Ontologia principal, sendo empregada nesse sistema a ontologia do domínio científico, uma ontologia reusada a partir da ontologia do projeto europeu (KA)<sup>2</sup> (*Knowledge Annotation Initiative of the Knowledge Aquisition Community*).
- **Ontologia da Web:** Contém definições de *hyperlink*, termo e frequência, e de página da Web em suas várias representações e atributos - como listas de palavras-chaves e suas frequências, ponteiros, e-mails, representações de páginas em PLN, etc. Contém ainda definições e instâncias de protocolos, tipos de arquivos e outros conceitos relativos à Web.
- **Ontologia de manipulação integrada de informação:** Classes, relações, restrições e instâncias empregadas na extração e classificação funcional e por conteúdo. Inclui:
  - *Templates* reconhecedores das categorias funcionais e páginas-conteúdo;
  - *Templates* extratores e classificadores de páginas-conteúdo (em eventos científicos, Conferência, *Workshop*, Jornal, etc);
  - Classes auxiliares, como definições de conceitos e seus sinônimos e palavras-chave, agentes e habilidades, etc;
  - Casos complexos que identificam atributos, classes de páginas, categorias funcionais e sugestões para outros agentes.
- **Ontologias auxiliares:** Conhecimento útil de outras áreas de conhecimento. Ontologias lingüísticas, como o WordNet (Mil95), de tempo e de locais, além de outras específicas de um agente (como dados bibliográficos para o agente de artigos científicos) são empregadas.

É possível perceber que a arquitetura do MASTER-Web usa vários tipos de reuso, facilitando a construção de novos agentes, que compartilham a mesma estrutura em termos de código, esquema de base de dados, serviços dos mecanismos de busca, e

até mesmo de boa parte do conhecimento que dispõem. Assim, na seção seguinte é apresentada a extensão do MASTER-Web para classificação de artigos.

## 3.2 MASTER-Web para classificação

A partir da arquitetura de sistemas multiagentes cognitivos descrita acima, uma possível extensão da arquitetura consiste na inclusão de técnicas de classificação de informação em artigos científicos. Dessa forma, as seções seguintes apresentam a extensão realizada na arquitetura do MASTER-Web para classificação de artigos.

### 3.2.1 Alterações no código

Para classificar artigos científicos no MASTER-Web, foi necessário estender a representação das páginas HTML existente no sistema, através de uma representação adicional para os dados contidos nas seções dos artigos. Para isso foi criada a subclasse chamada *ArticleWebPage*, que estende a classe *Web-Page* e possui atributos que representam seções encontradas em artigos científicos, como mostra a figura 3.4. O atributo *Abstract* foi adicionado para representar a seção resumo, já *Keywords-Article* representa as palavras-chave, *Conclusion* representa a seção conclusão e *Other-sections* representa outras seções consideradas relevantes, como explicado mais adiante.

Assim, na fase de pré-processamento o conteúdo de uma página é analisado pelo agente, visando o reconhecimento dessas seções. Em relação a relevância, as seções são assim classificadas:

- **Seção irrelevante:** Diz respeito a seção cujo conteúdo aborda assuntos gerais, citações das obras de referências consultadas ou sugestões de trabalhos a serem realizados futuramente, enfim, seção que não trata diretamente sobre o tema do artigo, e por essa razão pode interferir na precisão dos resultados, devendo ser ignorada. Alguns exemplos destas seções são introdução, bibliografia, trabalhos relacionados, trabalhos futuros, apêndices, entre outros.

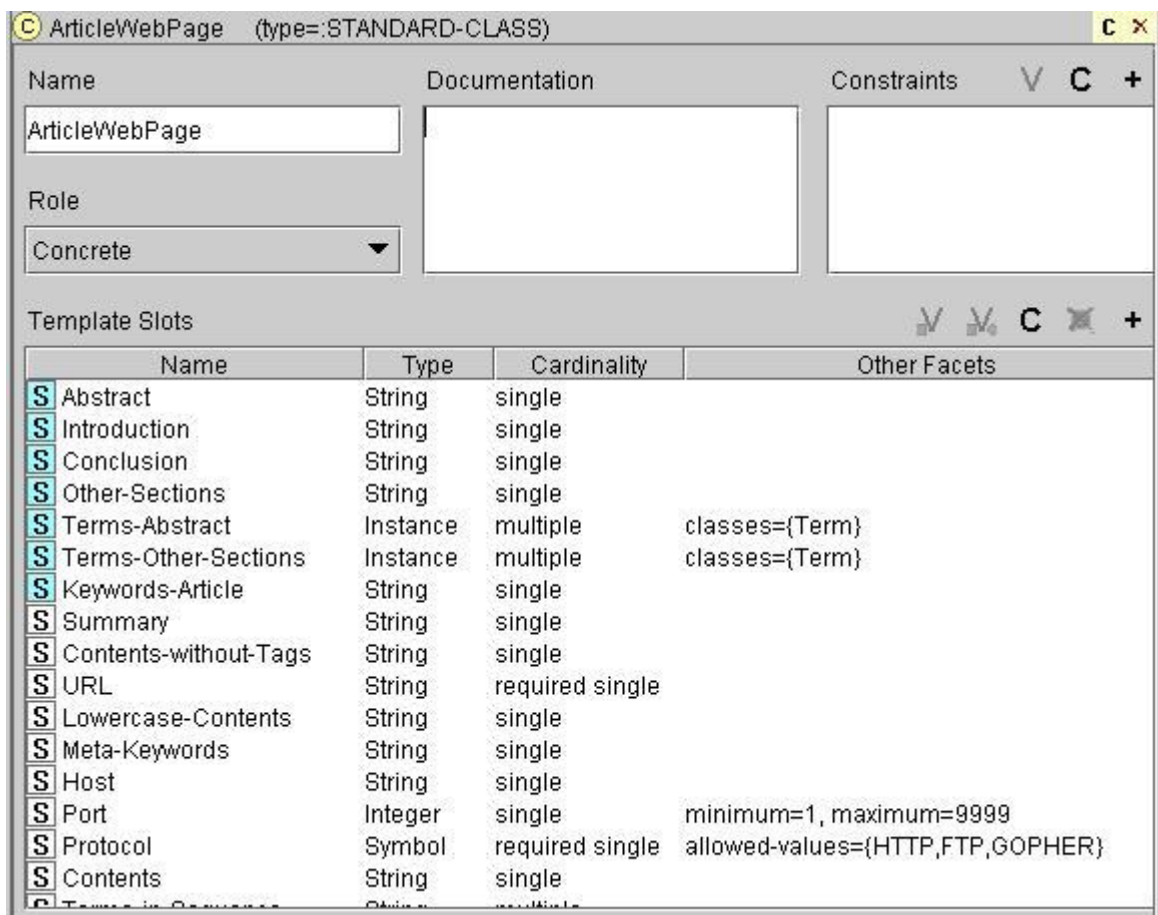


Figura 3.4: Classe *ArticleWebPage* e alguns de seus atributos e facetas.

- **Seção relevante:** Seção que trata especificamente sobre o tema do artigo, de modo a apresentar os aspectos teóricos, e por isso, é considerada significativa para classificação, devendo, portanto, ser extraída e inserida em atributos da classe *ArticleWebPage*. Como exemplo desse tipo de seção temos resumo, palavras-chave, conclusão e outras sessões (que não são consideradas irrelevantes).

Nas fases de categorização e classificação, o MASTER-Web utiliza o conceito de casos emprestado do sistema de extração AutoSlog (Ril94), para promover a genericidade e expressividade na especificação de condições em que uma página deva ser categorizada ou classificada. A fim de beneficiar a classificação de artigos, permitindo a criação de regras com maior expressividade e específicas para esse tipo de classificação, foram definidos novos atributos relacionados aos termos que podem influenciar na qualidade da classificação do artigo.

Abaixo, é apresentado o exemplo de caso com dois novos atributos (*Ignore-Class* e *Ignore-Synonyms*), para classificação de artigos:

```
( [Web_Instance_362] of Case
  (Importance HIGH)
  (Description "Article")
  (Ignore-Class)
  (Ignore-Synonyms))
```

O atributo *Ignore-Class*, define as classes que devem ser ignoradas no processo de classificação dos termos do domínio por se tratar de conceitos que não agregam significado semântico importante dentro do domínio investigado, assim todos os sinônimos dessas classes (vide 2.6.2), serão ignorados quando encontrados no texto. As classes *Problem* e *Application*, são exemplos de classes dentro do domínio de IA que devem ser ignoradas pois seus nomes são termos genéricos e, portanto, podem influenciar nos resultados.

O atributo *Ignore-Synonyms*, pela mesma razão que em *Ignore-Class*, determina quais termos não devem ser levados em conta dentro da lista de sinônimos de uma classe. Exemplos de sinônimos que devem ser ignorados são as siglas, como *SOM* e *MLP* usadas como abreviaturas para os termos *Self Organized Map* e *Multi-Layer Perceptron*, respectivamente, pois podem ser confundidas com abreviaturas de outras palavras.

Além da importância, que determinará a precedência do caso no disparo das regras, os casos podem especificar conceitos, atributos e palavras-chave presentes e/ou ausentes, na página ou no início dela, e número de atributos encontrados dentro um conjunto definido no caso.

### 3.2.2 Contribuições na base de conhecimento

Base de conhecimento representa o entendimento sobre o domínio e, utiliza uma representação declarativa onde, geralmente, uma coleção de regras forma expressões sintá-

ticamente corretas para derivar interpretações, do tipo *condição-ação*. Para a classificação de artigos, uma nova base de conhecimento foi adicionada ao MASTER-Web para se conhecer os assuntos tratados dentro do domínio em questão (subáreas, teorias, técnicas, métodos, aplicações), como mostram as próximas seções.

### 3.2.2.1 Estratégias de classificação

O processo de classificação do artigo quanto aos assuntos abordados, tem início com o reconhecimento das seções do artigo e dos subdomínios de IA. Uma visão geral das fases desse processo de classificação é ilustrada na figura 3.5.

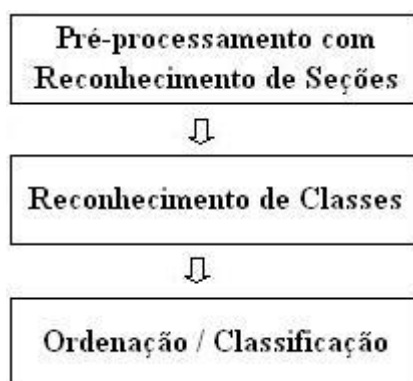


Figura 3.5: Visão geral da classificação de artigos.

- **Pré-processamento com Reconhecimento de seções.** O objetivo dessa tarefa é reconhecer as diferentes seções relevantes do artigo e delas extrair os termos do domínio. Esse reconhecimento se dá pela análise do artigo, reconhecendo as principais seções, identificando e extraíndo delas os termos presentes na ontologia do domínio. A figura 3.6 ilustra duas seções em HTML, *Abstract* e *Conclusion*, do artigo *Neural Networks*.
- **Reconhecimento de Classes.** Dentro do domínio de IA existem subdomínios, que chamamos de classes principais, ou seja, classes que representam assuntos do domínio. As classes principais são explicitamente definidas como classes a serem reconhecidas, conseqüentemente esse reconhecimento se estende às suas

```

<H2>Abstract</H2>
<P>This report is an introduction to Artificial Neural Networks. The various
types of neural networks are explained and demonstrated, applications of neural
networks like ANNs in medicine are described, and a detailed historical
background is provided. The connection between the artificial and the real thing
is also investigated and explained. Finally, the mathematical models involved
are presented and demonstrated. </P>

<H2>7. <A name=Conclusion>Conclusion</A></H2>
<P><FONT size=3>The computing world has a lot to gain from neural networks.
Their ability to learn by example makes them very flexible and powerful. Furthermore
there is no need to devise an algorithm in order to perform a specific task; i.e. there is
no need to understand the internal mechanisms of that task. They are also very well
suited for real time systems because of their fast response and computational times
which are due to their parallel architecture. </FONT></P> <P><FONT size=3>Neural
networks also contribute to other areas of research such as neurology and psychology.
They are regularly used to model parts of living organisms and to investigate the internal
mechanisms of the brain.</FONT></P> <P><FONT size=3>Perhaps the most exciting
aspect of neural networks is the possibility that some day 'conscious' networks might be
produced. There is a number of scientists arguing that consciousness is a 'mechanical'
property and that 'conscious' neural networks are a realistic possibility.</FONT></P><P>
<FONT size=3>Finally, I would like to state that even though neural networks have a
huge potential we will only get the best of them when they are intergrated with computing,
AI, fuzzy logic and related subjects.</FONT></P>

```

Abstract

Conclusion

Figura 3.6: Trecho do código HTML, destacando as seções *Abstract* e *Conclusion* do artigo *Neural Networks*.

subclasses, o que torna a classificação mais específica. Para isso foram criadas regras conforme as seguintes descrições:

- **Reconhecimento Direto de Classes Principais.** O reconhecimento das classes principais é efetuado quando o nome destas classes, ou suas subclasses, bem como seus sinônimos, desde que estes não estejam definidos em algum caso como sinônimos a serem ignorados, encontram-se presentes nas seções do artigo. No caso específico de IA, temos como exemplo de classes *Artificial Neural Networks*, *Knowledge Acquisition*, *Knowledge Engineering*, *Knowledge Representation Formalisms*, *Machine Learning*, *Ontology* e *Search*, definidas explicitamente como sendo classes a serem reconhecidas, como mostra a figura 3.7. Na figura 3.8 é mostrado o resultado desse reconhecimento para o artigo *Neural Networks* (vide Apêndice A.1), onde as classes *Knowledge Representation Formalisms* e *Artificial Neural Networks*



são reconhecidas no artigo.

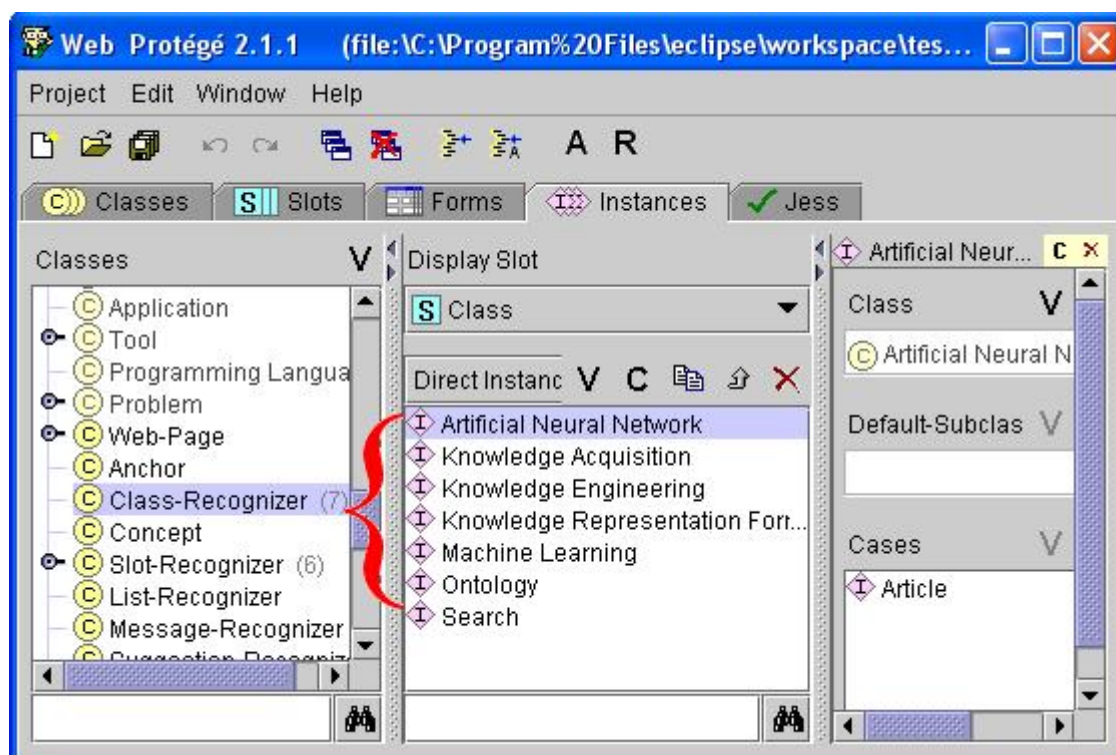


Figura 3.7: Classes principais definidas explicitamente para o domínio de IA.

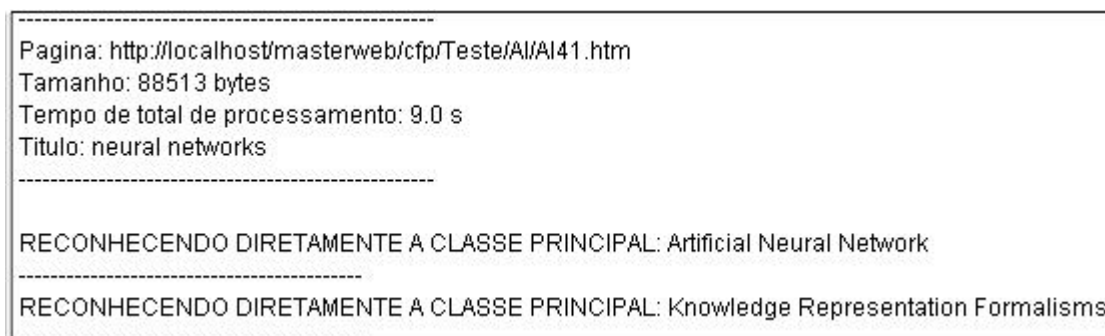


Figura 3.8: Resultado do reconhecimento direto de classes principais para o artigo *Neural Networks*.

- **Reconhecimento de Classes Através de Atributos.** Algumas relações diretas, ou seja, quando duas classes estão ligadas por meio de atributo, podem ser consideradas fortes ou fracas, dependendo do vínculo de dependência ou interação entre as partes. Por exemplo, a relação entre *Output Layer*

e *Artificial Neural Network* pelo atributo *Neurons*, é uma relação forte, enquanto que a relação *Inputs* entre as classes *Artificial Neural Network* e *Entry* é considerada uma relação fraca. Assim, essas ligações consideradas fortes devem ser explicitamente conhecidas dentro da base de conhecimento, como mostra a figura 3.9. Tendo posto essas relações, é possível encontrar no texto classes que conduzem ao reconhecimento de classes principais por meio dessas ligações. A figura 3.10 mostra o resultado desse reconhecimento para o artigo *Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models* (vide Apêndice A.1), onde a classe *Output Layer* está ligada a classe principal *Artificial Neural Network* pelo atributo *Neurons*.

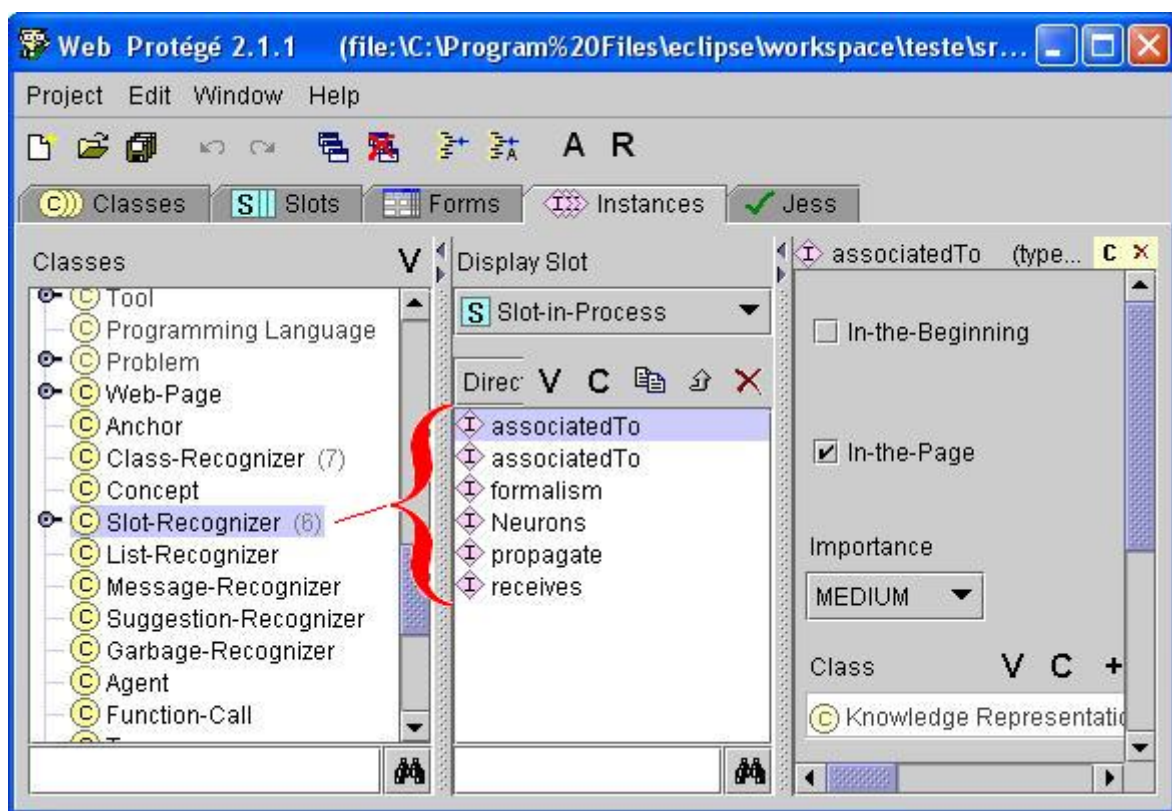


Figura 3.9: Atributos que identificam relações fortes entre as classes para o domínio de IA.

- **Reconhecimento de Classes Através de Relação Indireta.** Algumas classes não estão ligadas por meio de um relacionamento, mas por meio de

```
-----  
Pagina: http://localhost/masterweb/cfp/Teste/AI/AI20.htm  
Tamanho: 67342 bytes  
Tempo de total de processamento: 8.0 s  
Titulo: classification of binary document images into texture or non-textual data blocks  
        using neural network models  
-----  
  
RECONHECENDO CLASSE LIGADA DIRETAMENTE PELO SLOT:  
Output Layer esta ligada a Artificial Neural Network pelo slot Neurons  
-----
```

Figura 3.10: Resultado do reconhecimento de classes principais através de atributos no artigo *Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models*.

outras classes que atuam como ponte nesse relacionamento. Por exemplo, uma classe A está ligada a uma classe B que por sua vez está relacionada a uma classe C. Nesse caso dizemos que a classe A possui uma relação indireta com a classe C. Portanto, classes principais podem ser reconhecidas por meio dessas relações, quando classes que não são principais e possuem esse tipo de relação são identificadas no texto. Vamos exemplificar essa situação utilizando a figura 3.11, onde a classe *Knowledge Level* está associada à classe *Ontology* que por sua vez está relacionada, pelo atributo *formalism*, a *Domain-oriented KR Formalism* que é subclasse de *Knowledge Representation Formalisms*. Portanto, ao encontrar *Knowledge Level* estamos reconhecendo *Knowledge Representation Formalisms* e *Domain-oriented KR Formalism*, da mesma forma quando encontramos *Ontology* estamos reconhecendo *Knowledge Representation Formalisms*. A figura 3.12 mostra o resultado desse reconhecimento para o artigo *OWL Web Ontology Language Overview* (vide Apêndice A.1), onde a classe *Ontology* está ligada indiretamente a classe principal *Knowledge Representation Formalisms* por meio de dois relacionamentos, um com a classe *Language-oriented Formalism Knowledge Representation* e outro com *Domain-oriented KR Formalism*.

A última fase, ordenação e classificação, é descrita na próxima seção.

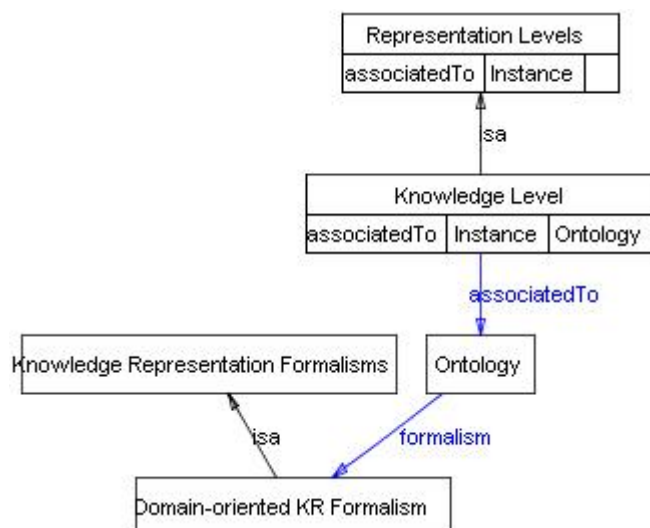


Figura 3.11: Relacionamento indireto entre as classes *Knowledge Level* e *Knowledge Representation Formalisms*.

```

-----
Pagina: http://localhost/masterweb/cfp/Teste/AI/AI21.htm
Tamanho: 73527 bytes
Tempo de total de processamento: 6.0 s
Titulo: owl web ontology language overview
-----

4- RECONHECENDO CLASSE LIGADA INDIRETAMENTE: Ontology
esta ligada a Language-oriented Formalism
que □ subclasse de Knowledge Representation Formalisms
-----

4- RECONHECENDO CLASSE LIGADA INDIRETAMENTE: Ontology
esta ligada a Domain-oriented KR Formalism
que □ subclasse de Knowledge Representation Formalisms
-----
  
```

Figura 3.12: Resultado do reconhecimento de classes principais através de relação indireta no artigo *OWL Web Ontology Language Overview*.

### 3.2.2.2 Ordenação e Classificação

É evidente que termos do domínio poderão ser encontrados por meio dos tipos de reconhecimentos apresentados, mas isso pode não garantir a classificação do artigo quanto à área que o termo pertence. O termo, por exemplo, pode aparecer no texto uma única vez ou ser reconhecido mais de uma vez através de um único tipo de reconhecimento. Por essa razão, objetivando aumentar a precisão da classificação, foram considerados a frequência dos termos no texto e o peso das seções onde os termos são encontrados.

Neste último caso, apenas a seção *resumo* possui um peso diferente das demais seções, por se tratar de uma síntese, onde o autor redige de forma breve a abrangência do artigo. Nessa seção devem estar ressaltados o tema da pesquisa, a idéia principal e detalhes importantes, por essa razão atribuímos um peso maior a seção *resumo*.

Desse modo, independente do tipo de reconhecimento, os termos identificados são extraídos, bem como suas freqüências e seções onde foram encontrados no artigo. Por conseqüência, adotou-se um esquema de pontuação onde cada classe principal reconhecida é pontuada em quatro itens, os quais posteriormente servirão para ordenar as classes reconhecidas e classificar o artigo quanto aos assuntos abordados. Os itens são:

- **Sinônimos** - Esse item é pontuado com a freqüência total dos termos encontrados no artigo referentes à classe principal em questão, isso inclui o número de sinônimos encontrados da classe principal, das suas subclasses e das classes relacionadas direta e indiretamente a classe principal, como mostra a fórmula abaixo.

$$\text{Sinônimos} = \sum \text{freqüência de cada termo reconhecido}$$

A tabela 3.1, mostra um exemplo para a pontuação do item Sinônimos referente a classe principal *Knowledge Representation Formalisms* para um artigo.

Reconhecimento	Sinônimos	Freqüência
Direto (Classes)	Knowledge Representation Formalisms	1
Direto (Subclasses)	Semantic Networks	6
Por atributo		
Relação indireta	Ontology	2
<b>Sinônimo</b>		<b>9</b>

Tabela 3.1: Exemplo da pontuação do item sinônimo referente a classe principal *Knowledge Representation Formalisms*.

- **Tipos de reconhecimento** - Classes principais podem ser reconhecidas de várias maneiras. Assim, uma classe principal pode ser reconhecida diretamente, através de atributos e/ou por relação indireta, como explicado na seção 3.2.2.1. Nesse caso, a pontuação é definida pelo somatório dos tipos de reconhecimento da classe principal. Quanto maior esse valor, maior a certeza da classificação. Para o exemplo apresentado pela tabela 3.1, a pontuação desse item é igual a 3, dado que dos quatro tipos de reconhecimento apenas o reconhecimento por atributos não apresentou nenhum termo.

$$\textit{Tipos de reconhecimento} = \sum \textit{tipos de reconhecimento da classe principal}$$

- **Tipos de seção** - Os termos do domínio podem ser encontrados em várias seções do artigo. Por conseguinte, o valor desse item é igual ao somatório do número total de seções onde cada termo é referenciado, como definido na fórmula abaixo.

$$\textit{Tipos de seção} = \sum \textit{número de seções onde cada termo é reconhecido}$$

Considerando o exemplo representado pela tabela 3.2, o valor deste item é igual a 4, pois o termo *knowledge Representation Formalisms* é encontrado na seção *other-sections*, já *Semantic Networks* aparece tanto na seção *other-sections* quanto em *conclusion*, e o termo *Ontology*, aparece duas vezes na seção *other-sections*.

- **Seção especial** - Se o artigo possui a seção *resumo* e pelo menos um termo for encontrado nesta seção, este item recebe como pontuação o número de vezes que os termos forem referenciados nesta seção.

$$\textit{Seção especial} = \sum \textit{frequência de cada termo reconhecido na seção resumo}$$

<b>Reconhecimento</b>	<b>Sinônimos</b>	<b>Frequência</b>	<b>Seções</b>
Direto (Classes)	Knowledge Representation Formalisms	1	other-sections
Direto (Subclasses)	Semantic Networks	6	(other-sections, conclusion)
Por atributo			
Relação indireta	Ontology	2	other-sections
<b>Tipos de seção</b>			<b>4</b>

Tabela 3.2: Exemplo da pontuação do item tipos de seção referente a classe principal *Knowledge Representation Formalisms*.

Os valores fornecidos em cada um dos itens, serão utilizados para determinar a ordenação das classes principais. A ordenação é iniciada pelo item *Sinônimos* e finalizada no item *Seção especial*, obedecendo ao nível de importância de cada um, que aumenta da esquerda para a direita, como ilustra a figura 3.13.



Figura 3.13: Níveis de importância das categorias para ordenação dos resultados da classificação.

A tabela abaixo é um exemplo do resultado da ordenação das classes principais reconhecidas em um determinado artigo:

<b>Classe Principal</b>	<b>Sinônimo</b>	<b>Tipos de reconhecimento</b>	<b>Tipos de seção</b>	<b>Seção especial</b>
Knowledge Representation Formalisms	4	3	2	1
Ontology	37	1	2	0
Search	1	1	0	0

Tabela 3.3: Exemplo do resultado da ordenação das classes principais reconhecidas em um artigo.

A área atribuída ao artigo é inferida a partir dos valores dos quatro itens, obedecendo a pelo menos uma das seguintes premissas:

- Sinônimos maior que 2;
- Tipos de reconhecimento maior que 1;
- Tipos de seção maior que 1;
- Seção especial maior ou igual a 1.

Pela análise de alguns artigos consideramos os valores acima como boas heurísticas para classificação, além disso privilegiamos as classes principais para garantir um bom ordenamento. Observe que a ordenação atribui o maior nível de importância a seção *resumo*, considerando que geralmente é possível encontrar nessa seção o assunto de que trata o artigo, ou seja, termos importantes para classificá-lo.

Analisando o exemplo da tabela 3.3, apenas *Search* não é atribuído como assunto tratado pelo artigo. A ausência de termos relacionados à classe *Search* na seção *resumo* e a insatisfação dos valores das demais categorias, de acordo com as três primeiras premissas acima, declaram que o artigo apenas cita a área *Search* e fala sobre *Knowledge Representation Formalisms* e mais especificamente sobre *Ontology*.

### 3.2.3 O Motor de Inferência Jess

O motor de inferência Jess (*Java Embedded Expert System Shell*) (FH03) é utilizado no MASTER-WEb, permitindo que os agentes possam derivar conclusões a partir das premissas. Todas as regras da base de conhecimento do MASTER-WEb estão escritas em Jess. Jess é implementado em Java e possui uma boa integração entre objetos Java e os formalismos de regras de produção. Segundo (Fre02), entre outras qualidades do Jess, podem ser citadas a capacidade de raciocinar e manipular diretamente sobre objetos, métodos e variáveis Java dentro de regras Jess, a possibilidade de criação de regras e fatos Jess dentro de código Java, a passagem de objetos nos dois sentidos entre Java e Jess. Além disso, o componente JessTab provê integração entre Jess e Protégé, formando uma ponte entre essas duas ferramentas, permite mapeamento de bases de conhecimento do Protégé em fatos no Jess e a manipulação dessas bases de



conhecimento (Eri04), habilitando o Jess a reusar, definir e refinar ontologias através do Protégé.

Regras em Jess tem duas partes separadas pelo símbolo " $=>$ ", é algo semelhante ao *se ... então* da programação procedural, mas usada de forma diferente, ou seja, enquanto declarações *se ... então* são executadas em um específico tempo e em uma específica ordem, regras em Jess são executadas todas as vezes que a primeira parte é satisfeita. Abaixo é mostrado um exemplo de uma regra feita no Jess (FH01):

```
Jess> (defrule lista
      (lista-de-compras $?list)
      =>
      (printout t "Eu preciso comprar: " $?list crlf))
TRUE
```

Nesse exemplo a regra será ativada quando o fato (lista-de-compras \$?list) aparecer na base de conhecimentos. Quando a regra dispara é impresso na tela uma frase informando os itens da lista de compras, onde \$?list se refere a um campo multivalorado.

Em seguida é mostrada a adição do fato *lista-de-compras* com seus itens, logo após o motor de inferência é executado através do comando run, por fim a regra lista dispara imprimindo na tela a lista de compras.

```
Jess> (assert (lista-de-compras ovos leite carne))
<Fact-0>
Jess> (run)
Eu preciso comprar (ovos leite carne)
1
```

### 3.2.4 Considerações

A arquitetura do MASTER-Web, ao utilizar uma abordagem baseada em conhecimento facilitou sua extensão, proporcionado reuso e compartilhamento da estrutura em ter-

---

mos de código e de boa parte do conhecimento de que dispõe, tanto em termos de ontologias e instâncias, como de regras. Assim, nesse modelo foi possível adicionar novas regras sem reescrever todo o sistema, significando uma importante vantagem comparada à programação procedural que implicitamente captura o conhecimento. A combinação de Regras de Produção, com sua forma efetiva de representação declarativa do conhecimento, e a utilização de uma linguagem ontológica que fornece uma representação estrutural concisa de relações e definição hierárquica, gera um ambiente expressivo e consistente.

# Capítulo 4

## Experimentos

Neste capítulo, apresentamos a metodologia usada para realizar os experimentos e os resultados obtidos. Descrevemos a base de artigos científicos utilizada nos experimentos e mostramos os resultados empíricos. Por fim, na seção 4.2, são realizadas as discussões a respeito dos resultados.

### 4.1 Metodologia

As regras para classificação de artigos foram construídas de forma que o conhecimento do domínio fosse capturado, considerando não apenas os conceitos, mas a hierarquia e os relacionamentos existentes entre as classes, independentemente do domínio. Para isso foram utilizados a ontologia de IA e artigos fictícios, ou seja, artigos que tiveram sua composição modificada pela alteração das seções e inserção de termos, com o propósito de ajustar as regras ao domínio. Após a construção das regras foram realizados testes cegos, acompanhados pela observação e análise dos resultados, como mostra a seção 4.2. A seguir fazemos uma descrição da base experimental usada e logo após apresentamos os resultados experimentais.

### 4.1.1 Base experimental

Com o propósito de avaliar o MASTER-Web para classificação de artigos científicos, utilizamos como dados experimentais uma base de texto composta por 406 artigos do domínio de IA e de outros domínios, como outros ramos da Computação, Medicina, Biologia, Economia, Filosofia entre outros. A tabela 4.1 descreve como estão distribuídos esses artigos por domínio, visto que alguns artigos tratam sobre temas que envolvem mais de um domínio:

<b>Domínio</b>	<b>Quantidade</b>
Inteligência Artificial	274
Outras áreas da Computação	65
Biologia	23
Direito	8
Economia	10
Estatística	1
Filosofia	9
Física	3
Lingüística	4
Matemática	4
Medicina	16
Psicologia	20
Química	2
Sociologia	3

Tabela 4.1: Distribuição dos artigos por domínio.

Os artigos que formam a base de textos possuem um caráter heterogêneo não apenas no que diz respeito aos assuntos abordados, mas também na sua divisão em seções. Alguns artigos apresentam o texto dividido em diferentes seções e subseções, por exemplo, resumo, sumário, introdução, trabalhos relacionados e bibliografia, enquanto outros possuem uma divisão menos preocupada com as diversas partes de construção do texto e consideram basicamente duas seções, a seção principal, de que trata o assunto do artigo, e a bibliografia.

Toda a base experimental é composta por documentos digitalizados em formato HTML, obtidos em sites de universidades, professores, revistas, entre outros, conforme estão listados no apêndice B.1. Esta linguagem não oferece nenhum padrão que permita a divisão do texto, como *tags* especiais que definam as seções e subseções. As *tags* HTML apenas descrevem como a página deve ser exibida, definindo por exemplo tipo, cor e tamanho da fonte, por essa razão a formatação das seções e subseções escritas em HTML, varia de artigo para artigo. Desse fato decorre que a base experimental é composta por documentos, também, heterogêneos na formatação usada para destacar as seções e subseções do artigo. Esse fato pode ser observado nas figuras 4.1 e 4.2, a primeira ilustra um trecho do artigo *Neural Networks* (vide Apêndice A.1), onde foi usado o par `<H2> </H2>` para destacar as seções, a segunda se refere ao artigo *Analysis of Three-Dimensional Protein Images* (vide Apêndice A.2), onde são usadas as *tags* `<STRONG> </STRONG>` para destacar a seção *Abstract*, e `<H1> </H1>` para a seção *Introduction*. Considerar a estrutura do documento é importante e se justifica pelo reconhecimento das seções relevantes do texto e posterior extração de informação nessas seções.

```
<H2>Abstract</H2> ←
<P>This report is an introduction to Artificial Neural Networks. The various
types of neural networks are explained and demonstrated, applications of neural
networks like ANNs in medicine are described, and a detailed historical
background is provided. The connection between the artificial and the real thing
is also investigated and explained. Finally, the mathematical models involved
are presented and demonstrated. </P>
<H2><A name=Contents>Contents</A>.</H2> ←
```

Figura 4.1: Trecho do código HTML, destacando as marcações usadas para destacar as seções *Abstract* e *Contents* do artigo *Neural Networks*.

```
<P><STRONG>Abstract:</STRONG><BR>
<P></P>
<P>A fundamental goal of research in molecular biology is to understand protein structure.
Protein crystallography is currently the most successful method for determining the
three-dimensional (3D) conformation of a protein, yet it remains labor intensive and relies on
an expert's ability to derive and evaluate a protein scene model. In this paper, the problem of
protein structure determination is formulated as an exercise in <l>scene analysis.</l> A
computational methodology is presented in which a 3D image of a protein is segmented into
a graph of critical points. Bayesian and certainty factor approaches are described and used to
analyze critical point graphs and identify meaningful substructures, such as <l>alpha</l> -helices
and <l>beta</l> -sheets. Results of applying the methodologies to protein images at low and medium
resolution are reported. The research is related to approaches to representation, segmentation and
classification in vision, as well as to top-down approaches to protein structure prediction. </P>
<P></P><BR>
<H1>Introduction</H1>
```

Figura 4.2: Trecho do código HTML, destacando as marcações usadas para destacar as seções *Abstract* e *Introduction* do artigo *Analysis of Three-Dimensional Protein Images*.

## 4.2 Estudo de caso e resultados

### 4.2.1 Estudo de caso

Antes de apresentarmos os resultados, precisamos primeiro fazer uma breve discussão a respeito do estudo de caso que conduziu aos resultados atingidos. Os experimentos realizados classificam o artigo com base na análise das seções do texto consideradas relevantes e no domínio ontológico. O artigo é classificado como sendo ou não do domínio em questão, no caso IA, atribuindo ao artigo as áreas de abrangência, ou seja, é feita uma classificação de menor granularidade. Assim, as próximas seções mostram esses resultados considerando também essa granularidade e percentuais de desempenho da classificação.

### 4.2.2 Resultados

Os resultados obtidos apresentaram excelentes percentuais de desempenho do sistema, como mostra a tabela abaixo. Para analisar os resultados apresentados, é preciso levar em conta o escopo da ontologia do domínio investigado. O escopo da ontologia usada compreende as áreas de Busca, Representação do Conhecimento, Redes Neurais e Aprendizado de Máquina. Os artigos que embora sejam da área de IA, mas não

tratam de áreas compreendidas na ontologia, não são classificados dentro deste domínio. Nesse contexto, a classificação garante um reconhecimento baseado na especificidade dos conceitos ontológicos, restringindo o domínio às classes e relações explicitamente definidas na ontologia. Visto que um artigo pode abranger mais de uma área, por exemplo, um artigo pode falar tanto de *Machine Learning* quanto de *Search*, a tabela 4.2 mostra os resultados da classificação dos artigos por área de reconhecimento.

Reconhecimento	Corretos	Falsos Positivos	Falsos Negativos	Acerto (%)
Artificial Neural Network	48	1	2	94,1
Knowledge Acquisition	17	0	1	94,4
Knowledge Engineering	3	0	0	100,0
Knowledge Representation Formalisms	56	9	1	84,8
Machine Learning	51	2	6	86,4
Ontology	19	0	0	100,0
Search	38	1	1	95,0
Outros domínios	228	7	11	92,7

Tabela 4.2: Percentuais de acerto quanto à classificação de artigos por área de reconhecimento.

Quanto ao reconhecimento das seções dos artigos, o sistema também apresentou resultados bastante significativos, apresentado um percentual de acerto igual a 82,3%.

Vejamos agora alguns resultados inferidos e exibidos pelo sistema. Primeiramente mostramos um artigo classificado como sendo de IA, o artigo *Neural Networks* (vide Apêndice A.1), cujo resultado é apresentado pela figura 4.3.

```

----- CLASSIFICATION ARTICLE -----
http://localhost/masterweb/cfp/Teste/AI/AI41.htm
Title: neural networks
--- The article is about:
Feed Forward ANN
Supervised Learning ANN
--- Citations:
Knowledge Representation Formalisms
Expert System
-----

```

Figura 4.3: Resultado da classificação do artigo *Neural Networks*

Para o artigo *Neural Networks*, o sistema afirma que o texto trata especificamente de *Feed Forward* e *Supervised Learning* e, ainda afirma que o artigo faz apenas uma referência a *Knowledge Representation Formalisms*. Realmente trata-se de um artigo sobre Redes Neurais, onde o autor fornece uma visão geral sobre o tema, investiga as similaridades entre neurônios artificiais e humanos e descreve a arquitetura das redes *Feed-forward* e métodos de aprendizado usados em redes neurais. De fato, o artigo faz uma pequena referência à representação do conhecimento.

Na figura 4.4, vemos um resultado onde o sistema faz uma afirmação veemente de que o artigo *Advances in Distributed Security* (vide Apêndice A.2) não pertence a IA. O sistema novamente acerta, pois esse artigo não fala de nenhuma subárea de IA nem mesmo faz alguma referência a esta área.

```

----- CLASSIFICATION ARTICLE -----
http://localhost/masterweb/cfp/Teste/NAI/NAI85.htm
Title: advances in distributed security
is NOT Artificial Intelligence article
-----

```

Figura 4.4: Resultado da classificação do artigo *Advances in Distributed Security*

A seguir temos o resultado da classificação do artigo *Analysis of Three-Dimensional Protein Images* (vide Apêndice A.2), onde o sistema não o considera como sendo do domínio de IA, mas percebe que o artigo faz citações às subáreas de IA. Este resultado é apresentado na figura 4.5.

```

----- CLASSIFICATION ARTICLE -----
http://localhost/masterweb/cfp/Teste/NAI/NAI73.htm
Title: analysis of three-dimensional protein images
--- The article is about:
--- Citations:
Machine Learning
Knowledge Representation Formalisms
Expert System
-----

```

Figura 4.5: Resultado da classificação do artigo *Analysis of Three-Dimensional Protein Images*.

Analisando o artigo *Analysis of Three-Dimensional Protein Images*, percebemos que o texto diz respeito a pesquisas na área de Biologia e usa técnicas de IA e de Química



para analisar estrutura de proteínas, possuindo características de multidisciplinaridade. Ainda assim, o método mais usado da área de IA para interpretar imagens de proteínas, segundo o texto, é Visão de Máquina (*Machine Vision*), do qual a ontologia de IA não aborda. Dessa forma os temas apontados pelo sistema como citações não passam, realmente, de referências feitas pelo autor.

Na figura 4.6, o sistema afirma que o artigo *AI Planning Resources on the Web* (vide Apêndice A.2) não é da área de IA, embora trate de *Planning*, tema que a ontologia de IA não cobre.

```
----- CLASSIFICATION ARTICLE -----  
http://localhost/masterweb/cfp/Teste/GAI/GAI4.htm  
Title: ai planning resources on the web  
is NOT Artificial Intelligence article  
-----
```

Figura 4.6: Resultado da classificação do artigo *AI Planning Resources on the Web*

Os resultados da classificação revelam, ainda, que os falsos positivos se caracterizam por serem artigos do domínio de IA, mas que apenas citam subáreas presentes na ontologia confundindo assim o sistema quanto a sua classificação. Como é o caso do falso positivo *Information Retrieval Issues* (vide Apêndice A.2). Os falsos negativos são provenientes de artigos que utilizam muitas abreviaturas ou termos ausentes na lista de sinônimos da ontologia, ou por empregar muitos termos genéricos, ou seja, termos usados em muitos domínios, podendo ter múltiplos significados, e por essa razão tais termos são ignorados no processo de classificação. Um exemplo de falso negativo é o artigo *Learning Sets of Related Concepts: A Shared Task Model* (vide Apêndice A.2).

Foi realizado um outro experimento utilizando como base de testes apenas os artigos classificados corretamente pelo MASTER-Web como sendo do domínio de IA. Neste experimento foi analisada apenas a seção *resumo*, ou seja, apenas a seção *resumo* foi considerada relevante, as demais seções foram desprezadas. O objetivo desse experimento é identificar se apenas essa seção é suficiente para classificar o artigo como um todo. Os resultados obtidos revelaram baixo percentual de acerto, classificando corretamente apenas 42,6(%). Analisando esses artigos foi observado que o conteúdo dessa

seção é em sua maioria abstrato, não retrata de forma concreta os temas abordados pelo artigo, dificultando a classificação da informação por parte do sistema, sobretudo a classificação em um menor nível de granularidade.

### 4.3 Discussão dos Resultados

Os resultados apresentados na seção anterior mostram que a ferramenta apresentou um desempenho bastante significativo, com percentuais de acertos relevantes. As subáreas atribuídas ao artigo buscam não apenas garantir a classificação do artigo mas identificar temas de interesse do usuário.

O sucesso dessa classificação deve-se, em parte, à escolha das seções do artigo, onde seções que não abordam especificamente o tema são ignoradas, enquanto as seções que desenvolvem o assunto, descrevendo técnicas e comunicando resultados são consideradas e delas são extraídos os termos do domínio. Por outro lado, fica evidente o papel da ontologia na classificação, a sua amplitude e especificidade têm implicações diretas nos resultados. Quanto maior a profundidade hierárquica da ontologia, maior poderá ser a especificidade da classificação. As relações existentes entre conceitos da ontologia favorecem a inferência do sistema, pois pela ligação de termos distintos pode-se chegar a um novo conceito. Por exemplo, a partir da relação entre *Expert System* e *Inference Engine* chega-se a *Knowledge Representation Formalisms*.

Alguns erros no reconhecimento das seções implicaram em falhas na classificação dos artigos. Esses erros foram ocasionados pela ausência de *tags* que destaquem seções como bibliografia, daí seções irrelevantes não foram reconhecidas e por consequência foram misturadas às seções relevantes. Outro fator de erro é a falta de termos usados por alguns artigos na lista de sinônimos da ontologia, o que evidencia a pluralidade de conceitos na área de IA. Quando considerada a abrangência da ontologia o percentual de acerto é elevado, assim é preciso levar em conta a classificação do artigo em relação à extensão compreendida pelo domínio ontológico e não por todo domínio de IA.

É possível concluir ainda, que a medida em que o vocabulário do domínio é ampliado

e a ontologia é refinada, a classificação por consequência se estende a esses novos conceitos, visto que o conhecimento é expandido e as regras de classificação são genéricas. Assim, novos subdomínios são reconhecidos aumentando o número de acertos.

# Capítulo 5

## Trabalhos Relacionados

### 5.1 Técnicas de Classificação

Várias técnicas têm sido aplicadas para problemas de classificação de texto. Neste capítulo, são brevemente descritas técnicas baseadas em conhecimento e técnicas baseadas em aprendizagem automática.

#### 5.1.1 Técnicas baseadas em conhecimento

Os sistemas de classificação baseados em conhecimento adotam uma abordagem apoiada em uma base de conhecimento existente e algum mecanismo de inferência associado. Possuem um forte propósito de classificação textual, com uma tentativa de incorporar semântica às informações, possibilitando que os documentos expressem significado e ao mesmo tempo estabeleçam regras para que as máquinas possam processar esse significado. O objetivo é dar significado semântico ao conteúdo dos documentos, criando um ambiente onde agentes de software e usuários possam trabalhar de forma cooperativa.

O principal componente de um sistema baseado em conhecimento é a sua base de conhecimento. A base de conhecimento é um conjunto de representação de fatos e regras sobre o domínio em questão (RN03). Assim, a engenharia de conhecimento, nome dado ao processo de construção da base de conhecimento, desempenha um papel

importante nesse tipo de técnica, uma vez que para realizar classificação automática de textos necessita-se de taxonomias dos domínios investigados.

Na base de conhecimento estão as sentenças em uma linguagem de representação do conhecimento que o agente de software possa entender. Por exemplo (JM02), o padrão:

```
(gold (&n (reserv ! medal ! jewelry))
```

significa detectar a palavra 'gold', mas ignorar as sentenças 'gold reserv', 'gold medal' e 'gold jewelry'.

A base de conhecimento deve ser clara, concisa, correta, não ambígua, devendo representar apenas os objetos e relações relevantes do mundo em questão.

Para (Ril94), abordagens tipicamente baseadas em conhecimento requerem um extensivo esforço para criar bases de conhecimentos, e todo processo de engenharia de conhecimento é repetido a cada novo domínio investigado. De fato, a utilização de especialistas para a construção manual da base de conhecimento é um processo custoso e lento, motivando o estudo de categorização automática de documentos utilizando técnicas de aprendizagem de máquina.

### 5.1.2 Técnicas baseadas em aprendizado de máquina

Os classificadores que utilizam técnicas de aprendizagem de máquina utilizam os mais diversos algoritmos de aprendizagem automática, dentre eles, destacam-se: Redes Bayesianas, Árvores de Decisão e Algoritmos Genéticos.

Uma alternativa à construção manual de classificadores baseados em conhecimento é usar técnicas de aprendizado de máquina para gerar classificadores, a fim de melhorar o desempenho e a precisão das respostas. Aprendizado de máquina pode ser definido como uma área que pesquisa métodos computacionais relacionados à aquisição automática de novos conhecimentos, novas habilidades e novas formas de organizar o conhecimento existente (CDF<sup>+</sup>98).

A idéia dos sistemas de aprendizagem automática é a criação de regras de conhecimento a partir de exemplos. Existem dois métodos principais utilizados para aprendizado de máquina. O primeiro é o aprendizado supervisionado, onde o conjunto de dados do qual se pretende extrair conhecimento já vem rotulado, isto é, a cada instância está associada sua classificação, a qual o algoritmo deve aprender a realizar (PN02). No aprendizado não supervisionado, também chamado de *clustering*, o algoritmo tenta agrupar os dados tal que dados em cada grupo tenham mais similaridade entre si que com outros dados em outros grupos, sob uma particular métrica de similaridade (JD88).

O método supervisionado é uma tarefa de classificação, e possui várias técnicas de aprendizagem dentre as quais se destacam: redes Bayesianas, árvores de decisão, algoritmos genéticos, redes neurais, inferência lógica, etc. Algumas destas técnicas são descritas brevemente:

- **Redes Bayesianas** - Seguindo o paradigma estatístico, o algoritmo faz uso de fórmulas estatísticas e cálculo de probabilidades para realizar a classificação (Mit97).
- **Árvores de Decisão** - Consiste de uma hierarquia de nós internos e externos que são conectados por ramos. O nó interno é a unidade de tomada de decisão que avalia através de teste lógico qual será o próximo nó descendente ou filho. Em contraste, um nó externo (não tem nó descendente), também conhecido como folha ou nó terminal, está associado a um rótulo ou a um valor. Em geral, no procedimento de uma árvore de decisão um conjunto de dados é apresentado ao nó inicial (ou nó raiz que também é um nó interno) da árvore; dependendo do resultado do teste lógico usado pelo nó, a árvore ramifica-se para um dos nós filhos e este procedimento é repetido até que um nó terminal é alcançado. A repetição deste procedimento caracteriza a recursividade da árvore de decisão (Sob03).

- **Algoritmos Genéticos** - São algoritmos de busca probabilística baseada no princípio de seleção natural, evolução e hereditariedade (MK00).

## 5.2 Recuperação de Informação

Recuperação de informação é um termo amplo. Para alguns autores é o processo de identificar, no conjunto de documentos (*corpus*), quais atendem à necessidade de informação do usuário. Para (All00), a recuperação de informação é um conjunto de técnicas que servem ao propósito de encontrar documentos que são relevantes para uma necessidade de informação. A tarefa de classificar documentos dentro de um conjunto de documentos relevantes e irrelevantes pode ser considerada como uma tarefa de classificação textual (JM02).

O processo de recuperação de informação utiliza modelos de acesso à informação que apresentam estratégias relevantes a uma consulta. Sendo a maioria dos modelos de natureza quantitativa, baseados em lógica, estatísticas e teoria dos conjuntos, porém outros modelos têm sido propostos como os modelos com bases de conhecimento e redes neurais.

Neste sentido são apresentadas a seguir algumas das técnicas e modelos utilizados no processo de recuperação pelos sistemas de buscas.

### 5.2.1 Técnicas de Recuperação de Informação

#### 5.2.1.1 *Stopwords*

*Stopwords* são termos muito genéricos sem significado específico, cujo conteúdo semântico é limitado, tais como preposições, artigos e conjunções. Exemplos de *stopwords* em Inglês são 'the', 'by', 'of' e 'this'. Qualquer linguagem natural contém muitas *stopwords* que ocupam um bom percentual de todas as ocorrências de palavras. A remoção desses termos significa a remoção de palavras consideradas não relevantes para uso na análise do texto, o que diminui o número de palavras a serem examinadas no documento. A

partir de listas contendo esses termos, as chamadas *stoplists*, são analisadas as palavras presentes em uma requisição, removendo as palavras que são iguais a alguma presente nas *stoplists*. A figura 5.1 ilustra um exemplo de remoção de *stopwords*.

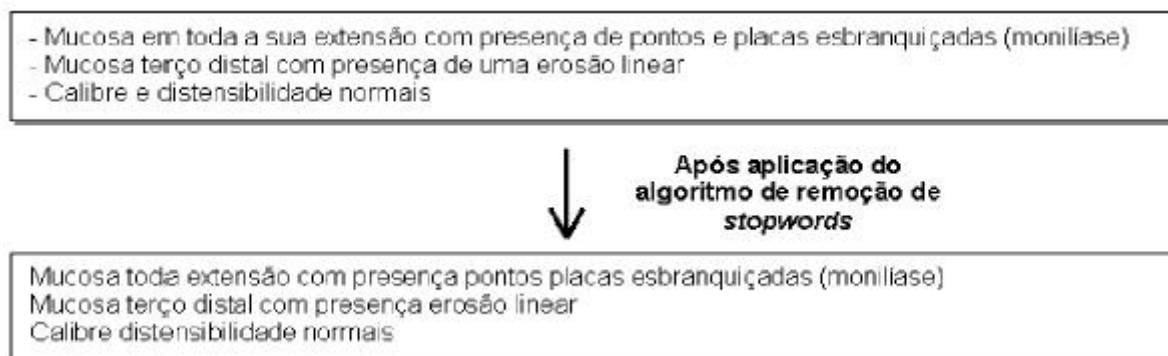


Figura 5.1: Exemplo de Remoção de *Stopwords* (HLM<sup>+</sup>04).

### 5.2.1.2 *Stemming*

*Stemming* ou radicalização é uma técnica para transformar diferentes flexões e derivações da mesma palavra em um *stem* (radical) comum. De acordo com (JW97), *stemming* consiste em reduzir as palavras ao mesmo *stem*, por meio da retirada dos afixos<sup>1</sup>, permanecendo apenas a raiz da palavra. O propósito, segundo (HLOS00), é chegar a um *stem* que captura uma palavra com generalidade suficiente para permitir sucesso na combinação de caracteres, mas sem perder muito detalhe e precisão. O *stem* de uma palavra é sua forma mais básica: o *stem* de um plural é o singular; o *stem* do verbo no tempo passado é o verbo no infinitivo. Um exemplo típico de um *stem* é 'conect' que é o *stem* de 'conectar', 'conectado' e 'conectando'.

O *stemming* reduz o tamanho do dicionário de termos (conjunto de termos indexáveis) resultantes no processo de indexação numa coleção de textos e permite pesquisar variantes morfológicos<sup>2</sup> dos termos especificados pelo usuário. Porém esta técnica apresenta a desvantagem de perder a informação correspondente às palavras radicalizadas

<sup>1</sup>Pré-fixos e sufixos.

<sup>2</sup>Modo que as palavras são construídas a partir de unidades menores chamadas morfemas (JM00)



e além da necessidade de espaço adicional para guardar tanto o radical como a palavra original. Essa desvantagem é decorrente de dois erros chamados *overstemming* e *understemming*. *Overstemming* se dá quando a cadeia de caracteres removida não é um sufixo, mas parte do *stem*. Por exemplo, a palavra gramática, após ser processada por um *stemmer*, é transformada no *stem* grama. Neste caso, a cadeia de caracteres removida eliminou parte do *stem* correto, a saber 'gramát'. Já *understemming* ocorre quando um sufixo não é removido completamente. Por exemplo, quando a palavra 'referência' é transformada no *stem* 'referênc', ao invés do *stem* considerado correto 'refer' (Cha03).

Para (CDHK01) a gravidade do problema depende tanto das regras de *stemming* quanto da coleção de documentos (mais precisamente dos termos usados para indexar a coleção de documentos). Ambos os conjuntos de regras e de termos indexáveis influenciam a quantidade de índices combinados sobre o mesmo *stem*.

Ao final da aplicação desse método as informações estão normalizadas conforme ilustra a figura 5.2.

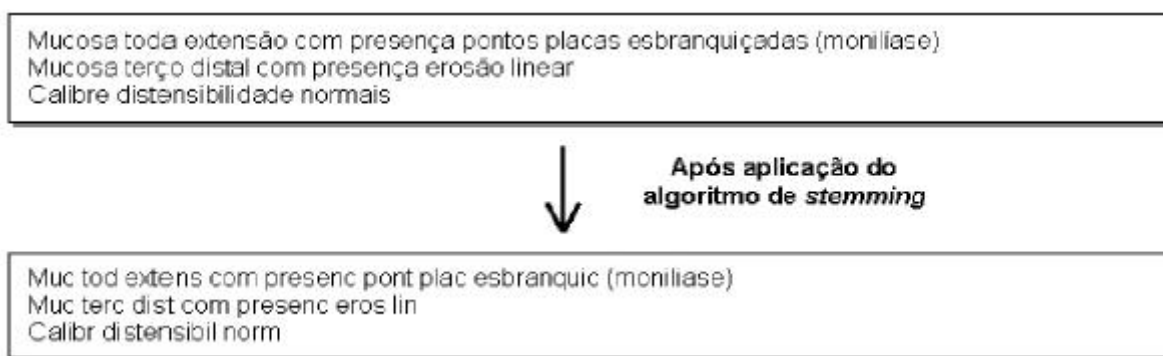


Figura 5.2: Exemplo de Aplicação de *Stemming* (HLM<sup>+</sup>04).

### 5.2.1.3 Indexação

Segundo (Hid01), a indexação é uma técnica que utiliza uma linguagem índice como sendo uma linguagem usada para descrever documentos e requisições. Onde os elementos da linguagem índice são termos de indexação, que podem ser derivados do texto do documento ou podem ser gerados independentemente. Assim, para essa técnica,

cada documento é descrito por um conjunto de palavras-chave, chamadas termos de indexação. De acordo com (Fer03), um termo de indexação é geralmente uma palavra que representa um conceito ou significado presente no documento.

A escolha dos termos de indexação e sua associação ao documento como forma de descrever o conteúdo não é uma tarefa fácil. Para (Fer03) existem termos mais representativos do assunto principal do documento e outros termos que representam assuntos periféricos à temática do mesmo.

Alguns fatores são empregados para avaliar o potencial representativo de um termo de indexação. Por exemplo, o número de ocorrências de uma palavra no documento, ou o aparecimento de uma palavra em milhares de documentos, ou ainda uma palavra que aparece em apenas três documentos podem indicar o grau de utilidade da palavra como termo de indexação e a importância do documento para uma dada requisição. Assim cada termo de indexação possui graus de relevância distintos, de acordo com os documentos e objetivos da requisição.

A figura 5.3 abaixo representa um exemplo de uma consulta em um sistema que utiliza a técnica de indexação.

- 1 - Expressão digitada  
*"pesquisa sobre processamento da língua natural na Espanha"*
  - 2 - Remover as *stop words*  
*"pesquisa processamento língua natural Espanha"*
  - 3 - Derivar os *stems* das palavras que restaram da etapa anterior:  
*"pesquis process língu natur Espanh"*
  
  - Índice
    - **524** **pesquis**
    - **258** **process**
    - **365** **língu**
    - **101** **natur**
    - **654** **Espanh**
  
  - 4 - Os *stems* são substituídos pelos números contidos no índice:  
*"524 258 365 101 654"*
  - 5 - O número de ocorrências de cada termo é contabilizado.  
*"101 | 258 | 365 | 524 | 654 |"*
- As etapas posteriores são dependentes do modelo de RI utilizado por cada máquina de busca.

Figura 5.3: Exemplo de uma consulta em um sistema que utiliza indexação (Cha03).

## 5.3 Sistemas Similares

Pesquisas com objetivos de recuperar e classificar informação têm usado ontologias para representar conhecimento e comparar conteúdo. Esta seção descreve sistemas que realizam tarefas similares ao uso de ontologias e classificação textual no MASTER-Web. O sistema OntoSeek (*Content-Based Access to the Web*), descrito na próxima subseção, usa ontologias para recuperar informação a partir de páginas amarelas e catálogos de produtos na Web, enquanto que o SARI (*A System for Semantical Information Retrieval*), descrito na subseção 5.3.2, utiliza ontologias para descrever a estrutura conceitual da informação em um sistema de recuperação de informação semântica.

### 5.3.1 OntoSeek

O OntoSeek (GMV99) é um sistema planejado para recuperação de informação baseado em conteúdo a partir de páginas amarelas e catálogos de produtos na Web. Combina um mecanismo de comparação de conteúdo dirigido por ontologia com um formalismo de representação moderadamente expressivo. Esse sistema utiliza uma ontologia ampla baseada em WordNet (Mil95). O WordNet é um banco de dados lingüístico que cobre palavras em Inglês. O OntoSeek usa uma ontologia lingüística acoplada a uma representação estruturada para acessar páginas amarelas.

Apenas pela adoção de simples técnicas de comparação de palavras usando uma pequena lista de sentidos de palavras para comparar conteúdo, o sistema consegue resultados com uma qualidade sensivelmente melhorada em relação a uma simples lista de palavra. Além da descrição codificada dos recursos, o sistema provê a formulação de consultas através de uma interface orientada pela ontologia, apresentando uma característica peculiar onde a ontologia não apenas afeta na relevância do conhecimento recuperado, mas também influencia na interação com o sistema. Assim, a ontologia torna-se um meio de comunicação entre o usuário e o sistema.

### 5.3.2 SARI

De acordo com (Tav98), SARI é um sistema destinado a agir como um *broker* entre usuários humanos, que necessitam de informação, e outros sistemas computacionais, que fornecem informações heterogêneas com seus motores de busca. A arquitetura do SARI reflete o papel do sistema como um *broker* entre seus usuários e fontes de informação, como mostra a figura 5.4. O SARI tem agentes dos seguintes tipos:

- Os Agentes de Aplicação representam os usuários (seres humanos ou outros sistemas computacionais) do sistema SARI;
- Agentes de Busca mediam fontes de informação. Traduzem requisições vindas de Agentes de Controle para a linguagem de suas fontes de informação, e envia o resultado de volta aos Agentes de Controle;
- Agentes de Controle agem como *broker* no sistema SARI. Cada Agente de Controle recebe mensagens dos Agentes de Aplicação contendo requisições de recuperação de informação, e decide para qual Agente de Busca serão enviadas as requisições. Envia mensagens contendo as requisições de recuperação aos Agentes de Busca apropriados, recebe dos Agentes de Busca as mensagens contendo resultados da busca, combina-os dentro do resultado de recuperação de informação, e envia os resultados de recuperação de volta aos Agentes de Aplicação;
- O Agente de Ontologia contém metadados na forma de ontologias que descrevem a estrutura conceitual da informação presente nas fontes de informação usadas pelo SARI.

Além desses, existem os Agentes Provedores de Conteúdo que representam provedores de conteúdo do sistema SARI. Provedores de Conteúdo são organizações ou indivíduos que possuem uma ou mais fontes de informação que são acessíveis no sistema SARI.

Os Agentes de Controle formam o coração do SARI. Tomam decisões com base no perfil do usuário, e dos metadados sobre a informação a ser recuperada nas ontologias.

O conteúdo de qualquer requisição de recuperação de informação oriunda de algum Agente de Aplicação é traduzido para a linguagem de requisição SAL (SARi query Language) antes de ser remetido ao Agente de Controle.

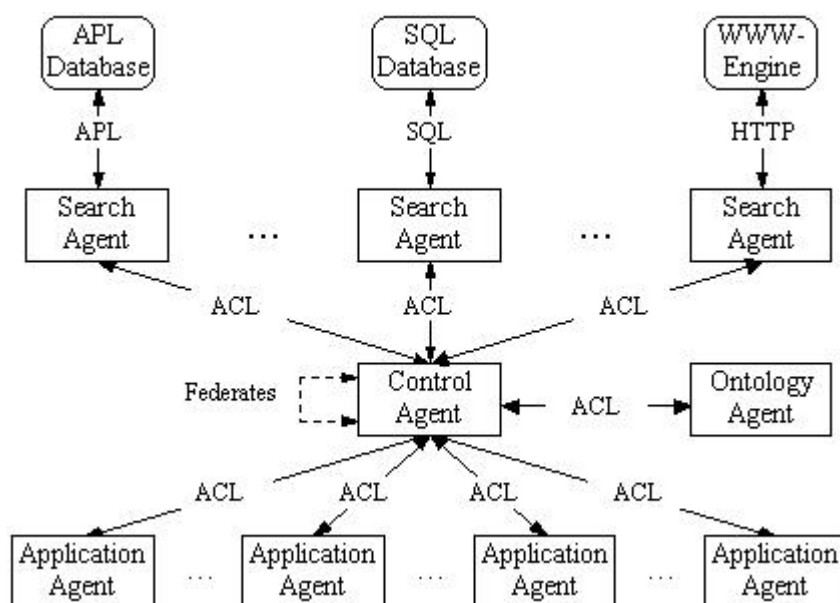


Figura 5.4: Arquitetura do SARI (Tav98).

A estrutura conceitual da informação contida nas fontes de informação disponível para o SARI é descrita por ontologias.

Um dos problemas mais importantes que precisa ser resolvido em recuperação de informação semântica de fontes heterogêneas é conciliar conceitualizações diferentes do mundo representado por fontes diferentes de informação. Em SARI os conceitos de diferentes ontologias são ligados uns aos outros através das noções de ponto de vista e ponte. As ontologias são interligadas de uma forma que a estrutura ontológica pode ser vista de perspectivas diferentes. Por exemplo, há uma ponte entre os conceitos de Mercadoria e Produto que são respectivamente as classes raiz das classificações sob o comércio exterior e pontos de vistas industriais.

Metas futuras para o SARI é a formação de pontes entre os conceitos de diferentes ontologias, e geração semiautomática de metadados a partir de recursos da Web.

## 5.4 Considerações

São muitas as técnicas utilizadas para resolver problemas de classificação de texto. A aplicação de novas técnicas mostra uma evolução dos sistemas especializados em recuperar, extrair e classificar informação. Os esforços são muitos na tentativa de obter aplicações inteligentes, capazes de processar a linguagem humana. Mas é preciso considerar tal como (JM02), que as tecnologias da informação são essenciais como a recuperação, extração e categorização informacional, constituem ferramentas úteis, desde que sejam entendidas suas atuais limitações.

Os sistemas apresentados na seção 5.3, assim como o MASTER-Web, são exemplos da evolução das técnicas usadas para classificar informação. Esses sistemas conseguem resultados satisfatórios e com qualidade aplicando técnicas simples para alcançar os objetivos que cada um se propõe. Porém, o OntoSeek para recuperar informação em páginas amarelas e catálogos de produtos na Web compara palavras usando basicamente uma ontologia baseada em WordNet, enquanto o MASTER-Web é ontologicamente mais rico, por abarcar o meio de pesquisa como um todo, com relações mais complexas e capacidade de inferência já durante a classificação e extração. O sistema SARI por sua vez, possui uma estrutura mais complexa com a confecção de vários agentes, todavia sua cobertura está voltada para o comércio.

A extensão do MASTER-Web com a classificação textual em artigos, ressalta a flexibilidade que o sistema proporciona na representação do conhecimento, podendo facilmente abarcar outros domínios sem sofrer grandes alterações no programa. Ainda há muito a ser feito no MASTERWeb, como o uso de aprendizado, mas a flexibilidade de seu projeto permite que esta e outras capacidades sejam adicionadas ao sistema num futuro próximo.

# Capítulo 6

## Conclusões e Trabalhos Futuros

### 6.1 Contribuições

A quantidade de recursos disponíveis na Web e seu contínuo crescimento apresentam desafios para encontrar informação desejada de modo fácil e eficiente. Apesar da diversidade dos sistemas que manipulam informação na rede, encontrar o que se pretende pode exigir do usuário paciência e por vezes uma seleção manual do resultado fornecido por esses mecanismos. Pesquisadores têm focalizado nestes desafios e utilizado técnicas de diversas áreas de pesquisa, como Recuperação de Informação, Inteligência Artificial e Processamento de Linguagem Natural, para solucionar o problema.

Este trabalho inicialmente pressupõe o uso de ontologias combinado a regras de classificação como um caminho para potencial solução de classificação textual. Um ponto chave para obter sistemas capazes de encontrar a informação que o usuário necessita é especificar formalmente conceitos do domínio, explicitando as relações semânticas existentes entre esses conceitos. É nesse contexto que se dá a importância das ontologias. Elas propiciam o compartilhamento e a captura de conhecimento consensual. Aliado a isso, deve-se prover os sistemas de um mínimo de inteligência para derivar conclusões e inferir sobre a relevância da informação.

Duas contribuições gerais são propostas nesse trabalho. A primeira contribuição diz respeito à construção de uma ontologia do domínio de IA para servir como repositório

semântico em sistemas de conhecimento, bem como em outros contextos, permitindo análise do conhecimento do domínio e o entendimento da estrutura da informação, através da estrutura interna semanticamente rica das ontologias que fornecem vocabulário sobre conceitos, seus relacionamentos e regras. Onde declaramos:

*Contribuição 1: Construir uma ontologia sobre o domínio de Inteligência Artificial.*

Assim, no capítulo 2, nós apresentamos a ontologia para o domínio de Inteligência Artificial. Desenvolvemos separadamente ontologias por subáreas e aplicações abrangendo as áreas de Busca, Representação do Conhecimento, Redes Neurais e Aprendizado de Máquina. A dimensão deste campo, a multiplicidade de conceitos e a falta de consenso em algumas subáreas impuseram dificuldades ao desenvolvimento da ontologia. Concluimos que para a construção de uma ontologia com qualidade e confiabilidade é preciso, além de fontes seguras e confiáveis das definições do domínio, capacidade para identificar nuances, especificar e posteriormente executar as transformações necessárias entre entidades ontológicas, impostas pelas mudanças ocorridas no próprio domínio. A construção de ontologias é, de fato, uma tarefa que requer tempo, e exige considerável esforço manual e habilidade para distinguir não apenas conceitos e suas relações, mas também a consistência do domínio.

O emprego de ontologias contribui para sistemas de informação de várias formas. Permite que esses sistemas possam beneficiar-se dos aspectos semânticos e da ausência de ambigüidade permitidos pelos relacionamentos explicitamente definidos entre as classes. Demonstra que informação pode ser extraída e categorizada em diferentes níveis, dependendo do grau de granularidade da árvore e do propósito da taxonomia. Além disso, favorece a inferência pela representação explícita do conhecimento e o processamento de texto.

Uma segunda contribuição diz respeito à classificação de informação. O objetivo é transpor os limites da ferramenta MASTER-Web, pela inclusão da classificação de in-



formação a partir de artigos científicos, ratificando a portabilidade da ferramenta para novos domínios através do uso da ontologia proposta no primeiro objetivo. Assim, a segunda contribuição desta dissertação é dirigida à classificação de informação:

*Contribuição 2: Modelagem para classificação de informação, a partir de artigos científicos, utilizando a ontologia de IA.*

Uma interpretação adequada de páginas HTML é indispensável para classificação de informação em artigos científicos que obedecem a essa estrutura de formatação. Foram realizadas alterações no código do MASTER-Web para representação dos dados contidos nas seções dos artigos escritos em HTML. A análise sobre as seções relevantes, considerando a influência do conteúdo dessas seções para o resultado da classificação do artigo, foi discutida e pode ainda se justificar pela generalidade de algumas seções, bem como o aspecto teórico apresentado por outras.

A contribuição é demonstrada pela adição de uma nova base de conhecimento juntamente com as estratégias de classificação inseridas no sistema. Os resultados dos experimentos evidenciam a contribuição desse trabalho e demonstram que a classificação textual baseada em ontologias pode obter altos níveis de precisão, com resultados significantes em relação a métodos tradicionais.

Em nossas considerações finais, entendemos que a diversidade de informação disponível na Internet e a falta de estrutura dos documentos carecem de soluções de softwares que, independentemente da estruturação do documento, consiga manipular eficientemente a informação e possa torná-la acessível satisfazendo as solicitações dos usuários. Para tanto, a aplicação de múltiplas técnicas, ao contrário de métodos tradicionais, mostra uma evolução dos sistemas especializados em recuperação, extração e classificação de informação.

O esquema apresentado neste trabalho, representa uma abordagem para classificação de informação em sistemas de domínios específicos sem requerer algoritmos

complexos. Pelo contrário, através da aplicação de técnicas simples e uma ontologia bem formada chega-se a resultados de classificação promissores, independentemente da estrutura do documento. Esta abordagem abre possibilidades para classificação textual automática pela simples troca ou adição de novos domínios, sem exigir praticamente nenhuma alteração no sistema e sem se preocupar em estabelecer padrões de estruturação dos documentos dispostos na Internet.

## 6.2 Trabalhos Futuros

Trabalhos futuros estão previstos para melhoria da classificação de informação. A principal extensão deste trabalho é a extensão da ontologia de IA, pela adição de ontologias que cubram subáreas como Agentes Inteligentes, Robótica, Visão de Máquina, entre outras. Assim, novas áreas de IA poderão ser reconhecidas nos textos, o que possibilitará classificar as informações contidas em um artigo em muito mais categorias.

O reconhecimento de outros tipos de formatos de artigos também pode ser adicionado, como arquivos em formato pdf, doc, entre outros. Possibilitando que o usuário possa submeter ao sistema a classificação de artigos sem se prender ao formato do documento.

Uma adição significativa é habilitar o sistema a classificar informações a partir de diversos tipos de textos, como e-mails, entrevistas, anúncios, trechos de texto e não apenas em artigos científicos. Essa nova funcionalidade é perfeitamente possível e viabilizada pelo reuso de muitas regras existentes, que por serem gerais facilitam esse tipo de extensão no sistema.

Melhorias podem ser feitas no que diz respeito ao reconhecimento de seções dos artigos escritos em HTML, passando a reconhecer os títulos das seções por alguma numeração presente no início de uma linha, quebra de linha ou início de parágrafos, além das *tags* geralmente usadas para destacar esses títulos.

Também estão previstas melhorias nas regras que ordenam a classificação dos artigos, fornecendo ao usuário um ranking baseado em percentual de certeza, agilizando a

localização da informação necessária ao usuário.

Melhorarias na integração do agente de classificação de informação em artigos com os demais agentes do MASTER-Web podem ser realizadas, promovendo a cooperação entre os agentes como propõe a arquitetura desse sistema.

Por fim, talvez a mais importante para que o sistema possa ser utilizado por inúmeros usuários, é a construção de uma interface que apresente os resultados da classificação dos artigos separados por categorias, e que permita a interação do usuário com o sistema facilitando suas buscas.

# Apêndice A

## Exemplos de Processamento

### A.1 Reconhecimento de seções e classes principais

A listagem abaixo, gerada pelo motor de inferência Jess, mostra a execução do sistema MASTER-Web no processamento da página da figura A.1, reconhecendo corretamente as seções existentes. Também é mostrado o reconhecimento direto de classes principais. O termo 'FIRE i' refere-se a regra i sendo disparada durante a execução.

```
Processing page from row with priority 2
Page: http://localhost/masterweb/cfp/Teste/AI/AI41.htm
FIRE 1 MAIN::i_900_start f-0,
FIRE 2 MAIN::v_314_valid f-536, f-534
FIRE 3 MAIN::i_906_filling f-547
FIRE 4 MAIN::i_908_fill-fields f-548
FIRE 5 MAIN::r_100_all-failed f-2073
FIRE 6 MAIN::a_001_instance_section_abstract f-2076,
Section: [ABSTRACT]
FIRE 7 MAIN::a_003_instance_section_othersections f-2076,
Section: [OTHER-SECTIONS]
FIRE 8 MAIN::a_004_instance_section_conclusion f-2076,
Section: [CONCLUSION]
...
FIRE 16 MAIN::a_011_recognize_direct_class_article_new f-71, f-2142, f-70,,
RECONHECENDO DIRETAMENTE A CLASSE PRINCIPAL: Artificial Neural Network
FIRE 17 MAIN::a_011_recognize_direct_class_article_new f-71, f-2121, f-58,,
RECONHECENDO DIRETAMENTE A CLASSE PRINCIPAL: Knowledge Representation
Formalisms
FIRE 18 MAIN::a_012_recognize_direct_class_article_modify f-71, f-2114, f-70,, f-2143
RECONHECENDO DIRETAMENTE A CLASSE PRINCIPAL: Artificial Neural Network
```

FIRE 19 MAIN::a\_012\_recognize\_direct\_class\_article\_modify f-71, f-2092, f-70,, f-2143  
RECONHECENDO DIRETAMENTE A CLASSE PRINCIPAL: Artificial Neural Network  
FIRE 20 MAIN::a\_013\_recognize\_sub\_class\_article\_new f-71, f-2141, f-70,  
RECONHECENDO SUBCLASSE: Supervised Learning ANN como subclasse de Artificial  
Neural Network  
FIRE 21 MAIN::a\_013\_recognize\_sub\_class\_article\_new f-71, f-2134, f-70,  
RECONHECENDO SUBCLASSE: Feed Forward ANN como subclasse de Artificial Neural  
Network  
...  
FIRE 48 MAIN::a\_037\_print\_result\_final\_article\_talk\_and\_citations f-2076,  
----- CLASSIFICATION ARTICLE -----  
Page: <http://localhost/masterweb/cfp/Teste/AI/AI41.htm>  
Title: neural networks  
The article is about:  
("Feed Forward ANN", "Supervised Learning ANN")  
Citations:  
("Knowledge Representation Formalisms")  
-----  
FIRE 49 MAIN::s\_303\_result\_rejected f-2076, f-2074  
rejected (REJECTED) ARTICLE  
Inserting article as classified...  
Saving Statistics... Done saving.

A seguir é mostrada a execução do sistema no processamento da página da figura A.2, onde a classe principal *Artificial Neural Network* é reconhecida através do atributo *Neurons*. Embora classifique corretamente o artigo, o sistema comete falha no reconhecimento das seções, reconhecendo todo o artigo como uma única seção, *Other-Sections*. Esse tipo de falha ocorre, em sua maioria, quando os títulos das seções não estão destacados em negrito, confundindo-se com o corpo do texto.

```

Processing page from row with priority 2
Page: http://localhost/masterweb/cfp/Teste/AI/AI20.htm
FIRE 1 MAIN::i_900_start f-0,
FIRE 2 MAIN::v_314_valid f-536, f-534
FIRE 3 MAIN::i_906_filling f-547
FIRE 4 MAIN::i_908_fill-fields f-548
FIRE 5 MAIN::r_100_all-failed f-1824
FIRE 6 MAIN::a_003_instance_section_othersections f-1827,
Section: [OTHER-SECTIONS]
...
RECONHECENDO CLASSE LIGADA DIRETAMENTE PELO SLOT: Output Layer esta
ligada a Artificial Neural Network pelo slot Neurons
...
FIRE 55 MAIN::a_037_print_result_final_article_talk_and_citations f-1827,
----- CLASSIFICATION ARTICLE -----
Page: http://localhost/masterweb/cfp/Teste/AI/AI20.htm
Title: classification of binary document images into texture or non-textual data blocks using
neural network models
The article is about:
("Self Organized Map", "Supervised Learning ANN", "Radial Basis Function Network")
Citations:
("Machine Learning", "Knowledge Representation Formalisms", "Expert System")
-----
FIRE 56 MAIN::s_303_result_rejected f-1827, f-1825
rejected (REJECTED) ARTICLE
Inserting article as classified...
Saving Statistics... Done saving.
```



# NEURAL NETWORKS

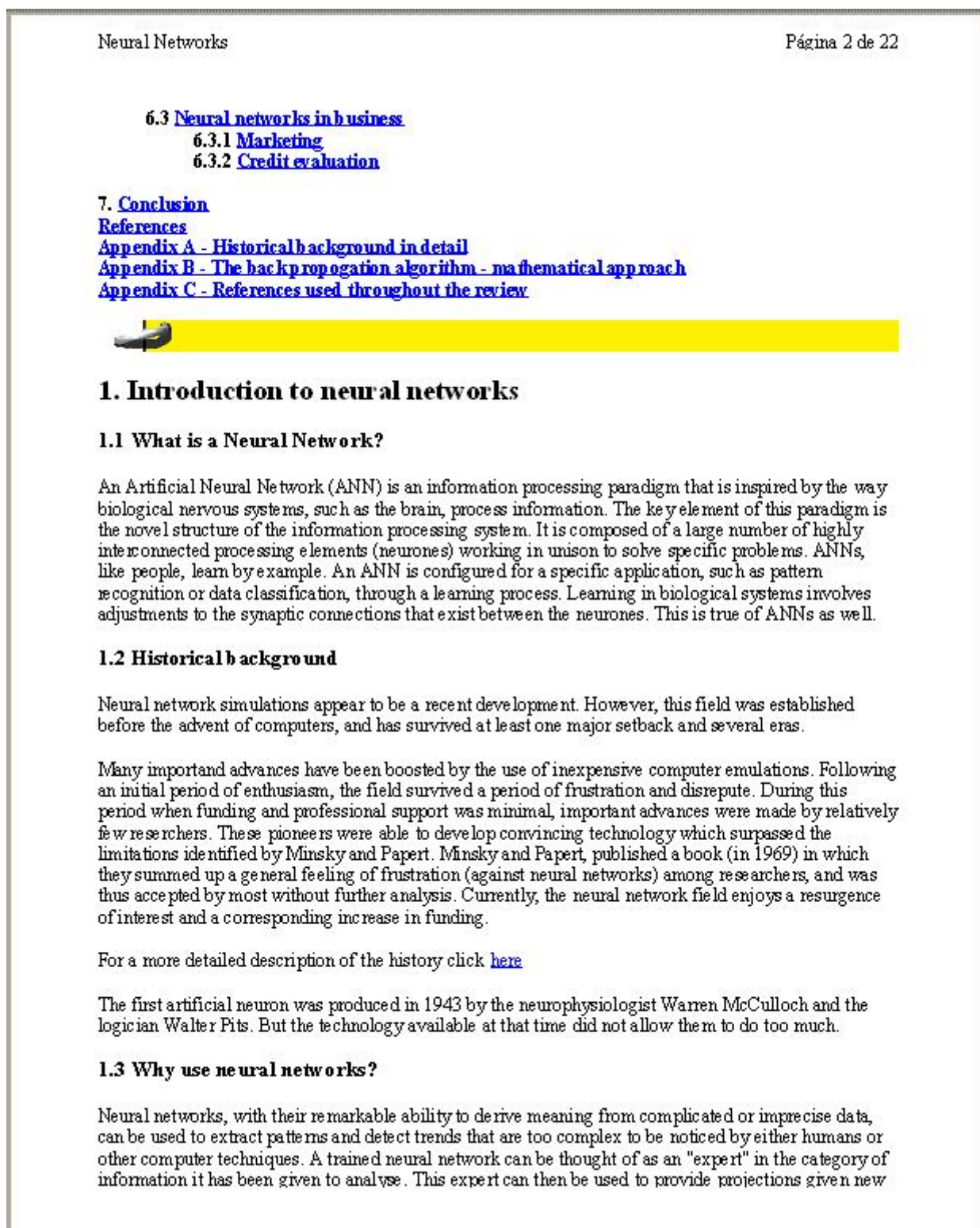
by Christos Stergiou and Dimitrios Siganos

## Abstract

This report is an introduction to Artificial Neural Networks. The various types of neural networks are explained and demonstrated, applications of neural networks like ANNs in medicine are described, and a detailed historical background is provided. The connection between the artificial and the real thing is also investigated and explained. Finally, the mathematical models involved are presented and demonstrated.

## Contents:

1. [Introduction to Neural Networks](#)
  - 1.1 [What is a neural network?](#)
  - 1.2 [Historical background](#)
  - 1.3 [Why use neural networks?](#)
  - 1.4 [Neural networks versus conventional computers - a comparison](#)
2. [Human and Artificial Neurones - investigating the similarities](#)
  - 2.1 [How the Human Brain Learns?](#)
  - 2.2 [From Human Neurones to Artificial Neurones](#)
3. [An Engineering approach](#)
  - 3.1 [A simple neuron - description of a simple neuron](#)
  - 3.2 [Firing rules - How neurones make decisions](#)
  - 3.3 [Pattern recognition - an example](#)
  - 3.4 [A more complicated neuron](#)
4. [Architecture of neural networks](#)
  - 4.1 [Feed-forward \(associative\) networks](#)
  - 4.2 [Feedback \(autoassociative\) networks](#)
  - 4.3 [Network layers](#)
  - 4.4 [Perceptrons](#)
5. [The Learning Process](#)
  - 5.1 [Transfer Function](#)
  - 5.2 [An Example to illustrate the above teaching procedure](#)
  - 5.3 [The Back-Propagation Algorithm](#)
6. [Applications of neural networks](#)
  - 6.1 [Neural networks in practice](#)
  - 6.2 [Neural networks in medicine](#)
    - 6.2.1 [Modelling and Diagnosing the Cardiovascular System](#)
    - 6.2.2 [Electronic noses - detection and reconstruction of odours by ANNs](#)
    - 6.2.3 [Instant Physician - a commercial neural net diagnostic program](#)

Figura A.1: Parte do artigo *Neural Networks*.



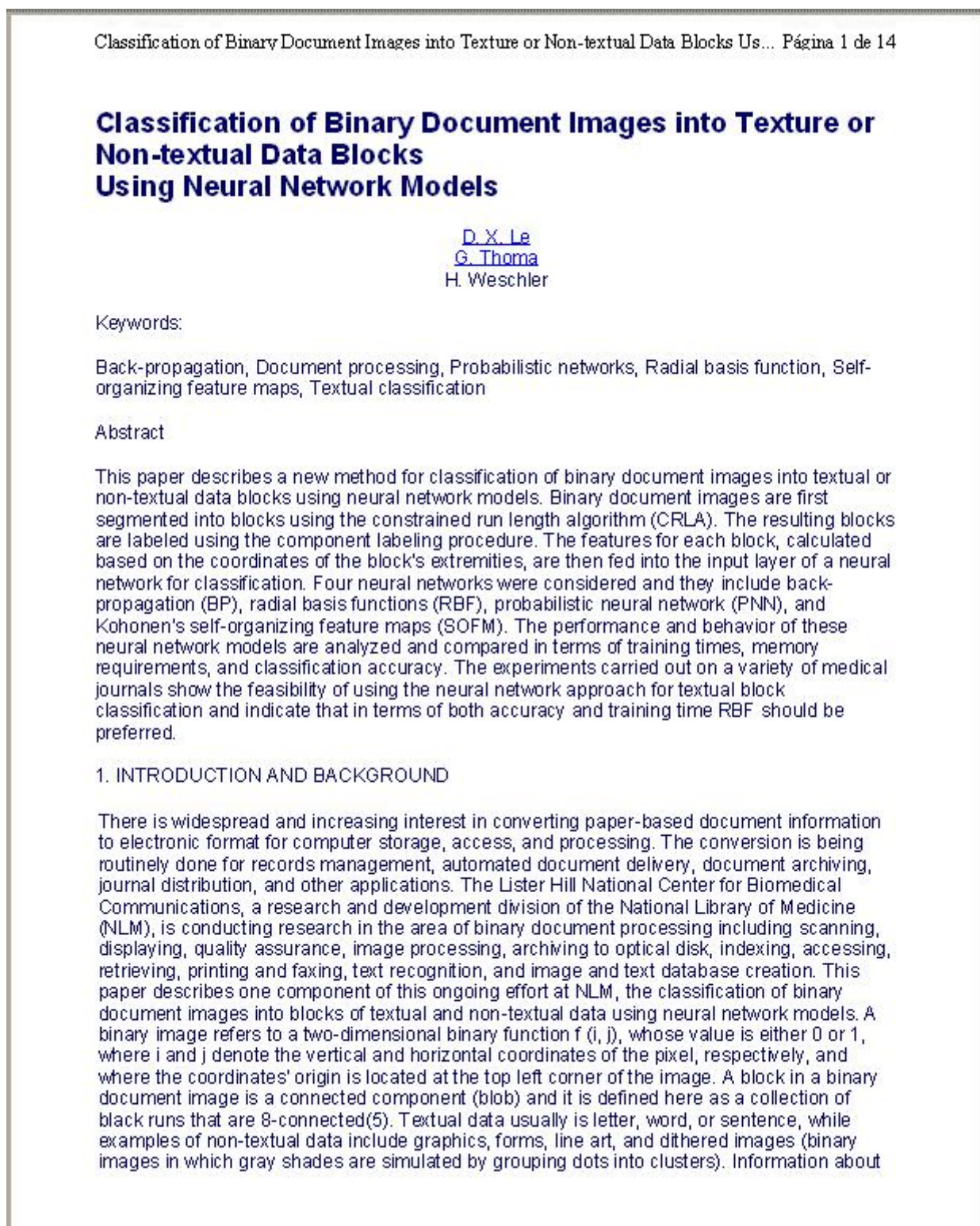


Figura A.2: Parte do artigo *Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models*.

Abaixo é mostrada a execução do sistema no processamento da página da figura A.3, onde a classe principal *Knowledge Representation Formalisms* é reconhecida através de relação indireta com a classe *Ontology*.

```
Processing page from row with priority 2
Page: http://localhost/masterweb/cfp/Teste/AI/AI21.htm
FIRE 1 MAIN::i_900_start f-0,
FIRE 2 MAIN::v_314_valid f-536, f-534
FIRE 3 MAIN::i_906_filling f-547
FIRE 4 MAIN::i_908_fill-fields f-548
FIRE 5 MAIN::r_100_all-failed f-2062
...
FIRE 19 MAIN::a_023_recognize_slots_indirect_class_article_fase_4_new f-71, f-2103,,
4- RECONHECENDO CLASSE LIGADA INDIRETAMENTE: Ontology esta ligada a Language-
oriented Formalism que é subclasse de Knowledge Representation Formalisms
FIRE 20 MAIN::a_025_recognize_slots_indirect_class_article_without_class f-2104,
FIRE 21 MAIN::a_023_recognize_slots_indirect_class_article_fase_4_new f-71, f-2102,,
4- RECONHECENDO CLASSE LIGADA INDIRETAMENTE: Ontology esta ligada a Domain-
oriented KR Formalism que é subclasse de Knowledge Representation Formalisms
...
FIRE 42 MAIN::a_037_print_result_final_article_talk_and_citations f-2065,
----- CLASSIFICATION ARTICLE -----
Page: http://localhost/masterweb/cfp/Teste/AI/AI21.htm
Title: owl web ontology language overview
The article is about:
("Ontology", "Knowledge Representation Formalisms")
Citations:
()
-----
FIRE 43 MAIN::s_303_result_rejected f-2065, f-2063
rejected (REJECTED) ARTICLE
Inserting article as classified...
Saving Statistics... Done saving.
```

OWL Web Ontology Language Overview Página 1 de 15

# OWL Web Ontology Language Overview

## W3C Recommendation 10 February 2004

This version:  
<http://www.w3.org/TR/2004/REC-owl-features-20040210/>

Latest version:  
<http://www.w3.org/TR/owl-features/>

Previous version:  
<http://www.w3.org/TR/2003/PR-owl-features-20031215/>

Editors:  
Deborah L. McGuinness (Knowledge Systems Laboratory, Stanford University)  
Frank van Harmelen (Vrije Universiteit, Amsterdam) [Frank.van.Harmelen@cs.vu.nl](mailto:Frank.van.Harmelen@cs.vu.nl)

Please refer to the [errata](#) for this document, which may include some normative corrections.

See also [translations](#).

Copyright © 2004 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) and [software licensing](#) rules apply.

---

### Abstract

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interoperability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

This document is written for readers who want a first impression of the capabilities of OWL. It provides an introduction to OWL by informally describing the features of each of the sublanguages of OWL. Some knowledge of [RDF Schema](#) is useful for understanding this document, but not essential. After this document, interested readers may turn to the [OWL Guide](#) for more detailed descriptions and extensive examples on the features of OWL. The normative formal definition of OWL can be found in the [OWL Semantics and Abstract Syntax](#).

### Status of this document

This document has been reviewed by W3C Members and other interested parties, and it has been endorsed by the Director as a [W3C Recommendation](#). W3C's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability of the Web.

Figura A.3: Parte do artigo *OWL Web Ontology Language Overview*.

## A.2 Classificação de artigos

Os resultados obtidos nesse trabalho mostra que o sistema apresentou elevado desempenho, considerando o domínio ontológico. Assim, artigos como o representado pela figura A.4 não são classificados como sendo de IA por tratar de assuntos não cobertos pelo domínio.

---

————— CLASSIFICATION ARTICLE —————  
Page: <http://localhost/masterweb/cfp/Teste/GAI/GAI4.htm>  
Title: ai planning resources on the web  
is NOT an Artificial Intelligence article

---

Um aspecto relevante são as citações apontadas pelo sistema, quando o mesmo não encontra elementos suficientes para inferir que o artigo trata sobre determinado tema. O artigo da figura A.5, é um exemplo típico, como mostra o resultado da classificação abaixo. Nesse artigo o tema principal não é IA, mas o texto faz várias referências às técnicas de Inteligência Artificial.

---

————— CLASSIFICATION ARTICLE —————  
Page: <http://localhost/masterweb/cfp/Teste/NAI/NAI73.htm>  
Title: analysis of three-dimensional protein images  
The article is about:  
( )  
Citations:  
("Machine Learning", "Knowledge Representation Formalisms", "Expert System")


---

O resultado exibido abaixo para o artigo da figura A.6, considera que o artigo não é de IA. O artigo realmente não aborda assuntos de IA, sequer faz alguma referência a área.

---

————— CLASSIFICATION ARTICLE —————  
Page: <http://localhost/masterweb/cfp/Teste/NAI/NAI85.htm>  
Title: advances in distributed security

ACM: Digital Library: intelligence Página 1 de 4



**intelligence**  
**Volume 12, Number 1 (2001), Pages 17-19**  
**Links: AI planning resources on the Web**  
Robert S. Amant, R. Michael Young

**Table of Contents**

- Introduction
- What is planning?
- Can I find out more about the details?
- Although I'm familiar with other areas of AI, I'm a novice to planning. Where does the field stand today? What kind of resources are available online?
- Suppose I'd like to try out some of these planning systems. Is that possible?
- Are there common problem sets that all these planners try to solve?
- Where do planning people congregate?
- References
- Authors

↑

With the introduction of new, efficient techniques for planning, interest in the field has risen sharply in recent years. Planning systems have come a long way since the days of Blocks World. Today, planning and scheduling techniques are being used to solve problems in military campaign planning, robot navigation, industrial equipment scheduling, human-computer interaction, and many other real-world domains. Planning, which is the process of finding a sequence of actions that meets an agents goal, is a thriving area of artificial intelligence (AI) research and is growing almost faster than one can keep track. Fortunately, an abundance of information about planning is available, in both paper and electronic forms; answers to most of the questions a newcomer to the field might ask can easily be found.

---

*Planning and scheduling techniques are being used to solve problems in military campaign planning, robot navigation, industrial equipment scheduling, human-computer interaction, and many other real-world domains.*

Figura A.4: Parte do artigo *AI planning resources on the Web*.

is NOT an Artificial Intelligence article

---

Analysis of Three-Dimensional Protein Images Página 1 de 20

**HOME HELP PREFERENCES**

search titles a-z authors a-z volumes phrases

**Analysis of Three-Dimensional Protein Images**

**Laurence Leherte, Janice Glasgow, Kim Baxter, Evan Steeg and Suzanne Fortier**

**Volume 7, 1997**

**Links to Full Text:**

- [Compressed PostScript format \(size: 513k\)](#)
- [PostScript format \(size: 3.0M\)](#)
- [PDF format \(size: 460k\)](#)

**NO HIGH DETACH LIGHTING**

**Abstract:**

A fundamental goal of research in molecular biology is to understand protein structure. Protein crystallography is currently the most successful method for determining the three-dimensional (3D) conformation of a protein, yet it remains labor intensive and relies on an expert's ability to derive and evaluate a protein scene model. In this paper, the problem of protein structure determination is formulated as an exercise in *scene analysis*. A computational methodology is presented in which a 3D image of a protein is segmented into a graph of critical points. Bayesian and certainty factor approaches are described and used to analyze critical point graphs and identify meaningful substructures, such as *alpha*-helices and *beta*-sheets. Results of applying the methodologies to protein images at low and medium resolution are reported. The research is related to approaches to representation, segmentation and classification in vision, as well as to top-down approaches to protein structure prediction.

---

**Extracted Text**

## Introduction

Crystallography plays a major role in current efforts to characterize and understand molecular structures and molecular recognition processes. The information derived from crystallographic studies provides a precise and detailed depiction of a molecular scene, an essential starting point for unraveling the complex rules of structural organization and molecular interactions in biological systems. However, only a small fraction of the currently known proteins have been fully characterized.

Figura A.5: Parte do artigo *Analysis of Three-Dimensional Protein Images*.

Advances in Distributed Security Página 1 de 17

---

[n.szabo@law.gwu.edu](mailto:n.szabo@law.gwu.edu)

[Nick Szabo's Papers and Concise Tutorials](#)

## Advances in Distributed Security

Copyright (c) 2003 by Nick Szabo  
rough draft – quoting or redistribution without permission of the author prohibited

---

### Table of Contents

- [Introduction](#)
- [Fault Tolerance and Security in Networks and Their Services](#)
- [Partial and Total Orders](#)
- [Some Cryptographic Primitives](#)
- [Physical Broadcasts and Global Clocks](#)
- [Secure Time-stamping](#)
- [The Byzantine Generals Problem](#)
- [Logical Broadcast Over the Internet](#)
- [Byzantine-Resilient Replication](#)
- [Some Applications](#)
- [Conclusion](#)
- [Appendix A – Implementations](#)
- [References](#)

### Introduction

The last decade has witnessed a revolution in distributed security. Old, pessimistic proofs that security and fault tolerance were "impossible", based on assumptions that protocols had to be deterministic and security and fault tolerance properties had to be absolutely certain, have given way to new proofs and implementations of provable security based on the assumption of cryptography and other randomized protocols that achieving security with very high probability is sufficient. The old view "proved" that the integrity properties of a wide variety of services on which civilization depends, whether [synchronized clocks](#), [public directories](#), [censorship-proof file sharing and publication](#), or [issuing money or securities](#) were "impossible" on asynchronous networks like the Internet unless we put unlimited faith in a third party to enforce many of the rules of the service. We now know how to provide such services with a high degree of integrity and availability, yet far more resilient to the possibility that any party might act in a malicious manner.

### Fault Tolerance and Security in Networks and Their Services

As a result of these new possibilities, we are witnessing a shift in the way we view trust. The old view in computer and network security was that trust was all-or-nothing – either we must place an essentially blind faith in a third party (for example a certificate authority or an issuer of digital cash) or we must

Figura A.6: Parte do artigo *Advances in Distributed Security*.

O resultado exibido abaixo se refere ao falso positivo *Information Retrieval Issues* representado na figura A.7.

————— CLASSIFICATION ARTICLE —————  
Page: <http://localhost/masterweb/cfp/Teste/GAI/GAI8.htm>  
Title: information retrieval issues  
— The article is about:  
("Expert System", "Inference Engine")  
— Citations:  
( )

---

Em seguida mostramos o resultado para o falso negativo da figura A.8.

————— CLASSIFICATION ARTICLE —————  
Page: <http://localhost/masterweb/cfp/Teste/AI/AI35.htm>  
Title: learning sets of related concepts: a shared task model  
The article is about:  
("Feed Forward ANN")  
— Citations:  
("Machine Learning", "Explanation-based Learning")

---



[\[Index\]](#) [\[Previous\]](#) [\[Next\]](#) [\[Postscript\]](#)

## CHAPTER 5

### INFORMATION RETRIEVAL ISSUES

#### 1. Introduction

Navigation or browsing is effective only for small hypertext systems. For large hypertext databases, information retrieval (IR) through queries becomes crucial. Conklin had suggested that search and query mechanisms can present information at a manageable level of complexity and detail [Conklin, 1987]. Halasz's view was that "navigational access itself is not sufficient. Effective access to information stored in a hypemedia network requires query-based access to complement navigation..... search and query needs to be elevated to a primary access mechanism on par with navigation." [Halasz, 1988].

#### 2. Query and Search Mechanisms

Conventional IR systems focus on keyword based automatic searching (in conjunction with Boolean operations), weighting of words based on their statistical properties, ranking of documents according to probability of relevance, automatic relevance feedback for query modification and query languages [Croft et al., 1990]. However, very few (or none) of these methods retrieve complete or accurate information. Too general a query may yield a lot of items and too specific a query may retrieve no items. Thus, traditional IR is an inherently uncertain process. Combining inference techniques could eliminate or minimize uncertainty. In hypertext systems, a weighted keyword search combined with hypertext links can improve IR by finding only a subset of nodes or "hits" whose links can then be followed to other semantically related nodes [Carlson, 1989].

According to Halasz, query and search mechanisms can be classified into content search and structure search [Halasz, 1988]. Content search is standard IR technique extended to hypertext systems. That is, all nodes and links are treated independently and examined for a match to the given query. On the other hand, structure search will yield the hypertext sub-network that matches a given pattern. Query facilities which combine aspects of both content search and structure search will be capable of acting as filters. Based on the user's query, the interface will display only those nodes and links that match the query, filtering out other parts of the network. Filtered browsers have been implemented both for NoteCards and Tektronix's Neptune. In NoteCards, a user can filter out information based on the node or link type. In Neptune, the query can be content-based; if the query is broad enough, a global view of the entire network is displayed; if the query is well refined, the viewing size will be manageable.

#### 2.1. Content Queries and Indexes

Bruza proposed a two-level architecture for hypertext documents, the top level called hyperindex (containing index information) and the bottom level hyperbase (containing content nodes and links) [Bruza, 1990]. The hyperindex consists of a set of indexes linked together. When an index term describing the required information is found, the objects from the underlying hyperbase are retrieved for examination. Navigating through the hyperindex (not the hyperbase) and retrieving information from the hyperbase is called "Query By Navigation" [Bruza, 1990].

An index is made of a set of index entries. Each index entry consists of a term descriptor or keyword and a locator (like a page number). Term descriptors lack specificity. Term phrases are made of term

Figura A.7: Parte do artigo *Information Retrieval Issues*.

Learning Sets of Related Concepts: A Shared Task Model Página 1 de 9

## Learning Sets of Related Concepts: A Shared Task Model

**Tim Hume**  
ICS Dept., University of California, Irvine. Irvine, CA 92717  
[hume@interplay.com](mailto:hume@interplay.com)

**Michael J. Pazzani**  
ICS Dept., University of California, Irvine. Irvine, CA 92717  
[pazzani@cs.uci.edu](mailto:pazzani@cs.uci.edu) (949)824-5888 <http://www.ics.uci.edu/dir/faculty/AI/pazzani>

### Abstract

We investigate learning a set of causally related concepts from examples. We show that human subjects make fewer errors and learn more rapidly when the set of concepts is logically consistent. We compare the results of these subjects to subjects learning equivalent concepts that share sets of relevant features, but are not logically consistent. We present a shared-task neural network model simulation of the psychological experimentation.

### Introduction

Researchers have investigated how the relevant background knowledge of the learner influences the speed or accuracy of concept learning (e.g., Murphy & Medin 1985, Nakamura 1985, Pazzani 1991, Wattenmaker *et al.* 1986). However, the psychological investigation to date has only explored problems where subjects learn a single concept and the relevant background knowledge is either brought to the experiment by the subject or given in written instructions. In contrast, research in machine learning has addressed issues that occur when learning a set of related concepts. For example, relevant background concepts might be learned inductively from examples before learning concepts that depend upon this knowledge (Pazzani 1990). Here, we report on two experiments in which subjects induce the relevant background knowledge from examples and use this background knowledge to facilitate later learning. The experiments illustrate the importance of learning the relevance of combinations of features, rather than individual features. We model this experiment with shared-task neural networks (Caruana, 1993).

In the first experiment, subjects first induce the relevant background knowledge and then have the opportunity to use this knowledge in later learning. To more closely simulate the real world, we ran a second experiment wherein the subjects induce the relevant background knowledge at the same time as learning the concept that depends on this knowledge. In both experiments, subjects were divided into two groups. One group, the "feature consistency" group, learned a complex concept that shared relevant features with previously learned related concepts, but was not logically consistent with those concepts. Another group, the "logical consistency" group, learned a complex concept that was logically consistent with previously learned related concepts.

<http://localhost/masterweb/cfp/Tests/AI/AI35.htm> 14/9/2006

Figura A.8: Parte do artigo *Learning Sets of Related Concepts: A Shared Task Model*.

# Apêndice B

## Base Experimental

### B.1 Base de Teste

Toda a base experimental utilizada neste trabalho é composta por documentos digitalizados em formato HTML, obtidos em diversos sites, como sites de universidades, professores, revistas, entre outros. Abaixo são listados os 406 artigos, por títulos e respectivas URLs, utilizados nos testes experimentais:

<b>Artigos</b>
A Biomedical Information System for Combined Content-Based Retrieval of Spine X-ray Images and Associated Text Information - <a href="http://archive.nlm.nih.gov/pubs/antani/icvgip02/icvgip02.php">http://archive.nlm.nih.gov/pubs/antani/icvgip02/icvgip02.php</a>
A Comprehensive Trainable Error Model for Sung Music Queries - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A Domain-Independent Algorithm for Plan Adaptation - <a href="http://www.cs.washington.edu/homes/weld">http://www.cs.washington.edu/homes/weld</a>
A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies - <a href="http://www.acm.org/sigchi/chi95/proceedings/papers/jl_bdy.htm">http://www.acm.org/sigchi/chi95/proceedings/papers/jl_bdy.htm</a>
A Formal Language for Analyzing Contracts - <a href="http://szabo.best.vwh.net/contractlanguage.html">http://szabo.best.vwh.net/contractlanguage.html</a>
A Knowledge Compilation Map - <a href="http://www.jair.org/papers/paper989.html">http://www.jair.org/papers/paper989.html</a>
A Layered Interaction Analysis of Direct Manipulation - <a href="http://www.useit.com/papers/direct_manipulation.html">http://www.useit.com/papers/direct_manipulation.html</a>
A Learning Model for Forecasting the Future of Information Technology - <a href="http://ksi.cpsc.ucalgary.ca/articles/BRETAM/FCS_IT/">http://ksi.cpsc.ucalgary.ca/articles/BRETAM/FCS_IT/</a>
A Market-Oriented Programming Environment and its Application to Distributed Multicommodity Flow Problems - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>

A Measure of Sacrifice - <a href="http://szabo.best.vwh.net/synch.html">http://szabo.best.vwh.net/synch.html</a>
A method of content-based image retrieval for a spinal x-ray image database - <a href="http://archive.nlm.nih.gov/pubs/krainak/spie-sd2002krainak/spie-sd2002krainak.php">http://archive.nlm.nih.gov/pubs/krainak/spie-sd2002krainak/spie-sd2002krainak.php</a>
A Model of Inductive Bias Learning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A Multivariate Quantitative-Genetic Analysis of Behavioral Development in Mice - <a href="http://cogprints.org/32/00/developm.htm">http://cogprints.org/32/00/developm.htm</a>
A New General Method to Generate Random Modal Formulae for Testing Decision Procedures - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A New Look at the Easy-Hard-Easy Pattern of Combinatorial Search Difficulty - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume7/mammen97a-html/ehe3.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume7/mammen97a-html/ehe3.html</a>
A New Technique for Combining Multiple Classifiers using The Dempster-Shafer Theory of Evidence - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A Personalized System for Conversational Recommendations - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume21/thompson04a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume21/thompson04a-html/</a>
A Prototype Client/Server Application for Biomedical Text - <a href="http://archive.nlm.nih.gov/pubs/proto-client/proto-client.php">http://archive.nlm.nih.gov/pubs/proto-client/proto-client.php</a>
A Prototype Content-based Image Retrieval System for Spine X-rays - <a href="http://archive.nlm.nih.gov/pubs/long/cbms2003/cbms2003.php">http://archive.nlm.nih.gov/pubs/long/cbms2003/cbms2003.php</a>
A Research Agent Architecture for Real Time Data Collection and Analysis - <a href="http://www.cs.indiana.edu/~leake/papers/p-01-05_dir.html">http://www.cs.indiana.edu/~leake/papers/p-01-05_dir.html</a>
A Semantics and Complete Algorithm for Subsumption in the CLASSIC Description Logic - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A System for Induction of Oblique Decision Trees - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A Temporal Description Logic for Reasoning about Actions and Plans - <a href="http://www.jair.org/papers/paper516.html">http://www.jair.org/papers/paper516.html</a>
A UMLS-based Knowledge Acquisition Tool for Rule-based Clinical Decision Support System Development - <a href="http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&amp;artid=130080">http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&amp;artid=130080</a>
A Unified Model of Structural Organization in Language and Music - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
A Uniform Framework for Concept Definitions in Description Logics - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Access to a Digital Xray Archive over Internet - <a href="http://archive.nlm.nih.gov/pubs/adxa-internet/adxa-internet.php">http://archive.nlm.nih.gov/pubs/adxa-internet/adxa-internet.php</a>
Achieving Focus, Engagement, and Acceptance: Three Phases of Adapting to Internet Use - <a href="http://www.soc.hawaii.edu/leonj/leonj/leonpsy/instructor/compedutext.html">http://www.soc.hawaii.edu/leonj/leonj/leonpsy/instructor/compedutext.html</a>
Acquiring Word-Meaning Mappings \for Natural Language Interfaces - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/thompson03a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/thompson03a-html/</a>
Active Learning with Statistical Models - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/cohn96a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/cohn96a-html/</a>
Adapting to a Highly Automated World - <a href="http://ksi.cpsc.ucalgary.ca/articles/Trust/">http://ksi.cpsc.ucalgary.ca/articles/Trust/</a>
Adaptive Load Balancing: A Study in Multi-Agent Learning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Adaptive Parallel Iterative Deepening Search - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/cook98a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/cook98a-html/</a>
Additive Pattern Database Heuristics - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>

Advances in Distributed Security - <a href="http://szabo.best.vwh.net/distributed.html">http://szabo.best.vwh.net/distributed.html</a>
Agent-Based Engineering, the Web, and Intelligence - <a href="http://www-cdr.stanford.edu/NextLink/Expert.html">http://www-cdr.stanford.edu/NextLink/Expert.html</a>
AI planning resources on the Web - <a href="http://portal.acm.org">http://portal.acm.org</a>
ais: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume13/cheng00a.html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume13/cheng00a.html/</a>
Allergies and Their Treatments - <a href="http://serendip.brynmawr.edu/biology/b103/f01/web2/manewitz.html">http://serendip.brynmawr.edu/biology/b103/f01/web2/manewitz.html</a>
An Application of Reinforcement Learning to Dialogue Strategy Selection in a Spoken Dialogue System for Email - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
An Empirical Analysis of Search in GSAT - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
An Empirical Approach to Temporal Reference Resolution - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/wiebe98a.html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/wiebe98a.html/</a>
An Evaluation of Ontology Exchange Languages for Bioinformatics - <a href="http://xml.coverpages.org/OntologyExchange.html">http://xml.coverpages.org/OntologyExchange.html</a>
An Expressive Language and Efficient Execution System for Software Agents - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
An Intelligent Information System Supporting the Virtual Manufacturing Enterprise - <a href="http://ksi.cpsc.ucalgary.ca/articles/SMC95MED/">http://ksi.cpsc.ucalgary.ca/articles/SMC95MED/</a>
An Interactive Visual Language for Term Subsumption Languages - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/">http://pages.cpsc.ucalgary.ca/~gaines/reports/</a>
An Introduction to Back-Propagation Neural Networks - <a href="http://www.seattlerobotics.org/encoder/nov98/neural.html">http://www.seattlerobotics.org/encoder/nov98/neural.html</a>
An Introduction to Neural Networks - <a href="http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html#why">http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html#why</a>
An Introduction to Quantitative Genetics - <a href="http://cogprints.org/67/00/QUANTGEN.htm">http://cogprints.org/67/00/QUANTGEN.htm</a>
An On-Line Interview with Noam Chomsky: On the nature of pragmatics and related issues - <a href="http://cogprints.org/126/00/chomswb_399.html">http://cogprints.org/126/00/chomswb_399.html</a>
An Ontology for Engineering Mathematics - <a href="http://www-ksl.stanford.edu/knowledge-sharing/papers/engmath.html">http://www-ksl.stanford.edu/knowledge-sharing/papers/engmath.html</a>
An Ounce of Knowledge is Worth a Ton of Data: Quantitative Studies of the Trade-Off between Expertise and Data based on Statistically Well-Founded Empirical Induction - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/">http://pages.cpsc.ucalgary.ca/~gaines/reports/</a>
Analysis of Three-Dimensional Protein Images - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Analyzing Linkage Structure in a Course-Integrated Virtual Learning Community on the World Wide Web - <a href="http://www.soc.hawaii.edu/leonj/leonj/leonpsy/instructor/inet95.html">http://www.soc.hawaii.edu/leonj/leonj/leonpsy/instructor/inet95.html</a>
Anatomical Shape Representation in Spine X-ray Images - <a href="http://archive.nlm.nih.gov/pubs/antani/396-131/396-131.php">http://archive.nlm.nih.gov/pubs/antani/396-131/396-131.php</a>
Answer Set Planning Under Action Costs - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
AntNet: Distributed Stigmergetic Control for Communications Networks - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
ap: Online Parallelization of Plans with Heuristic State Search - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Applying GSAT to Non-Clausal Formulas - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Artificial Intelligence - <a href="http://www.cs.indiana.edu/~leake/papers/p-01-07/p-01-07.html">http://www.cs.indiana.edu/~leake/papers/p-01-07/p-01-07.html</a>
Artificial Life: Synthetic vs. Virtual - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad</a>
Artificial neural network - <a href="http://en.wikipedia.org/wiki/Artificial_neural_network">http://en.wikipedia.org/wiki/Artificial_neural_network</a>

Aspartame: The REAL Story? - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/dyar2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/dyar2.html</a>
Assessing the Usability of a User Interface Standard - <a href="http://www.useit.com/papers/standards.html">http://www.useit.com/papers/standards.html</a>
attachbig: An Adaptive Autonomous Bidding Agent - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/stone01a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/stone01a-html/</a>
Autism and The Motor Theory of Language - <a href="http://cogprints.org/3188/01/autism.htm">http://cogprints.org/3188/01/autism.htm</a>
Autism II: The aetiology of autism. Neuroembryology and prefrontal-neocerebellum disconnection - <a href="http://cogprints.org/790/00/autism2.htm">http://cogprints.org/790/00/autism2.htm</a>
Automated Collaborative Filtering and Semantic Transports - <a href="http://www.lucifer.com/~sasha/articles/ACF.html">http://www.lucifer.com/~sasha/articles/ACF.html</a>
Automatically Training a Problematic Dialogue Predictor for a Spoken Dialogue System - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume16/walker02a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume16/walker02a-html/</a>
Axiomatizing Causal Reasoning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Balancing Formality with Informality: User-Centred Requirements for Knowledge Management Technologies - <a href="http://ksi.cpsc.ucalgary.ca/AIKM97/sbs/sbs-paper1.html">http://ksi.cpsc.ucalgary.ca/AIKM97/sbs/sbs-paper1.html</a>
Being Real - <a href="http://smg.media.mit.edu/papers/BeingReal/BeingReal.html">http://smg.media.mit.edu/papers/BeingReal/BeingReal.html</a>
Beyond Object Constancy - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad</a>
Bi- and Multivariate Analyses of Diallel Crosses: A Tool for the Genetic Dissection of Neurobehavioral Phenotypes - <a href="http://cogprints.org/33/00/diallel2.htm">http://cogprints.org/33/00/diallel2.htm</a>
Bidirectional Heuristic Search Reconsidered - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Binary Encodings of Non-binary Constraint Satisfaction Problems: Algorithms and Experimental Results - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/samaras05a-html/samaras.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/samaras05a-html/samaras.html</a>
Blackboard Pattern - <a href="http://chat.carleton.ca/~narhorn/project/patterns/BlackboardPattern-display.html">http://chat.carleton.ca/~narhorn/project/patterns/BlackboardPattern-display.html</a>
Bound Propagation - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume19/leisink03a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume19/leisink03a-html/</a>
Brains, Eyes and Computers - <a href="http://www.frc.ri.cmu.edu/~hpm/book97/ch3/retina.comment.html">http://www.frc.ri.cmu.edu/~hpm/book97/ch3/retina.comment.html</a>
Breadth-first search crawling yields high-quality pages - <a href="http://www10.org/cdrom/papers/208/">http://www10.org/cdrom/papers/208/</a>
Breast Cancer and the Environment - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/robbin2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/robbin2.html</a>
Brief History of Artificial Intelligence - <a href="http://www.aaai.org/AITopics/bbhist.html">http://www.aaai.org/AITopics/bbhist.html</a>
Building and Refining Abstract Planning Cases by Change of Representation Language - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Can We Get Too Clean? - <a href="http://serendip.brynmawr.edu/biology/b103/f01/web3/bond.html">http://serendip.brynmawr.edu/biology/b103/f01/web3/bond.html</a>
Can We Learn to Beat the Best Stock - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Cannabis: Passing the Pipe to Cancer Patients - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/rochelle2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/rochelle2.html</a>
CBR in Context: The Present and Future - <a href="http://www.cs.indiana.edu/~leake/papers/p-96-01_dir.html/paper.html">http://www.cs.indiana.edu/~leake/papers/p-96-01_dir.html/paper.html</a>
Censoring the Internet: The Situation in Turkey - <a href="http://www.firstmonday.dk/issues/issue7_6/altinta/index.html">http://www.firstmonday.dk/issues/issue7_6/altinta/index.html</a>
Challenges in Providing Access to Digitized Xrays over the Internet - <a href="http://archive.nlm.nih.gov/pubs/public-access/public-access.php">http://archive.nlm.nih.gov/pubs/public-access/public-access.php</a>
Chatterbots, Tinymuds, And The Turing Test: Entering The Loebner Prize Competition - <a href="http://robot-club.com/lti/pub/aaai94.html">http://robot-club.com/lti/pub/aaai94.html</a>

Class Library Implementation of an Open Architecture Knowledge Support System - <a href="http://ksi.cpsc.ucalgary.ca/articles/KBS/OOAI/OOAI.html">http://ksi.cpsc.ucalgary.ca/articles/KBS/OOAI/OOAI.html</a>
Classification of Binary Document Images into Texture or Non-textual Data Blocks Using Neural Network Models - <a href="http://archive.nlm.nih.gov/pubs/doc_class/mv.php">http://archive.nlm.nih.gov/pubs/doc_class/mv.php</a>
Client/Server Design for Fast Retrieval of Large Images on the Internet - <a href="http://archive.nlm.nih.gov/pubs/client-server/client-server.php">http://archive.nlm.nih.gov/pubs/client-server/client-server.php</a>
Coalition Design for Secure Protocols - <a href="http://szabo.best.vwh.net/coalition.html">http://szabo.best.vwh.net/coalition.html</a>
Cognitive Mechanisms Underlying the Creative Process - <a href="http://cogprints.org/2546/01/CandC.htm">http://cogprints.org/2546/01/CandC.htm</a>
Collective Intelligence, Data Routing and Braess' Paradox - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Combining Knowledge- and Corpus-based Word-Sense-Disambiguation Methods - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/montoyo05a-html/Montoyo05a.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/montoyo05a-html/Montoyo05a.html</a>
Combining Rules and Cases to Learn Case Adaptation - <a href="http://www.cs.indiana.edu/~leake/papers/p-95-09.html">http://www.cs.indiana.edu/~leake/papers/p-95-09.html</a>
Combining Spatial and Temporal Logics: Expressiveness vs. Complexity - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Comments on the specification for FIPA '97 AGENT COMMUNICATION LANGUAGE - <a href="http://www.cs.umbc.edu/kqml/papers/fipa/comments.shtml">http://www.cs.umbc.edu/kqml/papers/fipa/comments.shtml</a>
Committee-Based Sample Selection For Probabilistic Classifiers - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Communications Engineering Branch - <a href="http://archive.nlm.nih.gov/pubs/reports/annual01/report2001.php">http://archive.nlm.nih.gov/pubs/reports/annual01/report2001.php</a>
Comparing Conceptual Structures: Consensus, Conflict, Correspondence and Contrast - <a href="http://ksi.cpsc.ucalgary.ca/articles/KBS/COCO/COCO.html">http://ksi.cpsc.ucalgary.ca/articles/KBS/COCO/COCO.html</a>
Comparing Constructions through the Web - <a href="http://ksi.cpsc.ucalgary.ca/articles/CSCL95WG/">http://ksi.cpsc.ucalgary.ca/articles/CSCL95WG/</a>
Comparing Subjective Contour Perception in Artificial Neural Networks and Infants - <a href="http://www.cs.indiana.edu/event/maics96/Proceedings/Best/best.html">http://www.cs.indiana.edu/event/maics96/Proceedings/Best/best.html</a>
Computational Approach to Anaphora Resolution in Spanish Dialogues - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/palomar01a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/palomar01a-html/</a>
Computational Aspects of Reordering Plans - <a href="http://www.jair.org/vol/vol9.html">http://www.jair.org/vol/vol9.html</a>
Computational Grounding - <a href="http://www.ecs.soton.ac.uk/%7Eharnad/Papers/Harnad">http://www.ecs.soton.ac.uk/%7Eharnad/Papers/Harnad</a>
Connecting Object to Symbol in Modeling Cognition - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad92.symbol.object.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad92.symbol.object.html</a>
Connectionist Theory Refinement: Genetically Searching the Space of Network Topologies - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Cooperation between Top-Down and Bottom-Up Theorem Provers - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Crafting Papers on Machine Learning - <a href="http://www-csli.stanford.edu/icml2k/craft.html">http://www-csli.stanford.edu/icml2k/craft.html</a>
Creating a global knowledge network - <a href="http://people.ccmr.cornell.edu/~ginsparg/blurb/pg01unesco.html">http://people.ccmr.cornell.edu/~ginsparg/blurb/pg01unesco.html</a>
CS 395T: Intelligent Robotics - <a href="http://www.cs.utexas.edu/users/kuipers/cs395T-S06.html">http://www.cs.utexas.edu/users/kuipers/cs395T-S06.html</a>
Cue Phrase Classification Using Machine Learning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Data Storage and Organization for a General Access Xray Image Archive - <a href="http://archive.nlm.nih.gov/pubs/data-storage/data-storage.php">http://archive.nlm.nih.gov/pubs/data-storage/data-storage.php</a>

Data-intensive e-science frontier research - <a href="http://delivery.acm.org">http://delivery.acm.org</a>
Decentralized Control of Cooperative Systems: Categorization and Complexity Analysis - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Decentralized Markets versus Central Control: A Comparative Study - <a href="http://www.jair.org/papers/paper627.html">http://www.jair.org/papers/paper627.html</a>
Decentralized Supply Chain Formation: A Market Protocol and Competitive Equilibrium Analysis - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Decidable Reasoning in Terminological Knowledge Representation Systems - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Decision-Theoretic Bidding Based on Learned Density Models in Simultaneous, Interacting Auctions - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume19/stone03a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume19/stone03a-html/</a>
Decision-Theoretic Foundations for Causal Reasoning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Decision-Theoretic Planning: Structural Assumptions and Computational Leverage - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Desiderata for Agent Communication Languages - ( <a href="http://www.cs.umbc.edu/kqml/papers/desiderata-acl/root.html">http://www.cs.umbc.edu/kqml/papers/desiderata-acl/root.html</a> - university of maryland, baltimore county)
Design Issues for a Digital Xray Archive Accessed over Internet - <a href="http://archive.nlm.nih.gov/pubs/design-issues/design-issues.php">http://archive.nlm.nih.gov/pubs/design-issues/design-issues.php</a>
Design: Are agile methods good for design? - <a href="http://delivery.acm.org">http://delivery.acm.org</a>
Dinosaur Extinction - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/kwon2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/kwon2.html</a>
Distributed Knowledge Modeling through the World Wide Web - <a href="http://ksi.cpsc.ucalgary.ca/articles/AAAI96/KMWWW/KMWWW.html">http://ksi.cpsc.ucalgary.ca/articles/AAAI96/KMWWW/KMWWW.html</a>
Distributing Authorities and Verifying Their Claims - <a href="http://szabo.best.vwh.net/authorities.html">http://szabo.best.vwh.net/authorities.html</a>
Document Classification using Connectionist Models - <a href="http://archive.nlm.nih.gov/pubs/doc_class/ieee.php">http://archive.nlm.nih.gov/pubs/doc_class/ieee.php</a>
Do-I-Care: A Collaborative Web Agent - <a href="http://sigchi.org/chi96/proceedings/shortpap/Starr/sb_txt.htm">http://sigchi.org/chi96/proceedings/shortpap/Starr/sb_txt.htm</a>
Domain Filtering Consistencies - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Dual Modelling of Permutation and Injection Problems - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Dynamic Backtracking - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Dynamic Non-Bayesian Decision Making - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Eastern and Western Shingles Treatment - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/hayesconroya2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/hayesconroya2.html</a>
Efficient Methods for Qualitative Spatial Reasoning - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Efficient Solution Algorithms for Factored MDPs - <a href="http://www.jair.org/papers/paper1000.html">http://www.jair.org/papers/paper1000.html</a>
Eliciting Knowledge and Transferring it Effectively to a Knowledge-Based System - <a href="http://ksi.cpsc.ucalgary.ca/articles/KBS/KSS0/KSS01.html">http://ksi.cpsc.ucalgary.ca/articles/KBS/KSS0/KSS01.html</a>
Emergency Economics - <a href="http://szabo.best.vwh.net/emergencyeconomics.html">http://szabo.best.vwh.net/emergencyeconomics.html</a>
Ethical Issues in Advanced Artificial Intelligence - <a href="http://www.nickbostrom.com/ethics/ai.html">http://www.nickbostrom.com/ethics/ai.html</a>
Evolutionary Algorithms for Reinforcement Learning - <a href="http://www.jair.org/papers/paper613.html">http://www.jair.org/papers/paper613.html</a>
Existence of Multiagent Equilibria with Limited Agents - <a href="http://www.jair.org/papers/paper1332.html">http://www.jair.org/papers/paper1332.html</a>
Experience with the Learning Web - <a href="http://ksi.cpsc.ucalgary.ca/articles/LearnWeb/EM96Exp/">http://ksi.cpsc.ucalgary.ca/articles/LearnWeb/EM96Exp/</a>
Experiments with Infinite-Horizon, Policy-Gradient Estimation - <a href="http://www.jair.org/papers/paper807.html">http://www.jair.org/papers/paper807.html</a>



Exploiting Contextual Independence In Probabilistic Inference - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/poole03a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/poole03a-html/</a>
Extensions of Simple Conceptual Graphs: the Complexity of Rules and Constraints - <a href="http://www.jair.org/papers/paper918.html">http://www.jair.org/papers/paper918.html</a>
Finding a Path is Harder than Finding a Tree - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Finding the Way in the Age of Infoglut - <a href="http://www.ontopia.net/topicmaps/materials/tao.html#d0e140">http://www.ontopia.net/topicmaps/materials/tao.html#d0e140</a>
Flaw Selection Strategies For Partial-Order Planning - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume6/pollack97a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume6/pollack97a-html/</a>
FLECS: Planning with a Flexible Commitment Strategy - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Flexibly Instructable Agents - <a href="http://www.jair.org/papers/paper150.html">http://www.jair.org/papers/paper150.html</a>
Foresight Background 2, Rev. 1 - <a href="http://www.foresight.org/Updates/Background2.html">http://www.foresight.org/Updates/Background2.html</a>
Foresight Guidelines for Responsible Nanotechnology Development - <a href="http://www.foresight.org/guidelines/current.html">http://www.foresight.org/guidelines/current.html</a>
Formalizing and Securing Relationships on Public Networks - <a href="http://www.firstmonday.dk/issues/issue2_9/szabo/index.html">http://www.firstmonday.dk/issues/issue2_9/szabo/index.html</a>
Foundations for the Learning Web - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/LW/EM96Found/index.html">http://pages.cpsc.ucalgary.ca/~gaines/reports/LW/EM96Found/index.html</a>
From Hand To Mouth: Some Critical Stages In The Evolution Of Language - <a href="http://eprints.ecs.soton.ac.uk/3346/03/harnad76.hand.html">http://eprints.ecs.soton.ac.uk/3346/03/harnad76.hand.html</a>
Generalization of Clauses under Implication - <a href="http://www.jair.org/papers/paper194.html">http://www.jair.org/papers/paper194.html</a>
GIB: Imperfect Information in a Computationally Challenging Game - <a href="http://www.jair.org/papers/paper820.html">http://www.jair.org/papers/paper820.html</a>
Goal Composition: Extending Task Analysis to Predict Things People May Want to Do - <a href="http://www.useit.com/papers/goalcomposition.html">http://www.useit.com/papers/goalcomposition.html</a>
Goal Recognition through Goal Graph Analysis - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Graduality in Argumentation - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/cayrol05a-html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/cayrol05a-html</a>
Grounded Semantic Composition for Visual Scenes - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume21/gorniak04a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume21/gorniak04a-html/</a>
Grounding the Lexical Semantics of Verbs in Visual Perception using Force Dynamics and Event Logic - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/siskind01a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/siskind01a-html/</a>
Hindu Philosophy - <a href="http://san.beck.org/EC11-Hindu.html">http://san.beck.org/EC11-Hindu.html</a>
How Long Before Superintelligence? - <a href="http://www.nickbostrom.com/superintelligence.html">http://www.nickbostrom.com/superintelligence.html</a>
How to Conduct a Heuristic Evaluation - <a href="http://www.useit.com/papers/heuristic/heuristic_evaluation.html">http://www.useit.com/papers/heuristic/heuristic_evaluation.html</a>
Hybrid BDI-POMDP Framework for Multiagent Teaming - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/nair05a-html/nair05a.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/nair05a-html/nair05a.html</a>
Hydrophobicity Analysis of Amino Acids - <a href="http://www.mat.univie.ac.at/~neum/software/protein/amino-acids.html">http://www.mat.univie.ac.at/~neum/software/protein/amino-acids.html</a>
iCab: New Browser With Structural Navigation - <a href="http://www.useit.com/papers/icab.html">http://www.useit.com/papers/icab.html</a>
Identifying image structures for content-based retrieval of digitized spine x-rays - <a href="http://archive.nlm.nih.gov/pubs/long/spie-sd2003/spie-sd2003.php">http://archive.nlm.nih.gov/pubs/long/spie-sd2003/spie-sd2003.php</a>
IDL-Expressions: A Formalism for Representing and Parsing Finite Languages in Natural Language Processing - <a href="http://www.jair.org/vol/vol21.html">http://www.jair.org/vol/vol21.html</a>
If Uploads Come First - <a href="http://hanson.gmu.edu/uploads.html">http://hanson.gmu.edu/uploads.html</a>
Image Databases in Medicine - <a href="http://archive.nlm.nih.gov/pubs/nsf/nsf-wkshp-meddb.php">http://archive.nlm.nih.gov/pubs/nsf/nsf-wkshp-meddb.php</a>

Image Query and Indexing for Digital X-rays - <a href="http://archive.nlm.nih.gov/pubs/sdiut/ist99.php">http://archive.nlm.nih.gov/pubs/sdiut/ist99.php</a>
Imperfect Match: PDDL 2.1 and Real Applications - <a href="http://www.jair.org/papers/paper1994.html">http://www.jair.org/papers/paper1994.html</a>
Improved Heterogeneous Distance Functions - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume6/wilson97a.html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume6/wilson97a.html/</a>
Improving Connectionist Energy Minimization - <a href="http://www.jair.org/papers/paper130.html">http://www.jair.org/papers/paper130.html</a>
Improving the Efficiency of Inductive Logic Programming Through the Use of Query Packs - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/blockeel02a.html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/blockeel02a.html/</a>
IMS Abstract Framework: White Paper - <a href="http://www.imsglobal.org/af/afv1p0/imsafwhitepaperv1p0.html">http://www.imsglobal.org/af/afv1p0/imsafwhitepaperv1p0.html</a>
Incremental Recompilation of Knowledge - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Induction of First-Order Decision Lists: Results on Learning the Past Tense of English Verbs - <a href="http://www.jair.org/papers/paper148.html">http://www.jair.org/papers/paper148.html</a>
Induction of Ripple-Down Rules Applied to Modeling Large Databases - <a href="http://ksi.cpsc.ucalgary.ca/articles/Induct/JIIS95/JIIS95.html">http://ksi.cpsc.ucalgary.ca/articles/Induct/JIIS95/JIIS95.html</a>
Inferring Strategies for Sentence Ordering in Multidocument News Summarization - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Information Retrieval Issues - <a href="http://www.e-papyrus.com/hypertext_review/chapter5.html">http://www.e-papyrus.com/hypertext_review/chapter5.html</a>
Information Retrieval of Imperfectly Recognized Handwriting - <a href="http://www.useit.com/papers/handwriting_retrieval.html">http://www.useit.com/papers/handwriting_retrieval.html</a>
Integrative Windowing - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume8/fuernkranz98a.html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume8/fuernkranz98a.html/</a>
Intelligent Agents and Multi-Agents - <a href="http://www.asap.cs.nott.ac.uk/themes/ma.shtml">http://www.asap.cs.nott.ac.uk/themes/ma.shtml</a>
Interactive Execution Monitoring of Agent Teams - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume18/wilkins03a.html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume18/wilkins03a.html/</a>
Interactive Publication: Extending The American Physical Society's Discipline-Specific Model For Electronic Publishing - <a href="http://www.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad92.interactivpub.html">http://www.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad92.interactivpub.html</a>
International perspectives: Y2K international status and opportunitie - <a href="http://delivery.acm.org">http://delivery.acm.org</a>
Internet Encyclopedia of Philosophy - <a href="http://www.utm.edu/research/iep/c/carnap.htm">http://www.utm.edu/research/iep/c/carnap.htm</a>
Internet in the Lives of Turkish Women - <a href="http://www.firstmonday.dk/issues/issue7_3/sevdik/index.html">http://www.firstmonday.dk/issues/issue7_3/sevdik/index.html</a>
Interpreting Power: The Principle of Least Authority - <a href="http://szabo.best.vwh.net/interpretingpower.html">http://szabo.best.vwh.net/interpretingpower.html</a>
Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents - <a href="http://www.msci.memphis.edu/franklin/AgentProg.html">http://www.msci.memphis.edu/franklin/AgentProg.html</a>
Issues in Stacked Generalization - <a href="http://www.jair.org/papers/paper594.html">http://www.jair.org/papers/paper594.html</a>
Kelly's "Geometry of Psychological Space"and its Significance for Cognitive Modeling - <a href="http://www.ucalgary.ca/~dmjacobs/cpsc679/kelly.html">http://www.ucalgary.ca/~dmjacobs/cpsc679/kelly.html</a>
K-Implementation - <a href="http://www.jair.org/vol/vol21.html">http://www.jair.org/vol/vol21.html</a>
Knowledge Acquisition Tools based on Personal Construct Psychology - <a href="http://ksi.cpsc.ucalgary.ca/articles/KBS/KER/KER1.html">http://ksi.cpsc.ucalgary.ca/articles/KBS/KER/KER1.html</a>
Knowledge representation - <a href="http://en.wikipedia.org/wiki/Knowledge_representation">http://en.wikipedia.org/wiki/Knowledge_representation</a>
Knowledge Systematization in the International IMS Research Program - <a href="http://ksi.cpsc.ucalgary.ca/articles/SMC95GN/">http://ksi.cpsc.ucalgary.ca/articles/SMC95GN/</a>
KQML - A Language and Protocol for Knowledge and Information Exchange - <a href="http://www.cs.umbc.edu/kqml/papers/kbkshtml/kbks.html">http://www.cs.umbc.edu/kqml/papers/kbkshtml/kbks.html</a>
KQML as an Agent Communication Language - <a href="http://www.cs.umbc.edu/kqml/papers/kqml-acl.html/root2.html">http://www.cs.umbc.edu/kqml/papers/kqml-acl.html/root2.html</a>

Language of Thought Hypothesis: State of the Art - <a href="http://web.clas.ufl.edu/users/maydede/LOTH.SEP.html">http://web.clas.ufl.edu/users/maydede/LOTH.SEP.html</a>	-
Large Cranberry and High Bush Blueberry Cultivation - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web1/hayesconroya.html">http://serendip.brynmawr.edu/biology/b103/f00/web1/hayesconroya.html</a>	-
Learning Geometrically-Constrained Hidden Markov Models for Robot Navigation: Bridging the Topological-Geometrical Gap - <a href="http://www.jair.org/vol/vol16.html">http://www.jair.org/vol/vol16.html</a>	
Learning Membership Functions in a Function-Based Object Recognition System - <a href="http://www.jair.org/papers/paper236.html">http://www.jair.org/papers/paper236.html</a>	-
Learning Sets of Related Concepts: A Shared Task Model - <a href="http://www.ics.uci.edu/~pazzani/CogSci95.html">http://www.ics.uci.edu/~pazzani/CogSci95.html</a>	-
Learning to Order Things - <a href="http://www.jair.org/abstracts/cohen99a.html">http://www.jair.org/abstracts/cohen99a.html</a>	
Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction - <a href="http://www.jair.org/papers/paper1199.html">http://www.jair.org/papers/paper1199.html</a>	-
LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume22/erkan04a-html/erkan04a.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume22/erkan04a-html/erkan04a.html</a>	-
Lively Flights Of Fancy - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad97.thes.alife.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad97.thes.alife.html</a>	
Logarithmic-Time Updates and Queries in Probabilistic Networks - <a href="http://www.jair.org/papers/paper238.html">http://www.jair.org/papers/paper238.html</a>	-
Long-Term Growth As A Sequence of Exponential Modes - <a href="http://hanson.gmu.edu/longgrow.html">http://hanson.gmu.edu/longgrow.html</a>	
MAS-SOC: a Social Simulation Platform Based on Agent-Oriented Programming - <a href="http://jasss.soc.surrey.ac.uk/8/3/7.html">http://jasss.soc.surrey.ac.uk/8/3/7.html</a>	-
Mathematical Modeling - <a href="http://www.mat.univie.ac.at/~neum/model.html">http://www.mat.univie.ac.at/~neum/model.html</a>	
Mean Field Methods for a Special Class of Belief Networks - <a href="http://www.jair.org/papers/paper734.html">http://www.jair.org/papers/paper734.html</a>	
Mean Field Theory for Sigmoid Belief Networks - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>	
Mechanisms for Automated Negotiation in State Oriented Domains - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>	
Minds, Machines and Searle - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad89.searle.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad89.searle.html</a>	
Modeling and Forecasting the Information Sciences - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/MFIT/InfSci/index.html">http://pages.cpsc.ucalgary.ca/~gaines/reports/MFIT/InfSci/index.html</a>	
Motion Planning - <a href="http://www.cs.hut.fi/~evp/Res/MP/motion_planning.html">http://www.cs.hut.fi/~evp/Res/MP/motion_planning.html</a>	
Motion Planning for Dexterous Robots - <a href="http://www-users.cs.umn.edu/~gini/motion.html">http://www-users.cs.umn.edu/~gini/motion.html</a>	
Multi-group Analysis with LISREL: An Appendix to "Pragmatic Perspectives on the Measurement of Information Systems Service Quality"- <a href="http://www.misq.org/archivist/appendices/article2.html">http://www.misq.org/archivist/appendices/article2.html</a>	
Natural Language - <a href="http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/natlang.html">http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/natlang.html</a>	
Natural Language Processing... Understanding what you say - <a href="http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol1/ctt/#Introduction">http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol1/ctt/#Introduction</a>	-
Neoconstructivism: A Unifying Constraint for the Cognitive Sciences - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad82.neoconst.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad82.neoconst.html</a>	-
Networking in the Mind Age - <a href="http://www.lucifer.com/~sasha/mindage.html">http://www.lucifer.com/~sasha/mindage.html</a>	
Networks, Netwars, and the Fight for the Future - <a href="http://www.firstmonday.org/issues/issue6_10/ronfeldt/">http://www.firstmonday.org/issues/issue6_10/ronfeldt/</a>	-
Neural Networks - <a href="http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction">http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction</a> to neural networks	
Neurotheology: Which came first, God or the brain? - <a href="http://serendip.brynmawr.edu/biology/b103/f01/web3/ford.html">http://serendip.brynmawr.edu/biology/b103/f01/web3/ford.html</a>	-

NHK Laboratories Note No. 464 Speech Recognition Engine For Realime Broadcast News Captioning - <a href="http://www.nhk.or.jp/strl/publica/labnote/lab464.html">http://www.nhk.or.jp/strl/publica/labnote/lab464.html</a>
No Silver Bullet Essence and Accidents of Software Engineering - <a href="http://www.virtualschool.edu/mon/SoftwareEngineering/BrooksNoSilverBullet.html">http://www.virtualschool.edu/mon/SoftwareEngineering/BrooksNoSilverBullet.html</a>
OBDD-based Universal Planning for Synchronized Agents in Non-Deterministic Domains - <a href="http://www.jair.org/papers/paper649.html">http://www.jair.org/papers/paper649.html</a>
On Partially Controlled Multi-Agent Systems - <a href="http://www.jair.org/papers/paper318.html">http://www.jair.org/papers/paper318.html</a>
On Planning while Learning - <a href="http://iew3.technion.ac.il/Home/Users/SECOND.php?Moshet+Moshe+Tennenholtz+2+1">http://iew3.technion.ac.il/Home/Users/SECOND.php?Moshet+Moshe+Tennenholtz+2+1</a>
On Polynomial Sized MDP Succinct Policies - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
On the Informativeness of the DNA Promoter Sequences Domain Theory - <a href="http://www.jair.org/papers/paper161.html">http://www.jair.org/papers/paper161.html</a>
Ontologies as the Representation (and Re-Representation) of Agreement - <a href="http://www-ksl.stanford.edu/KR96/BillMarkPositionPaper.html">http://www-ksl.stanford.edu/KR96/BillMarkPositionPaper.html</a>
Ontologies: What are they, and where's the research? - <a href="http://www-ksl.stanford.edu/KR96/PanelIntro.html">http://www-ksl.stanford.edu/KR96/PanelIntro.html</a>
Ontology (computer science) - <a href="http://www.absoluteastronomy.com/">http://www.absoluteastronomy.com/</a>
Operations for Learning with Graphical Models - <a href="http://www.jair.org/papers/paper62.html">http://www.jair.org/papers/paper62.html</a>
Optimizing Dialogue Management with Reinforcement Learning: Experiments with the NJFun System - <a href="http://www.jair.org/vol/vol16.html">http://www.jair.org/vol/vol16.html</a>
Ordered Landmarks in Planning - <a href="http://www.jair.org/papers/paper1492.html">http://www.jair.org/papers/paper1492.html</a>
Origins of the Joint-Stock Company - <a href="http://szabo.best.vwh.net/jointstock.html">http://szabo.best.vwh.net/jointstock.html</a>
Other Bodies, Other Minds A Machine Incarnation of an Old Philosophical Problem - <a href="http://cogprints.org/1578/00/harnad91.otherminds.html">http://cogprints.org/1578/00/harnad91.otherminds.html</a>
OWL Web Ontology Language Overview - <a href="http://www.w3.org/TR/owl-features/">http://www.w3.org/TR/owl-features/</a>
Parameter Learning of Logic Programs for Symbolic-statistical Modeling - <a href="http://www.jair.org/papers/paper912.html">http://www.jair.org/papers/paper912.html</a>
Partial-Order Planning with Concurrent Interacting Actions - <a href="http://www.jair.org/vol/vol14.html">http://www.jair.org/vol/vol14.html</a>
Pattern Matching and Discourse Processing in Information Extraction from Japanese Text - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Patterns of Integrity – Separation of Duties - <a href="http://szabo.best.vwh.net/separationofduties.html">http://szabo.best.vwh.net/separationofduties.html</a>
PDDL2.1 : An Extension to PDDL for Expressing Temporal Planning Domains - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume20/fox03a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume20/fox03a-html/</a>
Perseus: Randomized Point-based Value Iteration for POMDPs - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
PHA*: Finding the Shortest Path with A* in An Unknown Physical Environment - <a href="http://www.jair.org/papers/paper1373.html">http://www.jair.org/papers/paper1373.html</a>
Planning - <a href="http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/planning.html">http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/planning.html</a>
Planning for Contingencies: A Decision-based Approach - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume4/pryor96a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume4/pryor96a-html/</a>
Planning Graph as a (Dynamic) CSP: Exploiting EBL, DDB and other CSP Search Techniques in Graphplan - <a href="http://www.jair.org/papers/paper655.html">http://www.jair.org/papers/paper655.html</a>
Policy Recognition in the Abstract Hidden Markov Model - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Popular Ensemble Methods: An Empirical Study - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volumel1/opitz99a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volumel1/opitz99a-html/</a>

Positive Feedback Processes Underlying the Formation of Expertise - <a href="http://ksi.cpsc.ucalgary.ca/articles/PosFB/PosFB.html">http://ksi.cpsc.ucalgary.ca/articles/PosFB/PosFB.html</a>
Post-Gutenberg Galaxy: The Fourth Revolution In The Means Of Production Of Knowledge - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html</a>
Probabilistic Inference from Arbitrary Uncertainty using Mixtures of Factorized Generalized Gaussians - <a href="http://www.jair.org/papers/paper533.html">http://www.jair.org/papers/paper533.html</a>
Propositional Independence: Formula-Variable Independence and Forgetting - <a href="http://www.jair.org/papers/paper1113.html">http://www.jair.org/papers/paper1113.html</a>
Protein design as a pathway to molecular manufacturing - <a href="http://www.imm.org/PNAS.html">http://www.imm.org/PNAS.html</a>
Provably Bounded-Optimal Agents - <a href="http://www.jair.org/papers/paper133.html">http://www.jair.org/papers/paper133.html</a>
Pure Nash Equilibria: Hard and Easy Games - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Quantitative Results Comparing Three Intelligent Interfaces for Information Capture: A Case Study Adding Name Information into an Electronic Personal Organizer - <a href="http://www.cs.washington.edu/research/jair/volume5/schlimmer96a-html/schlimmer96-0.html">http://www.cs.washington.edu/research/jair/volume5/schlimmer96a-html/schlimmer96-0.html</a>
Randomized Algorithms for the Loop Cutset Problem - <a href="http://www.jair.org/vol/vol12.html">http://www.jair.org/vol/vol12.html</a>
Real Time User Context Modeling for Information Retrieval Agents - <a href="http://www.cs.indiana.edu/~leake/papers/p-01-09_dir.html">http://www.cs.indiana.edu/~leake/papers/p-01-09_dir.html</a>
Reasoning about Action: An Argumentation-Theoretic Approach - <a href="http://www.jair.org/papers/paper1602.html">http://www.jair.org/papers/paper1602.html</a>
Reasoning on Interval and Point-based Disjunctive Metric Constraints in Temporal Contexts - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume12/barber00a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume12/barber00a-html/</a>
Reasoning within Fuzzy Description Logics - <a href="http://www.jair.org/papers/paper813.html">http://www.jair.org/papers/paper813.html</a>
Reinforcement Learning for Agents with Many Sensors and Actuators Acting in Categorizable Environments - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume23/porta05a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume23/porta05a-html/</a>
Reinforcement Learning: A Survey - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/kaelbling96a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/kaelbling96a-html/</a>
Replicability of cognitive imaging of the cerebral cortex By PET and fMRI: A survey of recent literature - <a href="http://cogprints.org/1903/00/1.html">http://cogprints.org/1903/00/1.html</a>
Representing and Aggregating Conflicting Beliefs - <a href="http://www.jair.org/papers/paper1206.html">http://www.jair.org/papers/paper1206.html</a>
Representing Hard-to-Formalise, Contextualised, Multidisciplinary, Organisational Knowledge - <a href="http://kmi.open.ac.uk/people/sbs/org-knowledge/aikm97/sbs-paper2.html">http://kmi.open.ac.uk/people/sbs/org-knowledge/aikm97/sbs-paper2.html</a>
Rerepresenting and Restructuring Domain Theories: A Constructive Induction Approach - <a href="http://www.jair.org/papers/paper129.html">http://www.jair.org/papers/paper129.html</a>
Rights, Remedies, and Security Models - <a href="http://szabo.best.vwh.net/rrsmodels.html">http://szabo.best.vwh.net/rrsmodels.html</a>
Risk-Sensitive Reinforcement Learning Applied to Control under Constraints - <a href="http://www.jair.org/papers/paper1666.html">http://www.jair.org/papers/paper1666.html</a>
Robotics - <a href="http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/robotics.html">http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/robotics.html</a>
Robust Agent Teams via Socially-Attentive Monitoring - <a href="http://www.jair.org/vol/vol12.html">http://www.jair.org/vol/vol12.html</a>
Rule-based Machine Learning Methods for Functional Prediction - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Sapa: A Multi-objective Metric Temporal Planner - <a href="http://www.jair.org/vol/vol20.html">http://www.jair.org/vol/vol20.html</a>
Scarce Objects - <a href="http://szabo.best.vwh.net/scarce.html">http://szabo.best.vwh.net/scarce.html</a>
Schemas - <a href="http://wik.ed.uiuc.edu/index.php/Schemas">http://wik.ed.uiuc.edu/index.php/Schemas</a>
Scholarly Skywriting And The Prepublication Continuum Of Scientific Inquiry - <a href="http://www.ecs.soton.ac.uk/%7Eharnad/Papers/Harnad/harnad90.skywriting.html">http://www.ecs.soton.ac.uk/%7Eharnad/Papers/Harnad/harnad90.skywriting.html</a>

Science and the Precautionary Principle - <a href="http://www.biotech-info.net/science_and_pp.html">http://www.biotech-info.net/science_and_pp.html</a>
Search in ai <a href="http://jsl.sourceforge.net/search.html">http://jsl.sourceforge.net/search.html</a>
Search Methods - <a href="http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/">http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/</a>
Searching for Bayesian Network Structures in the Space of Restricted Acyclic Partially Directed Graphs - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/acid03a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume18/acid03a-html/</a>
Seductive User Interfaces - <a href="http://www.useit.com/papers/seductiveui.html">http://www.useit.com/papers/seductiveui.html</a>
Segmentation and Feature Extraction of Cervical Spine X-ray Images - <a href="http://archive.nlm.nih.gov/pubs/sdiut/sd99.php">http://archive.nlm.nih.gov/pubs/sdiut/sd99.php</a>
Self-Replication: Even peptides do it - <a href="http://www.santafe.edu/sfi/People/kauffman/sak-peptides.html">http://www.santafe.edu/sfi/People/kauffman/sak-peptides.html</a>
Semantic Networks - <a href="http://www.jfsowa.com/pubs/semnet.htm">http://www.jfsowa.com/pubs/semnet.htm</a>
Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library?c=jair&amp;p=about">http://www.sadl.uleth.ca/nz/cgi-bin/library?c=jair&amp;p=about</a>
Semantic Web - <a href="http://en.wikipedia.org/wiki/Semantic_Web">http://en.wikipedia.org/wiki/Semantic_Web</a>
Semantic Web vision paper - <a href="http://www.lucifer.com/~sasha/articles/SemanticWeb.html">http://www.lucifer.com/~sasha/articles/SemanticWeb.html</a>
Semantics, Conceptual Role - <a href="http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ConceptualRoleSemantics.html">http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ConceptualRoleSemantics.html</a>
SHADE: Knowledge-Based Technology for the Re-Engineering Problem - <a href="http://www.ksl.stanford.edu/knowledge-sharing/papers/shade-overview.html">http://www.ksl.stanford.edu/knowledge-sharing/papers/shade-overview.html</a>
Shelling Out – The Origins of Money - <a href="http://szabo.best.vwh.net/shell.html">http://szabo.best.vwh.net/shell.html</a>
SHOP2: An HTN Planning System - <a href="http://www.jair.org/vol/vol20.html">http://www.jair.org/vol/vol20.html</a>
Single-Sweep Analysis of the Theta Frequency Band During an Auditory Oddball Task - <a href="http://cogprints.org/45/00/tetad.htm">http://cogprints.org/45/00/tetad.htm</a>
Situated Adaptive Design: Toward a New Methodology for Knowledge Systems Development - <a href="http://www.ifi.unizh.ch/staff/pfeifer/">http://www.ifi.unizh.ch/staff/pfeifer/</a>
SMOTE: Synthetic Minority Over-sampling Technique - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a-html/</a>
Social Information Filtering: Algorithms for Automating "Word of Mouth" - <a href="http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/us_bdy.htm">http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/us_bdy.htm</a>
Solving Multiclass Learning Problems via Error-Correcting Output Codes - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Solving Transition Independent Decentralized Markov Decision Processes - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Some Philosophical Problems From The Standpoint Of Artificial Intelligence - <a href="http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html">http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html</a>
Space Efficiency of Propositional Knowledge Representation Formalisms - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume13/cadoli00a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume13/cadoli00a-html/</a>
Spatial Aggregation: Theory and Applications - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Speeding Up the Convergence of Value Iteration in Partially Observable Markov Decision Processes - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume14/zhang01a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume14/zhang01a-html/</a>
Squeaky Wheel Optimization - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume10/joslin99a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume10/joslin99a-html/</a>
Steps Toward Artificial Intelligence - <a href="http://web.media.mit.edu/~minsky/papers/steps.html">http://web.media.mit.edu/~minsky/papers/steps.html</a>
Storing and Indexing Plan Derivations through Explanation-based Analysis of Retrieval Failures - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume7/ihrig97a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume7/ihrig97a-html/</a>

Structured and Unstructured Induction with EDAGs - <a href="http://ksi.cpsc.ucalgary.ca/articles/Induct/EDAG95/">http://ksi.cpsc.ucalgary.ca/articles/Induct/EDAG95/</a>
Structured Knowledge Representation for Image Retrieval - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/disciascio02a-html/disciascio02a.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/disciascio02a-html/disciascio02a.html</a>
Substructure Discovery Using Minimum Description Length and Background Knowledge - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Sun Microsystem's 1997 Web Design - <a href="http://www.useit.com/papers/sun/1997">http://www.useit.com/papers/sun/1997</a>
Symbol Grounding is an Empirical Problem: Neural Nets are Just a Candidate Component - <a href="http://www.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad93.cogsci.html">http://www.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad93.cogsci.html</a>
TALplanner in the Third International Planning Competition: Extensions and Control Rules - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Taming Numbers and Durations in the Model Checking Integrated Planning System - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Technical Abstract for Algernon and Access-Limited Logic - <a href="http://www.cs.utexas.edu/users/qr/algernon-technical.html">http://www.cs.utexas.edu/users/qr/algernon-technical.html</a>
Technical Paper Recommendation: A Study in Combining Multiple Information Sources - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Techniques used in Natural Language Processing - <a href="http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol2/">http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol2/</a>
Temporal Decision Trees: Model-based Diagnosis of Dynamic Systems On-Board - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Temporal Difference Learning and TD-Gammon - <a href="http://www.research.ibm.com/massive/tdl.html">http://www.research.ibm.com/massive/tdl.html</a>
Text Document Classification with Self-Organizing Maps - <a href="http://lipas.uwasa.fi/stes/step96/step96/hyotyniemi3/HH3.html">http://lipas.uwasa.fi/stes/step96/step96/hyotyniemi3/HH3.html</a>
The Abuse of Statistics in Cryptography - <a href="http://szabo.best.vwh.net/statistics.html">http://szabo.best.vwh.net/statistics.html</a>
The AIDS Epidemic in Sub-Saharan Africa - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/kennedy2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/kennedy2.html</a>
The Anatomy of a Large-Scale Hypertextual Web Search Engine - <a href="http://infolab.stanford.edu/~backrub/google.html">http://infolab.stanford.edu/~backrub/google.html</a>
The Anti-Mac Interface - <a href="http://www.acm.org/pubs/cacm/AUG96/antimac.htm">http://www.acm.org/pubs/cacm/AUG96/antimac.htm</a>
The Application of Neural Networks in the Modeling of Plate Rolling Processes - <a href="http://www.tms.org/pubs/journals/JOM/9704/Gorni/Gorni-9704.html">http://www.tms.org/pubs/journals/JOM/9704/Gorni/Gorni-9704.html</a>
The Ariadne's Clew Algorithm - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/mazer98a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume9/mazer98a-html/</a>
The Automatic Inference of State Invariants in TIM - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Collective Stance in Modeling Expertise in Individuals and Organizations - <a href="http://ksi.cpsc.ucalgary.ca/articles/Collective">http://ksi.cpsc.ucalgary.ca/articles/Collective</a>
The Design Process: A Computational Perspective - <a href="http://www.cs.iastate.edu/~honavar/dbd-position-paper.html">http://www.cs.iastate.edu/~honavar/dbd-position-paper.html</a>
The Deterministic Part of IPC-4: An Overview - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Divide-and-Conquer Subgoal-Ordering Algorithm for Speeding up Logic Inference - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Doomsday Argument is Alive and Kicking - <a href="http://www.anthropic-principle.com/preprints/ali/alive.html">http://www.anthropic-principle.com/preprints/ali/alive.html</a>
The Economic Approach to Artificial Intelligence - <a href="http://delivery.acm.org">http://delivery.acm.org</a>
The Effects of Alcohol on the Body - <a href="http://serendip.brynmawr.edu/biology/b103/f01/web1/chew.html">http://serendip.brynmawr.edu/biology/b103/f01/web1/chew.html</a>

The FF Planning System: Fast Plan Generation Through Heuristic Search - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Future of Hypertext - <a href="http://www.useit.com/papers/hypertextfuture.html">http://www.useit.com/papers/hypertextfuture.html</a>
The Gnm Phase Transition is Not Hard for the HC Problem - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The God Protocols - <a href="http://www.theiia.org/ITAAudit/index.cfm?act=itaaudit.archive&amp;fid=216">http://www.theiia.org/ITAAudit/index.cfm?act=itaaudit.archive&amp;fid=216</a>
The Good Old Davis-Putnam Procedure Helps Counting Models - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Great Filter - Are We Almost Past It? - <a href="http://hanson.gmu.edu/greatfilter.html">http://hanson.gmu.edu/greatfilter.html</a>
The GRT Planning System: Backward Heuristic Construction in Forward State-Space Planning - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/refanidis01a.html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume15/refanidis01a.html/</a>
The Harder Problem of Consciousness - <a href="http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/harder.htm">http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/harder.htm</a>
The Knowledge Argument Against Physicalism - <a href="http://www.iep.utm.edu/k/know-arg.htm">http://www.iep.utm.edu/k/know-arg.htm</a>
The Learning Web: A System View and an Agent-Oriented Model - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/LW/">http://pages.cpsc.ucalgary.ca/~gaines/reports/LW/</a>
The Mental Accounting Barrier to Micropayments - <a href="http://szabo.best.vwh.net/micropayments.html">http://szabo.best.vwh.net/micropayments.html</a>
The Metric-FF Planning System: Translating "Ignoring Delete Lists" to Numeric State Variables - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
The Mind as the Software of the Brain - <a href="http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html">http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html</a>
The Moron's Guide to Kerberos, Version 2.0 - <a href="http://www.isi.edu/gost/brian/security/kerberos.html">http://www.isi.edu/gost/brian/security/kerberos.html</a>
The Paper House of Cards (and why it's taking so long to collapse) - <a href="http://www.ariadne.ac.uk/issue8/harnad/">http://www.ariadne.ac.uk/issue8/harnad/</a>
The Role of Knowledge Level Models in Situated Design - <a href="http://arti.vub.ac.be/www/icons/previous.xbm">http://arti.vub.ac.be/www/icons/previous.xbm</a>
The Self-Organizing Map (SOM) - <a href="http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm.shtml">http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm.shtml</a>
The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion - <a href="http://cogprints.org/1584/00/harnad92.turing.html">http://cogprints.org/1584/00/harnad92.turing.html</a>
There is No Universal Security Architecture - <a href="http://szabo.best.vwh.net/universal.html">http://szabo.best.vwh.net/universal.html</a>
To Uplift or Not To Uplift: The Ethics of Genetic Manipulation - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web1/kaczmarek.html">http://serendip.brynmawr.edu/biology/b103/f00/web1/kaczmarek.html</a>
Toward Principles for the Design of Ontologies Used for Knowledge Sharing - <a href="http://ksl-web.stanford.edu/KSL_Abstracts/KSL-93-04.html">http://ksl-web.stanford.edu/KSL_Abstracts/KSL-93-04.html</a>
Towards a Multi-Agent Model for Visualizing Simulated User Behavior to Support the Assessment of Design Performance - <a href="http://www.ddss.nl/Eindhoven/publications/176">http://www.ddss.nl/Eindhoven/publications/176</a>
Towards a WWW information filtering and seeking agent - <a href="http://www.ics.uci.edu/~pazzani/Coldlist.html">http://www.ics.uci.edu/~pazzani/Coldlist.html</a>
Towards Adjustable Autonomy for the Real World - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Towards Flexible Teamwork - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Transforming Rules and Trees into Comprehensible Knowledge Structures - <a href="http://ksi.cpsc.ucalgary.ca/articles/Induct/EDAG94/EDAG94.html">http://ksi.cpsc.ucalgary.ca/articles/Induct/EDAG94/EDAG94.html</a>
Translating between Horn Representations and their Characteristic Models - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Translation of Pronominal Anaphora between English and Spanish: Discrepancies and Evaluation - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume18/peral03a.html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume18/peral03a.html/</a>



Trusted Third Parties Are Security Holes - <a href="http://szabo.best.vwh.net/ttps.html">http://szabo.best.vwh.net/ttps.html</a>
Turing Indistinguishability And The Blind Watchmaker - <a href="http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad98.turing.evol.html">http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad98.turing.evol.html</a>
Two Steps Closer on Consciousness - <a href="http://ase.tufts.edu/cogstud/papers/churchland.htm">http://ase.tufts.edu/cogstud/papers/churchland.htm</a>
Unifying Class-Based Representation Formalisms - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume11/calvanese99a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume11/calvanese99a-html/</a>
Using Computational Critics to Facilitate Long-term Collaboration in User Interface Design - <a href="http://delivery.acm.org">http://delivery.acm.org</a>
Using Memory to Transform Search on the Planning Graph - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/zimmerman05a-html/PEGG-zimmerman-kambhampati.htm">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume23/zimmerman05a-html/PEGG-zimmerman-kambhampati.htm</a>
Using Pivot Consistency to Decompose and Solve Functional CSPs - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Variational Cumulant Expansions for Intractable Distributions - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume10/jaakkola99a-html/paper.html">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume10/jaakkola99a-html/paper.html</a>
Variational Probabilistic Inference and the QMR-DT Network - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume10/jaakkola99a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume10/jaakkola99a-html/</a>
Verifying Machines' Minds - <a href="http://www.ecs.soton.ac.uk/~harnad/Temp/Kata/verifying.minds.html">http://www.ecs.soton.ac.uk/~harnad/Temp/Kata/verifying.minds.html</a>
Vernor Vinge on the Singularity - <a href="http://www.geocities.com/jhill314/Singularity.html">http://www.geocities.com/jhill314/Singularity.html</a>
Vertebra Shape Classification using MLP for Content-Based Image Retrieval - <a href="http://archive.nlm.nih.gov/pubs/antani/1031-791/1031-791.php">http://archive.nlm.nih.gov/pubs/antani/1031-791/1031-791.php</a>
VHPOP: Versatile Heuristic Partial Order Planner - <a href="http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume20/younes03a-html/">http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume20/younes03a-html/</a>
Virtual Symposium on Virtual Mind - <a href="http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.virtualmind.html">http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.virtualmind.html</a>
Vision - <a href="http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/vision.html">http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/vision.html</a>
Vision-Based Road Detection in Automotive Systems: a Real-Time Expectation-Driven Approach - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Visual Languages for Knowledge Representation - <a href="http://pages.cpsc.ucalgary.ca/~kremer/papers/KAW98/visual/kremer-visual.html#_Toc2.1">http://pages.cpsc.ucalgary.ca/~kremer/papers/KAW98/visual/kremer-visual.html#_Toc2.1</a>
We are the music makers, We are the dreamers of dreams - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web3/kaczmarek3.html">http://serendip.brynmawr.edu/biology/b103/f00/web3/kaczmarek3.html</a>
Web Hunting: Design of a Simple Intelligent Web Search Agent - <a href="https://secure.acm.org/crossroads/xrds5-4/webhunting.html">https://secure.acm.org/crossroads/xrds5-4/webhunting.html</a>
Web Page Downloading and Classification - <a href="http://archive.nlm.nih.gov/pubs/tran/webpageload.php">http://archive.nlm.nih.gov/pubs/tran/webpageload.php</a>
WebMap: Concept Mapping on the Web - <a href="http://pages.cpsc.ucalgary.ca/~gaines/reports/">http://pages.cpsc.ucalgary.ca/~gaines/reports/</a>
Well-Founded Semantics for Extended Logic Programs with Dynamic Preferences - <a href="http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/brewka96a-html/">http://www-2.cs.cmu.edu/afs/cs/project/jair/pub/volume4/brewka96a-html/</a>
What is Cloning - Is it Right or Wrong? - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web1/ali.html">http://serendip.brynmawr.edu/biology/b103/f00/web1/ali.html</a>
What is Unicode? - <a href="http://www.unicode.org/standard/whatisunicode.html">http://www.unicode.org/standard/whatisunicode.html</a>
What's All the Huff...Or Rather Puff...About? - <a href="http://serendip.brynmawr.edu/biology/b103/f01/web2/arima.html">http://serendip.brynmawr.edu/biology/b103/f01/web2/arima.html</a>
When do Numbers Really Matter? - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
When will computer hardware match the human brain? - <a href="http://www.transhumanist.com/volume1/moravec.htm">http://www.transhumanist.com/volume1/moravec.htm</a>

Why the future doesn't need us - <a href="http://www.wired.com/wired/archive/8.04/joy_pr.html">http://www.wired.com/wired/archive/8.04/joy_pr.html</a>
Wide Area Collaboration: a Proposed Application - <a href="http://www.unidata.ucar.edu/staff/caron/collab/wa_collab.html">http://www.unidata.ucar.edu/staff/caron/collab/wa_collab.html</a>
Wonder Girls and Steroids - <a href="http://serendip.brynmawr.edu/biology/b103/f00/web2/naimzadeh2.html">http://serendip.brynmawr.edu/biology/b103/f00/web2/naimzadeh2.html</a>
World Wide Web Platform-independent Access to Biomedical Text/Image Databases - <a href="http://archive.nlm.nih.gov/pubs/sd98.php">http://archive.nlm.nih.gov/pubs/sd98.php</a>
Wrapper Maintenance: A Machine Learning Approach - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>
Wrap-Up: a Trainable Discourse Module for Information Extraction - <a href="http://www.sadl.uleth.ca/nz/cgi-bin/library">http://www.sadl.uleth.ca/nz/cgi-bin/library</a>

## Referências Bibliográficas

- [All00] James Allan, *Natural language processing for information retrieval*, NAACL/ANLP Language Technology Joint Conference (Seattle, Washington), 2000.
- [Bit98] Guilherme Bittencourt, *Representação de conhecimento: da metafísica aos programas*, SBC, JAI'98 (1998).
- [BLG98] Kurt Bollacker, Steve Lawrence, and Lee Giles, *Citeseer: An autonomous web agent for automatic retrieval and identification of interesting publications*, International ACM Conference on Autonomous Agents, 1998.
- [BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila, *The semantic web*, Scientific American (2001).
- [BMM04] Antonio José S. Brandão, Luciana A. F. Martimiano, and Edson S. Moreira, *O uso de ontologia em alertas de vulnerabilidades*, 22º Simpósio Brasileiro de Redes de Computadores - IV Workshop em Segurança de Sistemas Computacionais (2004).
- [Car00] Olinda Nogueira Paes Cardoso, *Recuperação de informação*.

- [CDF<sup>+</sup>98] Mark Craven, Dam DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Seán Slattery, *Learning to extract symbolic knowledge from the world wide web*, 15th National Conference on Artificial Intelligence (USA), 1998.
- [CDHK01] Johan Carlberger, Hercules Dalianis, Martin Hassel, and Ola Knutsson, *Improving precision in information retrieval for swedish using stemming nada-kth*, In the Proceedings of NoDaLiDa-01 - 13th Nordic Conference on Computational Linguistics (2001).
- [Cha03] Marcírio S. Chaves, *Uma introdução à lingüística computacional*, Faculdade de Informática, Pontifícia Universidade Católica - PUC, Rio Grande do Sul (2003).
- [Con01] The Gene Ontology Consortium, *Creating the gene ontology resource: Design and implementation*, *Genome Research* **11** (2001), no. 8, 1425–1433.
- [Eri04] Henrik Eriksson, *Jesstab manual integration of protégé and jess*, Linköping University (2004).
- [Fap05] Fapesp, *Fapesp*, 2005.
- [Fen01] Dieter Fensel, *Ontologies: A silver bullet for knowledge management and electronic commerce*, Springer-Verlag (2001).
- [Fer03] Edberto Ferneda, *Recuperação de informação: Análise sobre a contribuição da ciência da computação para a ciência da informação*, Ph.D. thesis, Escola de Comunicação e Artes da Universidade de São Paulo, 2003.

- [FH01] Ernest Friedmann-Hill, *Jess, the expert system shell for the java platform*.
- [FH03] ———, *Jess in action rule-based systems in java*, Manning Publications Co., Greenwich, 2003.
- [Fre02] Frederico L. G. Freitas, *Sistemas multiagentes cognitivos para a recuperação, classificação e extração integradas de informação da web*, Ph.D. thesis, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2002.
- [GF92] Michael R. Genesereth and Richard E. Fikes, *Knowledge interchange format, reference manual, technical report logic-92-1*, California, 3 ed., 1992.
- [GMV99] Nicola Guarino, Claudio Masolo, and Guido Vetere, *Ontoseek: Content-based access to the web*, IEEE Intelligent Systems (1999).
- [GP99] Asunción Gómez-Pérez, *Ontological engineering: A state of the art*, Expert Update **2** (1999), no. 3.
- [GPFLC04] Asunción Gómez-Pérez, Mariano Fernández-López, and Oscar Corcho, *Ontological engineering*, Spring, Madrid, Espanha, 2004.
- [Gru93a] Thomas R. Gruber, *Toward principles for the design of ontologies used for knowledge sharing*, International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation (1993).
- [Gru93b] ———, *A translation approach to portable ontology specifications*, Knowledge Acquisition **5** (1993), 199–220.

- [Gua98] Nicola Guarino, *Formal ontology and information systems*, 1st International Conference on Formal Ontologies in Information Systems, FOIS'98 (1998).
- [Has03] Pim Haselager, *As múltiplas formas da inteligência*, Cienc. Cult. **55** (2003), no. 3, 14–15.
- [Hid01] José María G. Hidalgo, *Text representation for automatic text categorization*, Departamento de Inteligencia Artificial (Department of Artificial Intelligence) Universidad Europea de Madrid (2001).
- [HLM<sup>+</sup>04] Daniel F. Honorato, Huei D. Lee, Renato B. Machado, Feng C. Wu, and Antonio P. Neto, *Utilização da indexação automática para auxílio à construção de uma base de dados para a extração de conhecimento aplicada à doenças pépticas*, Laboratório de Bioinformática - LABI (2004).
- [HLOS00] Asunción Honrado, Ruben Leon, Ruairi O'Donnell, and Duncan Sinclair, *A word stemming algorithm for the spanish language*, Seventh International Symposium on String Processing Information Retrieval (SPIRE'00) (A Coruña, Spain), 2000.
- [JD88] Anil K. Jain and Richard C. Dubes, *Algorithms for clustering data*, Prentice Hall, New Jersey, 1988.
- [JM00] Daniel S. Jurafsky and James H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, NJ : Prentice Hall, Upper Saddle River, 2000.

- [JM02] Peter Jackson and Isabelle Moulinier, *Natural language processing for on-line applications; text retrieval, extraction and categorization*, John Benjamins, Amsterdam, Philadelphia, 2002.
- [JW97] Karen Spark Jones and Peter Willet, *Readings in information retrieval*, Morgan Kaufmann, San Francisco, 1997.
- [KGC<sup>+</sup>00] Andries Kruger, Lee Giles, Frans Coetzee, Eric Glover, Gary W. Flake, Steve Lawrence, and Christian W. Omlin, *Deadliner: Building a new niche search engine*, Ninth International Conference on Information and Knowledge Management, 2000.
- [KR03] James F. Kurose and Keith W. Ross, *Redes de computadores e a internet: Uma nova abordagem*, 1 ed., Addison Wesley, São Paulo, SP, 2003.
- [Kra97] Robert van Kralingen, *A conceptual frame-based ontology for the law*, Proceedings of the First International Workshop on Legal Ontologies (1997).
- [Lam00] Guirau de Lame, *Knowledge acquisition from texts towards an ontology of french law*, EKAW'2000 Workshop on ontologies and texts (2000).
- [LG90] Douglas B. Lenat and Ramanathan V. Guha, *Building large knowledge-based system: Representation and inference in the cyc project*, Addison Wesley, Boston, Massachusetts, 1990.
- [LL03] Leonardo C. Langie and Vera Lúcia S. Lima, *Classificação hierárquica de documentos textuais digitais usando o algoritmo knn*, TIL'2003 - 1º Workshop em Tecnologia da Informação e Linguagem Humana - XVI Bra-

- zilian Symposium on Computer Graphics and Image Processing - (SIB-GRAPI) (São Paulo, Brasil), 2003.
- [MDSC02a] Maria B. Marietto, Nuno David, Jaime S. Sichman, and Helder Coelho, *Proposta de uma nova classificação para a Área de inteligência artificial distribuída: Focos de interesse e sistemas computacionais*, 8th. Ibero-American Conference on Artificial Intelligence (IBERAMIA'02), 2002.
- [MDSC02b] Maria Bruno Marietto, Nuno David, Jaime Simão Sichman<sup>1</sup>, and Helder Coelho, *Requirements analysis of agent-based simulation platforms*, 2002.
- [MG03] Crediné S. Menezes and Tânia B. S. Gava, *Especificação de software baseada em ontologias*, In. III Escola Regional de Informática, 2003, pp. 167–205.
- [Mil95] G. Miller, *Wordnet: A lexical database for english*.
- [Mit97] Tom Mitchell, *Machine learning*, McGraw Hill, 1997.
- [MK00] Michael V. Mannino and Murlidhar V. Koushik, *The cost minimizing inverse classification problem: A genetic algorithm approach*, Decision Support Systems **29** (2000), no. 3.
- [MR05] Leora Morgenstern and Doug Riecken, *Snap: An action-based ontology for e-commerce reasoning*, IBM T.J. Watson Research Center (2005).
- [MVI95] Riichiro Mizoguchi, Johan Vanwelkenhuysen, and Mitsuru Ikelda, *Task ontology for reuse of problem solving knowledge*, In: Mars N (ed) Towards



- Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS'95), IOS Press, Amsterdam, The Netherlands, 1995.
- [Nav04] Sergio C. Navega, *Manipulação semântica de textos os projetos wordnet e lsa*, Infoimagem 2004 (2004).
- [NM01] Natalya F. Noy and Deborah L. McGuinness, *Ontology development 101: A guide to creating your first ontology*, Knowledge Systems Laboratory Technical Report KSL-01-05 (2001).
- [Par99] André Parente, *O virtual e o hipertextual*, Pazulin: Núcleo de Tecnologia da Imagem/ECO-UFRJ, Rio de Janeiro, RJ, 1999.
- [PN02] Thiago A. S. Pardo and Maria das Graças V. Nunes, *Aprendizado bayesiano aplicado ao processamento de línguas naturais*, Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional, NILC - ICMC-USP (São Carlos, SP), 2002.
- [PRO00] PROTÉGÉ, *The protégé project*, 2000.
- [Ril94] Ellen M. Riloff, *Information extraction as a basis for portable text classification systems*, Ph.D. thesis, University of Massachusetts Amherst, Department of Computer Science, University Massachusetts, 1994.
- [Rio05] Jocelma Almeida Rios, *Ontologias: alternativa para a representação do conhecimento explícito organizacional*, CINFORM - Encontro Nacional de Ciência da Informação VI (Salvador, Bahia), 2005.

- [RL94] Ellen M. Riloff and Wendy Lehnert, *Information extraction as a basis for high-precision text classification*, In ACM Transactions on Information Systems **12** (1994), no. 3, 269–333.
- [RN03] Stuart J. Russell and Peter Norvig, *Artificial intelligence*, 2 ed., Prentice-Hall, 2003.
- [SBF98] Rudi Studer, Richard Benjamins, and Dieter Fensel, *Knowledge engineering: Principles and methods*, Data Knowledge Engineering **25** (1998), no. 1-2, 161–197.
- [Sob03] Ana Paula Barbosa Sobral, *Previsão de carga horária - uma nova abordagem por Árvore de decisão*, Ph.D. thesis, Pontifícia Universidade Católica do Rio de Janeiro, 2003.
- [SX00] Amanda Spink and Jack L. Xu, *Selected results from a large study of web searching: the excite study*, Information Research: an international electronic journal **6** (2000), no. 1.
- [Tav98] Kuldar Taveter, *Sari - a system for semantical information retrieval*, ER-CIM News (1998), no. 35.
- [UG96] Mike Uschold and Michael Grüninger, *Ontologies: Principles, methods and applications*, Knowledge Engineering Review **11** (1996), no. 2.
- [UJ99] Mike Uschold and Robert Jasper, *A framework for understanding and classifying ontology applications*, IJCAI-99, ONTOLOGY WORKSHOP (1999).

- [XC96] Jinxi Xu and Bruce W. Croft, *Query expansion using local and global document analysis*, Proceedings of the 19th ACM Conference on Research and Development in Information Retrieval - SIGIR, 1996.