

# Do MASTER-Web ao AGATHE: a evolução de uma arquitetura de manipulação de informações para a web usando ontologias

DOI: 10.3395/reciis.v2i1.137pt



*Fred Freitas*

Centro de Informática,  
Universidade Federal de  
Pernambuco, Recife, Brasil  
fred@cin.ufpe.br



*Luciano Cabral*

Centro de Informática,  
Universidade Federal de  
Pernambuco, Recife, Brasil  
lsc4@cin.ufpe.br

*Rinaldo Lima*

Centro de Informática, Universidade Federal de Pernambuco, Recife, Brasil  
rjl4@cin.ufpe.br

*Eunice Palmeira*

Coordenação de Informática, Centro Federal de Educação Tecnológica de Alagoas, Maceió, Brasil  
eunice@cefet-al.br

*Guilherme Bittencourt*

Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, Brasil  
gb@das.ufsc.br

*Bernard Espinasse*

Laboratoire des Sciences de l'Information et des Systèmes Universités d'Aix-Marseille, Domaine Universitaire de St Jérôme, Marselha, França  
espinasse@univ-cezanne.fr

*Sébastien Fournier*

Laboratoire des Sciences de l'Information et des Systèmes Universités d'Aix-Marseille, Domaine Universitaire de St Jérôme, Marselha, França  
sebastien.fournier@isis.org

## Resumo

Este artigo apresenta duas arquiteturas para sistemas de manipulação de informações que se propõem a tratar domínios específicos da Web, como o domínio acadêmico ou o domínio biológico. Este tratamento de texto baseia-se no uso de ontologias sobre o domínio tratado, empregando-a como modelo semântico bem definido e compreensível para o software. Se a solução apresentada aqui não é escalável para toda a Web, por outro lado, os serviços oferecidos têm maior versatilidade e precisão, podendo combinar informações que estão distribuídas na web e que possuam relacionamentos bem definidos. Os sistemas apresentados conseguem ainda realizar inferências sobre as informações presentes na Web sobre estes domínios. Como prova de conceito, são apresentados experimentos com bons resultados em dois domínios distintos, mostrando que a solução é factível e portátil entre domínios, além de apresentar um alto grau de reuso durante a portabilidade.

## Palavras-chave

Multiagentes, agentes de informação, engenharia de software orientada a agentes, manipulação cooperativa de informação, classificação de informação

## Introdução

Atualmente, a quase totalidade dos artigos científicos produzidos encontra-se disponível também em (vários) formatos eletrônicos. A revolução tecnológica inegavelmente elevou o volume de informação e a tornou mais acessível, sobretudo aquela que se encontra na internet (PALMEIRA et al., 2006).

O resultado disto terminou por cunhar o termo “sobrecarga de informação” (do inglês “*information overload*”): uma enorme quantidade de documentos disponíveis coloca ao usuário a difícil tarefa de buscar informação útil. Para minimizar este problema, os mecanismos de busca - entre eles o Google, que praticamente dominou este mercado - foram então projetados, com base em técnicas desenvolvidas pela área de Recuperação de Informação (RI) (RIBEIRO-NETO et al., 1999).

Porém, os mecanismos de busca, como o Google (GOOGLE, 2008), possuem várias deficiências, essencialmente ligadas à forma como foram concebidos. Em primeiro lugar, quando um usuário solicita ao Google uma busca por palavra-chave, é bastante comum que no conjunto de páginas devolvido como resposta exista uma grande quantidade de endereços de páginas inúteis ou irrelevantes. Utilizando algoritmos estatísticos para atribuir relevância às páginas, os mecanismos de busca não conseguem dotar de semântica a busca, porque possuem capacidade de representar as páginas baseando-se apenas no nível léxico (FREITAS, 2002). Comparando-se os serviços de mecanismos de buscas com os de Sistemas Gerenciadores de Bancos de Dados, as limitações dos primeiros tornam-se evidentes: Bancos de Dados, diferentemente de páginas em texto (sejam na Web ou em bibliotecas digitais), podem ser facilmente consultados por armazenarem dados sobre um contexto restrito, de forma estruturada e sem ambigüidade. Sistemas Gerenciadores de Bancos de Dados podem, portanto, prover ao usuário consultas semanticamente claras e precisas sobre entidades e relacionamentos entre elas, inclusive combinando e totalizando dados, tarefas que os mecanismos de busca não conseguem realizar. Por exemplo,

um usuário poderia consultar o sistema sobre artigos da área de “redes neurais” publicados em eventos asiáticos depois do ano 2000.

Por outro lado, uma vez que estejamos tratando de páginas sobre um domínio restrito (como o domínio acadêmico, por exemplo) e usando conhecimento sobre este domínio, pode-se oferecer um tratamento mais profundo à informação em contrapartida à simples recuperação, com tarefas que fornecem benefícios adicionais ao usuário. Estas tarefas serão descritas na próxima subseção.

## Outras tarefas ligadas à recuperação de informação: classificação e extração de informação

As tarefas que podem ser desempenhadas em cooperação com a tarefa de recuperação de informação são:

- a classificação de páginas, processo que irá determinar a que classe de páginas uma determinada página pertence. Por exemplo, uma página que descreve um congresso, cairia na classe Evento Científico, mais especificamente em sua subclasse Congresso;
- a extração de informações. Segundo KUSHMERICK (1999), extração de informações “é a tarefa de identificar os fragmentos específicos de um documento que constituem o núcleo de seu conteúdo semântico”. Por exemplo, da frase “*The Parliament was bombed by the guerrillas.*” processada como sendo do domínio de terrorismo, foram extraídas três informações: o tipo de ato terrorista: bombardeio (*bombing*), o alvo do ato terrorista: o Parlamento (*The Parliament*), os autores do ato terrorista: as guerrilhas (*the guerrillas*).

## Complementaridade entre as tarefas de recuperação, classificação e extração de informação

Este trabalho teve como premissa inicial a hipótese de que as tarefas de recuperação, categorização e extração podem e devem ser integradas, o que pode ocasionar uma melhora de desempenho em todas elas, como ilustra a Figura 1.

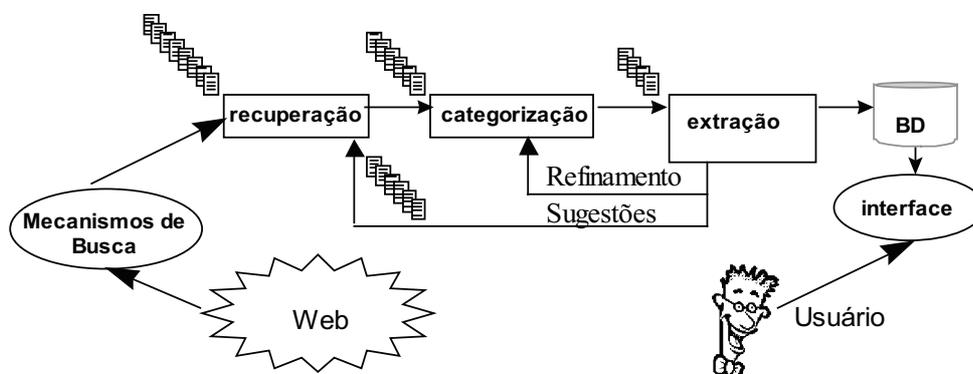


Figura 1 - Esboço da arquitetura de um sistema de extração, evidenciando a complementaridade entre as tarefas de recuperação, categorização e extração.

Fonte: FREITAS, F. Sistemas Multiagentes Cognitivos para a Recuperação, Classificação e Extração Integradas de Informação da Web. Tese de doutorado, Universidade Federal de Santa Catarina. 2002.

Sistemas de recuperação podem fornecer o acesso a um conjunto inicial de páginas sobre um determinado domínio, por exemplo, eventos científicos. Após isto, sistemas de categorização deveriam selecionar quais páginas pertenceriam às classes a serem processadas, e então extratores poderiam capturar a informação requerida, por exemplo, data e local de realização do evento. Durante o processo de extração, poderiam ainda ser encontrados, dentre as âncoras das páginas processadas, endereços de outras páginas que pertenceriam a outras classes também processadas, por exemplo páginas de pesquisadores. Estas sugestões são encontradas dentro de um contexto semântico e seguro, e, assim, o processo de extração estaria auxiliando a recuperação, se os agentes cooperassem (p. ex., o agente de eventos científicos envia endereços de páginas dos membros do comitê de programa ao agente de pesquisadores). A extração poderia ainda refinar a categorização, uma vez que páginas que não contivessem os dados que caracterizam a classe processada seriam desprezadas, por exemplo, eventos sem data.

Note-se que a execução integrada destas três tarefas só é possível graças a um requisito imprescindível: a restrição de domínios, ou seja, o sistema que as executa só é capaz de processar páginas de um determinado domínio. No entanto, pensando em aproveitar o desenvolvimento de uma solução desta natureza, seria interessante que esta solução pudesse ser portátil entre domínios distintos.

## Portabilidade entre domínios através de ontologias

A problemática envolvendo a portabilidade de domínios muitas vezes provoca alterações drásticas em sistemas computacionais. Por outro lado, o uso de soluções computacionais declarativas, - i.e., aquelas em que

o conhecimento necessário localiza-se fora do sistema e não em seu código executável - facilita a implementação da portabilidade.

Uma solução declarativa muito em voga atualmente é o termo ontologia, uma vez que elas estão presentes em muitos sistemas, ferramentas e produtos de manipulação de informação e comércio eletrônico, representadas como hierarquias de palavras-chave, conceitos, e muitas outras formas.

Apesar da palavra “ontologia” denotar uma teoria sobre a natureza do ser ou existência, em informática ela pode ser interpretada como o conjunto de entidades com suas relações, restrições, axiomas e vocabulário. Uma ontologia define um domínio, ou, mais formalmente, especifica uma conceitualização acerca dele (GRUBER, 1995). Normalmente, uma ontologia é organizada em hierarquias de conceitos (ou taxonomias).

Por exemplo, a Figura 2 mostra algumas classes da ontologia Ciência (FREITAS, 2001) e algumas de suas relações. A classe Membro do Corpo Acadêmico (*Academic-Staff*) é um subconjunto mais específico da classe Pesquisador (*Researcher*), o que caracteriza a relação conhecida como herança. Esta classe herda, portanto, todo o conhecimento associado a classe Pesquisador, como os relacionamentos, restrições e outros itens de conhecimento desta última classe. Na figura encontram-se ainda outras relações entre classes, como por exemplo a relação Coordenador (*HeadOf*) que liga as classes Membro do Corpo Acadêmico (*Academic-Staff*) e Projeto (*Project*), sinalizando que um projeto é coordenado por um membro do corpo acadêmico. Um exemplo de restrição seria a restrição de cardinalidade imposta nesta relação: apenas um membro do corpo acadêmico participa desta relação.

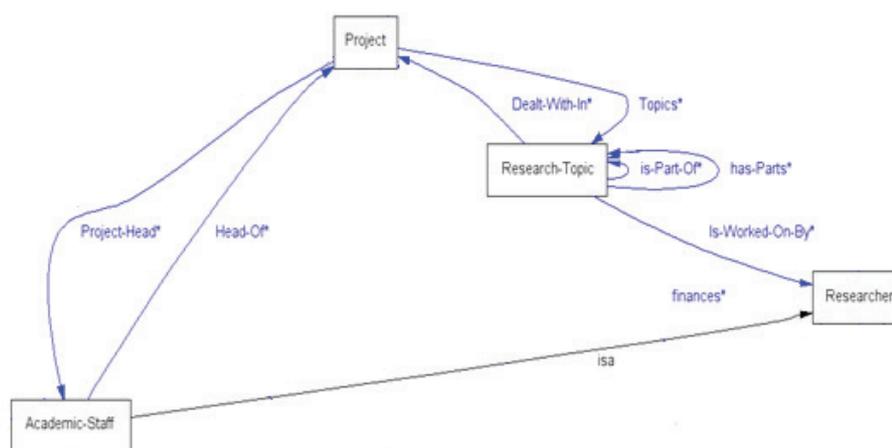


Figura 2 - Relacionamentos entre algumas das principais classes da ontologia Ciência.

Fonte: FREITAS, F. 2001. Ontology of Science. [http://protege.stanford.edu/plugins/ontologyOfScience/ontology\\_of\\_science.htm](http://protege.stanford.edu/plugins/ontologyOfScience/ontology_of_science.htm).

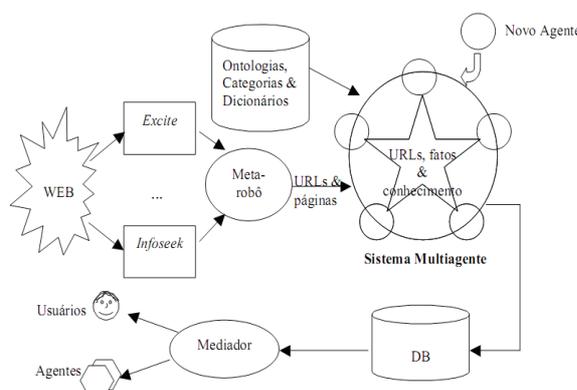
Postos estes requisitos, a adoção de manipulação integrada de informação na Web, que envolve recuperação, classificação e extração de informações da Web e o uso de ontologias para restringir o domínio a ser tratado, descreveremos, nas próximas seções, a pesquisa realizada pelos autores nesta área de pesquisa. Serão detalhados dois sistemas com arquiteturas diferentes, o MASTER-Web (FREITAS et al., 2003) e o AGATHE (ESPINASSE et al., 2007), bem como suas semelhanças e diferenças e alguns estudos de caso que confirmam a portabilidade destas soluções. No fim do artigo, apresentaremos trabalhos de pesquisa relacionados a estes tópicos e uma discussão comparativa denotando as qualidades das arquiteturas propostas.

## Descrição do Sistema MASTER-Web

O sistema MASTER-Web (*Multi-Agent System for Text Extraction, Classification and Retrieval over the Web*) (FREITAS et al., 2003) consiste de uma arquitetura de Sistemas Multiagentes Cognitivos para solucionar o problema da extração integrada de entidades pertencentes às classes que integram um grupo de páginas

(ou, em inglês, *cluster*). Este sistema apresentou bons resultados nas tarefas de recuperação, classificação e extração de informação, e permite a cooperação entre agentes para a realização destas tarefas. Utiliza-se uma abordagem baseada em conhecimento e apresentam-se vários tipos de reuso, já que os agentes compartilham a mesma estrutura em termos de código, serviços dos mecanismos de busca, e boa parte do conhecimento de que dispõem (ontologias e regras de produção), o que facilita e acelera a construção de novos agentes (PALMEIRA et al., 2006).

Este sistema multiagente manipula a informação referente a um conjunto de classes sobre um mesmo grupo, como, por exemplo, o grupo científico, com classes como artigos científicos, eventos, pesquisadores etc. A arquitetura em si, visa recuperar, classificar e extrair dados de páginas pertencentes às classes de um grupo, e 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.



**Figura 3 - Arquitetura de sistemas multiagentes para manipulação integrada de informação de grupos de classes de páginas.**

Fonte: FREITAS, F.; BITTENCOURT, G. An Ontology-based Architecture for Cooperative Information Agents. Proceedings of the Internacional Joint Conference on Artificial Intelligence - IJCAI'2003. Acapulco, México, 2003.

Os agentes, representados como círculos nesta figura, são habilitados a reconhecer, filtrar e classificar páginas que pertençam (ou não) à classe de páginas processada por eles (por exemplo, o agente PPR processa páginas de artigos científicos enquanto outros agentes do mesmo sistema multiagente são responsáveis por processar outras classes do domínio científico). Os agentes reconhecem uma página de sua classe principalmente quando reconhecem a existência de atributos referentes a esta classe (vide exemplo a seguir). Os agentes cooperam, trocando mensagens contendo regras de reconhecimento e fatos (conhecimento dos agentes), além das páginas sugeridas entre os agentes.

Os usuários podem beneficiar-se do acesso aos dados extraídos através de um agente especial, chamado

mediador (WIEDERHOLD, 1995). Ele possui a função de ajudar às consultas aos dados, provendo visões mais simples da base de dados, permitindo ao usuário formular consultas complexas envolvendo várias tabelas da base de dados. Por exemplo, um usuário poderia realizar a consulta já descrita na seção 1, procurando artigos da área de "redes neurais" publicados em eventos asiáticos depois do ano 2000. Digno de nota o fato de que os mecanismos de busca atuais, como o Google (GOOGLE, 2008), não possuem meios de realizar uma pesquisa como essa.

Ao entrar no sistema, os agentes registram-se e anunciam-se aos outros agentes, mandando fatos e regras de reconhecimento de páginas e *links* úteis a si próprio, que serão empregadas pelos outros para lhe indicarem sugestões de páginas. O novo agente receberá, em res-

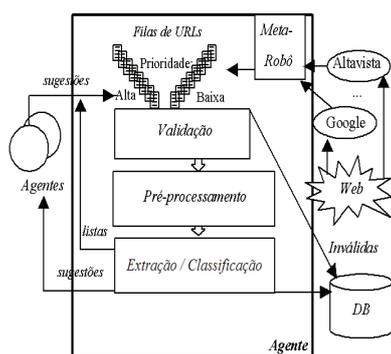
posta, regras úteis aos outros agentes. Assim, quando um agente acha informação que dispara alguma das regras referentes aos outros, este agente repassa a informação (*link* ou página) ao agente que lhe enviou a regra disparada, implementando desta forma a cooperação entre eles.

Na seção a seguir, descreveremos o funcionamento de um agente em detalhes.

## Funcionamento e estrutura de um agente no MASTER-Web

Cada agente, representado em detalhe na Figura 4, reconhece, filtra e classifica páginas que correspondem a instâncias da classe que ele processa (por exemplo, páginas de pesquisadores, chamadas de eventos científicos - *'Call for Papers'*), extraindo também seus atributos (por exemplo, áreas de pesquisa e instituição dos pesquisadores). Cada agente possui ainda um meta-robô, que pode conectar-se a múltiplos mecanismos de busca - como Google, Excite e outros. Ele consulta os mecanismos de busca com palavras-chave que garantem cobertura em relação à classe de páginas processada pelo agente. (e.g., os termos *'call for papers'* e *'call for participation'* para o agente CFP) (PALMEIRA et al., 2006).

**Figura 4 - Um agente e seus passos de processamento.**



Fonte: FREITAS, F.; BITTENCOURT, G. An Ontology-based Architecture for Cooperative Information Agents. Proceedings of the Internacional Joint Conference on Artificial Intelligence - IJCAI'2003. Acapulco, México, 2003.

Como pode ser observado na Figura 4, cada agente desempenha quatro tarefas consecutivas no processamento de cada página (PALMEIRA et al., 2006): validação, pré-processamento, reconhecimento e extração.

A **Validação** elimina as páginas inacessíveis, repetidas ou em formatos que os agentes não possam processar. A fase de **Pré-processamento** tem por meta representar as páginas de várias maneiras, tais como conteúdo com e sem HTML, palavras-chave e suas freqüências, dentre outras, com dados extraídos delas, aplicando, se necessário, recuperação de informação e processamento de linguagem natural. Esses dados são, então, passados ao motor de inferência. Nas fases de **Classificação** e **Extração de Atributos**, o sistema descobre se a página é do domínio tratado, reconhece de que classe a página tratada é instância e extrai atri-

butos que irão compor a instância da classe. Estas fases são executadas através de processos computacionais conhecidos como raciocínio automático ou inferência lógica. Na próxima subseção explicaremos como eles estão implementados no MASTER-Web.

## Raciocínio dos agentes

Cada uma destas classes possui um conjunto de atributos, que devem ser extraídos ou identificados, e cuja presença pode indicar se uma página é instância de uma classe ou não. Este processo é realizado através da uma combinação de casos, regras e ontologias, que será melhor explanada a partir de um exemplo.

Um caso particular muito comum para o agente de artigos científicos, é a situação em que este reconhece uma página se ela contém no seu início o resumo (abstract) e os atributos *First-Name* (primeiro nome de um dos autores), *Org-name* (nome da organização à qual um ou mais autores está afiliado) e *Location-Place* (país ou estado americano onde se localiza esta organização). Este caso está descrito abaixo:

```
([ppr_00356] of Case
  (Description "aff,1st,loc")
  (Concepts-in-the-Beginning [abstract])
  (Slots-in-the-Beginning
    [First-Name]
    [Org-name]
    [Location-Place]))
```

Este caso deve ser associado a uma classe a ser reconhecida através de um reconhecedor de classes (*Class-Recognizer*). Como quaisquer artigos, sejam eles de revistas ou eventos, normalmente seguem este padrão, associamos este caso à classe *Publicação-Parte* (*Part-Publication*), pois artigos são sempre partes de uma publicação divisível (e.g. um livro, ou anais de eventos). Segue abaixo esta associação:

```
([ppr_00528] of Class-Recognizer
  (Cases
    [ppr_00536]
    [ppr_00356])
  (Class [Part-Publication]))
```

O relacionamento é completado por regras reusáveis por instanciação, muitas delas comuns a vários agentes, como a descrita no próximo código.

```
Regra r_900_slots_hi_func
  Se tenho uma instância de página em estado AR-
  MAZENADA e
  Se tenho um caso que tem
    Uma lista de atributos que devem estar no começo
    da página
    E uma lista de conceitos que devem estar no co-
    meço da página
  E um reconhecedor de uma classe abstrata com uma
  lista de casos
```

Teste

Se o caso especificado está na lista de casos

E se algum dos conceitos do começo da página está lá

E se os atributos do caso estão na lista de atributos encontrados

Então

A página passa para o estado de RECONHECIDA como sendo da mesma classe do caso especificado

Assim, como a maior parte das regras de reconhecimento instância casos, elas podem ser incluídas em todos os agentes, sendo completamente reusáveis por agentes que tenham outros casos, o que facilita a construção de um novo agente. De fato, um novo agente só precisa definir novos casos de reconhecimento, classificação e extração.

O reconhecimento em geral recai primeiramente sobre uma classe abstrata, que não pode ter instâncias - como *Part-Publication* no agente de artigos, ou *Evento Científico ao Vivo* ou *Evento de Publicação* no agente CFP. A partir do reconhecimento da classe abstrata, regras de classificação incluindo casos tentam classificar a página entre as subclasses dela. Caso isso não ocorra, existe uma classe genérica à qual a classe será classificada, como *Parte de Publicação Genérica*, no agente de artigos e *Evento Científico Genérico ao Vivo* ou *Evento de Publicação Genérica*.

## Estudo de caso

Em sua primeira versão, o MASTER-Web foi desenvolvido usando, para as tarefas de raciocínio, o motor de inferência JESS e o ambiente PROTÉGÉ para especificação e manipulação das ontologias (PROTÉGÉ, 2008). O sistema foi inicialmente testado sobre o domínio científico (FREITAS et al., 2003), usando uma ontologia do meio científico (descrita em (FREITAS, 2001)). Para a produção desta ontologia, a ontologia do projeto (KA)<sup>2</sup> (*Knowledge Annotation Initiative of the Knowledge Acquisition Community*) (BENJAMINS et al. 1998) foi reusada e melhorada, principalmente, em granularidade. A melhoria principal foi a adição de classes visando reorganizar a ontologia sob a ótica das classes com características comuns. Como um exemplo prático, pode-se citar a classe *Evento-Científico*, que, foi dividida em duas subclasses, *Evento-Científico-ao-Vivo* (com subclasses *Conferência* e *Workshop entre outras*) e *Evento-de-Publicação-Científica* (com subclasses *Jornal* e *Revista*). Assim, aumenta-se a capacidade de reconhecimento e adiciona-se granularidade e coerência à ontologia (FREITAS et al., 2003). Desse modo, entidades do cluster (conhecimento do domínio) podem ser definidas com uma granularidade adequada, representando a classificação mesmo com diferenças sutis entre as entidades (ESPINASSE et al., 2007).

Com cada agente, foram realizados três testes para classificação de páginas (FREITAS et al., 2003). Para dois dos três primeiros testes, lançou-se mão de *corpus* de páginas recuperadas de consultas a mecanismos de busca (como o Google e AltaVista). O primeiro *corpus*

foi empregado no processo conhecido como aquisição de conhecimento, que consiste na definição de casos e regras que auxiliarão os processos de classificação e extração rodando interativamente o sistema até que ele atinja o desempenho aceitável. O segundo *corpus* foi usado para um teste cego. O terceiro teste foi efetuado coletando páginas candidatas diretamente da Web. Estes dois testes, portanto, avaliaram o desempenho dos agentes frente a novos *corpora*, não consultados durante a elaboração das regras e casos.

Dois agentes do grupo científico foram construídos. O agente CFP processa páginas candidatas de chamadas de trabalhos ("*Call for papers*") para eventos científicos, como conferências e jornais, classificando-as em oito classes de páginas (as quatro citadas acima, mais *Evento-Genérico-ao-Vivo*, *Evento-Genérico-de-Publicação* e *Edição-Especial-de-Jornal* e de *Revista*). O segundo agente processa páginas candidatas a artigos e documentos científicos, classificando-as como artigos de *workshop*, conferência, jornal e revista, capítulo de livro e artigos genéricos, além de teses, dissertações, relatórios técnicos e de projeto.

## Resultados

Os resultados a serem apresentados referem-se a dois conjuntos de testes. Um primeiro que compreende um total de quatro testes para avaliação de desempenho do sistema na classificação. Na Tabela 1 são exibidas as taxas de desempenho dos agentes CFP e de artigos. Nela, os resultados são categorizados pelo tipo de *corpus* nas colunas e o tipo de classificação nas linhas.

Os resultados parecem bastante promissores: a maior parte deles está acima de 90%. Os poucos casos de falha, no caso do agente CFP eram páginas longas, geralmente sobre um assunto ou de uma comunidade (ex: *Open Source*).

**Tabela 1 - Desempenho dos agentes CFP e de artigos**

	Agente "Call for Papers"				Agente Artigos		
	Corpus Aquisição	Corpus de Teste	Teste na Web	Web usando Listas	Corpus Aquisição	Corpus de Teste	Teste na Web
Reconhecimento	97,1	93,9	96,1	96,3	93,1	82,7	87,8
Classif. Conteúdo	94,9	93,3	92,9	91,7	97,0	93,0	81,4
Págs Processadas	244	147	129	188	190	150	184

Fonte: FREITAS, F.; BITTENCOURT, G. An Ontology-based Architecture for Cooperative Information Agents. Proceedings of the International Joint Conference on Artificial Intelligence - IJCAI'2003. Acapulco, México, 2003.

De forma análoga, o agente PPR, que processou artigos científicos, não conseguiu reconhecer artigos com poucos atributos, com atributos difíceis de identificar (ex.: uma afiliação a uma empresa desconhecida), ou com atributos no final do artigo. Porém, este fato

ocorre também em sistemas similares como o *CiteSeer* (BOLLACKER et al., 1998).

Na seção seguinte, veremos evoluções da arquitetura proposta, no sentido de provê-la com dois requisitos relevantes: a portabilidade para outros domínios e a escalabilidade, i.e., a possibilidade de executá-la com muitas páginas mais.

## Evolução

Um dos maiores benefícios de uma arquitetura declarativa é a sua genericidade; apenas trocando as ontologias e bases de conhecimento envolvidas, pode-se implementar uma profunda mudança de comportamento do sistema. A descrição do experimento a seguir comprova que este requisito é aplicável à arquitetura.

### Portabilidade para outros domínios: MASTER-Web no domínio de Inteligência Artificial

Este teste comprova que, a portabilidade de domínios em sistemas baseados em conhecimento para tarefas relacionadas à recuperação de informação é possível sem alterações bruscas no sistema.

As alterações para a execução deste teste foram as seguintes:

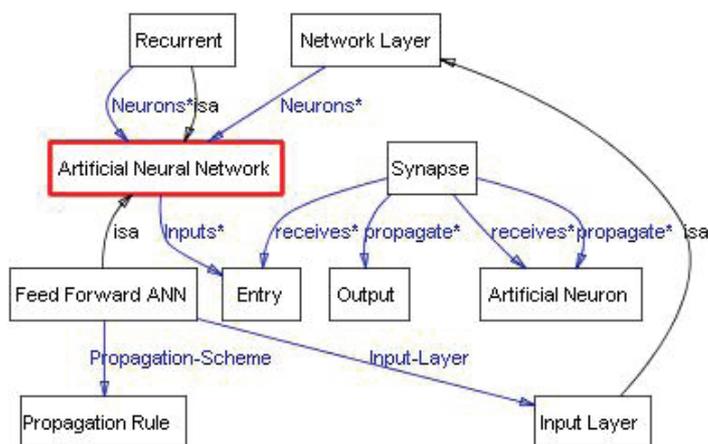
- retirou-se a ontologia de eventos científicos e adicionou-se uma ontologia da área de Inteligência Artificial (IA) e algumas de suas sub-áreas;
- estendeu-se no sistema a representação de páginas HTML para uma representação de artigos científicos, a fim de, facilitar a identificação das seções típicas ocorridas em artigos científicos, tais como introdução, trabalhos relacionados, conclusões, resumo, e outras. Na fase de pré-processamento, foi incluído o reconhecimento destas seções, fato que ajudou sobremaneira

durante o reconhecimento e classificação de artigos da área de IA;

- as seções foram divididas em duas classes: seções relevantes e irrelevantes. As seções relevantes tratam especificamente sobre os temas principais do artigo, sendo em geral mais concisas e apresentando aspectos teóricos (e.g. resumo, palavras-chave e conclusão). Já as seções irrelevantes para o reconhecimento e classificação apresentam conteúdos mais gerais e de várias áreas, possuindo, portanto menor especificidade. Citações, trabalhos relacionados e trabalhos futuros estão entre as seções consideradas irrelevantes para a classificação (PALMEIRA et al., 2006).

Os experimentos de reconhecimento e classificação de artigos de IA utilizaram um *corpus* bastante heterogêneo colhido da Web, com artigos científicos de diversas áreas de conhecimento, sendo ainda, heterogêneos quanto à divisão das seções. Para este teste, partimos da premissa de que o sistema deveria apresentar robustez frente a esta falta de padronização de seções dos artigos (PALMEIRA et al., 2006).

A base experimental foi composta por documentos em HTML, obtidos em diversos sítios de universidades, páginas de professores e de revistas científicas. Os experimentos realizados classificaram os artigos com base na análise das seções do texto consideradas relevantes e numa ontologia de IA construída especificamente para este fim. Foram desenvolvidas separadamente ontologias de sub-áreas e aplicações que posteriormente são agrupadas para compor a ontologia de IA. O escopo dessa ontologia compreende, a princípio, as áreas de Busca, Representação do Conhecimento, Redes Neurais e Aprendizado de Máquina. Ontologias de outras sub-áreas de IA poderão ser incorporadas estendendo a ontologia existente (PALMEIRA et al., 2006).



**Figura 5 - Algumas classes e seus relacionamentos no domínio de Redes Neurais da ontologia de Inteligência Artificial.**

Fonte: PALMEIRA, E. and FREITAS, F. Detailed Ontologies and text classification: a promising union, Proceedings of the Workshop on Building Applications with Ontologies for the Semantic Web, Encontro Português de Inteligência Artificial, 2006.

Na Figura 5, destacam-se alguns relacionamentos entre as principais classes definidas na ontologia de Redes Neurais. Dois tipos de Redes Neurais, *Recurrent* e *Feed Forward ANN*, bem como alguns conceitos relevantes, como *Synapse*, *Neuron* e *Layer*, são apresentadas nessa figura. Várias relações estão disponíveis entre estes conceitos.

As ontologias dos ramos de Busca, Representação do Conhecimento e Aprendizado de Máquina foram desenvolvidas com um nível de detalhes semelhante

a de Redes Neurais. Os artigos da área de IA, que não tratam de áreas compreendidas na ontologia, não serão classificados dentro deste domínio (PALMEIRA et al., 2006). Outra observação relevante é que um artigo pode abranger mais de uma área, como por exemplo, falar tanto de Aprendizado de Máquina (*Machine Learning*) quanto de Busca (*Search*) (PALMEIRA et al., 2006). Os resultados obtidos apresentaram promissores percentuais na classificação de artigos de IA, conforme mostra a Tabela 2.

**Tabela 2 - Percentuais de acerto de classificação de artigos por área de conhecimento**

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

Fonte: PALMEIRA, E. and FREITAS, F. Detailed Ontologies and text classification: a promising union, Proceedings of the Workshop on Building Applications with Ontologies for the Semantic Web, Encontro Português de Inteligência Artificial, 2006.

Provada a portabilidade entre domínios, o próximo requisito para a arquitetura passou a ser a escalabilidade. Este requisito foi atacado com uma concepção completamente nova do sistema em termos de funcionalidades e topologia dos agentes, que exigiu a confecção de outra sistema, o AGATHE (Agent GATHERing system, Sistema de agentes de manipulação (ESPINASSE et al., 2007), que será detalhado na próxima seção.

## Descrição do Sistema AGATHE

O AGATHE (ESPINASSE et al. 2007) é o resultado de uma reengenharia da topologia de agentes do MASTER-Web (FREITAS et al., 2003). O propósito específico deste sistema consiste em definir uma arquitetura multiagente que seja escalável, adaptável e extensível, a fim de apoiar o processo de recuperação inteligente de informação na internet. O AGATHE adiciona modularização e distribuição de tarefas em sua arquitetura, impedindo a sobrecarga de funções, que poderia ocorrer no sistema base, podendo-se desta forma obter um melhor desempenho na recuperação, extração e classificação.

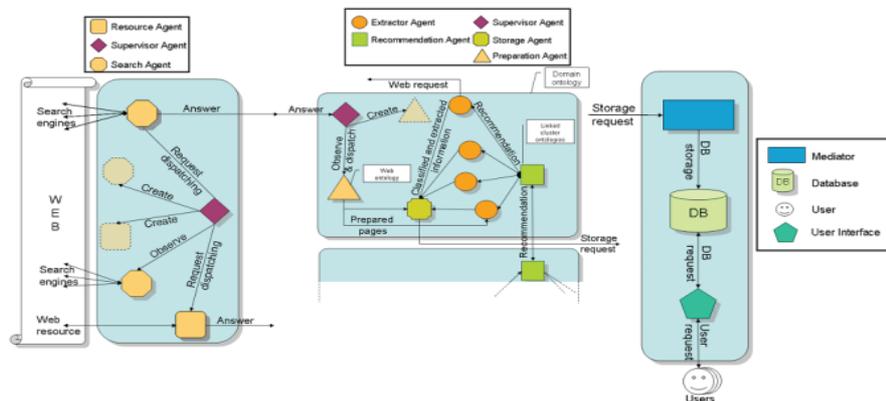
Esta nova arquitetura é baseada numa organização de agentes específica que define diferentes tipos de

agentes, que serão descritos na próxima subseção. Esta versão integra a plataforma JADE (Java Agent Development Environment – Ambiente de Desenvolvimento de Agentes em Java – TILAB, 2008), além do motor de inferência JESS, e do ambiente PROTÉGÉ.

Foram incluídas ainda algumas novas funcionalidades como, por exemplo, a cooperação entre sistemas multiagentes de domínios diferentes, como, por exemplo, os domínios acadêmico e turístico. Há diversos aspectos turísticos associadas a eventos científicos (viagens, acomodação, atividades e eventos turísticos etc), e que estão presentes em páginas de eventos científicos. É possível vislumbrar então a cooperação entre agentes destes domínios diferentes, em que o agente CFP enviaria a agentes do domínio turístico páginas para que estes classificassem e extraíssem informações que lhes fossem úteis. Este serviço não poderia ser oferecido pelo MASTER-Web em sua concepção original.

## Arquitetura

O AGATHE é composto por 3 subsistemas que interagem entre si, o subsistema de busca, subsistema de extração e o subsistema de acesso, descritos a seguir e apresentados na Figura 6.



**Figura 6 - Arquitetura do AGATHE com subsistema de extração em detalhes.**

Fonte: ESPINASSE, B., FREITAS, F. e FOURNIER, S. AGATHE: an Agent and Ontology based System for Restricted-Domain Information Gathering on the Web. Proceedings of the International Conference on Research Challenges in Information Sciences (IEEE-RCIS), April 23-26, Ouarzazate, Morocco, 2007.

a) O *subsistema de busca* é responsável por consultar os motores de busca externos (como o Google), selecionar e processar as páginas selecionadas repassando-as ao *subsistema de extração*.

b) O *subsistema de extração* é composto por diferentes agentes extratores de um mesmo domínio responsáveis por especializar as páginas processadas (i.e. domínio Acadêmico, Inteligência Artificial etc.).

c) O *subsistema de acesso* tem a função de organizar os dados extraídos das páginas processadas de forma amigável ao usuário ou para agentes inteligentes, usando técnicas de interface homem-máquina e de linguagem de comunicação entre agentes.

Todos os subsistemas citados são sistemas multiagentes, compostos de agentes de informação, que utilizam o domínio descrito pela ontologia para executar suas tarefas (ESPINASSE et al. 2007).

Desta forma podemos exemplificar usando o subsistema de extração, que contém diversos agentes de informação conforme descrito na Figura 7, diferentemente do MASTER-Web, onde todos os agentes são do mesmo tipo e efetuam todas as tarefas.

#### Agente Supervisor

Recebe os resultados das consultas do Subsistema de Busca e instancia um ou mais Agentes de Preparação que irão preparar estas páginas para processamento pelos Agentes Extratores e de Recomendação.

#### Agentes de Preparação

Estes agentes têm o seu ciclo de vida (criação e deleção) controlado pelo Agente Supervisor, recebendo páginas Web do Subsistema de Busca e realizando as tarefas de validação e pré-processamento, já explanadas anteriormente. Estas páginas são então entregues a Agentes Extratores e de Recomendação.

#### Agente Extrator

Estes agentes executam as tarefas de classificação e extração descritas, a partir de páginas recebidas dos Agentes de Preparação. Os resultados de seu processamento (dados extraídos e classificação das páginas) são enviados aos Agentes de Armazenamento.

#### Agente de Recomendação

Este agente recebe as páginas preparadas pelo Agente de Preparação e as envia para outros agentes no mesmo domínio (recomendação interna) ou de outros domínios (recomendação externa).

#### Agente de Armazenamento

Este agente tem a tarefa de armazenar as informações extraídas/classificadas de uma página no banco de dados do Subsistema de Acesso para ser consultado pelos usuários do sistema.

No teste da arquitetura AGATHE, foi utilizado o mesmo domínio do sistema MASTER-Web, restrito aos eventos científicos na investigação acadêmica.

No seu primeiro teste, de uma amostra de 310 páginas obtidas da web através de motores de busca, o AGATHE classificou corretamente cerca de 280, atingindo um desempenho bastante promissor, com taxa de acerto de 90,32%.

Na seção seguinte discutiremos sobre trabalhos relacionados aos sistemas aqui apresentados, comparando-os quando houver similaridade.

### Trabalhos relacionados

Muitas áreas estão envolvidas com estas arquiteturas: recuperação de informação, extração, classificação, sistemas multiagentes, processamento de linguagem natural e ontologias, entre outras. Porém, iremos focar em

algumas propostas de soluções similares à nossa (*WebKB* (CRAVEN et al., 1998), *CiteSeer* (BOLLACKER et al., 1999), *DEADLINER* (KRUGER et al. 2000) e *Ontoseek* (GUARINO et al., 1999) comparando estes sistemas com os sistemas aqui apresentados (MASTER-Web e AGATHE), bem como apresentar ferramentas de apoio na tarefa de anotação semântica (*KIM Plataforma*), que consiste na idéia de atribuir a entidades, links para suas descrições semânticas, isto é, fornecer informações de metadados para as instâncias de entidades referenciadas nos textos (POPOV et al., 2004).

### WebKB: classificação e extração baseadas em aprendizado e ontologia

O sistema WebKb (CRAVEN et al. 1998) aprende automaticamente regras de categorização e extração integrada de páginas na Web, empregando uma ontologia do domínio com classes e relacionamentos. As páginas da Web são representadas, com título, palavras-chave, freqüências e links.

O sistema emprega uma ontologia do domínio com apenas quatro entidades: atividades (subdivididas em projetos e cursos), pessoas (subdivididas em estudantes, professores, membro do corpo acadêmico etc.) e departamentos. Relações também estão presentes nesta ontologia, tais como: instrutores de cursos, membros de projeto e orientadores, entre outras.

O WebKB corresponde aproximadamente ao trabalho de três futuros agentes do sistema MASTER-Web ou do AGATHE - o de pesquisadores, o de projetos e o de organizações. Por outro lado, estes sistemas visam tratar o domínio científico a partir de pesquisas sobre a Web, enquanto o WebKB processa sítios de faculdades.

O MASTERWeb 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. Por isso, ele exige mais esforço na confecção de agentes para cada classe de páginas. O WebKB tem como vantagem adaptar-se rapidamente a novos domínios e usa heurísticas estatísticas de padrões de conexão entre páginas e de palavras-chave (não processando expressões) enquanto o MASTERWeb baseia-se em palavras-chave e expressões associadas a conceitos, contidas em links para sugerir-los a outros agentes, se estes processam classes de páginas que se relacionam semanticamente com os conceitos referidos.

Os autores do WebKB avaliam a classificação apenas através dos falsos positivos, reportando percentagens entre 73% e 83%, exceto para as classes membro do *Staff* e *outros* (rejeitadas). Contudo, se computados os falsos negativos, a classe *outros* tem boa performance (93,6%), a classe estudante tem 43% e as outras seis classes comportam-se abaixo de 27%, baixando a média de acerto para apenas cerca de 50%. Isto leva à hipótese de que a ontologia empregada no WebKB não tenha sido abrangente o suficiente. Já a ontologia de Ciência usada pelo MASTERWeb possui classes, como projetos e produtos, que não foram usadas por dois motivos: os agentes precisam destes conceitos para suas funções e

futuros agentes que tratem delas podem ser elaborados. Por outro lado, uma ontologia com muitas classes pode dificultar a generalização do aprendizado. Neste caso, seriam necessários mais agentes com aprendizado.

### Os Sistemas CiteSeer e DEADLINER

Estes sistemas perfazem uma eficiente recuperação, filtragem e extração da Web, usando métodos estatísticos e de aprendizado combinados com conhecimento *a priori*.

O *CiteSeer* (BOLLACKER et al., 1999) é um dos mais usados na busca de artigos científicos. O sistema monitora *newsgroups* e editores e mecanismos de busca a partir dos termos *publications*, *papers* e *postscript*. São extraídos dados bibliográficos do artigo e da bibliografia, que atua como lista, ajudando a achar outros artigos. O número de citações de um artigo por outros representa uma medida de sua relevância. Banco de dados de autores e revistas acadêmicas, bem como técnicas complexas são aplicadas para identificar co-referências de autores e artigos.

O *DEADLINER* (KRUGER et al., 2000) busca anúncios de conferências, extraíndo deles: data inicial, final e limite, comitê, afiliação de membros do comitê, temas, nome do evento e país. A performance de reconhecimento do *DEADLINER* está acima de 95%, contudo, sua definição de evento é mais restritiva: todos os atributos têm de estar presentes, exceto país, além de dados de submissão. O MASTER-Web oferece mais flexibilidade e cobertura, aceitando anúncios de capítulos de livros, jornais, revistas e concursos. Os requisitos estão em casos, que são mais flexíveis.

### OntoSeek: acesso baseado em conteúdo na Web

*OntoSeek* (GUARINO et al., 1999) é um sistema para recuperação de informações baseada em conteúdo de um repositório de páginas amarelas e catálogos de produtos. Combina um mecanismo de casamento de conteúdo orientado a ontologias com um formalismo de representação moderadamente expressivo. Ele também faz uso de termos em linguagem natural para obter descrições de recursos mais precisas, além de possuir uma completa flexibilidade terminológica para consultas, devido a um processo de casamento semântico entre as consultas e as descrições de recursos.

Este sistema faz uso do *WordNet*, que é um banco de dados lingüístico formado por *synsets* - termos agrupados em conjuntos de equivalência semântica, cada um atribuído a uma categoria léxica (substantivo, verbo, advérbio, adjetivo). Cada *synset* representa um sentido particular de uma palavra em inglês e é comumente expresso numa única combinação de sinônimos.

Geralmente, cada palavra é associada a mais de um *synset* e a mais de uma categoria léxica. Assim, para a desambiguação de sentidos de uma palavra, o *OntoSeek* trabalha interativamente com a ajuda da interface léxica do *WordNet*, permitindo que se selecione o *synset* e a categoria mais apropriados para a mesma. Ele também possui vários tipos de relações semânticas entres os *synsets* e que

são fundamentais no processo de desambiguação. Por exemplo, se desejamos procurar carros que possuem rádios, as descrições que devem ser selecionadas são apenas aquelas em que os conceitos “rádio” e “carro”, aparecem ligados pela relação “parte de”, portanto eliminando as lojas que vendem rádios e carros, por exemplo.

### **KIM: gestão de conhecimento e informação**

A plataforma KIM (POPOV et al., 2004) fornece uma infra-estrutura para gerenciamento de conhecimento e informação, além de serviços para anotação semântica automática, indexação e recuperação de documentos. Esta plataforma permite às aplicações fazerem uso dela para fins de anotação semântica automática, recuperação de conteúdo baseada em restrições semânticas e, finalmente, consulta e modificação das ontologias e da base de conhecimento.

Ela combina extração de informações baseada em outra madura plataforma para engenharia de textos, a GATE (Arquitetura geral para Mineração de Textos), que é uma abrangente plataforma de Processamento de Linguagem Natural e extração de informação desenvolvida na Universidade de Sheffield, Inglaterra e que tem estado em constante desenvolvimento desde 1995, tendo sido usada em uma grande variedade de projetos de pesquisa (MAYNARD et al., 2000).

Um dos pontos fortes da plataforma KIM é a geração automática de anotações sobre *entidades nomeadas* (*Named-Entities* - NE) - entidades do mundo real que são referenciadas por nome, tais como: *Pessoa*, *Organização*, *Empresa*, *Localidades*, etc. - com referências de classes e instâncias apontando para um repositório semântico. Através destas entidades, uma base de conhecimento com uma vasta cobertura de entidades do mundo real é mantida, usada e constantemente enriquecida, facilitando a interpretação de nomes próprios.

Esta plataforma emprega dois repositórios de conhecimento para o desempenho de suas tarefas: a Ontologia KIM (*KIM Ontology* ou *KIMO*) e uma base de conhecimento.

Mais especificamente, a KIMO é uma ontologia de topo com cerca de 250 classes de entidades e 100 atributos e relações. Este tipo de ontologia declara algumas distinções filosóficas básicas entre tipos de entidades, tais como: *Objetos* - entidades reais como localidades e agentes; *Acontecimentos* - definindo eventos e situações; e *Abstrações*, que não são nem Objetos nem Acontecimentos. Vale também notar que a KIM KB é pré-populada com instâncias da ontologia (mais de 80.000), bem como relações entre estas instâncias que auxiliam no processo de extração de informações para se ter bom desempenho em conteúdos inter-domínios na Web.

### **Discussão**

Os fatos apresentados revelam uma ferramenta promissora que, em todas as fases de seu ciclo evolutivo, contrasta positivamente em relação às ferramentas com propósitos similares, e.g., WebKB (CRAVEN et al., 1998), CiteSeer (BOLLACKER et al., 1999), DeadLiner

(KRUGER et al., 2000), KIM (POPOV et al., 2004) e Ontoseek (GUARINO et al., 1999), conforme descrito nos parágrafos a seguir.

A grande e inovadora idéia do MASTER-Web/AGATHE reside em sua arquitetura reusável e a portabilidade entre domínios, conforme evidenciado nos testes com os domínios acadêmico e de Inteligência Artificial.

Nos sistemas similares, o conhecimento está escondido dentro de seus algoritmos, não sendo possível o compartilhamento de conhecimento, nem a especificação de contextos em que seriam úteis. Assim nesta abordagem, para se processar uma nova classe, um novo sistema precisa ser elaborado, sem maiores reusos (FREITAS et al., 2003).

### **Considerações finais e trabalhos futuros**

A problemática referente à manipulação de informação na internet e em grandes bibliotecas digitais colocou questões de difícil solução para tornar efetivo o acesso ao grande manancial de informação disponível pelos usuários. O presente trabalho partiu da premissa de que sistemas baseados em conhecimento consistem numa alternativa de solução mais flexível e promissora do que as abordagens procedimentais tradicionais. O advento das ontologias, em particular, propicia a estruturação de conhecimento necessária para a especificação do conhecimento necessário às tarefas propostas. Procuramos, ao elaborar os dois sistemas apresentados, melhorar progressivamente a escalabilidade de nossa solução, e realizamos experimentos mostrando a viabilidade de explorar domínios novos com a mesma solução.

O presente trabalho e seus protótipos podem, ainda, ser estendidos em vários sentidos. Recuperação, extração e classificação de informação devem envolver o uso de tecnologias capazes de aprender a reconhecer padrões de informação de forma a recuperar acuradamente os itens de informação necessários. Pretendemos empregar estas técnicas basicamente por duas razões: elas aceleram o processo de aquisição de conhecimento necessário para as tarefas (e.g. a extração de citações de moedas, em que os padrões de reconhecimento e extração são aprendidos das páginas), e também porque melhoram o desempenho destas tarefas.

A habilidade para reconhecer padrões lingüísticos, presentes ou próximos a itens de dados a serem extraídos, permite um bom desempenho na tarefa de extração de informação mesmo diante de textos desestruturados, como só ocorre na Web.

A integração de ferramentas lingüísticas, bases de conhecimento e ontologias de topo e/ou de domínios específicos aponta para uma melhora expressiva de desempenho nas tarefas de manipulação de informação, particularmente extração e classificação. Para a língua inglesa, existe uma enorme gama de recursos disponíveis. Entre os recursos mais populares para o inglês, podemos citar a ontologia lingüística Wordnet (MILLER, 1995) e as ferramentas e recursos da arquitetura GATE.

Pretendemos ainda adequar as nossas soluções às linguagens de representação de conhecimento usadas na Web Semântica, tais como a linguagem de ontologias para a Web Semântica OWL (*Ontology Web Language*) (HERMAN, 2007) e a linguagem de regras SWRL (*Semantic Web Rule Language*) (HORROCKS et al., 2006).

## Agradecimentos

Agradecemos à Scopus Tecnologia pelo incentivo à pesquisa.

## Referências bibliográficas

BENJAMINS, V.R.; FENSEL, D.; GOMEZ-PÉREZ, A. Knowledge Management through Ontologies. In: INTERNATIONAL CONFERENCE ON PRACTICAL ASPECTS OF KNOWLEDGE MANAGEMENT, 2., 1998, Basel. **Proceedings**...Basel: Switzerland, 1998, p.1-5.

BOLLACKER, S.; LAWRENCE, C.; GILES, L. CiteSeer: an autonomous web agent for automatic retrieval and identification of interesting publications. In: INTERNATIONAL ACM CONFERENCE ON AUTONOMOUS AGENTS, 2., 1998, **Proceedings**... USA, 1998.

ESPINASSE, B.; FREITAS, F.; FOURNIER, S. AGATHE: an agent and ontology based system for restricted-domain information gathering on the web. In: INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCES (IEEE-RCIS), April 23-26, **Proceedings**... Ouarzazate, Morocco, 2007.

FREITAS, F.; BITTENCOURT, G. An ontology-based architecture for cooperative information agents. In: INTERNACIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE – IJCAI, 2003, Acapulco. **Proceedings**... México.

FREITAS, F. **Ontology of science**. Disponível em: <[http://protege.stanford.edu/plugins/ontologyOfScience/ontology\\_of\\_science.htm](http://protege.stanford.edu/plugins/ontologyOfScience/ontology_of_science.htm)>. Acesso em: 2001.

FREITAS, F. **Sistemas multiagentes cognitivos para a recuperação, classificação e extração integradas de informação da web**. 2002. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis.

FREITAS, F.L.G. Ontologias e a web semântica. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE

COMPUTAÇÃO, 23., 2003. **Anais**... Campinas: SBC, 2003.

GOOGLE, **Google Search Engine**, Disponível em: <<http://www.google.com>>. Acesso em: 2008

GUARINO, N. et al, OntoSeek: content-based access to the web. **IEEE Intelligent Systems**, v.14, n.3, p.70-80, May-Jun. 1999.

HERMAN, I. **Web ontology language (OWL)**, Disponível em: [www.w3.org/2004/OWL/](http://www.w3.org/2004/OWL/). Acesso em: 2007.

HORROCKS, I. et al. **SWRL: a semantic web rule language combining owl and ruleml: w3c member submission**, Disponível em: <<http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/>>. Acesso em: abr. 2006.

MAYNARD, D. et al. **A survey of uses of gate**. technical report cs-00-06. Department of Computer Science, University of Sheffield, 2000.

MILLER, G.A. Wordnet: a lexical database for English. **Communications of the ACM**, v.2, n.11, p.39-41, Nov. 1995.

PALMEIRA, E.; FREITAS, F. Detailed ontologies and text classification: a promising union. In: WORKSHOP ON BUILDING APPLICATIONS WITH ONTOLOGIES FOR THE SEMANTIC WEB, 1., 2006. ENCONTRO PORTUGUÊS DE INTELIGÊNCIA ARTIFICIAL, 12., 2006, Porto, **Proceedings**...Lisboa: APPIA, 2006.

POPOV, B. et al. KIM: a semantic platform for information extraction and retrieval, **Journal of Natural Language Engineering**, v.10, n.3-4, p.375-392, Sep. 2004.

PROTÉGÉ. Disponível em: <<http://Protege.stanford.edu>>. Acesso em: 2008.

RIBEIRO-NETO, B.; BAEZA-YATES, R. **Modern information retrieval**. Addison Wesley: ACM Press, 1999.

TILAB, JADE tutorial. Disponível em: <http://jade.tilab.com>. Acesso em: 2008

WIEDERHOLD, G. Mediation in information systems: research directions in software engineering. **ACM Computing Surveys**, v.27, n.2, p.265-267, June 1995. 

## Sobre os autores

### *Fred Freitas*

O autor é Ph.D (2002) em Engenharia Elétrica pelo Departamento de Automação e Sistemas da Universidade Federal de Santa Catarina (UFSC). Atualmente é Professor Adjunto do Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE), desde 2005, nos cursos de graduação em Ciência e Engenharia da Computação e membro do colegiado da Pós-Graduação em Ciência da Computação. Suas áreas de interesse são: Ontologias, Inteligência Artificial, Mediadores, Sistemas Multiagentes, Agentes Inteligentes e Recuperação e classificação de informação.

### *Luciano Cabral*

O autor está em fase de conclusão do Mestrado em Ciência da Computação pelo Centro de Informática da Universidade Federal de Pernambuco (2008), possui graduação em Bacharelado em Sistemas de Informação / Engenharia de Software pela Faculdade Integrada do Recife (2006). Atualmente é tutor virtual da Universidade Federal Rural de Pernambuco e Analista Trainee na Scopus Tecnologia. Tem experiência na área de Ciência da Computação, com ênfase em Inteligência Artificial, atuando principalmente nas áreas de interesse: Ontologias, Text Mining, KDT, Machine Learning, Extração/Classificação de Conhecimento e Informações.