

Busca de Documentos Distribuída com o Sistema MOTHRA

Paulo V. W. Radtke, Celso A. A. Kaestner
{radtke, kaestner}@ppgia.pucpr.br

PPGIA - PUCPR
Programa de Pós-Graduação em Informática Aplicada
Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição, 1155 – Prado Velho
80215-901 Curitiba (PR) – Brasil

Resumo

Este artigo apresenta o resultado de teste com o sistema MOTHRA, um sistema multi-agente que emprega a tecnologia de agentes móveis para a busca de documentos na *Web*. Neste ambiente, características específicas impossibilitam a aplicação direta de técnicas de *full text retrieval*, sendo necessárias abordagens que possibilitem contornar este problema ou partir para novas soluções. Os testes no sistema MOTHRA apresentados abordam questões de desempenho e de qualidade dos documentos retornados frente à consulta original do usuário.

Palavras-Chave: agentes móveis, recuperação de informações.

Abstract

This article presents the test results of the MOTHRA system, a multi-agent system which employs mobile agents for the retrieval of documents over the Web. Specific features of Web documents void the use of standard full text retrieval techniques, so specific approaches are required to solve this problem. The test results presented here are about the MOTHRA system performance and document relevance over the user query.

Keywords: mobile agents, information retrieval.

Introdução

Sob o ponto de vista de seu conteúdo, a *Web* consiste de um conjunto de documentos sem um padrão pré-estabelecido, distribuídos em uma rede numa topologia desconhecida. As informações nela contidas são voláteis, têm alta redundância e grau de confiança duvidoso. Além disto estão em um formato não-estruturado: o formato HTML define apenas a visualização do documento, mas não a estrutura da informação apresentada. No contexto da solução do problema clássico de Recuperação de Informações, quando se deseja recuperar os documentos relevantes a uma dada consulta, tais características representam uma barreira para a aplicação direta de técnicas de *full text retrieval* [19].

Para permitir a aplicação das técnicas de *full text retrieval* na *Web*, muitos dos sistemas de busca disponíveis atualmente (Altavista [1], Yahoo[20], etc) indexam os documentos cadastrados em uma base local, que serve de referência para fornecer o resultado de uma consulta ao sistema. Novas páginas devem ser diretamente cadastradas pelos seus *webmasters*, e não há uma preocupação em relação à volatilidade dos dados, sendo possível muitas vezes encontrar páginas que não existem mais como resultado de uma máquina de busca convencional. Algumas técnicas especializadas em busca na *Web* procuram resolver o problema da falta de confiança dos dados através do número de *hyperlinks* entre as páginas, usando a heurística que indica que quanto mais referenciada for uma página, maior será a confiança nas informações nela contida.

O sistema MOTHRA [14] – MOBILE Text and Hyperdocument Retrieval Architecture – tem por objetivo superar algumas das dificuldades acima apresentadas, é descrito a seguir.

1 O Sistema MOTHRA

O MOTHRA [14] é um sistema multi-agente [4], que emprega a tecnologia de agentes móveis [6] para realizar a busca e recuperação de documentos em formato HTML, em uma rede de topologia arbitrária. O objetivo primário da proposta é a de fornecer uma solução inovadora para a busca e recuperação de documentos em uma base de documentos desconhecida (a *Web*), sem que haja a necessidade da criação de um índice estático de todos os documentos. Desta forma a busca é realizada de forma dinâmica e em tempo-real, permitindo melhor performance na recuperação de documentos em bases cujo período de atualização é reduzido.

A solução é fundamentada numa característica intrínseca da *Web*, qual seja a possibilidade de um documento possuir *hyperlinks* (referências) para outro(s) documento(s), que possivelmente tratam do mesmo assunto ou de tópicos diretamente relacionados. Assume-se que a partir de um único documento é possível atingir parte considerável da base de documentos, bastando para isto seguir os *hyperlinks* encontrados (é possível que um grupo de documentos não possua referências para outros documentos, mas tais casos podem ser considerados raros). Desta forma pode ser estabelecida uma correspondência entre uma busca convencional realizada no espaço de estados de um problema [23] e a busca realizada pelo MOTHRA. Com efeito, há uma analogia direta entre o estado atual e o documento (*site*) que está sendo visitado em um determinado instante, e entre as transições entre estados e os *hyperlinks* que relacionam os documentos na *Web*. Uma vantagem sensível da abordagem adotada pelo MOTHRA, que atua com múltiplos agentes móveis, é a possibilidade de realizar estas buscas em paralelo.

A abordagem proposta segue a linha de filtragem de documentos, conforme proposto por Pazzani [11], constituindo-se num mecanismo não-convencional frente aos sistemas do tipo *full text retrieval*.

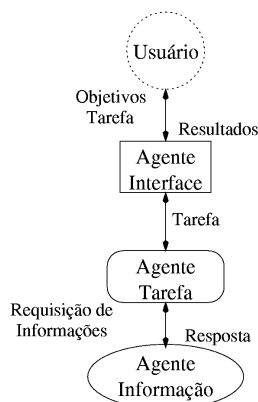


Figura 1 - Arquitetura do MOTHRA

A arquitetura do sistema MOTHRA é indicada na Figura 1. Cada agente da arquitetura é especializado em uma parte da tarefa global, da seguinte forma:

- Agente Interface: é o responsável pela interface do sistema com o usuário, utilizado especialmente na formulação das consultas e na apresentação dos resultados;
- Agente Tarefa: coordena as buscas em andamento no sistema e armazena as informações sobre buscas anteriores;
- Agente Informação: é o agente responsável pela busca em si, deslocando-se pela rede de servidor em servidor de forma a obter documentos que satisfaçam os critérios determinados pelo usuário.

Uma versão protótipo do sistema foi implementada em JAVA, atuando sobre a plataforma de agentes móveis Aglets SDK da IBM [22].

2 Avaliação do Sistema MOTHRA

Para os testes com o protótipo do MOTHRA foi empregado um mecanismo de avaliação que determina a relevância de um documento em relação a um questionamento realizado por meio de um conjunto de palavras. Em tal avaliação foram utilizadas várias técnicas de pré-processamento de textos, tais como a eliminação de *stop-words* – palavras de uso muito comum que devem ser ignoradas na busca, bem como a aplicação de *stemming* – a eliminação de prefixos / sufixos das palavras de forma a se obter a sua forma radical [13]. Tais procedimentos também são aplicados às palavras indicadas no questionamento, para permitir a compatibilidade entre as mesmas e os documentos em sua forma posterior ao tratamento.

A lista de *stop-words* e o algoritmo de *stemming* [13] utilizados são para a língua inglesa, tendo em vista a base textual que foi empregada nos testes do sistema. A utilização do MOTHRA para outras línguas – como para o Português – envolve a criação de uma lista de *stop-words*, e de um algoritmo de *stemming* adequados.

Após a etapa de pré-processamento, a relevância de um documento para uma consulta é determinada pela aplicação de duas medidas estatísticas calculadas sobre a distribuição das freqüências das palavras da busca no documento (tf_i : *term frequency* da palavra i). A

primeira medida é a média aritmética (\bar{X}) das frequências tf_i . Desta forma, um documento com $\bar{X} = 3$ é mais relevante que um documento com $\bar{X} = 1$. A segunda medida é o desvio padrão (S) da distribuição dos tf_i , usado como indicativo para a correlação entre as ocorrências das diversas palavras. Normalmente, as palavras da busca indicam um contexto e têm uma ocorrência em conjunto, o que não pode ser identificado usando-se apenas a média. O desvio padrão entra como uma medida auxiliar, evitando que documentos com contagem elevada em, por exemplo, apenas uma palavra, e baixa nas demais, tenha uma indicação de relevância elevada.

Assim, no MOTHRA a relevância de um documento em relação a uma consulta é determinada por um *Peso*:

$$Peso = \frac{\bar{X}}{S + 0.001}$$

O valor 0.01 é inserido apenas para permitir que o documento tenha um peso válido mesmo quando a contagem de todas palavras-chave forem iguais (desvio padrão igual a 0), aumentando o seu valor em relação aos demais sem perda da consistência.

Outros procedimentos mais sofisticados podem ser considerados, como o uso da frequência inter-documentos de um termo (*idf*), e a possibilidade de se avaliar a qualidade de uma página de acordo com os *hyperlinks* apontando para esta. O cálculo da frequência inter-documento de um termo no MOTHRA é não-trivial, já que vários agentes realizam a busca em paralelo, sem um ponto central de processamento (à exceção do retorno dos resultados). Mesmo com a possibilidade do agente levar consigo tais informações, a quantidade de dados trafegada pela rede seria muito grande, invalidando a proposta mais geral de se levar a computação aos dados para diminuir o tráfego na rede. Por outro lado não há conhecimento prévio de quantas páginas apontam para um dado documento, sendo no máximo possível determinar no conjunto de páginas já pesquisadas na busca que apontaram para o documento em questão, o que não reflete adequadamente a medida desejada.

3 Base de Testes: Tópicos e Relevâncias da TREC

A TREC (Text REtrieval Conference) [10] nos seus esforços para a melhoria dos mecanismos de busca e recuperação de informações, disponibiliza para pesquisadores uma base de documentos padronizada, a também um conjunto de questionamentos e as relevâncias de cada documento em relação aos mesmos. A avaliação destas relevâncias foi feita por seres humanos, que analisaram o conteúdo dos documentos frente a um questionamento.

Cada questionamento contém uma descrição do domínio de conhecimento envolvido (ex: *international economics*), um título (ex: *Airbus Subsidies*), um sumário e uma descrição, além de uma narrativa, que descreve o que deve conter um documento relevante à busca. Como um complemento, cada questionamento possui uma lista dos conceitos envolvidos; no exemplo de subsídios para a Airbus, encontram-se conceitos tais como tratados comerciais, subsídios federais, política *anti-dumping*, entre outros.

Como as buscas no protótipo do MOTHRA são feitas a partir de um conjunto simples de palavras, foi escolhido o sumário do questionamento para a consulta, por ser sucinto e não envolver diretamente o uso de conceitos. Para a comparação dos resultados obtidos, foram contrastados os documentos recuperados pelo MOTHRA com os dados como corretos pela TREC em suas avaliações de relevância.

Para a avaliação dos resultados obtidos nas buscas foram utilizadas as medidas de precisão, recobrimento e F-Measure [15]. *Precisão* é a relação entre o número de documentos corretos retornados e o número total de documentos retornados, e *recobrimento* é a razão

entre o número de documentos corretos retornados e a quantidade de documentos corretos existente na base. A última medida, a *F-Measure*, proposta por van Rijsbergen [15] relaciona as medidas de precisão e recobrimento por meio de uma média harmônica:

$$F - MEasure = \frac{2 * Rel * Prec}{Rel + Prec}$$

Assim, para obter-se um bom resultado, é preciso obter os melhores valores possíveis para a relevância e precisão [3], já que quanto mais ambos se aproximam de 1 (um), maior é o valor da F-Measure.

4 Testes de Relevância de Documentos

Com o intuito de verificar a qualidade dos resultados fornecidos pelo sistema MOTHRA, uma base de 15.534 documentos em formato HTML foi preparada a partir de documentos da base textual TIPSTER [17]. Esta base, por questões de velocidade dos testes, foi instalada em apenas uma máquina (Pentium Celeron 366, 64Mb RAM, 4,3Gb de disco), rodando Windows NT4.0 e Aglets 1.03b. Mesmo estando localizados fisicamente em uma única máquina, os documentos possuem relacionamentos por meio dos *links*, que foram utilizados como o único meio para realizar a busca.

Foi escolhida para os testes a base do jornal San José Mercury, documentos SJMN91-06300045 a SJMN91-06364024 (identificação única na base para cada documento). Como indicado, estão disponíveis na TREC [10] questionamentos nas bases da TIPSTER e suas relevâncias conforme o julgamento por seres humanos. Quatro questionamentos foram escolhidos, para a realização de 5 testes (um deles é repetido para demonstrar uma característica do mecanismo de busca), sobre os quais foram determinadas as medidas de precisão, recobrimento e F-Measure.

Não foi possível a realização de testes diretamente na Internet, devido à indisponibilidade de servidores Aglets capacitados a receber os agentes móveis, e por serem necessários, para a avaliação, a definição de um conjunto de questionamentos e os respectivos documentos relevantes. Os resultados apresentados nesta seção se aplicam então a uma situação específica e servem como uma projeção para os possíveis cenários da execução do MOTHRA no ambiente *Web*.

4.1 Questionamento 52

Discute sanções internacionais aplicadas a África do Sul devido a política de discriminação racial do *apartheid*. Deste tópico, o sumário fornecido pela TREC é “*Document discusses international sanctions against South Africa*”, a partir do qual foi criado o questionamento com as palavras de busca “*international sanctions against South Africa*”.

Na base do teste há 9 (nove) documentos que satisfazem o questionamento, sendo estes 9 retornados corretamente pelo sistema MOTHRA. Como se esperava, outros documentos que contém as palavras da busca, mas não se relacionam ao tópico, foram retornados, introduzindo um nível de ruído. Na Tabela 1, apresentam-se os valores obtidos de precisão, recobrimento, F-Measure a partir de um valor limite mínimo (*threshold*) para o peso.

Limite	Precisão	Recobrimento	F-Measure	Retornados	Corretos
1	0,321429	1	0,4864865	28	9
1,1	0,36	1	0,5294118	25	9
1,2	0,36	1	0,5294118	25	9
1,3	0,409091	1	0,5806452	22	9
1,4	0,4	0,888888889	0,5517241	20	8
1,5	0,4375	0,777777778	0,56	16	7
1,6	0,454545	0,555555556	0,5	11	5
1,7	0,333333	0,333333333	0,3333333	9	3
1,8	1	0,222222222	0,3636364	2	2
1,9	1	0,222222222	0,3636364	2	2
2	1	0,222222222	0,3636364	2	2
0	0,003535	1	0,007045	2546	9

Tabela 1 - Resultados do Tópico 52

4.2 Questionamento 51

Este segundo tópico escolhido discute disputas comerciais entre empresas de aeronáutica a partir de subsídios fornecidos pelo governo americano à Airbus Industrie. O sumário da TREC é “*Document will discuss government assistance to Airbus Industrie, or mention a trade dispute between Airbus and a U.S. aircraft producer over the issue of subsidies*”. A partir deste resumo foi criado o questionamento com as palavras “*trade dispute between airbus industrie and aircraft producer over government subsidies*”.

Em toda a base de teste há apenas um documento que satisfaz este questionamento, o que torna este teste especialmente interessante, já que a escolha das palavras de busca (como será mostrado a seguir nos dois últimos testes) é fundamental para a obtenção de um bom resultado. Este único documento resposta foi retornado pelo MOTHRA, como sendo o quarto mais relevante da busca (Tabela 2).

Limite	Precisão	Recobrimento	F-Measure	Retornados	Corretos
0,5	0,00102	1	0,002039	980	1
0,6	0,003663	1	0,007299	273	1
0,7	0,005525	1	0,010989	181	1
0,8	0,027027	1	0,052632	37	1
0,9	0,05	1	0,095238	20	1
1	0,2	1	0,333333	5	1
1,03	0,25	1	0,4	4	1
0	0,000211	1	0,000422	4740	1

Tabela 2 - Resultados do Tópico 51

4.3 Questionamento 59

Este questionamento se relaciona com mortes causadas por fenômenos meteorológicos, como furacões, enchentes, verões ou invernos intensos e outros. O sumário fornecido pela TREC é “*Document will report a type of weather event which has directly caused at least one fatality in some location*”, dando origem ao questionamento empregado no MOTHRA “*weather event which has directly caused fatalities in some location*”.

Tal busca apresenta um resultado regular, sendo que dos 42 documentos esperados apenas 33 foram encontrados. Isto deve-se ao fato de que as palavras-chave “*weather event*” e “*some location*” não implicam nos tipos de fenômenos meteorológicos esperados, como chuva, enchente, furacão e outros, ou os locais em que eles correram; ou seja, não há

conhecimento de domínio envolvido. Tal problema pode ser resolvido com o uso de redes semânticas ou assistentes de ontologias na formulação do questionamento, conforme proposto por Barros [3], o que teoricamente deve melhorar o resultado. Os próximos dois testes mostram este problema com mais clareza (Tabela 3).

Limite	Precisão	Recobrimento	F-Measure	Retornados	Corretos
0,4	0,013453	0,785714286	0,0264529	2453	33
0,5	0,058824	0,452380952	0,1041096	323	19
0,6	0,062092	0,452380952	0,1091954	306	19
0,7	0,060185	0,30952381	0,1007752	216	13
0,8	0,162162	0,142857143	0,1518987	37	6
0,9	0,291667	0,166666667	0,2121212	24	7
0	0,012962	0,785714286	0,0255023	2546	33

Tabela 3 - Resultados do Tópico 59

4.4 Questionamento 54 - Primeiro Teste

Este questionamento trata de reservas para o lançamento de um satélite comercial, sendo o sumário da TREC “Document will cite the signing of a contract or preliminary agreement, or the making of a tentative reservation, to launch a commercial satellite”. Deste sumário, foi criado o questionamento “contract or agreement or reservation to launch a commercial satellite”. Há 3 documentos relevantes na base. Apesar dos 3 terem sido encontrados na busca, e do primeiro documento encontrado como terceiro mais relevante, o segundo mais relevante foi encontrado apenas na posição 25, o que tomaria muito tempo do usuário para verificar os documentos esperados e identificar os corretos. Visando uma melhoria no resultado, o teste foi refeito com outras palavras de busca (Tabela 4).

Limite	Precisão	Recobrimento	F-Measure	Retornados	Corretos
0,6	0,005797	1	0,0115274	345	2
0,7	0,010363	1	0,0205128	193	2
0,78	0,08	1	0,1481481	25	2
0,8	0,055556	0,5	0,1	18	1
0,9	0,1	0,5	0,1666667	10	1
1	0,333333	0,5	0,4	3	1
0	0,000734	1	0,0014668	2725	2

Tabela 4 - Resultados do Tópico 54 - Primeiro Teste

4.5 Questionamento 54 - Segundo Teste

Para este teste, foram reformuladas as palavras chave de “contract or agreement or reservation to launch a commercial satellite” para “reservation to launch a commercial satellite”. Tal mudança foi feita baseada em uma inspeção visual na base de documentos.

Os documentos corretos nesta base parcial da TIPSTER tratam apenas da reserva para o lançamento de satélites comerciais, não havendo necessidade das palavras “*contract*” e “*agreement*” no questionamento, tornando-o mais restrito. É importante ressaltar que tal inspeção foi feita em caráter de avaliação dos resultados em dois questionamentos diferentes, e só foi possível na base de testes por haver avaliação prévia da mesma. Os resultados foram melhores do que em relação ao teste anterior (os documentos corretos figuraram nas posições 1 e 4), mostrando assim o impacto da melhoria das palavras de busca no resultado obtido (Tabela 5).

Limite	Precisão	Recobrimento	F-Measure	Retornados	Corretos
0,9	0,022472	1	0,043956	89	2
1	0,5	1	0,666667	4	2
1,1	0,5	1	0,666667	4	2
1,2	0,333333	0,5	0,4	3	1
1,7	0,333333	0,5	0,4	3	1
2	1	0,5	0,666667	1	1
0	0,001212	1	0,002421	1650	2

Tabela 5 - Resultados do Tópico 54 - Segundo Teste

5 Testes de Desempenho do Sistema

Visando validar o desempenho do sistema, foi criada uma base de documentos distribuída em três computadores (COMPAQ Pentium 200, 32Mb RAM e 2Gb de disco, Windows NT4.0, Aglets 1.03b e servidor HTTP Apache). A configuração de hardware das máquinas é idêntica para permitir uma comparação simples.

O sistema foi testado em três situações: (a) abordagem completamente centralizada; (b) abordagem centralizada, com base distribuída na rede; e (c) abordagem com busca e base distribuída na rede. A mesma consulta foi submetida às três situações para avaliar o desempenho do sistema e verificar se há ganho real no uso de busca distribuída com agentes móveis. Como os agentes interagem indiretamente entre si através de pegadas [21] deixadas no sistema (solução específica para a comunicação durante o deslocamento na rede), é interessante avaliar o impacto em termos de processamento deste mecanismo. Assim, a abordagem centralizada é dividida em suas situações: com e sem o mecanismo de interação.

Após executar o questionamento, foram obtidos os valores para uso de memória, processador e tempo total para execução, expressos na Tabela 6.

Teste	Uso de Memória (%)			Uso de Processador (%)			Tempo (mm:ss)
	Máquina 1	Máquina 2	Máquina 3	Máquina 1	Máquina 2	Máquina 3	
Agente Centralizado Não-Interativo, Base Centralizada ¹	70,08	-	-	63,01	-	-	33:58
Agente Centralizado, Base Centralizada	69,81	-	-	55,22	-	-	47:34
Agente Centralizado, Base Distribuída	69,98	47,52	58,13	59,48	0,24	1,36	36:43
Agente Distribuído, Base Distribuída	69,39	55,92	65,79	40,26	38,02	39,97	21:37

Tabela 6 - Resultados do Teste de Desempenho do MOTHRA

Nota-se claramente um ganho no tempo de execução da tarefa, o que valida a abordagem da busca com o uso de agentes móveis. Com estes resultados é possível inferir que se o número de documentos aumentar na coleção e forem mantidas o número de máquinas, a abordagem de busca centralizada e distribuída terão o seu tempo aumentado proporcionalmente. Porém, se o número de documentos crescer pela adição de novas máquinas na rede, o tempo da abordagem centralizada aumentará, mas a busca distribuída irá manter-se o mesmo. O mesmo aconteceria se a base fosse mantida e aumentado o número de máquinas, a abordagem centralizada manteria o tempo total da tarefa, ao passo que a busca distribuída seria executada em menos tempo.

¹ Neste teste, os mecanismos de interação entre os agentes não estão ativos.

6 Conclusões e Perspectivas Futuras

Os resultados apresentados indicam que a proposta MOTHRA, baseada em agentes móveis que realizam uma busca distribuída, se constitui numa interessante alternativa à realização de tarefas de recuperação de informações em ambientes distribuídos. Em relação à qualidade dos resultados, é possível observar que mesmo com um mecanismo simples de comparação entre o questionamento e o documento a recuperar, o sistema fornece resultados similares aos de uma máquina de busca convencional da Internet. Salienta-se novamente o aspecto paralelo da busca, principal característica do MOTHRA, que tira vantagem do aspecto distribuído dos agentes móveis na rede.

Ficam abertas outras possibilidades para o emprego de mecanismos mais sofisticados de comparação – tais como generalizações distribuídas da medida $tf*idf$ [2], que podem fornecer resultados ainda mais expressivos.

Outras possibilidade de pesquisa consiste no uso de ontologias na formulação do questionamento, de forma a fornecer palavras de consulta com um significado mais abrangente (como o caso de “*fenômeno meteorológico*”), ou com muitos sinônimos.

Como última consideração, é previsto como trabalho futuro uma investigação detalhada sobre o impacto do tráfego de dados nas abordagens tradicionais (máquinas de busca na *Web*) frente a abordagem distribuída proposta pelo MOTHRA.

7 Referências

- [1] Altavista. <http://www.altavista.com>.
- [2] Baeza-Yates, Ricardo e Ribeiro, Berthier de Araújo Neto. Modern Information Retrieval. Addison Wesley, Harlow, Inglaterra, 1999.
- [3] Barros, Flávia A., Gonçalves, Pedro F. e Santos, Thiago L. V. L.. Ontologies for Enhancing Web Searches Precision e Recall. Anais do XXV Seminário Integrado de Software e Hardware (SEMISH'98), 1998.
- [4] Ferber, Jacques. Les Systèmes Multi-Agents – Vers une intelligence collective. InterEditions, Paris, 1995.
- [5] Frankli, S. e Graesser, A.. Is it na Agent or Just a Program? – A Taxonomy for Autonomous Agents. <http://www.mscl.memphis.edu:80/franklin/Agent-Prog.html>, Março, 1996.
- [6] Lange, Danny B.. Mobile Agents: The Future of Distributed Computing? Lecture Notes in Computer Science 1445, 1998.
- [7] Lieberman, Henry. Letizia: Assisting Web Browsing. <http://lieber.www.media.mit.edu/people/lieber/Lieberary/Letizia-Intro.html>, 1997.
- [8] Luke, S., Spector, L. e Rager D. L. Ontology-based Knowledge Discovery on the World-Wide Web. Proceedings of the Workshop on Internet-based Information Systems/AAAI-96, 1996.
- [9] Maes, Paty e Lashkari, Yezdi. Webdoggie (webhound). <http://webhound.www.media.mit.edu/projects/webhound>, 1997.
- [10] NIST- National Institute of Standards & Technology. TREC: Text REtrieval Conference. <http://trec.nist.gov>.
- [11] Lieberman, Henry. Letizia: Assisting Web Browsing. <http://lieber.www.media.mit.edu/people/lieber/Lieberary/Letizia-Intro.html>, 1997.
- [12] Pazzani, Michael e Shoham, Y. LIRA. <http://robotics.stanford.edu/users/marko/lira/demo1.html>, 1997.

- [13] Porter, M. F. An Algorithm for Suffix Stripping. Readings in Information Retrieval, 1993.
- [14] Radtke, Paulo V. W.. MOTHR: Uma Proposta de Sistema para busca de Informações Baseado em Agente Móveis. Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática Aplicada da PUCPR, 2000.
- [15] Rijsbergen, K. V. Information Retrieval. Butterworths, Londres, 1979.
- [16] Starr, B., Ackerman, M. e Pazzani, M. DICA.
<http://www.uci.edu/ackerman>.
- [17] University of Pennsylvania. Linguistic Data Consortium.
<http://www ldc.upenn.edu>.
- [18] Utiyama, M. e Hasida, K. Bottom-Up Aligment of Ontologies. IJCAI-97 Workshop EP24 – Ontologies and Multilingual NLP, 1997.
- [19] Witten, Ian H., Moffat, Alistair e Bell, Timothy C. Managing Gigabytes – Compressing and Indexing Documents and Images. Van Nostrand Reinhold, Nova Iorque, 1994.
- [20] Yahoo! <http://www.yahoo.com>.
- [21] Radtke, Paulo V.W., Kaestner, Celso A.A., Scalabrin, Edson E. Pegadas – Uma Proposta para a Interação entre Agentes Móveis na Internet. ISADS 2000, México, 2000.
- [22] Tokyo Research Labs – IBM Corporation. Aglets Software Development Kit. <http://www.trl.ibm.co.jp/aglets>, 1998.
- [23] Norvig, Peter, Russel, Stuart. Artificial Inteligence – A Modern Approach. Prentice Hall Inc., New Jersey, 1995.